

# Conclusion générale

Nous sommes sans doute actuellement entrain de vivre une époque importante de l'intelligence artificielle. Le temps ou nous rêvions encore de converser avec une machine semble à la fois proche en terme de temps, mais loin en terme de progrès. Les utilisateurs réguliers de produits technologiques sont confrontés à une technologie nouvelle mais prometteuse. Ce secteur d'activités peut donc s'avérer très prolifique si les efforts fournis sont suffisamment conséquents.

La création de ce nouveau besoin a été la principale source de motivation pour ce projet. Cependant, et par faute de temps ainsi que de moyens techniques (surtout en ce qui concerne le coté matériel pour l'apprentissage automatique), certains modules n'ont pas été exploités à leur maximum. Cela ne nous a pas empêché de réaliser un travail dont nous sommes particulièrement fières. Mais nous gardons toutefois un esprit critique, ainsi qu'une objectivité envers le travail fourni.

Tout au long de la réalisation de ce mémoire, nous avons étudié l'état actuel des assistants personnels intelligents. Nous avons dû passer une majeure partie de cette étape à comprendre les fondements théoriques et conceptuelles de chaque composants de ces systèmes. Principalement a cause de la grande densité de techniques, concepts et théories qui sont nouvelles pour nous.

Après avoir assimilé la totalité des concepts, et qui font office d'état de l'art du domaine. Nous sommes arrivé à certaines conclusions. Tout d'abord, développer un système en partant de rien était un travail assez massif. Dépassant de loin le cadre d'un projet de fin d'études de Master. Nous avons donc mit l'accent sur certains modules. Ces modules, à savoir le module de compréhension du langage naturel et le module de gestion du dialogue, dépendaient énormément de notre problématique. Le module de reconnaissance automatique de parole a été sujet à une amélioration spécifique à nos besoins tout en exploitant un système de base déjà existant (à savoir DeepSpeech).

Avec une idée claire du travail à réaliser, nous avons pu entamer la conception de chaque module en y incorporant nos ajouts et modifications. Beaucoup de ces modifications sont le fruit de longues séances de débat et de discussions.

L'étape de réalisation est celle qui a pris le plus de temps. Entre la découverte de nouvelles technologies à utiliser, la collecte des données et l'apprentissage des différents modèles.

La partie de construction du corpus pour le modèle de langue du module de reconnaissance automatique de la parole a été soumise à beaucoup d'optimisations incrémentales, en tombant à chaque fois sur un nouveau problème, ou bien un obstacle matériel (manque de puissance de calcul). L'ajout de ce modèle de langue a amélioré les résultats. Cela s'accordait avec nos prédictions théoriques. Cependant, les résultats n'étaient pas assez encourageants. Surtout si le but est de détrôner les systèmes propriétaires comme celui de Google qui réalise un score quasi-parfait sans apprentissage supplémentaire. L'avenir de systèmes Open source pour la reconnaissance automatique de la parole reste cependant prometteur. Ces derniers offrent un moyen libre de mener des études et contribuer au développement à grande échelle de cette discipline. Une perspective future pour ce module serait de lancer notre propre plateforme de collecte de données. Cette idée a déjà été discutée dans la partie de l'étude de l'état de l'art. Malheureusement le temps a cruellement manqué. L'investissement de la communauté dans de telles initiatives n'en reste pas moins indéniable. Comme l'a démontré le projet CommonVoice de Mozilla.

La construction de l'ensemble d'apprentissage à partir de zéro était l'étape la plus longue de la réalisation du module de compréhension du langage naturel. Plus l'ensemble grandissait, plus il était difficile de maintenir sa validité sans l'intervention d'un soutien externe. Les ajouts faits au modèle d'apprentissage pour ce module ont été expérimentalement validés dans le chapitre Réalisations et résultats. L'ajout de l'information morphosyntaxique a permis de donner plus de valeur sémantique à chaque mots de la requête. L'introduction d'erreurs aléatoires a quad à elle permit la gestion d'éventuelles erreurs que pourrait engendré le module de reconnaissance automatique de la parole. Cependant, vu que la tâche à accomplir était relativement simple et limitée, le réel impacte de cet ajout ne peut pas être certifié et validé dans un cadre plus général. Une autre problématique est celle du manque de données d'apprentissage consacrées au domaine de la manipulation d'ordinateurs. Ces données sont généralement construites manuellement par les développeurs du système. Faire appel à une collecte massive de données est une des perspectives envisagées. Le développement d'un outil d'aide à l'annotation d'un corpus était elle aussi le sujet d'un long débat. La mise à l'échelle de cette plateforme pourrait grandement faire avancer la tâche fastidieuse qu'est la collecte de données. Un autre point à soulever est celui du manque de diversité dans les tâches réalisable par l'assistant. Cet ensemble de tâche reste facilement extensible. Il suffit d'ajouter des exemples assez exhaustifs à l'ensemble de données. Cela reste néanmoins une tâche lourde et manuelle à plus grande échelle. L'automatisation de cette étape n'a pas été discutée pendant la réalisation de ce travail.

En ce qui concerne le gestionnaire de dialogue, nous l'avons conçu pour qu'il soit facilement ajustable et mis à l'échelle. L'utilisation d'une architecture hiérarchisée d'agents de dialogue permet de grandement réduire la complexité de développement et d'ajout d'un nouveau gestionnaire de tâches dans le système. L'utilisation d'un système de ges-

tion d'état du dialogue plus riche que ce qui est proposé en temps normal a permis à l'agent apprenant de mieux comprendre les échanges effectués avec l'utilisateur. L'utilisation d'un graphe de connaissances au lieu de trames sémantiques est justifié par le fait que ces derniers offrent une plus grande flexibilité dans la représentation de l'état du dialogue. Pour ce qui est des agents de dialogue, ils sont entraînés en utilisant des techniques d'apprentissage par renforcement. Ils interagissent avec le simulateur d'utilisateur, qui est un programme permettant de simuler le comportement d'un utilisateur avec la machine. Ce simulateur est capable d'estimer à quel point l'agent apprenant est efficace dans la tâche qui lui est confiée. Le but de l'agent est d'estimer la satisfaction de l'utilisateur en utilisant un réseau de neurones comme fonction d'approximation. Son but final est d'atteindre le but de l'utilisateur à travers l'optimisation d'une politique d'actions basée sur un système de récompenses. Cependant, l'utilisation du graphe de connaissances et de l'apprentissage profond a un prix. Pour essayer de connecter ces deux composants, nous avons dû condenser le graphe tout entier en un vecteur de taille fixe. Il peut être ensuite injecté au réseau de neurones pour effectuer la prédiction de l'action à entreprendre. Nous avons exploré deux approches pour cela en les comparant expérimentalement. En effet, la codification de la totalité du graphe en un seul vecteur était une tâche assez difficile. La taille de ce dernier augmente exponentiellement avec l'ajout de nouvelles connaissances. Ce qui rend le processus d'encodage en un vecteur de taille fixe très compliqué. De plus, les nœuds du graphe ont été codifiés avec des entiers distincts comme identificateurs. Ce qui a pour effet de faire perdre de l'information sémantique au graphe encodé. Le décodage du graphe s'en trouve grandement affecté dans la phase d'apprentissage de l'estimateur d'actions. Comme perspectives futures, il est envisageable d'utiliser une méthode d'apprentissage semi-supervisée pour l'encodage des nœuds du graphes. Cette ajout pourrait permettre d'enrichir la valeur sémantique de ces nœuds, et facilitera la tâche au réseau de neurones de l'agent apprenant pour le décodage du graphe. Principalement grâce au fait que les nœuds dont les sens sont arbitrairement proches auront des codifications similaires.

Pour le dernier module, à savoir le module de génération du langage naturel. Il fût assez simple à réaliser. Son fonctionnement est très rudimentaires et intuitif. Il s'agissait d'utiliser un ensemble de phrases modèles et de remplacer les valeurs manquantes avec des valeurs réelles. Une méthode plus sophistiquée comme l'utilisation de modèles d'apprentissage automatique basés encodeur-décodeur, ou bien basés convertisseur de graphes de connaissances en texte pourraient être utilisées. Cependant, ces architectures requièrent un très grand volume de données d'apprentissage annotés et spécifique à notre problématique.

Pour ce qui est de l'application, nous avons fait le choix d'utiliser une architecture trois tiers basée web. Ce choix fût motivé par le fait que l'utilisation d'un serveur offre une puissance de calcul considérablement plus élevée que celle d'une machine personnel en local. L'interface reste assez simple et épurée. Le but est de minimiser l'interaction avec l'utilisateur avec tout autre moyen de communication que la parole. Cependant cette interface à aussi pour but de montrer les fonctionnalités du système. Elle est donc plus orientée vers le développement. Une amélioration possible serait le déploiement du serveur dans un service de Cloud Hosting. Cela permettrait de minimiser les temps d'inférence et de

post-traitement. Le développement d'une interface plus légère et plus orientée vers les cas pratiques est une amélioration possible. Garder les deux cas de figures (utilisation et développement) est aussi possible.

Pour conclure, nous estimons que la totalité du projet était une énorme occasion d'approfondir nos connaissances. Que ce soit celles qui nous ont été enseignées durant notre cursus, comme l'apprentissage automatique, le traitement automatique du langage naturel, le web-sémantique et la représentation de connaissances. Ou bien celles que nous avons apprises au cours de notre étude de la littérature, comme l'apprentissage par renforcement, le traitement automatique de la parole, le nettoyage des données, etc. Ce projet nous a aussi initié au travail en équipe pour la réalisation d'un projet assez conséquent. Tout en étant encadrés par nos supérieurs. Finalement, nous pensons que la plus grande satisfaction vient du fait que nous avons réalisé, dans un certain délai restreint, un travail qui traite d'un sujet récent et ambitieux. Et ainsi, de poser la première pierre à l'édifice pour, idéalement, encourager les chercheurs en Algérie à s'investir dans ce domaine.