

Il est en outre fondamental de conduire des travaux juridiques inhérents aux services numériques basés sur de l'IA pour garantir le cadre de confiance avec le corps enseignant, les élèves et les parents. Il est obligatoire de démontrer la conformité au RGPD²⁴, à travers, par exemple, des analyses d'impact sur la protection des données à caractère personnel et l'analyse des flux de données collectées, stockées et utilisées.

Recommandations et perspectives

Pédagogie et co-conception

Afin d'éclairer tous les acteurs de la société, l'IA doit être étudiée et connue selon tous les points de vue (mathématiques et informatique ; sciences économiques; philosophie, sociologie, sciences de l'éducation, sciences politiques, ...). Ses apports doivent être connus tant en termes d'efficacité technique, que les questions éthiques qu'elle soulève, les impacts sociaux et économiques qu'elle aura. Cela ne pourra passer que par une forte acculturation et un accompagnement autour des enjeux de l'éducation aux données ; de former les cadres et les enseignants aux potentiels de l'IA et sur la manière d'amener les questions éthiques que l'IA peut poser auprès de leurs différents publics (élèves, parents). En parallèle d'une médiation large grand public, comme le préconise Cédric Villani dans son rapport (Villani, 2018), l'Education nationale doit accentuer ses efforts d'acculturation par de la formation, des animations, expositions et mises à disposition de contenus pédagogiques pour les enseignants.

En parallèle, l'Education nationale doit accentuer et promouvoir toutes les actions permettant une bonne acculturation de l'IA, et continuer à interroger les représentations et besoins des enseignants.

Dans l'éducation, comme ailleurs, et dans une volonté d'acceptation la plus importante, il sera primordial d'impliquer dans la conception des outils et services utilisant de l'IA, les apprenants et les enseignants.

Parmi les recommandations et les perspectives, on s'intéresse ici autant au développement d'agents conversationnels et à leurs expérimentations en classe, qu'aux usages d'outils commerciaux, ou encore à leur appropriation par les élèves, les professeurs et les parents d'élèves.

Recommandations dans le développement d'agents conversationnels

Comment piloter un projet d'interface conversationnelle qui satisfasse l'utilisateur

Les études de la littérature sur les chatbots montrent qu'il est efficace d'utiliser les plateformes de messageries instantanées où les utilisateurs sont déjà présents pour

²⁴ Règlement Général sur la Protection des Données.

construire un chatbot personnalisé par-dessus ²⁵. TechCrunch par exemple envoie quotidiennement cinq articles via l'application Facebook Messenger. Comme le stack d'un chatbot est composée de trois parties (l'interface de chat, la couche de processus du langage et la couche pour accéder aux bases de données), si certaines composantes sont déjà présentes dans d'autres applications, il est possible de les réutiliser.

- **Adapter au type d'utilisateur** : selon le type d'utilisateur auquel la plateforme s'adresse, on doit adapter la solution : les baby boomer préfèrent les appels téléphoniques aux messageries instantanées²⁶, les plus jeunes ne veulent plus Facebook Messenger (des bots peuvent être conçus à travers cette plateforme), etc. La solution doit être adaptable selon que l'élève ou le professeur/parent l'utilise.
- **Garantir l'inclusion technologique et l'intégration des élèves handicapés** : il est fondamental de développer des outils inclusifs pour tous les élèves, en prenant en considération les différences de niveaux, les difficultés de certains élèves, ou encore les handicaps. Les assistants conversationnels doivent être dimensionnés tant sur la partie algorithme que sur l'interface utilisateur, pour permettre aux enfants souffrant de handicaps de les utiliser. À cela s'ajoute le fait que certains outils peuvent être spécialement imaginés pour aider les élèves souffrant de handicaps. Le design doit être imaginé ainsi dès le démarrage de la conception et du développement.
- **Établir les objectifs d'usages** : les objectifs de l'interface conversationnelle doivent aussi être bien établis : **adoption, engagement puis efficacité**. Parfois les trois ne peuvent être atteints dès le début. Choisir qui va répondre : l'humain ou la machine. Commencer par construire un MVP pour apprendre ce que les utilisateurs veulent vraiment si ce n'est pas complètement connu au début. Le chatbot va apprendre des conversations passées. Plus il en a, mieux c'est.²⁷
- **Garantir la réversibilité de la plateforme et des données** : ne pas se rendre dépendant d'un unique fournisseur. S'assurer que les services de l'interface peuvent évoluer indépendamment du fournisseur si besoin.

Évaluer les solutions clés en main

Aucun fournisseur ne peut avoir de succès sans un modèle d'intention robuste, une sécurité/authentification parfaite, et une capacité multicanal (réseaux sociaux, applis mobiles, applis de messagerie etc.).²⁸

Pour évaluer et comparer des fournisseurs de solutions d'agents conversationnels, il est essentiel de comprendre l'architecture globale des solutions ainsi que les différents services spécifiques proposés à chaque étape. Beaucoup de fournisseurs n'utilisent pas d'IA

²⁵ Peter Wannemacher. Bots Aren't Ready To Be Bankers. Etude Forrester, 2016

²⁶ Julie A. Ask et Andrew Hogan. Chatbot commerce Benchmark. Etude Forrester, 2016

²⁷ Michael Facemire, Julie A. Ask, and Andrew Hogan. Chatbots 101: Building Conversational Interfaces, Etude Forrester, 2016

²⁸ Ian Jacobs. The Top 10 Chatbots For Enterprise Customer Service. Etude Forrester, 2016

mais des connaissances qui ont été triées a priori²⁹. Beaucoup d'utilisateurs s'imaginent bénéficier de modèles d'IA très avancés et peuvent être déçus par une expérience qui ne soit pas à la hauteur de leurs espérances.

Les industriels ont adopté différentes approches pour offrir des services conversationnels. On peut catégoriser l'offre en fonction de différents critères :

Le type d'utilisateur :

- chatbots,
- virtual customer assistants (VCA, adaptée pour automatiser des canaux d'engagement textuel tels que le chat en direct, les applications de messagerie grand public ou les SMS),
- virtual personal assistant (VPA)
- virtual enterprise assistant (VEA)

Les usages³⁰:

- chatbot social (conversations non structurées "comme un humain discuterait"),
- bots orientés pour une tâche spécifique : soit plusieurs tâches simples du quotidien soit plus spécialisé comme la recherche d'information (commander à manger, réserver un vol, etc.).

La technologie :

- Bots basés sur des règles encodées strictement,
- Bots basés sur de la donnée (data-driven).
- Une combinaison des deux points précédents : utilisation de règles apprises et données.

L'architecture de base est constituée d'un connecteur (plateforme de messagerie voix ou texte), une brique de NLP pour comprendre la sémantique de la requête utilisateur et un arbre de décision ou un modèle simple pour sélectionner une réponse possible³¹. Ceci constitue une base mais les différents fournisseurs doivent se différencier en offrant davantage de services.

Pour la plupart des solutions de chatbots, il n'y a pas d'interface utilisateur spécifique et le service conversationnel est embarqué dans une messagerie où l'utilisateur est déjà présent : Messenger, WeChat, Kik par exemple. Ces supports permettent une diversification dans le type de requête de l'utilisateur comme le traitement d'images mais elle enferme la solution dans une plateforme de messagerie spécifique.

Pour VPA et VEA, une interface utilisateur spécifique est plus souvent développée.

Critères d'évaluation des capacités d'intégration d'une solution :

²⁹ Magnus Revang, Brian Manusama, Anthony Mullen. Architecture of Conversational Platforms. Etude Gartner, 2017.

³⁰ Maali Mnasri. Recent advances in conversational NLP : Towards the standardization of Chatbot building. Preprint, 2019

³¹ Magnus Revang, Brian Manusama, Anthony Mullen. Architecture of Conversational Platforms. Etude Gartner, 2017.

- configuration multi-utilisateurs,
- la diversité des requêtes possibles (textes, images, vidéos, une combinaison des 3, etc.),
- la longueur de la requête autorisée (la plupart des briques de NLP accepte une longueur maximale de texte en entrée. Cette longueur peut constituer un élément différenciant).

Critères pour évaluer la voix :

- reconnaissance de la voix de l'utilisateur,
- authentification biométrique,
- détection de la langue, etc.

Critères pour évaluer la brique de NLP :

- enrichissement sémantique des requêtes grâce à des knowledge bases internes,
- spécification du modèle entraîné (ou non) sur un jeu de données spécifique au cas d'usage (ou modèle générique par défaut),
- apprentissage en continu du modèle au cours du temps (ou pas).

Critères pour évaluer le matching de l'intentionnalité de la requête (lié avec la brique NLP) :

- contextualisation,
- possibilité de choix entre différents modèles en fonction du type d'intentionnalité,
- extraction de modèles (patterns) récurrents et bien identifiés ("play a song"),
- priorisation dans les éléments d'une requête "commander un livre" puis "livrer à la bonne adresse".

Critères pour évaluer la manière dont la réponse est générée :

- génération de langage naturel,
- dialogue,
- *text to speech*
- ...

Enfin, certaines solutions peuvent générer des rapports d'utilisations avec des métriques spécifiques.

Évaluer les chatbots qui utilisent du « machine learning »

Les systèmes ne dépendent plus de réponses déterministes de la correspondance de modèles basée sur des règles, comme ELIZA et ALICE. L'évaluation est donc plus complexe.

Répartir la responsabilité des tests entre fournisseur de services (qui est responsable du test des intrants, de l'exécution des actions et le réalisme des réponses) et le client (qui évalue la facilité d'utilisation et l'efficacité de l'accomplissement des tâches)³². Pour évaluer un nouveau chatbot ou une nouvelle version du chatbot (par exemple réentraîné sur des nouvelles données), les ingénieurs sélectionnent ainsi les critères d'évaluation de qualité

³² Nicole Radziwill and Morgan Benton. Evaluating Quality of Chatbots and Intelligent Conversational Agents. Preprint 2017

des actions (sans métrique business) et les utilisateurs sélectionnent des métriques business (efficacité pour aider à réaliser une tâche).

Dans les modèles génératifs, les décodeurs et la longueur de la réponse sont particulièrement importants. On peut fixer une longueur minimum ou prédire la longueur de la réponse grâce à des données de dialogue humain-humain³³. Le processus d'entraînement se fait souvent en effectuant un pré-entraînement sur un large volume de données puis en raffinant sur les données du cas d'usage généralement disponibles en moins grand volume. Les données de dialogues disponibles en grande quantité doivent néanmoins être choisies avec attention pour anticiper les problèmes de biais sexistes, discriminatoires, etc. Des moyens de réduire ces biais en choisissant de générer des mots plus neutres.³⁴ La qualité du générateur de texte est très importante pour le résultat final : des réponses trop courtes sont perçues comme ennuyeuses ou montrant un manque d'intérêt, et des réponses trop longues sont perçues comme un manque d'écoute de la part du bot. On a différents choix d'architecture de bots : ceux qui génèrent des réponses en sélectionnant la réponse pertinente dans un ensemble fixe de réponses candidates (retrieve), ceux qui génèrent du texte (seq2seq) ou une combinaison des deux pour pallier les problèmes des modèles générateurs qui peuvent être répétitifs ou peu appropriés pour apprendre des nouvelles connaissances au cours du temps.

Recommandations pour déployer une solution de chatbots

- **Évaluer la nécessité d'incorporer du machine learning dans un chatbot** : au début d'un projet, il n'est pas forcément nécessaire d'incorporer des algorithmes d'apprentissage dans les chatbots pour que ces derniers s'améliorent en continu avec l'accroissement du nombre de conversations et d'interactions. On peut utiliser une solution statique dans un premier temps car certaines tâches ne nécessitent pas une compréhension poussée du contexte et de la sémantique. Choisir une technique plutôt qu'une autre dépend de la donnée disponible au début du développement (volume, structurée ou pas, labellisée ou pas) et du cas d'usage final³⁵. Les techniques « rule based » sont simplistes mais efficaces dans les cas les plus basiques. Elles peuvent aussi compléter un modèle de machine learning. Les approches basées sur du machine learning sont maintenant dominantes. Le Reinforcement Learning apprend des feedbacks de l'utilisateur au cours du temps et l'expérience s'améliore en continue, ce qui la rend très naturelle. Mais il faut garder en tête une contrainte : la technologie a besoin de beaucoup de temps et de beaucoup d'interactions avant que l'agent conversationnel soit entraîné de façon

³³ Roller S, Dinan E, Goyal N, Ju D, Williamson M, Liu Y, Xu J, Ott M, Shuster K, Smith EM, Boureau YL. Recipes for building an open-domain chatbot. arXiv preprint

³⁴ Dinan, Emily, et al. Queens are powerful too: Mitigating gender bias in dialogue generation. Preprint, 2019

³⁵ Maali Mnasri. Recent advances in conversational NLP :Towards the standardization of Chatbot building. Preprint 2019

satisfaisante. Si l'entraînement se fait en ligne, cela peut poser des problèmes pour les premiers utilisateurs.

- **Évaluer les cadres actuels de bot pour une utilisation possible dans le contexte voulu**, à la fois pour leur applicabilité aux tâches (use cases) et pour les avantages qu'ils offrent aux utilisateurs en termes d'efficacité et de facilité d'utilisation³⁶.
- **Ne pas fragmenter le parcours utilisateur**. Les VCA permettent de mieux comprendre les utilisateurs ; lorsque les utilisateurs communiquent de manière naturelle et conversationnelle, ils en révèlent davantage sur leurs préférences, leurs opinions, leurs sentiments et leurs inclinations. Intégrer les multiples bases de connaissances au sein du VCA³⁷.
- **Déployer au moins un framework de chatbot en interne** et commencer à développer activement des briques d'analyse pour évaluer leur utilisation en mettant l'accent sur les points faibles actuels dans l'utilisation des cas d'usages³⁸.
- **Assurer un consensus fort et clair sur la justification et les objectifs d'investissement**. Se concentrer sur l'objectif du cas d'application et apports pédagogiques et scientifiques, en ne s'éloignant pas du facteur "tendance". Améliorer de manière incrémentale. Le développement d'une application doit fournir des résultats rapidement avec un faible coût de démarrage, un risque minimal et des opportunités supplémentaires pour étendre l'utilisation facilement³⁹.
- **Assurer le caractère inclusif de la technologie**. Il faut évaluer comment la technologie donnée considère l'ensemble des profils des élèves et des professeurs. Cela inclut les élèves en situation de handicap. Une liste de handicaps doit être utilisée pour valider si la technologie proposée inclut tous les élèves souffrant de ces handicaps. Il est fondamental d'inclure cela dans les tests à valider en cours de déploiement et en production.

Expérimentations en classe

Ici sont présentées quelques recommandations sur l'expérimentation en classe dans le but de les faciliter, de construire de la confiance avec le corps enseignant, les élèves et les parents, mais aussi pour en tirer un maximum de bénéfices pour l'École et les chercheurs.

- **Co-conception et co-construction**
 - Favoriser les phases de co-construction des équipes de recherche et de développement avec les équipes du terrain lors des itérations et/ou lors du déploiement des solutions en classe ;
 - Inclure les parents dans cette phase de co-construction,

³⁶ Van L. Baker, Magnus Revang. Four Use Cases for Chatbots in the Enterprise Now. Etude Gartner, 2016

³⁷ Brian Manusama. Market Guide for Virtual Customer Assistants. Etude Gartner, 2016

³⁸ Van L. Baker, Magnus Revang. Four Use Cases for Chatbots in the Enterprise Now. Etude Gartner, 2016

³⁹ Brian Manusama, Frances Karamouzis, Tom Austin. Seven Decision Points for Success With Virtual Customer Assistants. Etude Gartner, 2016.

- **Accompagnement pédagogique** - Guider les équipes pédagogiques en amont et pendant l'utilisation des services numériques à base d'IA ;
- **Collecte de retours d'informations** - Partager via des modalités de retour d'expérience les points d'amélioration entre les utilisateurs (équipes pédagogiques) et les concepteurs (fournisseurs/distributeurs des solutions) ;
- **Communication des points de vigilance** - Détailler les conseils pratiques et points de vigilance (information versus consentement, chaîne d'acteurs à impliquer, transparence, accompagnement, communication)
- **Expérimentation à 360 degrés pour l'inclusion** - Une expérimentation complète et inclusive, qui doit inclure des élèves en situation de handicap par exemple, ou des enfants en grande difficulté scolaire. Une telle expérimentation permet d'accroître la confiance des professeurs, des élèves et des parents sur l'utilisation de telles technologies d'IA en assurant leur caractère inclusif.

Mises à l'échelle

Ici sont présentées quelques recommandations sur la mise à l'échelle d'outils utilisant de l'IA, dans les écoles, pour faciliter le déploiement et le post-déploiement quand les outils sont massivement utilisés en classe.

- **Actualisation des programmes** - Mettre en cohérence régulièrement les programmes (en lien avec les évolutions rapides de l'IA) un curriculum IA de l'école élémentaire à l'université (en lien avec le DigComp Edu européen et le cadre de référence des compétences numériques (CRCN) évalué via l'outil PIX)
- **Mise à disposition de formations** -
 - Proposer des plans de formation dans toutes les académies sur l'IA pour les personnels de directions, les pédagogues, etc.
 - Mettre en place des plateformes ouvertes de mise à disposition de contenus explicatifs conçues par des personnes en dehors de l'Education nationale (ingénieurs, entrepreneurs, ...), et labellisées "validées par le ministère". Ces contenus sont complémentaires aux formations institutionnelles.
- **Valorisation du rôle des enseignants dans ces actions de recherche.** Engager et former les enseignants à la recherche-action. Engager les chercheurs dans une véritable pratique de collaboration autour de leurs enjeux de recherche en éducation. Valoriser au sein de la recherche française et européenne les démarches collaboratives entre les différents acteurs (recherche, edTech, monde académique)
- **Valorisation des expérimentations en cours** - identifier et valoriser les expérimentations connues (quelle que soit l'échelle) pour une factorisation et une diffusion entre territoires (exemple démarche des TNE)
- **Initialisation de nouveaux partenariats d'innovation IA pour l'éducation** - Utiliser cette modalité de marché public innovant pour faciliter un passage à l'échelle rapide du laboratoire à la classe (un an de recherche et développement et de co-construction en classe suivi de 6 mois de pré-industrialisation suivis d'une possible

acquisition et déploiement au niveau national) - cible école, collège et lycée : assister les enseignants pour améliorer l'apprentissage de leurs élèves en français, mathématiques, sciences ou langues vivantes.

- **Exploration des possibilités offertes par l'IA dans l'exploitation des données des services de l'administration centrale, des rectorats ou des opérateurs** - Utiliser les défis d'innovation dans le cadre des investissements d'avenir pour élaborer des POC facilitant l'utilisation de l'IA pour améliorer l'organisation des formations, des parcours professionnels des enseignants ou scolaires des élèves.
- **Accompagnement des startups EdTechs dans l'IA** - Proposer aux EdTechs des programmes d'accompagnement répondant à un cadre de confiance pédagogique, technique et éthique, afin de soutenir une activité économique et de recherche sur les apports de l'IA en éducation et pousser des solutions innovantes répondant aux besoins de la communauté éducative.
- **Influence et collaboration nationale et européenne** - Soutenir et promouvoir la nécessité de travailler la thématique "IA et éducation" à échelle nationale ou européenne en utilisant davantage le dispositif ERASMUS+ et en articulant les actions du MENJS avec la mise en œuvre du nouveau plan numérique européen pour l'éducation (DEAP 21-27)

Conclusion

Dans ce rapport, nous avons présenté un état de l'art avec des références représentatives mais non exhaustives, des travaux et des avancées dans le domaine des agents conversationnels et plus particulièrement dans le domaine de l'éducation. Pour cela, nous avons réalisé des interviews avec une dizaine de chercheurs et d'ingénieurs spécialistes du domaine afin d'orienter nos recherches bibliographiques.

Nous avons également réalisé un sondage auprès des professeurs, des directeurs d'écoles, et des élèves sur leur perception de l'intelligence artificielle. Ces auditions ont permis de souligner les défis d'adoption à venir et le travail pédagogique à réaliser tant du côté du personnel éducatif que des élèves, mais aussi des parents d'élèves.

Nous avons proposé des pratiques de mise en place, de développement et d'usage d'assistants intelligents pour l'éducation. Pour cela, nous nous sommes intéressés aux aspects techniques et pédagogiques pour le déploiement d'outils éthiques, responsables et inclusifs, afin d'augmenter le niveau d'adoption et permettre aux outils d'évoluer dans la bonne direction.