

Pour une pédagogie de l'intelligence artificielle



Docteur en philosophie et professeur à l'université de Milan, **Giovanni Landi** travaille depuis 25 ans dans le domaine de l'informatique avancée. Il s'intéresse à l'IA vue sous le prisme du savoir philosophique. Selon lui, l'IA est « une continuation de la philosophie par d'autres moyens », thèse présentée dans son ouvrage *Intelligence Artificielle* comme Philosophie (éd. Ovidia, 2022) et dans plusieurs articles publiés en italien, français et anglais. Il est expert à l'Institut EuroPIA et vice-président de EuroPIA Italia.

L'histoire de l'IA est faite de hype mirobolants suivis d'effondrements tout aussi soudains. Si nous voulons que cette fois-ci soit la bonne, la formation et le storytelling ne suffiront pas, nous avons besoin d'un regard philosophique pour comprendre le dilemme actuel de l'industrie informatique et les solutions possibles.

IA : le progrès à tout-va

L'IA est un magnifique *concept marketing* (ou *storytelling* si l'on préfère) pour promouvoir et décrire la pénétration continue des technologies de l'information et du numérique dans nos vies en tant que consommateurs, producteurs, citoyens etc. Ce *concept marketing* apparaît en fait avec une fréquence inversement proportionnelle au degré de spécialisation des médias utilisés, c'est-à-dire que plus le public est généraliste, plus la promesse est exagérée. Pour ne citer que deux exemples strictement authentiques, voici l'algorithme « *inséré dans chaque traitement musical* » qui aide à synchroniser l'utilisateur « *avec les ondes cérébrales typiques d'une séance de méditation, favorisant l'élimination du stress et des pensées répétitives qui y sont liées, en moins de 16 minutes* » : ou encore le « *robot sommelier, intelligence artificielle au service du vin...*

capable de proposer des expériences liées aux vins millésimés à travers des suggestions précises et intuitives ».

Et la liste est longue. Cette approche à mi-chemin entre le prédicateur américain et le vendeur de voitures d'occasion, qui semble ne connaître aucune limite, contamine également le métavers, la dernière invention futuriste des spécialistes marketing des technologies de l'information.

Ne nous moquons cependant pas de cette *hubris* technologique, certes maladroite, mais également nourrie de l'inspiration même qui a fondé l'IA, celle qui a conduit d'éminents mathématiciens et scientifiques de bonne foi à énoncer des promesses ridicules à relire aujourd'hui.

C'est ce premier élan, authentique, celui de Turing et de la conférence de Dartmouth qui permettent aujourd'hui de revendiquer la philosophie comme connaissance privilégiée pour comprendre l'IA¹.

IA faible, très faible, pratiquement... stupide

D'autre part, l'IA effraie, pas seulement à cause de films style *Terminator* mais surtout par ses aspects économiques très concrets, le souci légitime de nombreux travailleurs d'être un jour remplacés par des systèmes informatiques actifs 24h/7j/7, sans revendications salariales et pour qui le mot « grève » n'existe pas. Il existe aussi le mou-

vement inverse, où l'on évoque « *l'IA faible... qui ne remplacera pas l'humain, mais qui libère l'énergie et la créativité, pas quelque chose dont nous souffrons passivement* », des manifestes *techno-optimistes* prônant la cause d'un outil de *calcul simple*. Dans certaines conférences spécialisées, l'expression « Intelligence Artificielle » peut même faire l'objet de réclamations, jugée trop sémantiquement chargée, à laquelle seraient préférées des formulations plus rassurantes, comme les bons vieux « systèmes experts » ou le plus simple « *machine learning* » (après tout la *nouvelle vague* de l'IA, celle des systèmes experts des années 1980, se positionnait également en opposition à l'IA jugée trop futuriste et généraliste, et se revendiquait comme simple outil de support à l'activité humaine).

Cette deuxième « âme » de l'IA se veut rassurante, non invasive, complètement incluse dans les arcanes de la *computer science*. Elle aspire à une intelligence « *re-productive et non cognitive... efficace*

précisément parce que son action est séparée de l'intelligence... (une IA pour laquelle) il ne s'agit pas de la capacité de reproduire l'intelligence humaine, mais de la capacité de s'en passer² ». En forçant très légèrement le trait, une simple calculatrice électronique pourrait donc être définie comme un système d'IA !

L'IA effraie, pas seulement à cause de films style Terminator mais surtout par ses aspects économiques très concrets...

Philosophie, encore de la philosophie, toujours de la philosophie

De toute évidence, derrière tout cela se cachent des intérêts économiques précis et légitimes, ceux de la transformation numérique qui ne peut pas se permettre de faire douter les utilisateurs, clients et citoyens, sur les avantages de ce que l'industrie informatique promet, et qui doit donc ramener l'IA à un niveau compréhensible pour le grand public (surtout dans un moment historique où les fausses nouvelles et les théories du complot prévalent).

Il nous semble alors que seule la philosophie est l'outil approprié pour comprendre cette évolution. D'abord parce que la définition même de l'IA « faible » nous vient d'un philosophe, John Searle, et de son expérience en 1980 de « pensée de la chambre chinoise » qui réfutait la possibilité d'une intelligence artificielle « forte » et ouvrait ainsi la voie à la « faible ».

Ensuite parce que cette neutralisation de l'IA est ce qui explique l'explosion du débat « éthique » sur l'IA, qui fait couler tant d'encre aujourd'hui et crée tant d'opportunités d'emplois pour les « éthiciens de l'IA ». Si l'IA n'est qu'un outil, alors sa puissance doit être soumise à des règles précises, lesquelles se révèlent cependant abstraites et malheureusement inapplicables.

Pour comprendre ce mouvement, relisons le bon Kant, qui limitait la connaissance au monde des phénomènes, en niant la possibilité pour l'homme d'arriver à la connaissance du *nouménon* (la chose en soi). Cependant, en échange de cette limitation, Kant promettait la certitude des conclusions. De la même manière, l'IA

« faible » nie la possibilité pour une machine d'être consciente, d'avoir des sentiments, etc., et limite sa fonctionnalité à la faculté de l'intellect, du calcul (informatique).

Pour tout ce qui ne peut être calculé, par exemple pour le libre arbitre et donc la moralité, la solution de Kant est la foi, la loi morale, le « *je dois parce que je dois* » sans plus de précisions. Si la règle est « *ne mens pas* », dit Kant, je ne peux même pas mentir pour sauver une personne qui se réfugie chez moi poursuivie par un meurtrier, et à qui

je « dois » révéler la présence de la victime (je peux peut-être espérer qu'elle parvienne à s'échapper mais jamais couvrir sa présence par un mensonge). Une telle morale abstraite évidemment dans la vie réelle est de peu d'utilité parce qu'elle n'est que pure forme.

C'est ce qui est en jeu dans l'éthique

de l'IA, qui produit des méthodologies très valables (*ethics by design*, *data ethics*, *ethics based assessment*), mais complètement abstraites. Par exemple, elles pourraient servir en Europe à protéger les droits de l'individu contre la société mais peut-être en Chine à protéger les droits de la société contre l'individu. Le *scoring social*, par exemple, est selon les méthodologies de l'éthique de l'IA justifiable autant que le RGPD, selon le point de vue adopté.

Seule une pédagogie peut nous sauver

Revenons au dilemme de l'industrie informatique. Le problème avec l'IA est que le battage médiatique actuel ne peut nous faire oublier que dans

sa courte histoire, elle a déjà connu deux « hivers » (cette notion renvoie à des périodes de gel des financements dédiés à la recherche en IA) et qu'aujourd'hui nous assistons à sa troisième renaissance. Évidemment, il est dit que cette fois sera la bonne, mais le doute reste légitime.

La philosophie peut-elle aider à trouver une solution ? Oui, certainement, sous la seule forme que la philosophie authentique peut prendre quand elle devient *praxis*, c'est-à-dire sous la forme d'une *pédagogie*. Il est nécessaire d'enseigner, de faire comprendre, d'expliquer l'IA à tous les niveaux de la société, de démontrer sa pertinence, aux étudiants, professeurs, juristes, politiciens, entrepreneurs, citoyens. Mais pour que cette opération soit efficace, la pédagogie ne peut pas se situer uniquement dans le monde éducatif ; nous ne parlons pas ici de *computer literacy*, ni de faire de nous tous des informaticiens. La formation « forme », c'est-à-dire à pour objectif de façonner ceux qui apprennent à découvrir un environnement différent, tandis que la pédagogie (d'origine socratique, pour être clair) fait ressortir chez l'apprenti une connaissance et une conscience qui sont déjà là, et à laquelle il suffit de donner le bon éclairage. Pour cette raison, une pédagogie de l'intelligence artificielle ne peut réussir que si celle-ci est conçue non pas comme une discipline STEM mais comme une discipline humaniste à tous égards, comme une « continuation de la philosophie par d'autres moyens ».

Références :

1. Luigi Maiello le fait dans "Le débat sur l'intelligence artificielle risque de devenir improductif et autoréférentiel", *Il Fatto Quotidiano*, 28 mai 2022 ; David Deutsch le fait dans "Philosophy will be the key that unlocks artificial intelligence", *The Guardian*, 3 octobre 2012 ; et aussi à notre petite façon : Giovanni Landi, *Intelligence Artificielle comme Philosophie*, éditions Ovidia, 2022.
2. "All'Intelligenza Artificiale non serve l'intelligenza", *La Stampa*, 11 mars 2022.

Il est nécessaire d'enseigner, de faire comprendre, d'expliquer l'IA à tous les niveaux de la société, de démontrer sa pertinence.