

Intelligence Artificielle et Grand Public

Pourquoi le grand public exige-t-il une intelligence artificielle compétitive avec l'humain, et comment une technologie imparfaite par nature peut-elle satisfaire un tel challenge.

Vincent Vanhoucke, Ph.D. (99)

June 2, 2007

Jean-Louis Gassée, ancien chroniqueur de la Silicon Valley pour Libération et aujourd'hui investisseur de capital risque, m'a un jour demandé: "Achèteriez-vous une voiture qui ne démarrerait que huit fois sur dix?" Il exprimait ainsi son scepticisme quant aux perspectives de l'intelligence artificielle dans un contexte grand public. Il n'a pas tort: confrontés à des systèmes complexes et loin d'être infaillibles, le rapport qu'ont la plupart des utilisateurs avec les outils dits "intelligents" tend il est vrai à aller de mitigé à franchement catastrophique.

Qui ne se souvient du trombone de Microsoft Office, affectueusement appelé "Clippy", et sa capacité inouïe à vous interrompre au moment le plus inopportun au beau milieu de l'édition d'un texte? Pour ceux qui crient encore vengeance, j'ai eu le plaisir d'observer un des inventeurs de cette vermine exaspérante, aujourd'hui professeur à l'Université de Stanford, tenter d'exterminer à grands coups de souris rageurs le parasite qui venait d'apparaître sur sa feuille de calcul Excel...il y a tout de même une justice.

1 Le cas des Centres d'Appel Automatisés

Avez-vous jamais essayé d'obtenir, sans succès, l'assistance d'un opérateur humain en appelant un service téléphonique interactif à reconnaissance vocale? Après avoir travaillé dans cette industrie pour plusieurs années, je suis toujours stupéfait par le nombre de services qui dissimulent de manière parfois créative comment accéder à un opérateur. Il

existe maintenant des sites Internets d'utilisateurs destinés à exposer les séquences de touches secrètes qui vous permettent de vous extirper du labyrinthe des menus téléphoniques. Mon conseil: dites "choucroute" trois fois de suite: après trois essais infructueux, ces systèmes ont tendance à prendre pitié de leurs utilisateurs.

Pour le meilleur ou pour le pire, les fournisseurs de services vocaux tentent de se doter de nouvelles armes pour résoudre le problème des utilisateurs mécontents. Dernière innovation: le détecteur automatique de client exaspéré. La littérature académique sur le sujet est en pleine expansion: ces systèmes analysent votre hauteur de voix, sa prosodie et son rythme, de façon à déterminer si vous êtes juste légèrement ennuyé ou bien complètement hors de vous, auquel cas peut-être daigneront-ils vous transférer à un être humain, qui sera très certainement ravi de vous prêter assistance. Les systèmes plus anciens se contentent de lister les jurons les plus courants et traitent la conversation différemment quand l'un d'entre eux est détecté au cours de la conversation. La prochaine fois qu'un serveur vocal automatisé se montre particulièrement récalcitrant, vous savez quoi faire... mais n'oubliez pas que vous êtes enregistrés.

Pourquoi les entreprises sont-elles prêtes à faire subir ce type de traitement à leurs clients? En un mot: parce qu'en dépit de la rancoeur affichée de nombre d'utilisateurs, ces systèmes fonctionnent! C'est là le grand paradoxe de cette industrie, et il illustre la grande majorité des difficultés qu'ont les systèmes intelligents à s'implanter auprès du grand public: si vous comparez un opérateur humain à

un système de reconnaissance vocale automatisé en matière de saisie de données, quelle que soit la mesure de succès que vous choisirez, il y a de fortes chances que le système automatisé l'emporte. Bien entendu, pour ce qui est du coût d'opération, le bénéfice de l'automatisation est clair: pas de salaire, disponibilité 7 jours sur 7, 24 heures sur 24. Peut-être plus surprenant, la durée moyenne de l'interaction téléphonique est en général plus courte avec la machine, qui n'a pas à transcrire sur un clavier ce que l'utilisateur lui dicte. Encore plus surprenant, le taux de succès de la transaction est fréquemment plus élevé, et les utilisateurs qui utilisent ces systèmes de manière répétée sont également en général plus satisfaits du système automatisé. Considérez par exemple le service de pages jaunes de Google pour les Etats-Unis: 1(800)GOOG-411. En deux phrases: votre localité, suivie du nom du prestataire de service ou d'une catégorie générique ("garage auto", "restaurant"), vous êtes mis en ligne avec votre contact. La rapidité et la précision du système n'ont aucun mal à surpasser ma capacité à naviguer mon agenda électronique, dont je peux désormais me dispenser sans peine.

Notez bien que j'ai qualifié ma description de l'utilisateur satisfait des services automatisés. Plus précisément, j'ai fait référence aux "utilisateurs qui utilisent ces systèmes de manière répétée". La distinction n'est pas innocente: les utilisateurs qui interagissent avec le système pour la première fois sont en général unanimes quant à leur expérience: déplaisante, troublante, voire traumatisante. Certains facteurs spécifiques aux interfaces vocales expliquent cette tension initiale: l'absence de support visuel est l'un d'entre eux, et l'exploration d'un nouvel espace logique sans appui visuel est considérée déroutante par la majorité du public. D'autres facteurs sont complètement universels, et s'appliquent à l'industrie de l'intelligence artificielle dans son ensemble.

2 Mon IA et moi: je t'aime, moi non plus

Deux facteurs essentiels définissent la relation du grand public avec les systèmes intelligents, l'un attractif, l'autre répulsif. Le premier pousse les créateurs d'interfaces utilisateur vers plus d'humanité dans le dialogue, l'autre vers plus de systématisation.

Le premier facteur a été notamment disséqué par Cliff Nass dans son livre "The Media Equation: How People Treat Computers, Televisions, and New Media Like Real People and Places". Son hypothèse: les humains traitent les ordinateurs exactement comme leurs propres congénères. Sa méthode: transposer des expériences de psychologie interpersonnelle en substituant un ordinateur à un des participants à l'expérience. De manière quasi-universelle, les résultats connus de psychologie humaine se confirment lorsque l'interaction homme / homme est transposée en interaction homme / machine: les utilisateurs mentent à leurs machines (et en éprouvent de la culpabilité), ils en disent du mal dans leur dos (mais jamais en face à face), ils réagissent positivement quand l'ordinateur les complimente, etc, etc... Toutes ces expériences ont été conduites en 1996, avant l'avènement des interfaces graphiques, et démontrent une anthropomorphisation inconsciente de l'interface utilisateur, aussi primitive soit-elle. Plus récemment, les sites web d'e-commerce ont observé par exemple que le fait de donner un visage à leur site web, en insérant une photo sur une page, change dramatiquement la tonalité du site et force l'attention de l'utilisateur sur cette personne avant tout autre forme de contenu.

La conséquence pour tout outil dit "intelligent" est que l'utilisateur s'attend par défaut à interagir avec un outil essentiellement humain, avec des capacités et une intelligence émotionnelle qui vont de pair avec cette attente. Une performance sous-humaine sera regardée fondamentalement comme une rupture de convention sociale, voire une insulte. Pourquoi un système de reconnaissance vocale avec un taux de réussite de 80% n'est-il pas considéré comme un système au moins au 4/5^e humain? C'est parce-qu'en prétendant être humain dans son mode d'interaction,

le système se positionne en acteur social et crée implicitement un niveau d'attente équivalent.

Vous êtes peut-être circonspect lorsque j'affirme que vous, l'utilisateur, traitez votre ordinateur comme un être humain. C'est là le second facteur qui gouverne l'adoption de l'AI auprès du grand public: consciemment, l'être humain préfère dans la majorité des cas que l'ordinateur se comporte comme une machine, et rejette intellectuellement l'idée d'une interface artificielle qui prétendrait être humaine. Alors que toutes les études qui analysent nos comportements supportent l'idée qu'un ordinateur est traité comme un agent social, l'utilisateur, lorsqu'il a à faire un choix conscient, préférera fréquemment un ordinateur qui se comporte comme une machine. Les utilisateurs fréquents de services vocaux ont internalisé le fait qu'ils interagissent avec une machine, et ont compris implicitement son fonctionnement en termes de menus de dialogue au travers desquels ils sont guidés. Pour les utilisateurs novices, l'ambiguïté quant au "degré d'humanité" l'interface à laquelle ils sont confrontés constitue une barrière psychologique déroutante. "Êtes-vous un ordinateur?" est une des questions que l'on entend souvent sur les enregistrements de centres d'appel: l'utilisateur qu'on a trompé et réalise qu'il parle à une machine est un utilisateur en détresse. Les intelligences artificielles utilisées dans les jeux vidéos utilisent le plein potentiel de cette tension: en réagissant de manière imprévisible, déroutante, elles ajoutent une dimension de danger et de surprise aux situations rencontrées par les joueurs.

Mon équipe à Like.com a construit un moteur de recherche pour les visages, qui permet de trouver à partir d'une photo toutes les personnes qui ressemblent à une personne donnée. C'est une tâche que le cerveau humain a perfectionné au point d'avoir développé des circuits spécialisés pour l'accomplir: certaines personnes atteintes de "prosopagnosie" sont incapables d'identifier les visages, sans qu'aucune autre fonction mentale ne soit affectée. C'est également une tâche pour laquelle l'intelligence artificielle a un taux de succès très élevé, et de plus en plus de forces de police exploitent les bases de données faciales à des fins d'identification. Cela veut-il dire que la reconnaissance de visages est prête

pour une utilisation grand public? Les sites webs de rencontre sont très demandeurs de technologies qui permettraient de mesurer si deux personnes se ressemblent, ou même si une personne est plus attractive qu'une autre (autre tâche que l'AI n'a aucune difficulté à accomplir, d'après une étude récente d'un groupe de chercheurs australiens). Google vient également d'incorporer une technologie proche de détection de visages dans son moteur de recherche d'images, vous permettant ainsi de trouver des photos de Paris Hilton sans vous encombrer d'images de l'hôtel Hilton de l'avenue Suffren... ou l'inverse.

La difficulté n'est pas dans la performance des algorithmes, mais dans la perception de l'intelligence artificielle, et de sa place en tant que "machine". Considérez l'expérience suivante, que nous avons conduite sur la base de notre détecteur de similarité visuelle: afin de calibrer nos algorithmes, nous avons construit un système factice, qui offre des résultats "parfaits", créés manuellement à partir de photos très similaires: ces résultats scorent un 20 / 20 sur notre échelle de qualité. Nous avons alors demandé à des utilisateurs d'évaluer ce système factice, en leur suggérant que les résultats qui leur étaient présentés avaient été créés par un ordinateur. Leur verdict: 13 / 20, à peine passable! Lorsque vous interrogez les participants, les défauts qui sont cités sont révélateurs: sont souvent cités la qualité du rasage de la personne, la qualité de la peau, la coupe de cheveux: ces attributs, qui n'auraient que peu de poids lorsqu'un être humain compare deux visages (on peut oublier de se raser, revenir bronzé de vacances, et toujours être soi-même...), prennent une importance de premier plan lorsque l'utilisateur s'attend à ce qu'une machine ait été la source de ce jugement. Pour évaluer cet effet, nous avons ensuite remplacé les visages par des objets qui sont moins à même d'élucider un jugement affectif. En remplaçant les visages par des articles de mode, les résultats qualitatifs sont immédiatement passés au dessus de 16 / 20: il est acceptable pour une machine de comparer des chaussures, mais plus déroutant de comparer des visages. De là est né notre site de lèchivitrine visuel (<http://like.com>): une AI appliquée à la reconnaissance d'objets, utilisée dans un contexte où la vision par ordinateur est exploitée de manière compatible avec les attentes des utilisateurs.

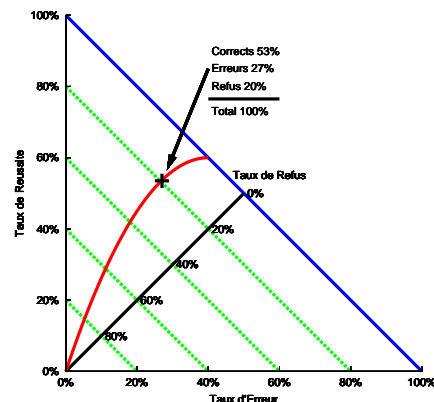
3 Apprendre à dire: "je ne sais pas"

Face à des contraintes opposées, systématisation ou humanisation, en fonction du niveau, conscient ou inconscient, auquel l'interaction avec l'utilisateur se produit, comment l'intelligence artificielle peut-elle espérer conquérir le grand public?

Une des façons les plus simples est de supprimer l'interface utilisateur. Il n'est pas étonnant qu'un des systèmes AI grand-public les plus populaires est un aspirateur autonome, le Roomba, sans autre interface qu'un bouton marche/arrêt. Le Roomba, de la société iRobot, s'oriente de manière autonome, navigue autour des obstacles, et retourne se charger automatiquement. Plus récemment, les appareils photos ont commencé à incorporer des techniques avancées comme la détection de visage pour améliorer l'autofocus et la réduction de l'effet "yeux-rouges". Ces systèmes sont totalement transparents et ainsi contournent la difficulté d'avoir à interagir avec utilisateur, mais le champ d'applications de telles technologies est néanmoins très limité.

Une approche bien plus générale, mais trop délaissée par les ingénieurs habitués aux systèmes déterministes, est de doter l'AI d'un certain niveau d'introspection: le système doit tout simplement apprendre à dire "je ne sais pas" de manière opportune. Dans la majorité des cas, c'est une tâche qu'une intelligence artificielle, si faillible soit-elle, est capable de conduire facilement, notamment grâce à une approche bayésienne. Le bénéfice en terme d'ergonomie est immense: imaginez si "Clippy" n'avait jamais tort, s'il apparaissait plus rarement sur votre feuille de calcul, mais jamais de manière intrusive. Imaginez si l'interface vocale vous disait: "désolé, je ne comprends vraiment pas, je vous passe à un opérateur..." L'équation change dramatiquement: le taux de succès du système est affecté de manière marginale, alors même que le taux d'échec est réduit à zéro.

Je vais tenter d'illustrer ce dernier point de manière graphique. La performance de tout système d'intelligence artificielle peut généralement se résumer à ceci:



En abscisses, le taux d'erreur du système, en ordonnées, son taux de réussite. En troisième coordonnée, le "taux de refus", qui représente la fréquence à laquelle le système répond "je ne sais pas". La performance d'un système peut être représentée par une courbe sur ce graphique, qui représente les divers niveaux auxquels le système est capable d'opérer.

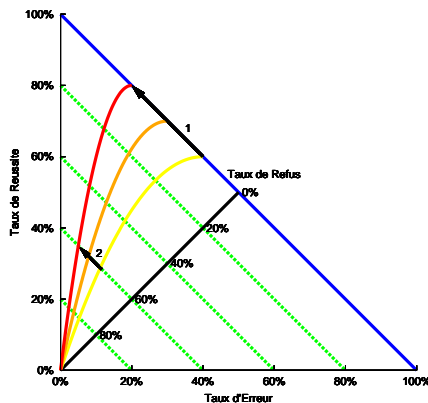
Dans un contexte grand public, c'est en général le taux d'erreur qui doit être strictement contrôlé. Le processus de découverte et d'apprentissage fait partie de toute conversation humaine, et l'utilisateur transpose facilement ce processus à l'interface homme / machine. Les frustrations émergent, comme dans le cas de "Clippy", quand le système commet une erreur de jugement au lieu d'admettre sa confusion. Ouvrir les portes de l'apprentissage en demandant de l'aide à l'interlocuteur, c'est déjà le rallier à sa cause, comme le sait intuitivement tout bon politicien.

Deux actions parallèles peuvent être entreprises pour contrôler le taux d'erreur d'un système:

- 1- L'amélioration de la précision du système:

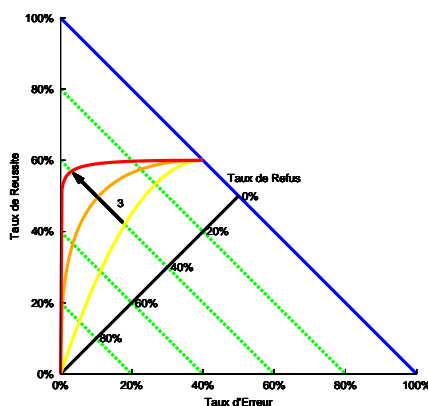
C'est la quête de la plupart des chercheurs en intelligence artificielle: améliorer la performance intrinsèque du système. Il s'agit souvent d'une entreprise lente et difficile, qui est au centre de la recherche scientifique. Comme le montre le graphique, le gain marginal d'une telle recherche dépend du point opérationnel sur la courbe de performance: il est possible d'améliorer un système considérablement à un taux de refus de 0% (flèche 1 sur

le graphique), tout en ayant un impact comparative-ment faible à un taux d'erreur bas (flèche 2): la valeur de l'investissement dépend énormément du contexte d'utilisation.



2- L'amélioration des mesures de confiance:

C'est une ligne d'étude souvent négligée en recherche, et paradoxalement à fort retour sur investissement pour l'ingénieur. C'est ce que j'entends par "apprendre à savoir quand on ne sait pas". Le bénéfice est le plus grand dans un contexte opérationnel où le coût marginal d'une erreur est élevé (flèche 3), ce qui est en général le cas dans un contexte grand public.



En général, cette ligne de recherche s'appuie sur

des techniques qui sont communes à toutes les formes d'intelligence artificielle: il s'agit de mécanismes de vote, ou de techniques de fusion des sources de données indépendantes. Les systèmes de reconnaissance de la parole sont parmi les plus avancés dans le domaine, et utilisent des sources acoustiques (traitement du signal), phonétiques (règles de prononciation), grammaticales (modèles de langage) et sémantiques (langage naturel) pour générer un modèle de probabilité. Ce modèle est souvent complété par une identification du sexe du locuteur, du type de microphone, et du type de ligne de transmission (numérique ou analogique). Cette combinaison de différentes sources d'information est essentielle pour non seulement déterminer ce que l'utilisateur dit, mais aussi quelle est la probabilité que le système ait commis une erreur dans son interprétation.

4 Un public en mouvement

L'industrie de l'IA dans son ensemble évolue vers une plus grande ouverture, et se cantonne de moins en moins aux niches traditionnelles que constituent les utilisateurs à haut degré de sophistication technique. A mesure que cette ouverture progresse, le niveau d'acceptabilité des agents intelligents s'étend à un public de plus en plus large. Les publicités contextuelles de Gmail, par exemple, dérangeaient il y a encore quelques années à cause de l'impression qu'elles donnent d'avoir été générées par quelqu'un qui analyserait le contenu de chaque courriel. Celle-ci sont maintenant mieux comprises et acceptées pour ce qu'elles sont, c'est à dire des produits d'algorithmes aveugles de traitement du langage naturel. De même, l'idée qu'Amazon soit capable de nous suggérer un bon livre en fonction de nos achats passés sur le site n'est plus généralement reçue comme une atteinte à notre vie privée.

La valeur et le rôle économique des systèmes prédictifs intelligents n'a cessé d'augmenter avec la quantité de données accessibles au travers de l'Internet. La demande en termes d'ingénieurs qualifiés dans le domaine est également en pleine expansion, et les qualités les plus recherchées vont bien au delà des simples compétences informatiques. Une

bonne compréhension du rapport à l'utilisateur est nécessaire à tous les niveaux de la conception logicielle, et l'on observe – enfin – une convergence des diverses activités de recherche, ingénierie, design, et ergonomie à travers toute une nouvelle gamme de produits intelligents destinés au grand public. "Clippy" ne s'en remettra jamais !