

Retour vers demain : en quoi l'intelligence artificielle peut-elle révolutionner nos usages et pratiques ?

Hugues Bersini pour TGS France

16/05 Lille 23/05 Nantes 30/05 Le Havre 13/06 Angers

Un peu d'histoire

Nous sommes en 1950, Maître Alan Turing sur son génie perché n'a pas encore subi l'horrible «traitement de faveur» que l'on réservait aux homosexuels à cette époque.

Il a trente-huit ans, sa créativité est au zénith et encore inimaginable est le morceau de pomme empoisonnée qui mettra fin à ses jours. Autour de lui, et grâce à lui, de nombreux chercheurs prennent conscience qu'un ordinateur, dénommé par le maître « machine universelle », pourrait, par son câblage et sa programmation, produire le type de processus cognitif dont jusqu'à présent les humains se réservaient la primeur.

À l'instar d'une forêt neuronale inextricable d'où jaillissent souvenirs, prévisions, inférences et raisonnement, les mêmes souvenirs, prévisions, inférences et raisonnements pourraient jaillir d'une forêt inextricable, mais cette fois de transistors en silicium et de portes logiques.

Ce cerveau machine, Turing l'ambitionne aussi infiniment programmable que l'est notre cerveau, dont les mêmes neurones se voient sollicités pour des tâches cognitives d'une variété infinie.

Entre deux marathons dont les photos d'époque témoignent, il perce le mystère de ces niveaux empilés d'abstraction qui font de l'informatique aujourd'hui la technologie la plus extraordinaire, la plus envahissante et la plus déconcertante jamais produite par l'humain : *machine universelle, machine à tout faire, universellement programmable et universellement intelligente.*

Quand on l'interroge sur sa définition de l'intelligence dès lors que c'est d'ordinateur qu'il s'agit (il faudra attendre la conférence de Dartmouth en 1956 pour que le terme « intelligence artificielle » prenne vraiment son envol), il se trouve à ce point dépourvu qu'il s'en tire par une pirouette : plutôt qu'une définition, c'est un test qu'il propose, que dit-on un test, une performance.

Bien évidemment, et il en est conscient déjà, c'est la notion d'intelligence qui fait débat. D'ailleurs, aujourd'hui, on ne compte plus les nombreux types d'intelligence qui se disputent le sacre : rationnelle, émotionnelle, sociale, animale, en essaim.... Sans doute, une définition audacieuse et quelque peu immodeste aurait pu être : « moi ».

Néanmoins, tout génie qu'il est, Turing entrevoyait déjà l'horizon très lointain et sans doute inatteignable de telles fulgurances pour sa créature. Il choisit alors de se cantonner à une tâche ô combien banale mais à la fois si emblématique de notre humanité : **le dialogue.**

Son test prend la forme suivante : « *si, lors d'un dialogue par écran interposé, un ordinateur est capable de vous confondre sur sa nature première, c'est-à-dire de se faire passer pour un interlocuteur humain, l'ordinateur ou l'algorithme en question que ce dernier exécute pourra être qualifié d'intelligent* ».

Remarquez combien notre homme fut prévoyant, car il évacua toute la problématique de l'apparence humaine, de celle qui fait tant défaut à nos machines. Il suffit de voir aujourd'hui un ordinateur/robot, aussi humanoïde qu'il tente de l'être, pour ne pas douter un seul instant de sa nature véritable.

Turing choisit d'épargner les chercheurs du futur et de faciliter la vie à nos machines. C'est bien de manifestations cognitives et seulement de cela qu'il s'agit ici, les deux interlocuteurs, le cobaye et l'ordinateur, devisant uniquement par écran interposé (on parlerait aujourd'hui, et dans un mauvais français, de chat ou de chatbot).

Ces jours-ci, ce sont les productions prétendument artistiques de l'IA qui nous forcent à nous poser à nouveau le même type de questionnement que Turing. À la simple vision des dessins, à la lecture des vers, à la simple écoute des mélodies, trouve-t-on à l'origine de ceux-ci des artistes authentiques ou des disciples logiciels, rattrapant du terrain, jusqu'à se confondre avec les maîtres humains ?

En effet, en proposant un tel test, Alan Turing, sciemment ou pas, nous confronta dès la genèse de l'informatique à une ambiguïté algorithmique fondamentale. S'il nous proposa une performance en effet, la capacité au dialogue d'une machine à ce point indistinguable de celle d'un être humain banal, il passa complètement sous silence la manière d'y parvenir. Or, il existe de fait plusieurs manières d'y arriver.

En voici une première, stupide on vous l'accorde, difficilement imaginable à l'époque de Turing, mais qui devrait pourtant vous sensibiliser à tout ce qui va suivre. Elle est très en phase avec l'IA d'aujourd'hui, mais pas celle de l'époque de Turing évidemment.

Imaginons que toutes vos conversations, sur WhatsApp, Messenger, SMS et autres réseaux sociaux, les vôtres et celles de l'humanité tout entière, aient été sauvegardées sur un même serveur informatique. Imaginons ensuite que le programme sommé de réussir le test de Turing procède de la façon suivante : à chaque question reçue, il fouille dans cette gigantesque base de données à la recherche de la réponse qui lui fut le plus souvent donnée.

Vu le gigantisme de cette base, on peut tout à fait supposer que chaque extrait du dialogue s'y retrouvera et même de nombreuses fois. Ainsi, « comment ça va ? » sera majoritairement suivi d'un « pas mal et toi », « cela te dit un cinéma ce soir ? » pourrait l'être d'un « chouette idée, quel film ? ». On simplifie à l'extrême (la supercherie pourrait être facilement déjouée en posant « quel jour sommes-nous ? » ou « quel temps fait-il ? »), mais vous devriez saisir l'idée.

Qualifieriez-vous vraiment un tel programme « d'intelligent », même si pourtant il parle plutôt bien, jusqu'au point de confondre son interlocuteur ? Ainsi, il ne fallut pas attendre très longtemps, quinze ans seulement, pour qu'un programme réussisse le test de Turing, mais qu'il le réussisse dans des circonstances quelque peu singulières : le cabinet d'un psychologue.

Voici un extrait de dialogue entre le logiciel dénommé Eliza, s'improvisant psychothérapeute, et une patiente. La patiente : « C'est mon ami qui m'a dit de venir vous parler. » Eliza : « Votre ami vous a dit de venir me parler ? » La patiente : « Il dit que je suis souvent déprimée. » Eliza : « Je suis vraiment navré d'apprendre que vous êtes souvent déprimée. » La patiente : « C'est vrai. Je suis plutôt malheureuse. » Eliza : « Pensez-vous que me parler vous aidera à ne plus être malheureuse ? »

La ficelle informatique avait beau être très grosse (le logiciel de quelques centaines de lignes, pas plus, se bornant à associer des bouts de phrase stéréotypées à des réponses construites à partir de ces bouts de phrase, mais tout aussi stéréotypées), les patients furent pour la plupart convaincus d'avoir affaire à un thérapeute véritable.

De vilaines langues pourraient argumenter que cela en dit long sur la profession de psychologue, mais, surtout, cela perturba considérablement Joseph Weizenbaum, le concepteur de ce logiciel idiot, au point qu'il décida par la suite de consacrer l'essentiel de ses réflexions à l'impact sociétal des technologies du numérique et à la crédulité des humains.

C'est aussi la preuve qu'un logiciel d'une grande stupidité, comme de nombreuses productions de l'IA aujourd'hui et surtout de l'apprentissage profond, perroquet obèse, obsessionnel du plagiat, peuvent nous bluffer sur leur soi-disant intelligence ou créativité.

Un autre philosophe américain aussi imaginatif qu'incisif, John Searle, dès les années quatre-vingt et alors que de premiers logiciels se montraient capables de traduire très approximativement, de résumer des textes simples et de répondre à quelques questions portant sur ces textes, se révéla très sceptique quant à la définition de l'intelligence artificielle proposée par Alan Turing.

Il n'écrivit pas un logiciel, mais présenta une expérience de pensée qui, à l'époque, suscita un formidable émoi dans la communauté des informaticiens et provoqua des milliers de réactions et de critiques. La voici très succinctement relatée. Elle reste toujours d'actualité.

Supposez que, lors d'un voyage en Chine, vous soyez emprisonné (rien de trop surprenant jusqu'à présent). Vous ne comprenez bien évidemment rien à la langue chinoise, ni parlée, ni écrite. Votre voisin de cellule, isolé depuis très longtemps, se voit soudain comblé d'aise par votre arrivée et se propose d'échanger avec vous par messages interposés qu'il vous glisse sous la porte.

On rappelle que vous ne saisissez rien de l'écriture chinoise. Cependant, ô astuce suprême, secours radical, délivrance, imagination philosophique débridée, vous avez à votre disposition un manuel gigantesque, associant sur sa page de gauche des expressions chinoises avec d'autres à relever sur sa page de droite.

Vous en déduisez qu'il vous suffit de feuilleter les pages de ce livre colossal (c'est une expérience de pensée, ne l'oublions pas) afin de retrouver sur la page de gauche le contenu du message transmis par votre voisin et de reproduire aussi adroitement que possible ce que vous retrouvez sur la page de droite.

Ensuite, après un laborieux griffonnage, c'est à votre tour de glisser la réponse sous la porte. Pour votre voisin de cellule, à qui ce compagnonnage apporte soudain quelque réconfort, il ne fait pas l'ombre d'un doute que vous parlez le chinois.

Mais vous ? Si, comme le questionnait John Searle avec quelque perfidie, vous affirmez comprendre le chinois, quand bien même vous vous limitez à tourner les pages d'un immense livre et à associer des symboles abstraits avec d'autres symboles abstraits, alors, en effet, les logiciels aujourd'hui, capables de mener un dialogue, traduire ou résumer un texte, répondre à des questions à son sujet, peindre des tableaux et rédiger des poèmes, ces logiciels comprennent ce dont ils parlent, saisissent ce dont ils traitent.

Car, dicit John Searle, les algorithmes d'hier et d'aujourd'hui, qui manipulent langages et textes, ne font pas plus pas moins que tourner les pages d'un immense manuel à la recherche de symboles qu'ils reconnaissent, et associer ceux qu'ils reçoivent en entrée à d'autres qu'ils produisent en sortie.

Pour le philosophe, cela n'a rien à voir avec le sentiment de compréhension, comme nous le vivons, comme nous le ressentons. À ses yeux, les manipulations syntaxiques ne conduisent pas directement au sens.

L'univers sémantique reste partiellement délaissé, inaccessible. Une compréhension hors de portée et qui trouve, selon lui, son origine dans l'expérience et le vécu des choses parlées, puise sa source dans notre incarnation et notre être au monde, tout simplement.

Tant Joseph Weizebaum que John Searle réussirent à démontrer, chacun à sa manière, que des algorithmes idiots, naïfs, efficaces mais sans finesse, bénéficiant de leur seule force brute, peuvent produire des comportements d'une grande sophistication. En substance, des algorithmes stupides peuvent apparemment se comporter intelligemment, prouvant que l'intelligence artificielle n'est pas toujours là où on la croit.

Les deux IA : inconsciente et consciente

Ainsi, pour de nombreux types de tâches cognitives que les logiciels aujourd'hui réussissent très décevantement et même avec tous les honneurs (comme les jeux de société, la conduite automobile et la traduction), il existe en effet deux manières très différentes d'accomplir ces tâches.

La première que nous taxerons « **d'inconsciente** », c'est-à-dire sans réelle explicitation de l'expertise humaine, sans règle, sans nul besoin de décomposer la tâche pour la simplifier et la soumettre au jugement humain, offre une très large place à l'apprentissage. Cet apprentissage s'effectue soit au départ de myriades de réalisations humaines passées, soit par simples essais/erreurs, conduisant graduellement à des logiciels qui se complexifient et s'opacifient d'autant.

On qualifiera **la deuxième** de « **consciente** ». Cette fois, l'expertise humaine reste toujours de mise, essentielle, la tâche se trouvant décomposée et découpée en séquences pour y glisser justement cette expertise.

Cette IA-là est à l'origine d'algorithmes dont la transparence et la patte humaine permettent encore et dans une large partie d'en comprendre le fonctionnement. C'est la bonne vieille IA d'alors, qui ne pouvait se passer complètement de nous ; mieux encore, qui gagnait en rapidité et en performance lorsqu'elle suivait le guide.

Malgré le succès croissant et l'engouement de plus en plus marqué au vu des performances décoiffantes de l'IA inconsciente, il n'en reste pas moins que cette tension entre les deux IA est vieille comme l'intelligence artificielle elle-même.

Dès la conférence de Dartmouth en 1956, son acte de naissance, on trouvait, à côté des premiers réseaux neuronaux et des logiciels apprenant par eux-mêmes à jouer au jeu de dames, des logiciels capables de raisonner (l'un d'entre eux s'appelait GPS pour General Problem Solver), de planifier, de résoudre des problèmes mathématiques et de diagnostiquer des situations complexes par la force de la raison humaine et la transposition logicielle de celle-ci.

On connaît l'histoire, ce sont les approches les plus cognitives qui l'emportèrent à l'époque et donnèrent véritablement naissance à l'IA et ses premiers succès pratiques. Le premier livre d'intelligence artificielle qu'il me fut donné de lire il y a plus de quarante ans, et un des premiers livres disponibles en français traitant du sujet, conviait Heidegger, Merleau-Ponty et Wittgenstein pour expliquer et cerner les limitations de l'intelligence artificielle de type « consciente », son incapacité à simplement s'interfacer au monde, faire corps avec le monde et ce, de par sa désincarnation et sa désolidarisation du substrat biologique.

D'où, de fait la contre-attaque continue des chercheurs de l'autre IA, voyant dans les réseaux de neurones (et leur version relookée ces temps-ci rebaptisée « apprentissage profond »), ainsi que dans la prise en charge informatisée des processus sensori-moteurs, dans leur apprentissage par imitation ou par renforcement, la meilleure manière d'affronter ces critiques.

Il n'y a sans doute dans cette distinction rien de très surprenant, ces deux « intelligence artificielle » se bornant à simplement reproduire la distinction bien connue des psychologues cognitifs entre les processus essentiellement conscients, lents, séquentiels, assez laborieux, et ceux inconscients, automatisés, parallèles, quasi-réflexes et pour lesquels nous excellons. Voilà bien une distinction archiconnue et détaillée dans tous les manuels de première année de psychologie.

Ces deux types de processus cognitifs ont été consacrés par le célèbre prix Nobel d'économie Daniel Kahneman et qualifiés de « système 1 » et « système 2 » de la pensée humaine². Dans les écrits des philosophes phénoménologues, tels Heidegger et Merleau-Ponty, une tentative est faite pour expliquer la transition entre ces deux systèmes, la sortie des processus inconscients et l'activation d'une réflexion plus lente, plus découpée, par l'occurrence de « ruptures » dans ces automatismes.

Ces ruptures sont, par exemple, la présence d'alternatives soudaines, d'un blocage imprévu, la survenue d'une attente non satisfaite ou d'un événement totalement inattendu. En réaction à ces ruptures, il nous faut éveiller notre conscience, déclencher les processus délibératifs qu'elle accompagne afin de nous sortir de ce mauvais pas, contourner l'obstacle, aller voir ailleurs. Actuellement, nous vivons cependant les succès spectaculaires essentiellement de l'IA inconsciente, celle qui s'ingénie à se passer de nous, celle qui met toute la force brute de nos ordinateurs à son service, celle qui se nourrit de la myriade de données que nous lui abandonnons dans les réseaux sociaux et autre carrefours informatiques incontournables sur Internet.

Méfions-nous de l'IA inconsciente

L'IA inconsciente gagne chaque jour, chaque heure, chaque microseconde davantage en autonomie. Elle doit une large part de son succès à la puissance de nos ordinateurs, leur parallélisme croissant, tel le recours aux cartes graphiques plutôt qu'aux processeurs classiques, ou l'utilisation de clusters (et demain l'ordinateur quantique ?).

À même d'éviter de plus en plus les ruptures évoquées précédemment, elle n'a plus jamais recours à l'introspection, à s'interroger sur son comportement, ses productions et à solliciter le cas échéant notre expertise, aussi modeste soit-elle. Nous vivons en ce moment un véritable paradoxe.

Chez nous, la complexité et le caractère inhabituel des situations à affronter nous obligent à sortir de nos bien confortables recettes inconscientes, à nous servir péniblement de notre conscience à la recherche de nouvelles solutions (par la simplification ou la découpe du problème, la recherche de situations quelque peu analogues).

En IA, ce même type de situation incite à contourner toute explicitation biaisante et peu satisfaisante, afin de s'en remettre au seul apprentissage, l'essai/erreur, et à la compilation de myriades d'expériences humaines ayant dû précédemment affronter ces problèmes.

Restons pourtant vigilants. Malgré ce que beaucoup aimeraient penser, fournir des données brutes à un algorithme pour lui faire apprendre une tâche n'a rien de vraiment objectif. Il est rassurant de penser que le processus d'apprentissage machine pourrait supprimer les biais que le programmeur met naturellement dans un programme qu'il fabriquera « à la main », de par les choix techniques et logiques qu'il doit opérer durant sa conception.

En fait, il n'en est rien ; on a plus que souvent simplement déplacé le problème du biais du concepteur aux données d'apprentissage.

Le réseau de neurones ainsi entraîné avec un jeu de données biaisé ne sera que l'image de ce dernier : un programme imparfait et biaisé, mais peut-être encore plus pernicieusement, sans qu'on l'ait forcément remarqué et avec la fausse impression du travail accompli. Il ne faut guère s'étonner que les pionniers de l'IA consciente, comme Marvin Minsky ou Noam Chomsky, se sont émus de l'importance prise par l'IA inconsciente qui, selon eux, donne beaucoup trop d'importance à la qualité des résultats au détriment de la manière d'y parvenir.

Or, pour ces pionniers, toute science exige une explicitation de la démarche, des étapes et des modèles mis en œuvre. Si les sciences exactes se détachent de l'ingénierie, c'est bien dans leur exigence de compréhension. Elles se doivent de comprendre autant que de prédire. Si la prédiction s'effectue au détriment de la compréhension, il y a tout lieu de faire marche arrière, de revenir aux fondamentaux.

Imaginons un physicien dont le seul apport serait de saturer le monde de caméras très précises et d'enregistrer toutes les évolutions de ce même monde. Afin de prédire l'évolution de ce dernier à partir de l'instant, et ce à partir de ce qu'il fut aux instants précédents, il suffirait de fournir à un réseau de neurones profond toutes les images du monde d'avant et de lui apprendre à prédire le monde d'après.

Il pourrait exceller à cette prédiction, mais la physique, elle, que gagnerait-elle à cette nouvelle manière d'exercer sa discipline ? Prédiction parfaite, d'accord, mais compréhension absente !

Pour ces pionniers, un logiciel qui traduit en se nourrissant d'une quantité gigantesque de traductions précédentes, quelle qu'en soit la qualité acquise, n'explique en rien sa démarche de traduction, comme un logiciel qui conduit en singeant de multiples conducteurs.

Ces perroquets logiciels n'ont plus grand-chose de scientifique car ils n'expliquent en rien leur performance. Un réseau de neurones, ingurgitant et digérant toutes les chansons des Beatles, peut produire une nouvelle chanson, originale bien sûr, mais sans rien maîtriser du génie créatif des Beatles.

Ces réseaux peuvent être à la fois excellents et ignorants. Il n'y a plus aucun objet conceptuel ni chaîne causale explicitée, formalisée, permettant de clarifier la raison de ces comportements, juste une banale performance d'ingénieur, aussi impressionnante soit-elle. Si cela n'est nullement condamnable pour les jeux de société, les linguistes (ce qu'est Chomsky à l'origine), eux, ont raison de regretter que ces nouveaux systèmes se détachent de la manière toute humaine de mener par exemple pareille traduction, rendant le résultat d'une grande sécheresse et dénué de toute nuance.

Pourquoi un réseau de neurones, aussi profond soit-il, devrait-il réapprendre les règles syntaxiques proposées par les linguistes et qui collent à notre maîtrise du langage, comme les accords ou les règles de grammaire ? Pourquoi réapprendre la loi d'Ohm ou celle des gaz parfaits pour un réseau de neurones censé prédire l'intensité d'un courant ou la mesure d'une pression ?

Bien sûr, il est de ces règles difficiles à expliciter, comme celles de nature statistique (dans la plupart des cas – mais combien – ce mot est suivi de celui-là), ou celles dont les prémisses sont trop complexes (quand le verbe auxiliaire est précédé du complément, où qu'il se situe dans les mots qui précèdent, alors il s'accorde...) et qui gagnent à être apprises, mais pas toutes...

Bien sûr, un réseau de neurones pourrait aussi prédire quelques écarts par rapport à ces lois parfaites de la physique (comme les non-linéarités), mais il le fera d'autant mieux qu'on lui a préinstallé dans ses connexions, à la main, les lois en question. Il est d'ailleurs tout aussi erroné de considérer les réseaux de neurones convolutifs comme de simples boîtes noires, car les filtres convolutifs qui les composent résultent de très longues années d'expérience et de connaissance en matière de traitement d'image.

La structuration subtile de ces réseaux de neurones profonds témoigne d'une ingénierie fine de ce champ d'étude et qui ne doit pas grand-chose au hasard. Les ingénieurs sont présents dans la préstructuration de ces réseaux.

On tend à penser qu'ils pourraient l'être encore plus. Avec le temps, il deviendra sans doute possible de mieux comprendre le fonctionnement, tout excellent qu'il soit, de ces opulents réseaux de neurones multiparamétrés, leur partie vraiment boîte noire, et de déceler en leur sein de nouveaux objets de sciences.

Cela aiderait quelque peu à la réconciliation de ces deux traditions qui tendent pour l'instant à s'observer en chien de faïence.

L'IA : Ultime recours face à un monde devenu trop complexe

Le monde est devenu bien trop complexe pour en confier la gestion aux seuls gouvernants en chair et en os. Les lourds nuages noirs chargés d'orage qui nous menacent, comme :

- l'économie mondialisée et les crises qui s'ensuivent régulièrement,
- l'explosion des inégalités,
- le réchauffement climatique et la dégradation de l'environnement,
- la détérioration de l'habitat, la transition énergétique inéluctable,
- l'agriculture qui s'intensifie en tout sauf en bienfaits pour la santé,
- les fractures communautaires et la flambée du terrorisme,

exigent une meilleure compréhension et interprétation des phénomènes, suivies de délibérations approfondies pour en trouver les parades et les issues.

De son côté, l'IA – meilleur joueur d'échecs, meilleur joueur de go, champion de poker, du jeu télévisé *Questions pour un champion*, conducteur irréprochable, médecin infallible, meilleur prévisionniste des soubresauts économiques et des variations climatiques, traducteur le plus sollicité, meilleur décideur – est capable, par sa boulimie de données multiples à très haute fréquence d'acquisition et son pouvoir de simulation pour en faire sens, d'appréhender ce monde, son évolution et les meilleures façons de le remettre sur les rails.

Cette même IA, que l'on magnifie ces jours-ci de super-intelligente et qui, de surcroît, s'immisce dans tous les recoins de nos existences, est capable d'explorer, à partir de ces simulations, ces états des lieux d'une précision encore jamais atteinte, les manières les plus efficaces d'affronter l'ensemble de ces difficultés et d'en découvrir les remèdes les plus appropriés.

Et si elle parvient à cela, c'est en exécutant à la vitesse de la lumière de multiples algorithmes dont les fonctions sont nombreuses et variées :

- planifier et tester des futurs alternatifs,
- recueillir des myriades de données,
- les stocker intelligemment pour mieux les exploiter,
- s'en servir en effet pour la conception de modèles au pouvoir prédictif,
- exploiter ces modèles jusqu'à plus soif pour tester à partir de ceux-ci les options possibles,
- ajuster, optimiser ces dernières
- et aller jusqu'à la prise de contrôle de nos vies et le tracé de nos existences vers davantage d'harmonisation sociale dans un monde préservé et pacifié.

Il n'y a plus guère un seul objet de notre quotidien qui ne recèle quelque intelligence algorithmique, qui ne soit capable de percevoir son environnement et de décider de manière autonome de la meilleure action à entreprendre.

Il décide en effet tout seul ou de façon coordonnée, dialoguant avec ses partenaires distribués à travers le globe. Nous vivons à la fois l'ère de la super-intelligence informatique, de la puissance décuplée des processeurs, des objets connectés et de l'informatique ubiquitaire. Pourtant, c'est cette même ère qui se trouve menacée d'extinction.

Et c'est surtout à vous que nous pensons, les plus jeunes, les programmeurs en herbe, afin de saisir les opportunités extraordinaires que l'IA vous offre pour faire face à tous ces défis et d'y faire face au plus vite.