Tous les mois, un nouveau livre nous avertit contre les dangers de l'intelligence artificielle, le plus précis d'entre eux, L' Intelligence artificielle ou l'enjeu du siècle: Anatomie d'un antihumanisme radical (par Eric Sardin) est un brûlot anti technologie. Voici quelques réflexions pour aller plus loin que les poncifs en la matière.

Toute technologie, l'intelligence artificielle peut être utilisé pour le meilleur et pour le pire. Le meilleur du meilleur, cela peut être de permettre à la médecine de progresser en interprétant mieux que le plus expérimenté des cancérologue l'IRM de votre cerveau, de détecter les effets de secondaire de médicaments ou d'adapter le programme d'apprentissage au profil de chaque élève. Cela peut être à un plus bas niveau, de délivrer l'entrepreneur de bases des tracasseries comptables et administratives (en automatisant la création d'entreprise, la saisie de pièces comptables grâce à des systèmes qui extraient TVA et type de dépenses et en détectant automatiquement les numéros de comptes associés), le tout étant intégré de façon invisible dans des services d'expert comptable en ligne comme on peut en trouver sur les systèmes de l-expert-comptable.com, un site sur la création d'entreprise .

Mais une partie des applications d'intelligence artificielle s'appuient sur l'analyse du comportement pour en tirer des règles permettant de modéliser des comportements/raisonnements complexes.

Internet a largement facilité cela dans la mesure où il représente à la fois un gisement de données et à la fois le medium permettant d'accéder aux données. Bien que ce lieu recèle d'énormes quantités de données, ce n'est pas le seul moyen d'apprendre et de s'améliorer, tant pour l'humain que pour les robots que nous créons. Ce type d'apprentissage peut être appelé inversion. Il s'appuie sur le principe que le praticien - l'expert du domaine dans lequel il exerce ses activités - accomplit sa tâche à 100% de façon parfaite et que l'étudiant ou l'intelligence artificielle apprend les étapes à suivre. et croyez que ce que le praticien fait est le moyen le plus optimisé d'accomplir cette tâche particulière. Si le praticien fait une erreur, elle ne sera pas remarquée.

L'intelligence artificielle est compliquée et facilement implémentée ou créée à tort - en raison de nombreux facteurs imprévisibles - que nous n'avions pas prévue. Principalement parce que cela semble facile ou que le résultat semble être juste - exactement comme nous le pensions.

Nous devons éviter cette hypothèse

Lorsque je parle d'intelligence artificielle, cela peut faire référence à quelque chose de physique - un robot - ou à quelque chose de moins physique - du code. À la fin, les deux se rejoignent. Comme le code a besoin de quelque chose qui puisse l'héberger.

Le bouton arrêt

Nous savons donc qu'Internet peut être d'une grande utilité dans la courbe d'apprentissage et la compréhension du cerveau humain. Aujourd'hui, plusieurs implémentations d'IA effectuent déjà une tâche unique, dans leur propre petit monde. Par exemple, une intelligence artificielle aux échecs est difficile à vaincre, mais lui demander de conduire une voiture est impossible. Elle ne sait tout simplement pas ce qu'est une voiture ni même comment fonctionnent les routes. Il ne connaît que les pièces sur le tableau, les étapes que chaque pièce peut faire et le résultat de chacune de ces étapes. Mais, selon certain, le Saint Graal de l'intelligence artificielle ne consiste pas à améliorer l'IA des échecs, mais à créer une "IA générale", qui possède les mêmes paramètres que nous, les humains.

Lorsque nous travaillons à la création de cette intelligence artificielle générale, aussi proche que possible de l'homme, nous devrions créer des mécanismes qui permettent à l'IA d'agir de manière à faire les choses, car c'est dans le meilleur intérêt de l'homme. Les choses que nous voulons. Par exemple, si nous demandons à un robot de faire du thé, il voudra effectuer cette tâche, quoi qu'il arrive. Si le robot qui se dirige vers la cuisine croise son chemin avec votre bébé rampant sur le sol, il ne changera pas de chemin. Il va directement à la cuisine. La tâche est la chose la plus importante à ce stade. Nous pouvons construire un bouton d'arrêt pour éviter la tragédie du robot écrasant le bébé sur son chemin. Mais le robot ne vous laissera pas appuyer sur le bouton, car la tâche ? - ? et donc la récompense (au sens algorithmique) passe par le fait de vous préparer une tasse de thé. Comment pouvons-nous contourner cela?

Faisons en sorte que la pression sur le bouton d'arrêt génère la même récompense que la préparation du thé. Le résultat sera probablement que le robot pousser le bouton tout de suite.

Intégrez secrètement du code afin de pouvoir arrêter le robot à distance. Mais finalement, le robot pourra apprendre ce secret et modifier son propre code.

Menace 1: Paramètres

Avant de pouvoir fabriquer un appareil, un robot ou même un humain, nous utilisons des paramètres, des facteurs qui définissent des règles ou des conditions pour exécuter la tâche donnée. Facile à dire quand tout fonctionne à partir d'une grande fonction qui s'exécute automatiquement lorsque vous démarrez un système. Les fonctions peuvent se recouper les unes les autres et peuvent transmettre des paramètres tels que des nombres, du texte ou même d'autres fonctions. Et ces paramètres sont cruciaux pour le résultat obtenu. Plus la fonction est petite ? - ? moins le code est important ? - ? plus il est facile de déboguer et de penser à l'avance à tous les scénarios possibles. Dès que nous construisons un cadre de fonctions, les choses se compliquent et un seul paramètre erroné peut entraîner la rupture du code, l'exécution d'une tâche erronée ou même une boucle sans fin.

La saisie est donc très importante et vous devriez avoir tous les résultats possibles couverts avant d'exécuter le code. Dans l'exemple du bouton d'arrêt, si vous n'avez pas codé la possibilité de la traversée de bébé et l'importance du bébé dans la portée de la tâche. Le robot ne comprendra pas pourquoi vous essayez d'appuyer sur le bouton d'arrêt et le robot tentera de vous arrêter.

Menace 2: apprentissage

L'apprentissage contrôlé est quelque chose que nous devrions encourager, mais cela n'empêchera pas d'autres phases d'apprentissage. Dès que la fonction devient compliquée, il est difficile de déboguer et de penser à tous les scénarios possibles. La quantité de données est trop difficile à contrôler. Nous savons ce que nous y mettons , mais nous ne savons pas avec certitude ce que cela va apprendre de plus ni comment cela affectera l' apprentissage futur. Comparez-le avec l'éducation. Vous essayez d'enseigner à vos enfants les éléments essentiels de la vie, mais le long de la route, vous ne pouvez pas contrôler chacune des étapes du processus d'apprentissage. Dès que l'enfant grandit, il contrôle mieux ses propres décisions , intérêts et apprentissages. Ses camarades peuvent faire bugguer l'algorithme pour lequel vous les avez programmés (éduqués).

Même chose pour l'IA, nous savons ce que nous y mettons, mais dès que cela deviendra plus intelligent, nous perdrons le contrôle. Nous pouvons essayer de persuader l'IA, comme nous le faisons avec notre propre espèce ? - ? d'autres humains. Mais nous ne pouvons pas non plus prédire le résultat, comme lorsque nous avions notre première machine à café à base d'IA avec une tâche simple et simple.

Menace 3: portée

La portée est la portée de laquelle le code ? - ? intelligence artificielle ? - ? évolue. Un code qui vit dans une portée ne doit jamais pouvoir entrer dans la portée parente (au sens informatique du terme). Alors que la portée d'un parent doit pouvoir interférer avec son enfant. Cela ressemble à une bonne pratique sécurisée . Mais que se passe-t-il s'il n'y a pas vraiment une portée parente ou si une portée peut entrer dans d'autres portées qui vivent dans le même niveau de hiéarchie ou si nous construisons des points d'accès pour connecter une portée à plusieurs niveaux de hiérarchie, afin qu'elle puisse effectuer les tâches nécessaires transfert ou contrôle de données.

Menace 4: les mythes

Tout d'abord , qu'est-ce qu'un mythe? C’est croire en quelque chose qui n’existe pas, mais uniquement dans notre esprit et notre imagination. Bien qu'un mythe puisse ressembler à un objet physique car il peut avoir des atouts, cela ne signifie pas pour autant qu'il est réel.

Certains mythes auxquels nous participons aujourd'hui:

1. Italie, États-Unis (des pays)

2. Islam, christianisme (des religions)

3. Heineken, Apple (des entreprises)

4. Euro, Dollar, Bitcoin (des monnaies)

5. Capitalisme, communisme (système socio-économique»)

Dès que nous cesserons de croire aux mythes susmentionnés, leur existence disparaîtra. Il y a des choses dans un pays, la religion "a" des atouts, une entreprise pourrait avoir un bureau, des voitures, des ordinateurs, de l'argent pourrait être imprimé sur du papier et un système

social "économique" pourrait avoir une empreinte sur l'histoire. Il est difficile de croire que votre pays est un mythe, comme vous pouvez le voir sur la carte. Mais si nous arrêtons tous de croire en un pays particulier , qu en vaut-il vraiment la peine? La raison pour laquelle il est difficile de comprendre que les éléments ci-dessus de la liste sont tous des mythes est qu'ils sont volumineux. Beaucoup de gens croient en eux. Et il faut beaucoup de mécréance pour le décomposer. Plus les gens croient en un mythe, plus il est difficile de le remettre en question et plus la révolution doit être dure pour le démolir.

Qu'est-ce que les mythes ont à voir avec l'IA? Il y a deux pensées à ce sujet. L'existence physique de l' intelligence artificielle peut être utilisée pour maintenir ou détruire les mythes, ce qui conduirait à de dures révolutions. Parce que la plupart des mythes ont une portée énorme. Deuxièmement, l'IA pourrait devenir un mythe par elle-même. Et je pense que c'est déjà le cas. Les gens croient que l'humanité a besoin de ce réseau autonome pour franchir une nouvelle étape dans l'évolution. Il se développe dans une foi basée sur la technologie.

Dernière menace: les érudits

Il y a quelques temps, je regardais une émission de télévision sur les technologies d'avenir. À la fin, un groupe d'érudits s'est présenté et a donné son point de vue sur l'IA et ce qu'elle nous apportera à l'avenir. Ce que j'ai remarqué à propos de leurs réponses aux questions du public, c'est que ce qu'ils disaient n'était ni nouveau, ni novateur , ni susceptible de répondre aux questions. Ils présentaient un point de vue commercial sur leur petit monde avec une largesse d'analyse très étroite.

Un universitaire a besoin de financement et un universitaire ne peut travailler que dans son domaine d'expertise (connaissances). Et ce domaine n'est rien de plus alors une petite partie de l'histoire complète. Vous pouvez comparer chaque chercheur à une seule fonction, au sein du grand logiciel de la science dans son ensemble. Un érudit ou un groupe ne peut jamais vraiment comprendre leurs résultats dévastateurs sur le monde quand ils insèrent leurs découvertes dans la vie quotidienne.

Bien que la science soit aussi claire qu'une formule ? - ? mathématiques pures, un érudit est en revanche un être subjectif. A la Dutch Design Week, de nombreux grands inventeurs de nouvelles implémentations d'IA étaient présents et présentaient tous la même caractéristique : une vision si égoïste de l'utilisation ou de ce qui est acceptable et cela uniquement pourrait constituer une menace pour le monde et la création d'IA.

Le plus dangereux avec une nouvelle technologie, c'est moins son potentiel que le manque de vision de ses promoteurs. Comme de promoteurs, il y a foule actuellement et que la mode est à la mission et au why des entreprises, nous avons face à nous, un nouveau défi: faire gagner une vision dans laquelle l'intelligence artificielle se met au service de l'homme au lieu de servir l'homme, avertit Raphael Richard, formateur en intelligence artificielle .