Session Lionel Lourdin - GEnIAI @ CERN

Contact Lionel Lourdin

27/062018

PRESENTATION

Introduction des notions d'intelligence artificielle et de machine learning

Dernièrement on a eu l'occasion avec la présentation de Google Duplex de voir comment l'utilisation du Machine Learning (ML) sur des corpus d'information cumulés depuis des années permettait de proposer une solution de type chatbot vocal comprenant des tournures de langage très sophistiquées. Nous sommes donc dans une époque où il est possible de proposer un réel interfaçage conversationnel, vocal ou écrit, avec un humain pour proposer un service de qualité.

L'utilisation du ML permet de ne pas avoir à programmer tous les cas de figure possibles du langage humain (impossible !), mais d'utiliser l'expérience humaine cumulée afin que les algorithmes et le ML puissent faire tout seuls cet apprentissage.

Dans la démonstration plutôt bluffante de Google Duplex (cf. https://www.youtube.com/watch?v=D5VN56jQMWM) on se rend compte qu'il est possible aujourd'hui pour un bot d'arriver à fixer un rendez-vous avec un humain sans que celui-ci ne s'aperçoive qu'il parle pas à une machine.
Cela n'est possible que grâce à l'approche "big data" de Google de ces dernières années qui a permi d'avoir les énormes volumes de données indispensables pour "nourrir le ML" et arriver à cet exploit technique que l'on appelle intelligence artificielle (IA).

Nécessité d'une intelligence artificielle compatible avec la notion de "service public" du projet Genial (projet d'utilité publique)

Alors comment peut-on imaginer une IA, et en l'occurrence dans le cadre du projet Genial un chatbot (qu'il soit écrit ou vocal, nous utiliserons le terme chatbot), qui puisse proposer un service s'inscrivant dans une démarche d'utilité publique, avec une qualité acceptable ?
Quel que soit le service spécialisé que l'on peut imaginer pour un chatbot de "service public" (culturel, juridique, intégration professionnels…), ils auront tous en commun le fait de devoir maîtriser le "langage humain" (ce qu'on appelle généralement la fonctionnalité de langage naturel, le terme technique anglais étant Natural Language Processing) et d'intégrer des informations de type "culturelles".

Au delà de l'indispensable intelligence permettant à la machine de comprendre la langage humain et de s'adresser de manière compréhensible à ses utilisateurs, il est indispensable qu'elle dispose d'un autre type d'intelligence que l'on pourrait qualifier davantage d'intelligence sociale et culturelle.

L'importance de ce point s'illustre très bien avec l'exemple de Microsoft et de son expérimentation de bot sur Twitter qui en l'espace de quelques heures en était arrivé à avoir un comportement inacceptable "en société", avec des propos clairement racistes/facistes (cf. Racisme, conspi et insultes : quand le bot millenial de Microsoft dérape sur Twitter).
C'est exemple parmi d'autres a montré comment le ML pouvait dangereusement dérivé et l'absolu nécessité de prendre les précautions dans la façon de penser l'architecture des IA.
C'est à ce moment qu'il est intéressant dans le cadre du projet Genial d'introduire les notions de patrimoine commun et de responsabilité numérique deviennent.
En effet, si l'exemple cité plus haut a été quelque peu dommageable pour l'image de Microsoft, on peut imaginer que les conséquences seraient tout autre si demain une IA mise en place par une institution publique se mettait à avoir ce type de comportement. Dans le cadre de projets d'utilité publique, ayant vocation d'aider et conseiller les citoyens, on se doit de considérer une certaine responsabilité culturelle, sémantique, rhétorique. L'IA sur laquelle repose le chatbot doit permettre de garantir une conversation dans le respect des lois publiques et de la bonne séance avec son interlocuteur.

Imaginer un corpus commun open source permettant d'entraîner une intelligence artificielle "service public compatible"

On en vient donc à considérer pour le projet Genial la mise en place d'un corpus d'information pensé de manière à pouvoir entraîner l'IA en la mettant à l'abri de tout risque de déviance, corpus que nous appellerons corpus commun.
L'idée est alors de penser l'architecture Genial basée sur des technologies open source et un corpus commun de manière à créer un socle ayant déjà la capacité de communiquer avec l'humain.
Le corpus commun est par ailleurs l'occasion, via un développement collaboratif, de travailler sur la façon de s'exprimer que l'on souhaite donner à cette IA afin de pouvoir considérer qu'elle "a le droit de parler au public". Les réflexions pouvant aussi porter sur des questions comme la différence de façon de parler lorsqu'elle s'adresse à des seniors ou à des enfants.
De par les enjeux de responsabilité et l'implication qu'il peut avoir, il nous apparaît indispensable que ce corpus commun soit ouvert et open source.
On peut alors considérer que l'on dispose d'un terrain de concertation, d'une sorte de consortium entre les acteurs publics, mais aussi les acteurs privés qui ont ce même besoin.

À titre de référence, Google, pour obtenir ce type d'intelligence commune nécessaire à ses projets d'IA, a développé le projet open source TensorFlow (cf. TensorFlow).
TensorFlow est un "système d'IA" (que l'on peut considérer pour simplifier comme un outil et une méthode de ML) qui, ayant été mis à disposition en open source, permet d'avoir un commun entre différents acteurs au niveau de leurs besoins similaires en matière de fonctionnement et de capacités.

Il est intéressant de noter que ainsi TensorFlow est autant utilisé, par exemple, par l'armée américaine pour contrôler ses drones que par des chatbots d'Amazon. Aussi, les bases du fameux Google Duplex évoqué plus haut reposent sur des technologies TensorFlow.
On se rend compte ainsi qu'il existe déjà sur le "marché de l'IA" des socles qui sont là, et qui sont déjà basés sur l'open source.
En dehors de TensorFlow d'autres briques open source sont aussi déjà disponibles, et pour Genial, il est intéressant de faire un état des lieux de l'existant et de déterminer ce qui pourrait être utilisé (on ne veut pas réinventer la roue !).

À noter qu'il est impératif que la qualité de la conversation que propose un service comme Genial soit irréprochable, car sinon il n'y aura pas d'adoption de la part des utilisateurs qui par ailleurs ont l'expérience de la qualité des IA de Google ou Amazon. Il faut garder à l'esprit la notion de concurrence, car même si le projet s'inscrit dans le "service public" et propose un service "vraiment gratuit" (c'est à dire un service qui ne se paie pas en donnant des données qui seront exploitées par la suite; cf. "si c'est gratuit c'est que c'est vous le produit"), le principe de concurrence reste présent.

Penser une architecture qui permet d'ajouter les "corpus spécialisés"

Une fois que l'on a une IA capable de converser correctement en utilisant le langage humain, on peut s'occuper du contenu spécialisé, spécifique aux sujets qui devront être couverts par le service (culturel, juridique, intégration professionnelle… pour reprendre nos exemples précédent). À tous ces sujets correspondent des corpus spécialisés de savoir dépendant de multiples acteurs (pour la culture ça sera les acteurs de la culture, pour le juridique ça sera les avocats et la sphère des juristes etc.).
Cette architecture permet de différentier les corpus informationnels communs qui vont permettre d'entraîner le langage humain de ceux spécifiques qui vont permettre de donner une spécialisation à ce langage humain.
Genial avec cette approche et avec son "intention d'utilité publique", peut proposer ce socle commun qui est essentiel à tous ceux, publics ou privés, qui aimerait pouvoir développer des interfaces de dialogues humains avec une garantie sur la qualité du "ton" de la conversation et sur l'impossibilité pour cette IA de tomber dans des dérives qui ne serait pas acceptables dans le cadre d'un service public.

Pour comprendre ce principe on peut prendre aussi l'exemple d'études d'avocats à Londres qui arrivent aujourd'hui à proposer pour plus de la moitié des cas qu'ils traitent des conseils issus directement de leur solution d'intelligence artificielle. Cette solution est basée sur un corpus commun et a été nourrie de l'ensemble des informations de l'historique de leur étude qui correspond à leur corpus spécialisé.

Pour Genial il y a donc d'un côté cette nécessité de développer ensemble ce corpus commun afin d'avoir ce langage humain "public compatible", mais aussi d'autre part de fournir des outils qui utilisent ces corpus pour permettre à des acteurs publics ou

privés de spécialiser le comportement de leur chatbot. On peut considérer que c'est déjà ce que fait TensorFlow aux USA, mais sans la dimension "public compatible".

Genial pensé comme écosystème technologique plutôt que comme service centralisé, et le modèle économique qui en découle

Deux options sont possibles pour Genial :

1. Un service centralisé, avec un portail qui va délivrer les services à l'ensemble des acteurs du marché (c'est l'approche des GAFAM).
2. Un écosystème technologique qui permet à chacun des acteurs du marché de délivrer une solution de chatbot pour améliorer son service aux utilisateurs.

Le projet Genial s'inscrit clairement dans l'option #2, et il est donc indispensable de proposer ce fameux corpus commun.
On se retrouve dans une dynamique de marché intéressante avec, d'une part le développement d'un corpus commun et d'un socle technologique et informationnel, et d'autre part cette possibilité pour chaque acteur du territoire de développer des services spécialisés.
Ces services peuvent être privés et payants ou publics et gratuits, ce qui offre des pistes de réflexion quant au modèle économique.
Ce point n'est pas anodin, car il faut savoir que afin de pouvoir proposer une solution qui puisse "rivaliser" avec celle des géants déjà en place, il va falloir des moyens. Au-delà du besoin de "moyens" (techniques, financiers…), il faut surtout absolument être en mesure de pouvoir rassembler l'information, les données, et pour cela il est indispensable de réussir à fédérer les acteurs du marché tout en leur donnant la liberté d'appliquer leur mission.
Les modèles de l'open source et du développement contributif sont particulièrement inspirants pour un tel projet : ils offrent la possibilité de différencier la partie commune (qui peut faire l'objet notamment de travaux universitaires ou de développement public), de la partie spécifique qui peut être source de création d'emplois et de valeur.

Contrôle, gouvernance et certification du projet : les possibilités offertes par l'open source et la blockchain

Si on veut que les IA respectent une responsabilité informationnelle, publique, nous avons l'obligation, de par le fonctionnement technique de l'IA, de faire en sorte que le corpus commun qui permet le langage et la conversation avec l'individu, soit public et soit régi par une "autorité" afin que les contenus ne puissent pas conduire à des comportements "déviants". Les IA, selon leur conception, vont avoir tendance à prendre la couleur de la mémoire collective qui a été cumulée par l'expérience. L'architecture de ce type de système est primordiale afin de ne pas se retrouver avec une boîte de Pandore et que rapidement on ne puisse plus maîtriser son comportement.

Ici, la blockchain avec sa mécanique permettant de sécuriser l'information est intéressante à envisager.
Imaginons des IA avec un logiciel open source, et que quiconque puisse avoir la possibilité d'utiliser ce logiciel, de charger un corpus commun et de commencer à parler comme un humain.
Comment pouvoir garantir que le corpus qui a été utilisé pour éduquer l'IA n'a pas été corrompu, modifié ? Il faut avoir en tête que l'on parle de corpus qui sont constitués de gigas de données, et donc impossible à vérifier "à la main" ! Et il faut aussi bien garder en tête qu'il peut suffire de quelques informations "toxiques" pour "contaminer" l'ensemble de l'IA (par exemple : glisser quelques paragraphes d'un livre comme 'Mein Kampf' dans le corpus pourrait être suffisant pour que l'IA devienne fasciste !).
C'est donc là que la blockchain et l'open source peuvent permettre de donner une garantie d'utilité publique pour ce socle.

Les possibilités offertes par les modèles de licences existants en Europe et la CERN Open Hardware Licence

Nous avons la chance d'avoir déjà en Europe des modèles de licences qui correspondent très bien aux besoins d'un tel projet.
Nous sommes d'ailleurs aujourd'hui au CERN, on peut donc citer la CERN Open Hardware Licence (OHL) qui permettrait même d'imaginer que ces chatbots puissent tourner sur une électronique qui est elle-même vérifiée et open source (voir plus bas dans les questions les précisions concernant la pertinence de la licence OHL et l'importance d'un hardware open source).

Comment le projet Genial peut être l'occasion de (re)mettre Genève au centre des débats sur Les Droits de l'Homme à l'heure de l'intelligence

Aussi, comme nous sommes à Genève, il faut aussi voir qu'il y a une carte à jouer sur la responsabilité numérique.
On parle en effet d'IA et de chatbots qui doivent parler avec l'humain : ne devrait-on pas faire en sorte que cette IA respecte "Les Droits de l'Homme ?".
Avec la position de Genève à l'international (siège du Haut-Commissariat des Nations Unies aux Droits de l'Homme), un tel projet pourrait être au coeur d'une réflexion sur la nature, la qualité, la validation de ces corpus qui permettent à des intelligences de communiquer avec l'humain en regard des Droits de l'Homme...

QUESTIONS

Comment la qualité du corpus influence l'intelligence artificielle

Jusqu'à quel point est-on capable de prédire/garantir le lien entre la qualité d'un corpus et l'effet que cela va avoir sur l'IA qui fera son apprentissage à partir de ce corpus ?

Cela va dépendre de l'architecture technique de l'IA.
Il faut imaginer qu'il y a plusieurs corpus informationnels avec différents niveaux de prédominance.
Prenons comme exemple un projet de chatbot juridique destiné à conseiller des entrepreneurs, avec une IA qui serait basée sur 3 corpus :

- Un corpus issu de l'expérience commune cumulée qui va faire que plus l'IA va avoir de personnes qui auront échangé avec elle, plus elle va améliorer son discours, plus elle aura de finesse dans son langage.
- Un corpus informationnel stricte (le "savoir" spécialisé) qui sera celui des textes de loi, du code des obligations suisses, des jurisprudences...
- Un corpus "existentiel", culturel, de bienséance, qui correspond à la manière dont l'IA va gérer la politesse, le comportement général avec l'individu.

L'architecture sera construite de manière à mettre en prédominance le corpus culturel/politesse, ensuite celui du savoir et seulement après le corpus d'expérience. Par les algorithmes il est possible de faire en sorte que jamais le corpus d'expérience puisse "prendre le dessus". C'est la programmation de ces algorithmes qui va définir ce type de responsabilités.

Les enjeux autour des licences et du hardware (cas de la licence CERN OHL); rôle de la blockchain; rôle que peut jouer Genève en matière de "responsabilité informationnelle"

Comment envisager le rôle de la licence CERN Open Hardware Licence (OHL) dans le cadre d'un tel projet ? (Question posée par une participante qui travaille au CERN avec le groupe en charge du transfert de connaissance et de technologie du CERN)

Il est important d'avoir à l'esprit que le "hack" de demain va consister à corrompre les corpus informationnels pour contrôler les réponses proposées par les IA et pouvoir ainsi contrôler le comportement des utilisateurs !
Les interfaces conversationnelles sont accessibles via des éléments hardware : smartphones, ordinateurs, assistants vocaux de type Alexa ou Google Home pour citer les plus courants, mais aussi pourquoi pas des poupées connectées ou tout autre équipement connecté et interfacé avec une IA.
Il est intéressant d'envisager que, avec la notion de responsabilité publique, ce type

de hardware puisse suivre la logique d'ouverture open source du software (logiciel) afin de garantir une intégrité technique et logique sur le fonctionnement de ces IA. C'est la raison pour laquelle la licence OHL était mentionnée.
Pour se prémunir des potentiels "hack de société" auxquels nous expose cette nouvelle ère de l'IA, il faut une sécurité optimale aussi bien au niveau du software que du hardware. Et, si hardware et software ne sont pas ouverts (open source), on ne peut pas bénéficier de l'intelligence collective qui permet de trouver où sont les risques, les failles, et de les corriger rapidement grâce à la puissance de la "communauté". L'histoire nous a montré que l'open source est le meilleur moyen pour assurer la sécurité des systèmes, mais aussi permettre une dynamique d'amélioration rapide, agile et constante.
Ce point est aussi important à avoir à l'esprit plus globalement quand on réfléchit à la smart city car l'IoT (cf. Internet des Objets) est de plus en plus présent, et cette approche du hardware avec ce type de licence est primordial pour envisager le développement et la stabilité des infrastructures concernées.
On a aussi la chance au niveau logiciel d'avoir l'European Union Public Licence (cf. Licence publique de l'Union européenne), qui existe depuis 2007 (11 an de jurisprudence maintenant !), qui permet de garantir l'ouverture du code source des algorithmes de l'IA : on a donc un cadre légal existant qui nous permet d'aller dans cette voie.
Enfin, la licence Creative Commons CC BY SA 4.0 internationale (cf. CC BY SA 4.0), qui a une compatibilité mondiale, nous permettrait de gérer une partie de ces corpus.
Maintenant se pose la question du fait qu'un corpus d'utilité publique pour former une IA ne devrait plus être modifiable.
Il nous manque des licences pour gérer ce genre de choses.
Des licences permettant de définir que c'est "proche du domaine public" mais sans l'être; de pouvoir dire : ce contenu est stricte, valable, quiconque peut l'utiliser, le produire et le distribuer, il fait partie d'un bien commun ...mais personne n'a le droit de le modifier. Ou du moins, si il est modifié, on en perd une certification, on en perd "quelque chose".
C'est là où la blockchain peut avoir un rôle à jouer.
Alors il est possible de parler d'un patrimoine informationnel stable, vérifiable.
Il est possible aujourd'hui d'arriver à certifier ces éléments là.
Imaginons nous dans quelques années, avec la démocratisation de ce type de services : comment un utilisateur pourra savoir si l'agent conversationnel à qui il compte s'adresser est bienveillant ou pas ?
Seuls ces systèmes de certification évolués permettront de garantir cela, et il y a probablement encore de nouveaux systèmes technico-légaux à inventer sur ces bases pour sécuriser ces points.
Un organisme d'institution privée mais qui oeuvre avec une vocation d'utilité publique, pour pouvoir faire par exemple un chatbot qui va parler avec des jeunes en réinsertion, devra avoir des contraintes, devra respecter certaines règles de langages, et ceci pourra alors se placer sur des corpus communs qui auront été validés.

Genève a un rôle intéressant à jouer en regard de sa position au niveau des droits de l'homme, mais aussi de la Déclaration de principes qui avait été faite suite au Sommet Mondial sur la Société de l'Information entre 2003 et 2005 (cf. Déclaration de principes - Construire la société de l'information: un défi mondial pour le nouveau millénaire).
Si on regarde aujourd'hui cette déclaration, on se rend compte que la notion d'information inclusive, contributive, permettant au citoyen de s'émanciper et de se positionner dans la société prend tout son sens à l'heure de l'IA.

Il faudrait ramener cette déclaration au cœur du débat et la relire avec un nouveau regard : il n'y avait pas à l'époque les effets des GAFAM et le constat qu'on a aujourd'hui sur leur "emprise", mais il y avait déjà cette conscience là, indirecte, cachée derrière la notion de responsabilité informationnelle et de société de l'information inclusive.
Un projet comme Genial devrait pouvoir, en regard de cette déclaration, répondre à chacun des principes en les validant un par un pour s'assurer que rien n'est oublié et que l'on s'inscrit vraiment dans la meilleure des responsabilité informationnelle (ou "la meilleure des responsabilité informationnelle *possible*" ...petit clin d'œil à Voltaire, qui a lui aussi marqué la région genevoise, et à sa critique de Leibniz ;)).

Les points de différenciation de Genial vs. solution des GAFAM : comment l'approche locale/spécialisée peut humaniser les services et être gage d'une expérience utilisateur optimale

Sur un territoire local, en l'occurrence le territoire de Genève où s'inscrit le projet Genial, nous ne sommes pas dans le même contexte que celui d'un service global proposé par un Google ou un Amazon : la relation d'échelle n'est pas la même.
Où placer l'IA et ce qui va caractériser le chatbot : à la fin de la "chaîne", juste au dernier point avant l'intervention des acteurs publics, ou plutôt en amont, avec les personnes pour pousser les contenus qualifiés vers les acteurs publics en permettant une approche plus "humanisée" ?
Le fait de pouvoir penser le service localement offre des possibilités de spécification au niveau de l'identité du chatbot et du ton conversationnel particulièrement adapté au public visé, possibilités qui ne sont pas imaginables pour les services globaux cités auparavant. On pourrait alors vraiment imaginer que le résultat des échanges soit beaucoup plus facilement pensé et construit comme "de humain à humain", avec une expérience utilisateur optimisée et une qualité de service idéale pour les acteurs territoriaux.

Cet argument va tout à fait dans le sens du choix pour Genial d'être pensé comme un service décentralisé.
Au delà des arguments de sécurité évoqués auparavant liés à l'option centralisée, permet de proposer d'un côté le fameux corpus commun vérifié et garanti, et d'un

autre côté un ensemble de briques logicielles open source et diverses autres ressources qui vont permettre aux différents acteurs voulant proposer des services de chatbot de localiser de manière optimale leurs projets en prenant en compte leurs diverses spécificités territoriales et/ou thématiques.
Le fait que le socle primaire sur lequel reposera leur IA soit ce fameux corpus commun permet alors d'offrir sécurité et garantie au niveau de la déontologie, sans pour autant entraver leur créativité et la possibilité d'adapter très précisément leur service à leur contexte.
Un autre point intéressant à noter aussi dans ces réflexions concernant cette différence d'échelle entre les géants et un projet comme Genial concerne le machine learning (ML). Les géants disposent d'un volume de données considérable pour entraîner leurs IA et améliorer continuellement la qualité de leurs services.
Il est indéniable que Genial ne pourra pas rivaliser sur ce point et qu'il n'est pas judicieux d'orienter le projet dans ce sens.
Par contre, le ML évolue beaucoup et très vite en ce moment, et on s'aperçoit que le volume critique de données nécessaire à élaborer une IA de qualité se réduit énormément. On se rend compte qu'avec des "matrices informationnelles", ce que l'on pensait par exemple devoir nécessiter 1 million d'échanges/conversations afin de permettre à l'IA de pouvoir converser convenablement, peut désormais se faire avec moins de mille échanges "seulement". Cela va dépendre de la façon dont on couple les données strictes et les bases de langage. Il y a désormais des manières de programmer qui permettent de mettre en place des algorithmes évolutifs qui s'adaptent en fonction du contexte, on parle aussi d'algorithmes neuromorphiques (cf. Neuroinformatique et modélisation sur Wikipédia et L'ingénierie neuromorphique pour les nuls).
On peut donc imaginer que le projet Genial, en bénéficiant de ces avancées, pourra se contenter d'un volume limité de données suffisant pour permettre une bonne gestion du langage au socle de son IA, et se concentrer dans la mise en place de solutions pointues et performantes pour permettre de localiser les services avec des contenus et une éditorialisation qui permettra la meilleure contextualisation possible.