

L'intelligence artificielle à la lumière de la mythologie grecque : rendre compréhensible les impacts de l'IA pour le grand public

Fabrice Muhlenbach

Université Jean Monnet, UJM-Saint-Etienne, CNRS,
Laboratoire Hubert Curien UMR 5516, F-42023 Saint Etienne, France

fabrice.muhlenbach@univ-st-etienne.fr

Résumé

Depuis les années 2010 et l'arrivée des modèles d'apprentissage profond [11], l'intelligence artificielle (IA) est un domaine qui s'est fait connaître du grand public. Face à une mauvaise compréhension de ce que peut faire l'intelligence artificielle, de ses limites et ses dangers, mais aussi comment utiliser cette technologie à bon escient, une manière originale de présenter l'IA est proposée sous la forme d'une analogie avec des légendes issues de la mythologie grecque afin d'être plus accessible aux personnes plutôt tournées vers les humanités que vers les sciences du numérique.

1 Introduction

L'intelligence artificielle (IA) peut se concevoir comme une discipline scientifique ayant pour objet de simuler sur une machine les processus de la pensée. Par glissement sémantique, l'intelligence artificielle en est venue à désigner aussi une technologie et tout type de productions de cette discipline : algorithmes, codes, programmes et autres applications informatiques, pourvu que ces derniers disposent de fonctionnalités qui peuvent être qualifiées d'« intelligentes » parce que mimant une tâche intellectuelle, de l'ordre de la perception, de l'apprentissage, du raisonnement ou du langage, de la prise de décision ou d'action sur l'environnement.

L'IA ne laisse pas indifférent. Enthousiastes, certains voient en l'intelligence artificielle une fabuleuse histoire composée de toute une série d'inventions témoignant du génie humain [16] alors que d'autres, critiques, la considèrent comme la nouvelle barbarie [5] ou une offensive technologique résolument antihumaniste [19]. En effet, l'utilisabilité de l'IA peut s'avérer être une réelle menace pour les valeurs humaines [14] mais peut aussi apporter une aide précieuse pour trouver des solutions aux grands problèmes que rencontre l'humanité [13]. Après avoir connu diverses périodes de déclin ou de renaissance dans son histoire, l'intelligence artificielle se voit dotée de capacités extraordinaires amenant des fantasmes avec une vision partisane, allant de la conception transhumaniste qui voit en l'IA un avantage compétitif permettant à une portion réduite de l'humanité d'accélérer son évolution et, à l'autre bout du spectre, à la prise de position de technophobes qui assimilent l'IA au mal ultime dont l'humanité doit se préserver à tout prix.

Face aux incompréhensions que l'intelligence artificielle suscite, notre approche se veut un clin d'œil à l'ouvrage du philosophe américain Hubert L. Dreyfus [6] publié dans les années 1970, et qui, dans sa version française, avait pour titre « L'Intelligence Artificielle : Mythes et Limites ». À l'époque, cette parution avait eu pour effet de doucher l'enthousiasme excessif qui régnait au sujet des potentialités de l'IA et avait fourni des arguments à l'origine du désamour pour cette discipline appelé « le premier hiver de l'IA ».

Nous souhaitons ici illustrer au moyen des mythes de la mythologie grecque certaines caractéristiques de l'IA. L'histoire d'Œdipe et de sa fille Antigone a inspiré des artistes de l'Antiquité comme les trois grands tragiques grecs Sophocle (avec *Œdipe roi*, *Œdipe à Colone* et *Antigone*) [22], Euripide (*Les Phéniciennes*) et Eschyle (*Les Sept contre Thèbes*) au V^e siècle av. J.-C. ou le Romain Sénèque au I^{er} siècle, à l'Époque moderne avec des pièces de Pierre Corneille ou Voltaire¹, jusqu'à nos jours au théâtre avec Jean Cocteau, Bertolt Brecht ou Jean Anouilh, à l'opéra (*Œdipe à Colone* d'Antonio Sacchini en 1786 ou *Œdipus rex* d'Igor Stravinsky en 1927) ou au cinéma (comme *Œdipe roi* de Pier Paolo Pasolini en 1967 ou *Antigone* de Sophie Deraspe en 2019). Au tournant du XIX^e et du XX^e siècle, Sigmund Freud donnera même une interprétation simplifiée du mythe grec d'Œdipe afin d'illustrer les relations entre parents et enfants et il fera du « complexe d'Œdipe » un concept fondamental de la psychanalyse [9].

L'intérêt d'un mythe est de parler à un grand nombre de personnes afin de partager des références communes, mais il existe de nombreuses variantes d'un même mythe et de multiples interprétations de ces mythes suivant ce que cherche à mettre en avant l'auteur. Pour Sophocle qui, en plus d'être dramaturge et poète, était aussi un homme impliqué dans la vie politique grecque, mettre en scène les mythes permettait de questionner la légitimité de la souveraineté. De la même façon, en suivant le destin d'Œdipe, de ses origines à la fin tragique qui toucha sa descendance, nous souhaitons éclairer des non-spécialistes en informatique sur les principaux impacts sociétaux liés à l'IA afin de dépasser une vision partisane et d'apporter des lumières sur la façon d'employer cette technologie avec raison.

1. Un grand nombre de ces pièces sont disponibles en libre accès, notamment sur le site <https://www.theatre-classique.fr/>

2 Les origines des Labdacides : de la révolution numérique aux biais algorithmiques

2.1 La fondation mythique de Thèbes

Comme bien souvent dans la mythologie grecque, tout commence par une intervention des dieux de l'Olympe qui s'éprennent à l'occasion des mortels [25]. Selon la légende, Zeus découvrit la ravissante Europe sur une plage des côtes asiatiques de la mer Méditerranée et il tomba sous son charme. Se métamorphosant en taureau blanc afin de ne pas effrayer la princesse phénicienne et d'échapper à la jalousie de son épouse, la déesse Héra, le roi des dieux s'approcha de la jeune femme, la laissa monter sur son dos et l'enleva en sautant dans l'eau pour traverser la mer jusqu'à arriver en Crète.

Agénor, roi de Tyr et père d'Europe, demanda à son fils Cadmos de retrouver sa fille disparue : tant que cette mission n'était pas remplie, il serait interdit à Cadmos de retourner en Phénicie. Pendant des années, avec l'aide de ses hommes, Cadmos parcourut le monde à la recherche de sa sœur, sans succès, et il vint en Grèce consulter l'oracle de Delphes où la prêtresse lui annonça que son destin n'était pas de revenir à Tyr mais de fonder une nouvelle cité là où se coulerait une vache.

L'interprétabilité des oracles et des modèles prédictifs

En 1962, l'auteur de science-fiction Arthur C. Clarke avait indiqué que : « toute technologie suffisamment avancée est indiscernable de la magie » [4]. Dans notre quotidien, nous sommes utilisateurs d'objets technologiques, mais nous n'avons le plus souvent qu'une très vague idée du mode de fonctionnement de ces objets. Pour nombre d'entre nous, le rapport entretenu avec l'intelligence artificielle est équivalent à celui de nos ancêtres de l'Antiquité lorsqu'ils s'adressaient aux dieux pour recueillir leurs oracles. Ils s'adressaient à une entité qui les dépassaient pour avoir des conseils ou des réponses à leurs questionnements, tout comme nous nous fions actuellement aux résultats des moteurs de recherche et des systèmes de recommandations auxquels nous accédons depuis nos téléphones portables et nos ordinateurs. Dans l'Antiquité, face aux grandes questions auxquelles ils ne parvenaient à trouver de réponses, les hommes s'adressaient aux dieux dont ils cherchaient à interpréter les signes à travers diverses manifestations observées dans la nature. La Pythie, au temple d'Apollon à Delphes, était l'un des plus célèbres oracles de cette époque. L'oracle rendu par la Pythie n'est cependant pas expliqué et peut donner lieu à des interprétations contradictoires. À titre d'exemple, Crésus, roi de Lydie, entreprit une guerre contre Cyrus, le roi des Perses, car il avait reçu de

l'oracle comme prophétie qu'il détruirait un grand empire... mais bien mal lui en a pris car l'empire qu'il détruisait fut le sien et non l'empire perse.

De la même manière, les prédictions apportées par l'intelligence artificielle peuvent donner l'impression d'avoir des propriétés quasi-divines. Les systèmes intelligents dont les modèles ont été produits à partir d'apprentissage automatique sur des réseaux de neurones profonds, par exemple, sont capables de réaliser très peu d'erreurs. Ils nécessitent cependant une grande part de confiance de la part de leurs utilisateurs car ces modèles sont des « boîtes noires » dont les raisonnements sont impossibles à interpréter par un être humain.

Cadmos se conforma aux recommandations de l'oracle et trouva une vache qu'il suivit jusqu'en Béotie et décida de poser les fondations de la ville où s'était couché l'animal. En envoyant ses hommes chercher de l'eau pour sacrifier la vache à Athéna, ces derniers rencontrèrent un dragon qui appartenait à Arès, le dieu de la guerre, et celui-ci les tua tous. Cadmos découvrit ce qu'avait fait la terrible créature et parvint à venger ses compatriotes mais fut désolé de se retrouver seul. Athéna apparut alors à Cadmos et lui conseilla de planter dans le sol les dents du dragon, ce que fit le prince phénicien. Des hommes armés surgirent alors de l'endroit où les dents du dragon avaient été semées et ils se combattirent les uns les autres. Seuls les cinq meilleurs guerriers survécurent, épargnés par Arès en jouant au dé, et ces champions appelés les Spartes (c'est-à-dire « les hommes semés ») devinrent les nouveaux compagnons de Cadmos, les autochtones à l'origine des cinq grandes familles, et l'aidèrent à bâtir sa cité : Thèbes.

L'apprentissage par la compétition, ou comment devenir un champion sans aucune connaissance humaine

Depuis que le mathématicien Alan Turing avait présenté, au milieu du XX^e siècle, son célèbre « jeu de l'imitation » proposant un protocole destiné à répondre à la question philosophique de savoir si les machines peuvent ou non penser [23], de nombreux chercheurs en intelligence artificielle ont employé le jeu comme terrain d'expérimentation. Pour Herbert Simon, le jeu d'échecs permettait d'observer et d'analyser les comportements de prise de décision des joueurs. C'est en 1997, lors de la revanche d'une compétition d'échecs contre celui qui était considéré comme le meilleur joueur de l'époque, Garry Kasparov, que le superordinateur d'IBM *Deep Blue* avait montré qu'une machine était capable de battre l'être humain le plus fort dans ce domaine. On s'était alors mis à reconsidérer la difficulté du jeu d'échecs, et on était parvenu à admettre qu'une machine capable d'évaluer 200 millions de positions par seconde pouvait dépasser les capacités de réflexion d'un homme.

On avait alors vu dans le jeu de go, avec son

nombre de parties possibles immensément plus important que celui du nombre de parties du jeu d'échecs, un nouveau seuil dans la difficulté à résoudre des problèmes. Lorsque l'entreprise britannique DeepMind, spécialisée dans la combinaison des techniques d'apprentissage automatique par réseaux de neurones profonds et d'apprentissage par renforcement, s'est faite racheter par Google/Alphabet, le géant américain lui avait fourni la puissance de calcul nécessaire pour développer un programme de jeu de go baptisé *AlphaGo* et, en octobre 2015, *AlphaGo* était parvenu à battre, dans des parties sans handicap sur un goban de taille normale, le champion européen de go Fan Hui. En mars 2016, dans une compétition diffusée et commentée en direct sur Internet, ce même programme était parvenu à battre le joueur sud-coréen Lee Sedol, considéré comme le meilleur joueur du monde de la décennie 2000, et en mai 2017, il avait réussi à battre le champion du monde en titre Ke Jie. Cette démonstration de force montra que les algorithmes tels que ceux présents dans *AlphaGo* étaient en mesure d'aborder des problèmes nécessitant de fouiller d'énormes quantités de solutions, comme le difficile problème du repliement des protéines.

L'algorithme d'*AlphaGo* combine des techniques d'apprentissage automatique et de parcours de graphe, associées à de nombreux entraînements avec des humains, avec d'autres ordinateurs, et en faisant des parties contre lui-même. En octobre 2017, DeepMind a même présenté une version améliorée d'*AlphaGo* appelée *AlphaGo Zero* [21]. Dans cette version, le programme n'avait pas réalisé d'apprentissage à partir de données provenant de parties jouées entre humains, ni de connaissances autres que les règles du jeu. Cette version était devenue très rapidement plus forte que n'importe quelle version précédente simplement en faisant jouer le programme contre lui-même. Deux mois plus tard sortait *AlphaZero*, une version généraliste d'*AlphaGo Zero*, surpassant non seulement *AlphaGo Zero* en très peu de jours d'apprentissage mais parvenant aussi à battre des programmes de haut niveau dans d'autres jeux : les échecs ou le shōgi.

Ainsi, sans aucune connaissance humaine, ces programmes de jeux sont nés pour se combattre et la sélection par les combats n'a laissé survivre que les versions les plus compétitives et les plus efficaces de ces programmes d'intelligence artificielle, tout comme des dents du dragon plantées en terre se sont dressés des hommes qui se sont affrontés pour ne laisser comme survivants que des champions se mettant au service de Cadmos.

Cadmos fut obligé de servir le dieu Arès pendant 8 ans car le dragon était sa progéniture. Mais cette période écoulée,

Cadmos reçut d'Athéna le royaume de Thèbes et de Zeus une épouse : Harmonie, fille d'Arès et d'Aphrodite. Les dieux, présents au mariage, firent divers présents aux jeunes époux, dont Héphestos qui offrit un collier que Cadmos remit à sa femme. Le collier d'Harmonie était cependant un collier maléfique. En effet, Héphestos est un dieu inventeur capable de créer des objets magiques. Or Harmonie est la fille de son épouse Aphrodite, déesse de l'Amour, et d'Arès, dieu de la guerre. Elle personnifie de manière évidente l'infidélité de son épouse. C'est ainsi que le collier d'Harmonie jouera un rôle fatal dans l'histoire de Thèbes car, en même temps qu'il assurera à la personne qui le porte la beauté d'une jeunesse éternelle, il lui portera aussi malheur. Ce collier sera transmis d'Harmonie à son fils Polydore, qui lui-même le transmettra à son fils Labdacos, Labdacos donnera le collier à son fils Laïos, Laïos l'offrira à sa femme Jocaste, puis le collier reviendra à Polynice, fils de Jocaste, et Polynice l'offrira à son épouse Argie.

Le collier d'Harmonie et les conflits de valeurs

Le bijou offert par le dieu Héphestos possède un double pouvoir : il apporte certes beauté et jeunesse éternelles à qui le porte, mais il est aussi source de malheur.

Lorsque l'intelligence artificielle est intégrée dans des systèmes de recommandation afin de rechercher, trier et proposer à un utilisateur du contenu sur une application numérique (que ce soit un moteur de recherche pour trouver de l'information sur le *web*, un produit ou un service sur un site *web* marchand, des informations sur un journal en ligne, des musiques ou vidéos sur des sites de diffusion en *streaming*, du contenu sur des réseaux sociaux), le critère que cherchera à optimiser ces systèmes intelligents est la pertinence, c'est-à-dire une recherche personnalisée des éléments les plus proches des centres d'intérêt déjà identifiés de l'utilisateur. Cependant, en faisant cela, les systèmes de recommandation vont procéder de manière statistique en se focalisant sur les tendances majoritaires, ce qui va avoir pour effet de proposer le contenu le plus populaire, avec une tendance à tirer l'utilisateur vers le bas, en réduisant fortement le champ des possibles pour ne proposer plus qu'une toute petite fenêtre ouverte sur le monde, à la manière d'une « bulle filtrante » [15]. Les contenus proposés par ces systèmes seront ainsi attractifs, comme la jeunesse et la beauté éternelles apportées le collier d'Harmonie, mais apporteront le malheur en allant contre les intérêts des utilisateurs en leur cachant la diversité du monde, que ce soit au niveau informationnel ou culturel.

Cadmos et ses compagnons fonderont la grande cité de Thèbes. De sa contrée natale, Cadmos apportera l'alphabet phénicien dont sera dérivé l'alphabet grec.

L'intelligence artificielle a connu un saut qualitatif majeur au cours des décennies 2000 et

2010 avec une augmentation importante des performances des systèmes d'IA. Elle est l'une des composantes majeures de ce que l'on appelle la « révolution numérique », cette nouvelle forme de travailler avec l'écrit, entamée dans les années 1980 avec la diffusion de l'ordinateur personnel, dans les années 1990 avec l'explosion du phénomène Internet et la mise en place du *web*, et dans les années 2000 avec l'apparition du *web 2.0* et du smartphone facilitant l'accès aux informations avec un ordinateur tenant dans la main et pouvant être utilisé pratiquement partout sur Terre. Cette évolution technologique, par la traduction de toutes choses sous forme de nombres [1], a changé les rapports de force dans notre société car il devient désormais plus important de posséder et savoir traiter l'information que de détenir les biens physiques qui nous entourent [2]. Avec l'hyper-connectivité apportée par la révolution numérique, notre manière de voir le monde et d'interagir avec lui se trouve radicalement transformée par rapport aux générations qui nous ont précédées [20]. Il en résulte que l'intelligence artificielle engendre ainsi des changements d'ordre économique, mais cette technologie a aussi des impacts inédits dans notre vie de tous les jours [8] en touchant aux sphères psychologique et sociale, redéfinissant notre rapport aux autres, au monde du travail, au droit, à la culture, à la santé, au savoir et à la prise de décision.

2.2 La descendance de Cadmos et la malédiction des Labdacides

Le règne de Cadmos fut brillant, la cité de Thèbes se développa et, avec son épouse Harmonie, ils eurent plusieurs enfants, dont un fils nommé Polydore. Lorsque Polydore succéda à son père sur le trône de Thèbes, son règne ne dura pas longtemps. Il épousa Nyctéis, fille de Nyctée, qui lui donna un fils nommé Labdacos. Sentant qu'il allait bientôt mourir, Polydore recommanda le royaume et son fils Labdacos à Nyctée, son beau-père, et quand celui-ci mourut à son tour, la régence revint à Lycos, le frère de Nyctée.

Labdacos, dont le nom signifie « le boiteux », est l'ancêtre éponyme de la lignée des Labdacides. Devenu adulte, Labdacos succéda à son père sur le trône. Afin de garantir les frontières du royaume de Thèbes, il engagea une guerre contre Pandion, roi d'Athènes, mais il fut vaincu car Térée, roi de Thrace, vint se battre avec son armée aux côtés des Athéniens.

Âgé seulement d'un an au décès de son père, Laïos (le « gaucher »), le fils de Labdacos, ne put monter sur le trône de Thèbes et la régence échut à nouveau à Lycos, son arrière grand-oncle. Cette période fut fortement troublée, avec beaucoup de violence. Lorsque Laïos parvint à la majorité, il ne put monter sur le trône de Thèbes car il avait été occupé par les jumeaux Amphion et Zéthos, fils d'Antiope et petits-neveux de Lycos, qui, pour venger leur mère maltraitée par Lycos, avaient assassiné leur grand-oncle. Les

deux frères fortifièrent la ville en érigeant autour d'elle un rempart percé de sept portes. Laïos dut s'enfuir dans le Péloponnèse où il trouva refuge auprès de Pélops, à la cour de Pise, non loin d'Olympie, en Élide.

Pélops est le fils de Tantale, roi de Phrygie et ami des dieux. Enfant, Pélops avait été tué par son père qui, pour tester l'omniscience des dieux de l'Olympe qui le recevaient à leur table divine, leur avait servi des morceaux de son fils lors d'un banquet. Les dieux, qui avaient découvert la supercherie, avaient ressuscité Pélops et condamné Tantale au célèbre supplice consistant à l'affamer et lui présenter de la nourriture pour la lui retirer au moment où il pensait enfin pouvoir assouvir sa faim. Devenu adulte, Pélops avait pu obtenir la main d'Hippodamie après avoir battu dans une course de chars son père Enomaos, le roi de Pise en Élide. Selon certaines versions du mythe, cette course aurait été gagnée parce qu'il avait fait saboter le char du roi en remplaçant une pièce par un morceau de cire qui fondit durant l'épreuve, ce qui provoqua la dislocation du char et la mort du roi, entraîné par ses chevaux. C'est ainsi que Pélops devint le nouveau roi de Pise et qu'il pu épouser Hippodamie. Il est l'ancêtre des Atrides à Mycènes et il donna son nom à la péninsule de Grèce du Péloponnèse.

L'efficacité des machines et leurs doubles usages

Dans la mythologie grecque, quand il y a un trouble dans l'ordre cosmique, il faut des générations pour le réparer. De plus, quand il y a une punition, un châtiment, quel qu'il soit, il a toujours un rapport symbolique avec le crime. Or le symbole, étymologiquement *symbolon*, est un signe de reconnaissance constitué d'une pièce d'argile séparée en deux parties dont les deux morceaux s'ajustent parfaitement.

Le char est un véhicule attelé à un ou plusieurs animaux pourvu de roues assemblées par paire via un essieu. Le char, qui permet de se déplacer plus vite, a facilité les échanges, le commerce, et il a été l'une des premières machines employées par l'Homme. Au fil des siècles, les améliorations technologiques ont rendu les chars encore plus efficaces, comme l'évolution des roues qui, initialement, étaient faites de bois plein et qui se sont transformées en roues à rayons et à jante, réalisées en bois cerclé de bronze, ce qui les rendaient à la fois bien plus légères et plus stables.

Toutes machines et toutes technologies, comme l'intelligence artificielle, connaissent un double usage. L'efficacité du char a provoqué son utilisation aussi bien dans le monde civil que dans le monde militaire. L'invention du char a ainsi marqué l'histoire militaire en permettant de réaliser des attaques très rapides qui terrorisaient les ennemis.

L'arrivée au pouvoir de Pélops étant associée à une ruse employée durant une course de chars, il sera frappé de malheur au niveau de sa descen-

dance à cause d'un char, et l'homme qui avait été l'instrument de ce coup du sort connaîtra à son tour une mort associant à la fois les éléments symboliques du char et de la descendance.

Le roi Pélops accueillit Laïos, le prince de Thèbes. Laïos fut chargé par son hôte d'apprendre à son fils Chrysippe à conduire un char, une aptitude que devaient connaître tous les princes. Laïos tomba amoureux de son jeune élève mais cet amour ne fut pas réciproque. Des années plus tard, lorsqu'Amphion et Zéthos décédèrent, Laïos décida de rentrer à Thèbes pour y régner et il enleva Chrysippe sur son char et le viola. Le malheureux Chrysippe, après ce crime commit contre sa personne, se suicida par pendaison et son père Pélops, dévasté par le chagrin et la colère, lança une imprécation en demandant aux dieux que la lignée des Labdacides soit vouée à l'anéantissement.

L'origine des biais algorithmiques

Il existe de nombreuses similitudes entre la façon de se déplacer dans l'espace et la manière dont un code informatique suit son cours. Un programme informatique suit un cheminement représenté par un algorithme traduit en code dans un certain langage de programmation, avec des chemins qui sont empruntés plutôt que d'autres en fonction de certaines conditions (les tests) et des chemins qui sont repris à de nombreuses occasions (les boucles). L'instruction « run » que tout programmeur exécute pour faire fonctionner son programme signifie littéralement « courir », un programme peut « marcher » ou non, et il peut être exécuté dans son ensemble ou suivant un mode « pas à pas ».

Le nom des Labdacides est à rapprocher de la lettre grecque λ (*lambda*), en grec ancien « labda », une lettre qui est penchée. Labdacos et sa descendance sur trois générations seront marqués par « un déséquilibre de la démarche, un manque de symétrie entre les deux côtés du corps, un défaut à l'un des deux pieds. Labdacos, c'est le boiteux, celui qui n'a pas les deux jambes pareilles, de même taille ou de même force ; Laïos, le dissymétrique, le tout gauche, le gaucher ; *Oidipous*, celui qui a le pied enflé » [26]. Il y a donc un problème qui frappe cette lignée qui ne marche pas droit, qui est penchée, qui présente donc un biais. De façon symbolique, la question d'un problème lié aux pieds concerne le rapport à la terre, c'est-à-dire l'ancrage dans la réalité.

Selon diverses sources, le mot « biais » est probablement arrivé en français à partir de l'ancien provençal *biais*, via le latin vulgaire (*e*)*bigassius* depuis le grec *epikarsios*, signifiant « en biais, en angle » à partir de *epi* « sur » et *karsios* « oblique », et cela depuis les colonies grecques de la côte provençale [3]. Ainsi, analogie et étymologie permettent de rapprocher la notion de biais, que l'on retrouve dans les programmes

d'intelligence artificielle, avec la lignée « maudite » des Labdacides.

Popularisés par Cathy O'Neil [14] qui avait relaté différents cas où l'emploi des systèmes artificiels reposant sur des bases mathématiques avait pour effet d'amplifier les inégalités sociales et constituait une véritable menace pour la démocratie, les biais algorithmiques peuvent être la conséquence de la sur- ou de la sous-représentation de certaines catégories d'exemples parmi les jeux de données servant à l'apprentissage (biais statistiques), par le fait – pour les concepteurs des programmes – d'avoir des préjugés dans la manière de tenir compte de tels ou tels paramètres dans la mise au point de leurs modèles (biais cognitifs), ou de se focaliser sur le seul profit à réaliser comme variable à optimiser dans le modèle, aux dépens d'autres valeurs (biais économiques). Les biais algorithmiques sont ainsi des problèmes présents parmi les programmes d'intelligence artificielle qui vont avoir des conséquences malheureuses pour les êtres humains, à la manière dont les Labdacides et leurs manières de ne pas « marcher droit » vont avoir des penchants funestes pour eux-mêmes, leurs descendances, leurs entourages et les cités dont ils seront les maîtres.

3 Œdipe, ou l'incarnation des programmes d'intelligence artificielle

3.1 *Οιδιπους*, le prince aux pieds enflés

Arrivé à Thèbes, Laïos devint roi et épousa Jocaste à qui il offrit le collier d'Harmonie, son arrière-grand-mère. Bien qu'ayant des penchants pour les garçons et jeunes hommes, Laïos partageait la couche de Jocaste mais ils ne parvenaient pas à avoir d'enfant. Laïos alla alors consulter l'oracle à Delphes pour savoir ce qu'il devait faire afin d'avoir un fils, un héritier, destiné à conforter sa situation sur le trône. L'oracle lui annonça alors une terrible prophétie : s'il devenait père d'un héritier mâle, celui-ci tuerait son père et épouserait sa mère. Inquiet, Laïos rentra à Thèbes et se garda d'avoir des relations charnelles avec sa femme, se contentant de la compagnie d'éphèbes. Cependant après une soirée où il s'était enivré, Laïos retrouva le lit de sa femme et Jocaste accoucha quelques mois plus tard d'un garçon.

Pour empêcher la réalisation de la prophétie de l'oracle, le roi et la reine confièrent le bébé à un serviteur chargé de l'exposer, c'est-à-dire que leur nouveau-né serait abandonné aux dieux dans un endroit où il serait livré à la merci des forces naturelles. Le serviteur se rendit jusqu'au mont Cithéron où il perça les chevilles du bébé pour y passer une corde et l'accrocher à un arbre. Un berger de Corinthe, témoin de la scène, délivra le nouveau-né qui survécut mais qui, à la suite de cette mutilation, garda les pieds enflés, d'où lui vint son nom : Œdipe. Le berger rentra dans son pays et confia le bébé à Polybe, le roi de Corinthe, qui se

désolait de ne pas parvenir à avoir d'enfant.

Polybe et Mérope, le roi et la reine de Corinthe, élevèrent Œdipe comme leur fils sans jamais lui dire qu'il avait été adopté. Avec les années, Œdipe devint beau, grand, fort et intelligent, il faisait la joie de ses parents et il vécut heureux en tant que prince héritier de Corinthe. Lors d'un banquet, cependant, Œdipe fut troublé par un homme éméché qui l'insulta et le traita d'étranger indigne du trône de Corinthe. Œdipe alla trouver son père qui lui indiqua de ne pas se préoccuper de la parole d'un ivrogne. Comme cette réponse ne parvenait pas à le satisfaire, il monta sur son char et se rendit auprès de l'oracle de Delphes afin connaître la vérité. La prophétie ne lui révéla néanmoins rien sur ses origines. Elle lui annonça par contre qu'il tuerait son père, épouserait sa mère et laisserait une descendance honteuse.

Œdipe prit peur et, pour que cette tragédie ne puisse pas se réaliser, il décida de ne pas retourner à Corinthe pour éviter de tuer son père bien-aimé et de prendre sa mère pour épouse. Alors qu'il errait sur les routes de Grèce, il rencontra un jour à un carrefour situé près du mont Parnasse un homme sur un char accompagné de ses serviteurs. La route étant étroite, les deux chars ne pouvaient pas se croiser. L'homme qui faisait face à Œdipe lui demanda de s'écarter car il était très pressé. Chaque équipage se sentant prioritaire, aucun des deux ne voulut s'écarter pour laisser passer l'autre et une querelle se déclencha : après les mots, ils en vinrent aux mains, puis aux armes, et bien qu'en infériorité numérique, Œdipe parvint à tuer l'homme et ses serviteurs. Parmi l'équipage qui s'était opposé au prince de Corinthe, il n'y eut qu'un survivant qui se précipita chez lui, à Thèbes, afin d'annoncer la terrible nouvelle : le roi Laïos avait trouvé la mort sur la route, tué par des voleurs. En effet, l'homme que venait de tuer Œdipe n'était autre que son père biologique qui se déplaçait incognito avec une faible escorte, or les deux hommes ne se connaissaient pas : Laïos imaginait son fils tué par son serviteur et Œdipe se croyait l'enfant des roi et reine de Corinthe.

Le discours de Laïos et les textes produits par l'intelligence artificielle générative

Le roi Laïos portait sur lui la malédiction lancée par Pélopos à la suite de l'enlèvement et du viol de son fils Chrysippe. Afin de sauver la cité de Thèbes que terrorisait un monstre, il souhaitait se rendre à Delphes pour y consulter l'oracle accompagné de quelques gardes. Lorsque son équipage voulut passer sur cette route étroite et ne pas céder la priorité à Œdipe, il prononça un discours sans expliquer ni qui il était ni les raisons de son empressement, ce qui ne parvint pas à convaincre l'orgueilleux prince de Corinthe de lui laisser la place. La dispute entre les deux hommes dégénéra et ils finirent par se battre jusqu'à la mort.

L'intelligence artificielle générative (ou « gen AI ») désigne des programmes d'IA capables de produire des contenus originaux (qu'il s'agisse de texte, d'images, de vidéo, d'audio ou de code informatique) en réponse à l'invite ou à la re-

quête d'un utilisateur. Pour cela, l'IA générative se base sur des modèles d'apprentissage profond (*deep learning*) réalisés à partir d'énormes quantités de données et exploite les informations issues de ces modèles pour parvenir à répondre, au moyen d'un procédé appelé « mécanisme d'attention » qui permet de traiter efficacement les séquences – et donc les suites de mots dans des phrases – afin de parvenir à répondre de façon pertinente à des questions posées en langage naturel. En 2022, la société OpenAI a mis en ligne *ChatGPT*, un agent conversationnel basé sur cette technologie, engageant les « géants de la Tech » (Google/Alphabet, OpenAI, Baidu, Facebook/Meta, Amazon et d'autres) à se lancer dans une fuite en avant pour produire leurs propres grands modèles de langage dans une compétition acharnée afin de proposer des modèles toujours plus puissants manipulant une quantité de paramètres gigantesque.

Cependant les productions réalisées par cette technologie sont issues de propriétés statistiques de relations entre les mots, ce qui permet de produire des textes dont la structure est correcte au niveau syntaxique (l'ordre des mots suit les règles de grammaire de la langue), ainsi qu'au niveau sémantique (les phrases exprimées ont du sens et elles donnent des résultats plausibles), mais qui ne donne aucune garantie au niveau pragmatique (au niveau où le sens ne peut être compris qu'en connaissant le contexte de leur emploi).

En effet, les textes produits ne sont pas issus de bases de données qualifiées, il s'agit du résultat d'assemblages de termes issus de relations statistiques entre les mots, et ce type de relation ne garantit en rien la valeur de vérité des énoncés produits. Il en résulte ainsi des « hallucinations », c'est-à-dire des productions fausses ou absurdes, que les utilisateurs peuvent parfois considérer comme étant justes.

Les productions de textes issues de l'IA générative, par leur manque d'« inspiration », ont des points communs avec les rédactions des candidats au concours d'entrée de l'École (impériale) Polytechnique au début du XIX^e siècle. En plus d'épreuves de mathématiques et de physique, les candidats avaient à réaliser une composition de français qui consistait à imaginer la réponse de Laïus (nom latinisé de Laïos) à Œdipe lorsqu'ils s'étaient croisés sur le chemin. Les étudiants, peu inspirés, avaient rédigé des pages de discours pompeux et sans intérêt, des textes particulièrement rébarbatifs à corriger. Depuis lors, le mot *laïus* est resté dans la langue française pour décrire tout discours particulièrement long, creux et ennuyeux, à la manière dont on peut souvent qualifier des textes issus d'IA générative.

À noter enfin que ces productions d'IA générative peuvent représenter des menaces si elles sont mal employées et peuvent avoir des conséquences funestes, tout comme le discours de Laïos – autrement dit, le « laïus » – s'était soldé par un échec : Œdipe ne s'était pas écarté de la route et Laïos s'est fait tuer à la suite de cette rage routière.

Après ce malheureux événement dont il sortit victorieux, Œdipe poursuivit son errance en Grèce et il se rendit en direction de Thèbes quand il apprit qu'un monstre faisait régner la terreur dans la Cité. Créon, le frère de la reine Jocaste, devint le nouveau régent et il promit le trône de Thèbes et la main de sa sœur à celui qui parviendrait à débarrasser les Thébains du Sphinx qui posait une énigme à ceux qu'il rencontrait et qui étaient dévorés s'ils ne parvenaient pas à trouver la solution. Selon les auteurs antiques, le Sphinx avait été envoyé par Héra contre Thèbes dont il ravageait les champs et terrorisait les populations pour punir la cité du crime de Laïos qui avait aimé Chrysippe, le fils de Pélopes, d'un amour coupable ayant provoqué sa mort. Le Sphinx, appelé aussi « la Sphinge », était une créature chimérique ayant un corps de lion, des ailes d'aigle, et une tête et une poitrine de femme. L'énigme qu'il posait à ceux qui le rencontraient était la suivante : « Quel est l'animal qui n'a qu'une voix, et qui marche à quatre pattes le matin, deux pattes à midi et trois le soir ? » Œdipe arriva auprès du Sphinx et lui expliqua que l'animal dont il parlait était l'homme qui est quadrupède au matin de sa vie, se traînant sur ses pieds et sur ses mains quand il est petit enfant, puis est bipède à l'âge adulte, et qui devient enfin tripède lorsque, au soir de sa vie, il est obligé de prendre un bâton pour se soutenir. Le Sphinx, surpris que quelqu'un parvint à résoudre cette énigme impossible, se jeta alors du haut d'une falaise.

Les décisions de l'intelligence artificielle, ou résoudre des problèmes sans les comprendre

Nous avons vu plus haut que les programmes informatiques présentent des analogies avec la façon de se déplacer dans l'espace. La rencontre entre Œdipe et le Sphinx est l'incarnation même de l'intelligence artificielle et de ses limites.

En effet, Œdipe n'est pas capable de comprendre la réponse donnée même s'il trouve la bonne solution. L'énigme du Sphinx porte sur la façon de marcher, or Œdipe, par son apprentissage, a déjà fait l'expérience de la solution : bien que jeune adulte, il a déjà vécu l'ensemble de ces modes de déplacements puisque, à la suite de la mutilation subie au niveau des pieds alors qu'il était nouveau-né, il éprouve des problèmes à se déplacer et a besoin d'une canne pour marcher, comme s'il était déjà une personne âgée.

Ensuite Œdipe est capable d'indiquer que la solution est l'homme alors que lui-même ne comprend pas sa propre nature humaine : il ne connaît pas ses origines véritables. De cette incapacité à comprendre réellement qui il y est découlera

entre Œdipe et Jocaste un rapport qui cumulera les trois âges de la vie : Œdipe est à la fois le bébé de Jocaste, son mari et le père de ses enfants, et la personne qui vieillira à ses côtés alors que Jocaste, portant le collier d'Harmonie que lui avait offert Laïos, gardera sa jeunesse et sa beauté jusqu'à sa mort.

3.2 Οιδιπους τυραννος, le roi de Thèbes

Ayant terrassé le Sphinx grâce à son intelligence, Œdipe entra en triomphe dans la cité de Thèbes, libérée de ce fléau. Le régent Créon lui remit la couronne du royaume et lui donna la main de la reine Jocaste, sa sœur, plus âgée qu'Œdipe mais ayant conservé la beauté de sa jeunesse grâce au collier d'Harmonie. Ainsi, sans qu'Œdipe le sut, la prophétie de l'oracle était en train de s'accomplir.

Pendant deux décennies, Œdipe fut roi de la cité de Thèbes. Bien que « tyran », c'est-à-dire qu'il n'était pas soumis à la loi des hommes, il fut un souverain bienveillant et apprécié du peuple de Thèbes.

La gouvernamentalité algorithmique

Les programmes informatiques n'ont pas de conscience. Ils n'ont ainsi pas de notion de la responsabilité.

En arrivant à Thèbes, Œdipe voit la société se délester du poids de la responsabilité sur sa personne, tout comme notre société éprouve une certaine défiance vis-à-vis de la classe politique et peut être tentée par la gouvernamentalité algorithmique, imaginant que des décisions issues de programmes d'intelligence artificielle seraient toujours justes et incorruptibles.

C'est en raison de son savoir qu'Œdipe accède au pouvoir, malgré l'ignorance de sa responsabilité dans la mort de Laïos, l'ancien roi. Les liens entre savoir et pouvoir ont été interrogés de la même manière par Antoinette Rouvroy et Thomas Berns au sujet de l'enregistrement massif et systématique des données (qui est à la base de la puissance de l'intelligence artificielle) et de ses conséquences sur la façon de gouverner. Selon eux, « la *gouvernamentalité algorithmique* [...] confirme [...] l'indistinction radicale entre modes de production du savoir et modalités d'exercice du pouvoir, dans la mesure précisément où ce nouveau mode de gouvernement se présente à la fois comme un ensemble d'interfaces cognitives et comme moyen de prédire, de prévenir et d'orienter certains comportements humains » [18].

Créon remet à Œdipe le pouvoir, tout comme nous suivons parfois aveuglément les décisions issues de programmes d'intelligence artificielle et nous remettons à l'IA des responsabilités qu'elle n'est pas à même d'assurer, en nous reposant sur une prétendue objectivité des données. Or, comme l'explique Rouvroy, « les algorithmes

ne “gouvernent” pas. Ils sont programmés, mais n’ont pas de programme politique, ne font pas de projet, ne font pas de choix : ils optimisent l’état de fait » [17]. Ce mode de fonctionnement a des conséquences importantes sur la façon dont se prennent les décisions. Ainsi, les performances de certains systèmes informatiques intelligents reposent sur le principe de l’apprentissage profond, or le modèle à la base de ce principe est issu de la détection de corrélations entre les nombreux exemples ayant servi à l’apprentissage et la solution à trouver. Pourtant, la détection de corrélations dans le monde réel n’est pas suffisante pour fournir une source d’explications, ce qui peut avoir des conséquences néfastes pour la société. Malgré cela, certains pays comme la Chine n’ont pas hésité à franchir le pas en développant des « robots juges » avec l’adoption d’une intelligence artificielle en tant que procureur au tribunal, alors que ce système informatique n’est pas en mesure de fournir des justifications aux décisions qu’il a prises.

Œdipe fut un bon mari pour Jocaste et eut d’elle quatre enfants, deux garçons nommés Étéocle et Polynice et deux filles nommées Antigone et Ismène.

Au bout d’une vingtaine d’années, toutefois, la cité de Thèbes fut marquée par un nouveau fléau : une épidémie de peste ravageait la ville. Cette épidémie se manifestait par l’apparition de choses effroyables touchant à la vie et à la fécondité : les fausses couches étaient nombreuses ou, quand les grossesses étaient menées à jusqu’à leurs termes, les femmes accouchaient d’enfants mort-nés ; les plantes ne poussaient plus ; les animaux venaient au monde avec de monstrueuses difformités...

Œdipe chargea alors son beau-frère Créon d’aller consulter l’oracle à Delphes. Créon en revint avec une explication : l’épidémie de Thèbes était le résultat d’une souillure, d’une pollution. Le crime du roi Laïos était resté impuni et l’épidémie ne quitterait la ville qu’avec la mort ou le bannissement du meurtrier du précédent roi.

Le remplacement et le renversement de l’autorité

Lorsque Œdipe croise son père sur la route reliant Delphes à Thèbes, se dispute avec lui et le tue, symboliquement, il bafoue sans le savoir deux formes d’autorité. Tout d’abord, en tuant son père Laïos, Œdipe commet un parricide, c’est-à-dire qu’il remet en question l’autorité paternelle, autrement dit l’ordre au sein de la cellule familiale. Ensuite, comme son père est également le roi de Thèbes, Œdipe est aussi régicide, c’est-à-dire qu’il annihile le représentant du pouvoir sur les hommes, Laïos étant le maître de la cité. Dans la légende, non seulement Œdipe n’est pas inquiet par ses crimes, ignorant lui-même la gravité des actes qu’il a commis, mais il est fêté à Thèbes comme un héros et il va remplacer l’ancien roi

à la fois sur le trône et dans le lit de Jocaste : il prend la place de Laïos à tous les niveaux : il devient le nouveau maître de la Cité et le chef d’une famille comportant quatre enfants.

Cette forme de remplacement se rencontre de la même manière avec l’intelligence artificielle. Nous assistons à l’occupation de l’IA dans de nouveaux domaines, ce qui nous procure, souvent non sans raison, un certain sentiment de malaise associé à la crainte de la voir occuper des postes occupés par les êtres humains et d’être relégués à des positions secondaires, que ce soit dans la vie de la cité, par des activités professionnelles dont le champ d’application est bouleversé par l’IA, mais aussi au sein de notre intimité car l’IA s’invite dans nos loisirs ou nos sources d’information, en modulant notre rapport au monde, que ce soit à travers les recommandations apportées au sein des réseaux sociaux, des moteurs de recherches, ou des contenus diffusés sur les plateformes numériques.

Après avoir appris les propos de Créon, Œdipe se mit à rechercher le meurtrier du roi Laïos. Soucieux du bien de sa Cité, il mena son enquête en envoyant ses meilleurs limiers explorer toutes les pistes possibles, mais sans succès, le crime ayant été accompli depuis 20 ans. Il fit alors annoncer partout que si le meurtrier se dénonçait, il ne serait pas inquiété mais juste condamné à l’exil, ce qui n’eut pas plus de résultat. Il s’en prit alors à Tirésias, le devin aveugle, qui semblait lui cacher des choses, et il le menaça. Tirésias, ne cédant aux demandes du roi que sous la contrainte, lui apprit que c’était lui-même, Œdipe, qui était le meurtrier du roi. Œdipe ne put croire Tirésias car il lui avait toujours été dit que des voleurs s’en étaient pris à Laïos et à son équipage, et il imaginait que cela n’était qu’un stratagème pour l’évincer du pouvoir et laisser le trône à son beau-frère. Œdipe accusa alors Créon qui se justifia en expliquant qu’il ne voulait pas de la couronne : aux côtés d’Œdipe, il bénéficiait déjà de tous les avantages de la royauté sans avoir à assumer les lourdes charges d’un roi.

Un peu plus tard, on annonça à Œdipe qu’une délégation arrivait de Corinthe. Œdipe reçut les ambassadeurs qui lui annoncèrent la mort du roi Polybe et qui lui proposaient de revenir dans la cité de son enfance afin de succéder à son père.

Attristé par cette nouvelle, Œdipe était néanmoins heureux d’apprendre qu’il n’avait pas été à l’origine du décès de Polybe qu’il imaginait être l’auteur de ses jours et ainsi avoir échappé à la prophétie de l’oracle de Delphes. Le roi de Thèbes refusa cependant l’offre des Corinthiens car il gouvernait déjà un royaume et il avait surtout peur de retourner à Corinthe et devoir épouser la reine Mérope, celle qu’il croyait être sa mère. Un vieil ambassadeur voulut alors le rassurer : il lui expliqua que, plus jeune, il était un berger, et il avait fait paître ses moutons sur le mont Citérion où il avait trouvé Œdipe exposé à un arbre alors qu’il n’était encore qu’un nouveau-né. Œdipe était ainsi adopté, Polybe

et Mèrope n'étaient pas ses parents biologiques...

Le messager de Corinthe reconnut alors parmi les serviteurs d'Œdipe l'homme qui avait abandonné le nouveau-né. Le serviteur prit peur et chercha à s'enfuir mais Œdipe l'arrêta et l'obligea à rapporter les faits qui s'étaient produits et d'indiquer qui étaient les parents qui avaient voulu se débarrasser de leur enfant. Témoin de la scène, Jocaste eut les larmes aux yeux et s'enferma dans une pièce de son palais. Le serviteur, sous la menace, confessa que les parents n'étaient nul autres que le roi Laïos et la reine Jocaste. Œdipe comprit alors l'insoutenable vérité : conformément à la prophétie, il avait bien tué son père et épousé sa mère.

La transgression des interdits civilisationnels

Pour l'anthropologue Claude Lévi-Strauss, la prohibition de l'inceste marque le passage de la Nature à la Culture [12]. Les interdits, comme celui de ne pas prendre sa mère pour épouse, constituent des règles à la base des civilisations humaines. Associés aux interdits se trouvent les lois, régissant les relations entre les hommes, ainsi que le phénomène religieux, avec la distinction entre le sacré et le profane : « les choses sacrées sont celles que les interdits protègent et isolent ; les choses profanes, celles auxquelles ces interdits s'appliquent et qui doivent rester à distance des premières » [7].

Œdipe parvient à lui seul à cumuler le viol de trois interdits : il a tué son père, il a tué son roi et il épousé sa mère. Ramené aux systèmes d'intelligence artificielle, ce destin tragique remet en perspective l'emploi de l'IA : est-ce que cette technologie, avec les pouvoirs qui lui sont délégués, ne risque pas de remettre en cause les bases mêmes de la civilisation ?

Horrifié, Œdipe alla retrouver Jocaste mais, lorsque la porte du palais de sa chambre put être ouverte, il découvrit que son épouse et mère s'était tuée par pendaison. Hurlant de douleur, Œdipe coupa la corde, prit les broches du vêtement de Jocaste et les utilisa pour se crever les yeux.

La quête de la vérité et l'interprétabilité des systèmes intelligents

Lorsque Œdipe prend conscience que la prophétie qu'il cherchait à tout prix à éviter s'était pourtant bien déroulée, il semble voir la lumière pour la première fois, mais apprendre la vérité est si difficile à accepter qu'il refuse cette lumière et va se mutiler pour ne plus rien voir désormais. Sophocle lui fait dire : « Hélas, hélas ! Tout apparaît clairement. Ô lumière, je te vois pour la dernière fois, moi qui suis né de ceux dont il ne fallait point naître, qui me suis uni à qui je ne devais point m'unir, qui ai tué qui je ne devais point tuer ! »

Au moment de les accomplir, Œdipe n'était pas conscient de la gravité de ses actions car il avait agi en état de légitime défense sur la route entre

Delphes et Thèbes à la suite de cette dispute qui avait dégénéré avec cet homme qu'il ne savait pas être son père, le roi Laïos, et il ignorait aussi que son épouse Jocaste était sa mère (pour revenir à ce que disait Jean-Pierre Vernant [24], Œdipe n'a donc pas de « complexe d'œdipe » au sens où l'avait théorisé Sigmund Freud [10]). Pourtant Œdipe est un homme intelligent, il est capable de trouver la solution à l'énigme du Sphinx et il était au courant de la prophétie de l'oracle de Delphes. Il était simplement incapable d'interpréter correctement ces informations.

De la même façon, les systèmes informatiques intelligents capables d'apporter une aide à la décision ne sont pas conscients des décisions proposées même s'il est possible de suivre la trace des raisonnements réalisés, comme c'est le cas dans les systèmes experts, les systèmes à base de règles. Dans le cas des modèles issus d'apprentissage profond, malgré les travaux encourageants menés actuellement dans le domaine de l'intelligence artificielle explicable (ou XAI), l'interprétabilité des résultats reste difficile, voire impossible, car elle se traduit sous la forme d'une quantité considérable de connexions plus ou moins fortes entre une multitude d'unités élémentaires reliées en réseaux, des neurones formels, ressemblant à la circulation des informations au sein du système nerveux, de neurones à neurones.

Mais l'analyse de ces unités élémentaires ne parvient pas à apporter du sens dans la globalité du modèle, tout comme Œdipe n'était pas parvenu à interpréter correctement le sens de la prophétie et s'en préserver en n'ayant que des informations partielles sur le destin qui lui était réservé.

3.3 L'exil et la fin de vie à Colone

Œdipe comprit qu'il représentait la souillure qui devait être expulsée de la ville et se bannit lui-même de la cité. Ses enfants le conduisirent jusqu'aux portes de la ville mais seule sa fille Antigone décida de l'accompagner. Les garçons Étéocle et Polynice, furieux de représenter le fruit d'un inceste, chassèrent leur père de Thèbes et réclamèrent le pouvoir suprême. Œdipe les mit alors en garde : s'ils humiliaient leur père, c'est l'épée à la main qu'ils devraient se partager son royaume.

Œdipe partit en errance dans la Grèce et finit par se réfugier à Colone, près de la cité d'Athènes, gouvernée par Thésée qui lui offrit asile. En accueillant Œdipe, Thésée cherchait à protéger son royaume car il avait appris que la terre sur laquelle serait enterrée la dépouille d'Œdipe serait bénie des dieux et connaîtrait la prospérité.

Au bout d'une année, Œdipe et Antigone furent rejoints par Ismène, l'autre fille d'Œdipe, qui vint leur donner des nouvelles. Ismène demanda à son père de revenir à Thèbes car ses deux frères Étéocle et Polynice se disputaient le pouvoir à Thèbes. Polynice, qui s'estimait injustement écarté du pouvoir par son frère aîné, souhaitait que son père puisse

trancher en indiquant quel était celui de ses fils qui était le plus digne de régner sur Thèbes. Cependant Œdipe refusa de prendre parti car il ne se sentait ni assez fort ni assez légitime pour intervenir encore dans les affaires des hommes. Puis Œdipe, sentant son heure venue, alla se purifier. Il demanda à ses filles de s'éloigner afin de ne rester qu'avec la seule compagne du roi Thésée à qui il révéla un secret, puis il disparut de façon merveilleuse, comme avalé par la terre et la fécondant par son expérience de la souffrance, ce qui apporta à Athènes la prospérité.

Les limites des systèmes intelligents

La fin de la vie d'Œdipe est marquée par la sagesse. On reconnaît en lui un pouvoir de décision mais il refuse d'en faire usage, il connaît ses limites. Œdipe considère que sa vie est arrivée à son terme, il confie une dernière information aux hommes, en la personne de Thésée, et livre son âme aux dieux, emmené par un envoyé divin ou accueilli par la terre. Cette mort fabuleuse apportera une bénédiction à Athènes, qui rayonnera pendant des siècles.

Œdipe, dont la vie a été marquée par la tragédie, se réconcilie avec l'univers à la fin de son existence et connaît une belle mort dont la bénédiction s'étend à la terre qui l'a accueilli. De la même façon, il ne faut utiliser l'intelligence artificielle que dans les domaines où elle est à même de s'appliquer et renoncer à l'employer dans des cadres qui dépassent son mode de fonctionnement. Un dicton populaire indique que : « quand on a pour seul outil un marteau, tous les problèmes ressemblent à des clous »... Croire en l'usage de l'intelligence artificielle comme solution à tous les problèmes, c'est vivre dans l'illusion qu'il s'agit d'un marteau universel et ne pas comprendre sa portée et ses limites.

4 Antigone, ou les principes éthiques dans l'intelligence artificielle

4.1 Antigone, *Ἀντιγόνη*, ou la compensation de la généalogie

Après la mort de leur père, Antigone et Ismène décidèrent de rentrer à Thèbes.

La réparation par la vertu

L'étymologie du nom *Antigone*, la fille d'Œdipe, apporte des informations intéressantes sur la façon dont elle va remettre de l'ordre dans la famille des Labdacides, maudite depuis plusieurs générations. « Antigone » est composé de *anti*-qui signifie « en opposition à » ou « en compensation de » et de *gone*, dérivé de « *genos* », « la génération », « la semence », ou « *gonos* », *γονός*, « la progéniture ».

Face à la malédiction qui frappait les Labdacides depuis le crime de son grand-père Laïos, le violeur homosexuel, et Œdipe, le parricide, régicide

et incestueux, Antigone se présente comme une figure qui remet de l'ordre, qui répare, qui arrive en compensation de l'action de ses géniteurs sur plusieurs générations. Antigone n'est que loyauté envers les membres de sa famille qu'elle cherche à protéger, au péril de sa vie.

Les systèmes intelligents peuvent ainsi être porteurs de biais pouvant entraîner des conséquences fâcheuses sur la vie individuelle ou la société, mais il est possible de les guider pour corriger leur trajectoire, à la manière d'Antigone guidant son père aveugle après qu'il fut banni de Thèbes.

Alors que Œdipe était parti sur les routes de l'exil, Étéocle et Polynice se disputèrent le pouvoir. Étéocle trouva cependant une idée pour conjurer la malédiction de leur père : avec son frère, ils se partageraient le pouvoir à tour de rôle, une année chacun. Comme Étéocle était l'aîné, ce fut lui qui porta en premier la couronne en faisant à son frère la promesse de la lui remettre au bout d'un an. Au cours de cette année, Polynice quitta Thèbes et se rendit à Argos, dans le Péloponnèse, où il épousa Argie, la fille du roi Adraste, et à qui il remit comme cadeau le collier de sa défunte mère Jocaste, le fameux collier d'Harmonie.

L'année étant passée, Polynice revint à Thèbes, mais Étéocle, estimé du peuple qui le préférait à son frère belliqueux, et ayant de plus apprécié l'exercice du pouvoir durant cette année, refusa de céder le trône à son frère et chercha à l'enfermer dans une prison.

Polynice s'échappa de justesse de Thèbes. Furieux d'avoir été dupé, il envoya sa sœur Ismène à Colone plaider sa cause auprès de son père, mais sans succès. Polynice retourna à Argos où il expliqua les raisons de son malheur à son beau-père le roi Adraste en lui demandant de l'aide. Il obtint de lui une armée conduite par sept capitaines qui se mirent en route pour la capitale de la Béotie afin de rétablir Polynice dans son droit. Thèbes était surnommée « la ville aux sept portes » car elle était protégée par de hautes murailles et on ne pouvait entrer dans la cité qu'à partir d'une des sept portes qui étaient toutes bien gardées.

Arrivée à Thèbes, l'armée de Polynice prit position devant chacune des sept portes de la Cité. Antigone essaya alors de raisonner ses frères afin qu'ils ne s'engageassent pas dans une guerre fratricide, mais ses efforts pour les réconcilier s'avérèrent vains. Les combats furent rudes et les sept capitaines périrent devant Thèbes avec de nombreux morts dans les deux camps, sans qu'aucune armée ne parvint à prendre le dessus sur l'autre. Pour éviter un carnage, il fut décidé que les deux frères s'affronteraient dans un combat singulier. Polynice et Étéocle se battirent avec férocité et finirent par s'entretuer, périssant chacun par l'épée de l'autre.

Les Argiens survivants retournèrent dans le Péloponnèse afin de rendre un dernier hommage aux combattants disparus. Après la mort d'Étéocle, Thèbes fut sans souverain, et Créon, l'oncle des frères ennemis, redevint le régent.

Considérant que le roi Étéocle était mort en héros en défendant la cité de Thèbes, Créon décida d'ensevelir son neveu dans une cérémonie grandiose. En revanche, considérant

que Polynice était un traître qui avait pris les armes contre Thèbes en l'attaquant avec une armée ennemie, Créon interdit à quiconque de donner une sépulture à son autre neveu, sous peine de mort, et le corps de Polynice devait être abandonné sur le champ de bataille, à la merci des chiens et des oiseaux, afin de servir d'exemple à quiconque voudrait s'en prendre à la cité de Thèbes. Pour les Grecs de l'Antiquité, ne pas avoir de sépulture est considéré comme le plus horrible de tous les châtements. En effet, si les honneurs funèbres n'étaient pas rendus à un mort, avec les chants rituels et l'obole déposée sur les yeux du défunt destinée au passeur Charon, son âme était condamnée à errer pour l'éternité sans jamais trouver le repos aux Enfers.

Antigone ne put accepter que son frère Polynice eut à subir une telle condamnation et vint voir son oncle pour le supplier de revenir sur sa décision. Créon refusa de céder à la demande de sa nièce.

Pour Antigone, les dieux mais non les mortels avaient le pouvoir de condamner un homme à voir son âme errer sans repos hors du royaume d'Hadès. Elle ne put ainsi se résoudre à accepter cette loi jugée injuste décida de braver l'interdit de Créon. Elle demanda de l'aide à sa sœur Ismène, mais celle-ci refusa, n'osant défier Créon et ses lois et craignant pour sa vie.

À la nuit tombée, Antigone prit une torche pour retrouver la dépouille de son frère, accompagnée de sa belle-sœur Argie, et les deux femmes rendirent à Polynice les derniers devoirs. Elles pratiquèrent les rituels requis sur le défunt mais à peine celui-ci avait-il été enterré qu'Antigone et Argie se firent surprendre par la garde de Thèbes. Sans opposer de résistance, les deux femmes furent conduites devant Créon qui leur rappela la gravité de leur acte et les condamna à être emmurées vivantes dans le tombeau des Labdacides.

Éthique dans les décisions politiques : au-delà de la valeur d'utilité

Redevenu maître de la cité de Thèbes, pour le bien du royaume, Créon imagine qu'il est de son devoir d'être ferme envers sa nièce et n'accepte pas de revenir sur sa décision. Cependant Antigone lui désobéit au nom de quelque chose de supérieur. Antigone prend soin de sa famille, que ce soit son père, condamné à quitter Thèbes et qu'elle va accompagner et guider tout au long de son exil, ou son frère, condamné à ne pas avoir de sépulture. Elle obéit à des valeurs plus élevées que celles édictées par un seul homme, et ceci au péril de sa vie. Antigone se révolte contre son oncle Créon, le frère de sa mère. Elle oppose la loi des dieux, avec la solidarité familiale, à la loi des hommes représentée par l'interdit prononcé par son oncle, le nouveau roi de Thèbes.

Cependant l'action de Créon, destinée à protéger les Thébains, n'eut que peu d'efficacité dans la durée : en se montrant dur avec ses opposants, il ne parvint qu'à attiser la rancœur des Argiens, et en campant sur ses positions vis-à-vis de sa nièce, il s'engagea dans un drame familial.

4.2 Antigone, ou la réalisation des modèles d'IA éthiques par conception

Apprenant la sentence qui avait frappé Antigone et Argie, Hémon, fils de Créon et fiancé d'Antigone, voulut retrouver sa promise.

Tirésias, le devin respecté de tous, prophétisa alors à Créon son malheur prochain pour ne pas avoir voulu respecter les lois naturelles édictées par les dieux, ce qui poussa le régent à revenir sur sa décision. Ce fut cependant trop tard car, devant le tombeau où était supplicié sa fiancée, il découvrit que son fils s'était donné la mort et il ne trouva plus âme qui vive en faisant ouvrir le caveau funéraire des Labdacides car, pour abrégier leur supplice, Antigone et Argie s'étaient pendues avec les ceintures de leurs robes. Avec ses lois et ses décrets, Créon se retrouva ainsi dans la plus tragique des solitudes. Ismène ne survécut pas longtemps à sa sœur, et avec elle s'éteignit la lignée maudite des Labdacides.

Dix ans plus tard, les Épigones (dont le nom signifie « les descendants »), c'est-à-dire les fils des sept chefs qui avaient péri en combattant contre Thèbes pour aider Polynice dans sa guerre de succession, décidèrent de venger leurs pères. Sous le commandement du roi Adraste, le beau-père de Polynice, ils menèrent une nouvelle expédition contre Thèbes et leurs alliés et parvinrent à les vaincre à la bataille de Glisas, en Béotie.

Les valeurs éthiques et la réalisation de programme d'IA éthiques par conception

Tout comme Antigone, il faut guider de façon éclairée les programmes aveugles et boiteux d'intelligence artificielle qui pourraient autrement nous envoyer dans le fossé, nous mener à notre perte. Il faut apporter cette béquille permettant à l'IA de marcher droit, il faut la guider dans le respect des valeurs humaines qui sont celles partagées et protégées au sein d'entités comme l'UE, ainsi que les droits et principes fondamentaux.

Emmurée vive, Antigone avait péri pour avoir obéi à un principe plus haut que toutes les lois : la morale. Antigone, c'est l'incarnation de la légitimité sociale (c'est-à-dire ce qui est fondé en droit, ce qui est juste et justifié) face à la légalité (c'est-à-dire le respect des lois de l'État). Ce qui est légal n'est pas nécessairement légitime. Il existe ainsi des lois injustes auxquelles on doit s'opposer au nom de la justice conçue comme valeur morale.

Le temps du droit est bien plus lent que celui de la technologie. Les bouleversements de la société qu'occasionne l'intelligence artificielle sont parfois amoraux, notamment quand ils se basent sur des fragilités de nos cerveaux qui se font piéger par des applications addictives, ils opèrent ainsi au-delà de la légitimité, mais ils s'effectuent malgré tout dans le cadre de la loi. Pourtant dans différents endroits du monde, et à différents niveaux, émergent désormais des réglementations comme la « Législation sur l'intelligence artificielle » (AI

Act) qui a été proposée par la Commission européenne depuis 2021 et que la France a ratifié en février 2024.

Légitimité et légalité se rejoignent pour les citoyens de l'Union Européenne : cette législation européenne établit une classification des risques sous forme d'une échelle à 5 niveaux allant de *risque inacceptable* à *risque minimal* et elle cherche à assurer un équilibre entre la promotion de l'innovation représentée par les technologies de l'intelligence artificielle et la protection des droits et des intérêts des citoyens européens face aux menaces associées à ces technologies.

5 Conclusion

Le lien entre le théâtre et l'intelligence artificielle a déjà plus d'un siècle et, dès ses origines, l'IA est un sujet qui a alimenté des fantasmes de remplacement ultime de l'espèce humaine par des machines. Dans la pièce de théâtre *R.U.R.* de Karel Čapek jouée à Prague en 1921 et à New York l'année d'après, on voit pour la première fois des créatures artificielles ayant le nom de « robots ». Ces robots sont des androïdes dénués de sensibilité et de sentiments, mais afin d'être plus robustes et plus efficaces, ils sont améliorés par l'ingénieur de l'usine R.U.R. qui les dote de plus d'intelligence. Cette amélioration aura pour conséquence la révolte des robots qui, en 10 ans, anéantiront l'humanité...

Le travail présenté ici a comme objectif de proposer de quoi aider à comprendre, via des similarités opérées avec la destinée mythique des Labdacides et en particulier du personnage d'Œdipe, certaines caractéristiques de l'intelligence artificielle. La mythologie grecque sert ainsi de fil rouge pour avoir une vision plus informée sur cette technologie qui touche autant à l'intime qu'à l'ensemble de la société. L'IA impacte notre rapport au monde d'une façon profonde, elle doit par conséquent être employée par tout un chacun en connaissant suffisamment bien ses avantages et ses limites, et ceci afin de rester utile à l'humanité.

Références

- [1] G. Berry. *Pourquoi et comment le monde devient numérique*. Collège de France / Fayard, 2008. Leçon inaugurale N° 197.
- [2] G. Berry. *L'hyperpuissance de l'informatique. Algorithmes, données, machines, réseaux*. Éditions Odile Jacob, 2017.
- [3] O. Bloch and W. von Wartburg. *Dictionnaire étymologique de la langue française*. Presses Universitaires de France, 4^e édition, 1964.
- [4] A. C. Clarke. *Profiles of the Future : An Inquiry into the Limits of the Possible*. Harper & Row, 1977.
- [5] M. David and C. Sauviat. *Intelligence artificielle. La nouvelle barbarie*. Éditions du Rocher, 2019.
- [6] H. L. Dreyfus. *Intelligence artificielle : Mythes et limites*. Flammarion, 1984. Trad. de "What Computers Can't Do : The Limits of Artificial Intelligence".
- [7] E. Durkheim. *Les formes élémentaires de la vie religieuse. Le système totémique en Australie*. Quadrige / Presses Universitaires de France, 1912.
- [8] A. Elliott. *The Culture of AI : Everyday Life and the Digital Revolution*. Routledge, Taylor & Francis Group, 2019.
- [9] S. Freud. *L'interprétation du rêve*. Quadrige / Presses Universitaires de France, 2010. Trad. de „Die Traumdeutung” paru en allemand en 1900.
- [10] S. Freud. *Totem et Tabou*. Éditions Points, 2010. Première parution en allemand en 1912.
- [11] Y. Le Cun. *Quand la machine apprend*. Éditions Odile Jacob, 2019.
- [12] C. Lévi-Strauss. *Les structures élémentaires de la parenté*. Mouton & Co, 1947.
- [13] F. Mazzi and L. Floridi, editors. *The Ethics of Artificial Intelligence for the Sustainable Development Goals*. Springer, 2023.
- [14] C. O'Neil. *Algorithmes : la bombe à retardement*. Les Arènes, 2018. Première parution en anglais en 2016.
- [15] E. Pariser. *The Filter Bubble : What the Internet Is Hiding from You*. Penguin Press, 2011.
- [16] C. A. Pickover. *La fabuleuse histoire de l'intelligence artificielle*. Éditions Dunod, 2021.
- [17] A. Rouvroy. L'art de ne pas changer le monde. *La Revue Nouvelle*, 8(8) :44–50, 2016.
- [18] A. Rouvroy and T. Berns. Le nouveau pouvoir statistique. ou quand le contrôle s'exerce sur un réel normé, docile et sans événement car constitué de corps « numériques »... *Multitudes*, 1(40) :88–103, 2010.
- [19] É. Sadin. *L'intelligence artificielle ou l'enjeu du siècle*. Éditions L'Échappée, 2018.
- [20] M. Serres. *Petite Poucette*. Collection « Manifestes ». Éditions Le Pommier, 2012.
- [21] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. P. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis. Mastering the game of go without human knowledge. *Nature*, 550(7676) :354–359, 2017.
- [22] Sophocle. *Tragédies. Les Trachiniennes. Antigone. Ajax. Œdipe roi. Electre. Philoctète. Œdipe à Colone*. Le Livre de Poche. Classique. Société d'édition "Les Belles Lettres", 1962.
- [23] Alan M. Turing. Computing machinery and intelligence. *Mind*, 49 :433–460, 1950.
- [24] J.-P. Vernant. Œdipe sans complexe. *Raison présente*, 4 :3–20, 1967. Pour une théorie de la laïcité.
- [25] J.-P. Vernant. *L'univers, les dieux, les hommes. Récits grecs des origines*. Points. Éditions du Seuil, 1999.
- [26] J.-P. Vernant. Le tyran boiteux : d'Œdipe à Périandre. In J.-P. Vernant and P. Vidal-Naquet, editor, *Mythe et tragédie en Grèce ancienne – Tome II*, pages 45–72. Éditions La Découverte / Poche, 2001.