



Et si les algorithmes d'IA évoluaient également suivant des lois darwiniennes ?

Par Thierry Caminel

Nous vivons une révolution technologique, amenée par des systèmes basés sur des techniques d'Intelligence Artificielle (IA) qui sont sur le point de faire aussi bien que les humains pour une large gamme d'activités, telles que conduire une voiture, répondre aux demandes des clients, recommander un livre, ou diagnostiquer des maladies. Ces systèmes sont constitués de centaines de milliers d'algorithmes informatiques enchevêtrés, qui évoluent rapidement, tellement rapidement que cela suscite de nombreuses inquiétudes quant aux impacts sur les sociétés humaines. Mais quelles sont les lois qui régissent ces évolutions ?

Dans cet article, nous proposons une réponse provocante à cette question : nous suggérons que les lois darwiniennes, qui régissent l'évolution biologique et culturelle, régissent aussi l'évolution des algorithmes informatiques. Les mêmes mécanismes sont à l'œuvre pour créer de la complexité, des cellules aux humains, des humains aux robots.

Plus précisément, nous nous concentrerons sur la théorie du [Darwinisme Universel](#), une version généralisée des mécanismes de variation, de sélection et d'hérédité proposés par Charles Darwin, déjà appliquée pour expliquer l'évolution dans une grande variété d'autres domaines, comme l'anthropologie, la musique, la culture ou la cosmologie. Selon cette approche, de nombreux processus évolutifs peuvent être décomposés en trois composantes : 1/ Variation d'une forme ou d'un modèle donné, typiquement par mutation ou recombinaison ; 2/ Sélection des variantes les plus adaptées ; 3/ Hérédité ou rétention, permettant de conserver et transmettre ces variations.

Nous commencerons par donner des exemples simples, pour illustrer comment ce mécanisme s'intègre bien dans l'évolution des algorithmes « open source ». Cela nous amènera à



considérer les grandes plateformes numériques comme des organismes, où les algorithmes sont sélectionnés et transmis. Ces plateformes ont un rôle important dans l'émergence des techniques d'apprentissage automatique, et nous étudierons cette évolution du point de vue darwiniste dans la section 2. Dans la troisième section, nous élargirons notre champ d'action et nous ferons nôtre la vision selon laquelle les gènes et certains éléments de la culture peuvent être considérés comme des algorithmes, de même que la concurrence pour les emplois entre l'homme et les machines peut être assimilée à un processus darwinien, mû par le passage des algorithmes humains aux algorithmes numériques. Dans la dernière section, nous étudierons plus en détail certaines des boucles de rétroaction qui sous-tendent l'évolution des algorithmes biologiques, culturels et numériques. Ces boucles de rétroaction aident à comprendre la fantastique rapidité d'évolution des algorithmes que nous voyons de nos jours, et mènent à des aperçus troublants sur le transhumanisme et les conséquences potentielles de l'IA dans nos sociétés.

1 – L'Évolution Darwinienne des Algorithmes Informatiques

Commençons par un exemple. De nombreux chercheurs et entreprises travaillent à [l'amélioration du noyau du système d'exploitation open source 'Linux'](#). Qu'ils travaillent sur de nouveaux concepts, ou sur une légère amélioration des algorithmes, le processus allant d'une idée d'amélioration au changement introduit dans la version principale du noyau (le 'tronc') est long, et beaucoup d'idées ou de propositions n'auront pas de suites. Mais si une modification est acceptée, alors le changement sera conservé pendant longtemps dans la base de code Linux. Elle servira de référence pour les travaux futurs, et pourrait être à un moment donné déployé [sur des millions d'ordinateurs](#) sous Linux, des smartphones aux supercalculateurs.

Nous voyons que la description du darwinisme universel ci-dessus est valable pour notre exemple: 1/ Des variantes des nombreux algorithmes Linux sont continuellement créées, typiquement en changeant une partie d'un algorithme existant ou en combinant différents algorithmes (souvent développés dans un autre contexte) 2/ Les meilleures variations apportant un bénéfice sont sélectionnées, et mises dans une version du noyau Linux 3/ Cette



version du noyau est intégrée dans des milliers de produits, et deviennent la base de nouvelles évolutions. Tout comme les gènes survivent et continuent d'évoluer après la mort des cellules, les algorithmes de votre smartphone continuent d'évoluer après que vous ayez décidé de la changer pour un nouveau avec des algorithmes améliorés ou plus de fonctionnalités.

Plus généralement, tout le mouvement open-source suit ce processus évolutif. Sur Github par exemple - la plus grande plate-forme de développement open-source, [des millions de projets existent](#) et sont continuellement copiés, modifiés et recombinaés par des millions de développeurs de logiciels. Certains changements sont sélectionnés pour être introduits dans la ligne principale d'un produit et deviennent stables et susceptibles d'être réutilisés dans de nombreux projets, tandis que la plupart d'entre eux sont oubliés. L'analogie avec l'évolution des gènes est forte : le contenu génétique des organismes mute et se recombine avec d'autres continuellement, et, parfois, des changements apportent un avantage et sont conservés dans de longues lignées de descendants.

Les codes des projets open-sources les plus intéressants sont intégrés en combiné dans de grandes plateformes de Cloud Computing, comme celle détenue par Amazon (Amazon Web Services) ou ses concurrents tels que Microsoft, Google ou Alibaba, qui proposent des services facilitant la composition de milliers d'algorithmes qu'elles mettent à disposition. Ces services sont utilisés par ces entreprises pour leurs besoins propres, mais aussi par d'autres entreprises, notamment les startups pour développer de nouvelles idées à moindre coût. Ces startups, à leur tour, peuvent mettre en open-source leurs propres variantes, qui pourront éventuellement être intégrées à ces plateformes. Ces plateformes servent ainsi de catalyseurs : elles font appliquer le processus évolutif darwiniste décrit ci-dessus en accélérant la combinaison, la sélection et la rétention des algorithmes.

2 - Évolution darwiniste des algorithmes de l'IA

Les services fournis par ces plateformes permettent de collecter beaucoup de données, ce qui conduit également à une percée majeure dans la dynamique d'évolution des algorithmes. En effet, les techniques d'apprentissage automatique permettent de créer des algorithmes à partir de données. Historiquement, la plupart des algorithmes ont été créés méticuleusement



par les humains. Mais de nos jours, les systèmes d'apprentissage automatique utilisent des méta-algorithmes (par exemple, les réseaux de neurones artificiels) pour créer de nouveaux algorithmes. Ces algorithmes peuvent être appelés 'modèles' ou 'agents'. Ils sont généralement créés en étant alimentés par de très grands ensembles de données d'exemples (« Big Data ») mais pas seulement.

Les progrès des algorithmes pour jouer au jeu de Go illustrent cette évolution. Pendant des décennies, les programmeurs ont développé et amélioré manuellement des algorithmes pour évaluer les positions et les mouvements du jeu. Mais de nos jours, les meilleurs logiciels utilisent des techniques d'apprentissage automatique, et en particulier les réseaux neuronaux profonds (« Deep Learning »). Ces réseaux ont, dans une première étape, été entraînés à évaluer des positions du jeu Go à partir d'un catalogue d'un million de parties jouées par des humains. Ceci a permis à Google AlphaGo de gagner contre des champions du monde humains. Puis, une nouvelle version appelée AlphaZero a supprimé cette contrainte, et a appris à évaluer la position en jouant des milliers de parties contre elle-même. Le processus peut être considéré comme darwiniste : après chaque jeu, quelques changements sont introduits dans la version qui a gagné pour produire d'autres versions, qui jouent les unes contre les autres, etc. Et juste en changeant les règles initiales, AlphaZero a appris en quelques heures à devenir aussi un grand maître aux échecs !

Aux niveaux inférieurs, il existe un mécanisme similaire appelé « apprentissage par renforcement » (Reinforcement learning). C'est l'une des méthodes clés à l'origine de l'explosion de l'IA que nous observons actuellement, utilisée notamment pour la traduction automatique, les assistants virtuels ou les véhicules autonomes. Le but est d'optimiser une séquence d'action pour atteindre un but de manière optimale (par exemple dans le jeu de Go, une séquence de placement de pierres pour maximiser l'encerclement du territoire adverse). Dans cet algorithme, un agent (c'est-à-dire un algorithme) sélectionne une action sur la base de ses expériences passées. Il reçoit une récompense numérique qui mesure le succès de son résultat, et cherche à apprendre à sélectionner les actions qui maximisent cette récompense. Au fil du temps, les agents apprennent ainsi comment atteindre un but. Pour améliorer le



processus, un grand nombre de réseaux de neurones programmés au hasard peuvent être testés en parallèle. Les meilleurs sont sélectionnés et dupliqués avec de légères mutations aléatoires, et ainsi de suite pendant plusieurs générations. Cette approche est appelée "neuroévolutive" et basée sur des "algorithmes génétiques", illustrant la similitude avec l'évolution biologique. La fonction qui mesure le progrès vers les objectifs peut aussi être remplacée par une mesure de nouveauté, qui récompense un comportement différent de celui des individus précédents. Les chercheurs qui travaillent sur ces techniques d'IA, comme Ken Stanley, sont explicitement inspirés par la propension naturelle de l'évolution naturelle à découvrir perpétuellement la nouveauté.

3 - Le darwinisme et la numérisation des algorithmes humains

On discute beaucoup des impacts de l'IA sur les activités humaines, et il y a un consensus sur le fait que les robots et les agents virtuels, alimentés par des algorithmes d'IA, pourraient remplacer les humains dans de nombreuses activités. Or, ces activités sont permises par nos capacités cognitives, qui sont la conséquence de centaines de milliers d'années d'évolution darwinienne. On peut donc considérer qu'il y a une certaine équivalence entre les algorithmes et ces fonctions cognitives, et que celles-ci peuvent être considérées comme des algorithmes biochimiques. Par exemple, des algorithmes numériques comme ceux de la vision par ordinateur sont comparables, et potentiellement interchangeables, avec des fonctions codées dans les gènes. Cette analogie a amené l'expert en robotique Gill Pratt à poser la question suivante : "[Est-ce qu'une explosion cambrienne est imminente pour la robotique?](#)", parce que l'invention de la vision au Cambrien a été la clé de voûte de l'explosion des formes de vie, et que quelque chose de semblable pourrait se passer si la vision est donnée aux ordinateurs. Les machines pourraient par exemple apprendre comment le monde physique fonctionne en "voyant" des milliers de vidéos, tout comme un bébé apprend la gravité et la conservation de l'inertie en observant le monde qui l'entoure. C'est actuellement un sujet de recherche actif pour les chercheurs en IA.

L'interchangeabilité des algorithmes entre l'humain et les machines est abordée par Yuval Noah Harari dans son best-seller "[Homo Deus: Une brève histoire du futur](#)". L'un des concepts



clés qu'il développe est que les organismes ont des algorithmes biochimiques pour "calculer" ce que l'on considère généralement comme des "sentiments" ou des "émotions", comme la meilleure décision à prendre pour éviter un prédateur, ou pour choisir un partenaire sexuel. Bien que ce ne soit pas explicitement mentionné dans le livre, ces algorithmes suivent un processus darwinien : ils se sont améliorés au cours de millions d'années d'évolution, et, si les sentiments de certains ancêtres ont conduit à une erreur, les gènes qui façonnent ces sentiments ne sont pas passés à la génération suivante. Harari soutient que les algorithmes numériques et les algorithmes biochimiques ont la même nature, et que les premiers sont susceptibles de devenir plus efficaces pour la plupart des tâches que les derniers.

Certains chercheurs, comme Richard Dawkins, considèrent également que les comportements ou les pratiques qui se propagent au sein d'une culture suivent des processus de Darwinisme universel. Les pratiques culinaires, les techniques agricoles, les croyances religieuses ou les stratégies guerrières sont des exemples typiques, mais les processus et le savoir-faire des entreprises entrent également dans cette catégorie. Le fait est que la plupart des mêmes sont des algorithmes : les recettes de cuisine sont de fait un exemple courant pour expliquer ce qu'est un algorithme. On peut donc considérer la transformation numérique des processus et des compétences dans les entreprises comme une transformation de la nature des algorithmes sous-jacents, depuis des algorithmes culturels vers des algorithmes numériques. Ceux-ci ont d'abord été développés par des informaticiens, qui ont codé explicitement des processus d'entreprises devenus de plus en plus automatisés. Mais ils sont maintenant complétés par l'apprentissage automatique, ce qui permet d'augmenter encore le niveau d'automatisation (par exemple en utilisant des algorithmes d'intelligence artificielle qui comprennent le langage naturel ou optimisent des activités complexes), ou d'améliorer les processus, en analysant le flux de données toujours croissant. Les entreprises n'ont d'autre choix que de s'adapter. Alors que le monde est de plus en plus bouleversé par les bouleversements numériques, la phrase de Charles Darwin « **ce n'est pas l'espèce la plus forte qui ne survit ni la plus intelligente ; c'est celle qui s'adapte le mieux au changement** » est plus que jamais vraie.



Cette transformation numérique de l'entreprise conduit également à une compétition entre les humains et les machines, comme l'expliquent Erik Brynjolfsson et Andrew McAfee dans leur livre "[Race Against The Machine](#)". C'est en effet une compétition darwiniste entre des formes d'algorithmes. Par exemple, les algorithmes basés sur l'apprentissage automatique d'Amazon sont en compétition avec ceux acquis par les libraires pour conseiller un client. Amazon développe aussi des systèmes basés sur l'IA pour le contrôler ses centres de données, gérer ses entrepôts et systèmes logistiques et automatiser les processus métier, ce qui leur permet de diminuer inexorablement l'intervention humaine. Quant aux algorithmes sous-jacents, ils sont continuellement améliorés par les développeurs, mais beaucoup ne sont même pas embauchés par Amazon grâce aux communautés Open Source. Les algorithmes sont également améliorés par de l'apprentissage automatique, en utilisant l'énorme quantité de données générées acquises par la plate-forme. Les humains peuvent difficilement rivaliser.

4 - Fermeture de la boucle et autocatalyse des algorithmes

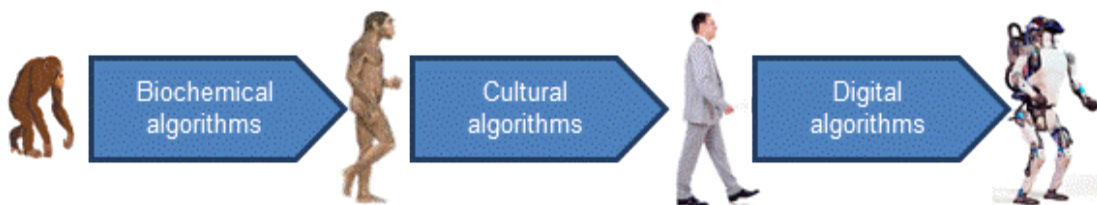
Nous avons vu que les algorithmes numériques ont tendance à remplacer ou à compléter les algorithmes biochimiques et culturels (y compris les processus d'entreprise), et qu'il existe des boucles de rétroaction positives entre les deux cas. Mais d'autres boucles de rétroaction méritent d'être étudiées.


Par exemple, la diffusion des connaissances est accélérée par les technologies de l'information comme Internet, les smartphones, les cours en ligne, les moteurs de recherche, les assistants personnels, etc. Or ces technologies de l'information s'appuient sur de nombreux algorithmes numériques. Les algorithmes numériques accélèrent donc le rythme de diffusion des algorithmes culturels (par exemple, la formation en informatique), qui peuvent en retour accélérer le développement des algorithmes numériques.

Un autre exemple, assez différent, est le progrès récent dans la manipulation de l'ADN, tels que les techniques appelées [CRISPR/Cas9](#). Ces progrès sont la conséquence des avancées en algorithmique numérique pour séquencer l'ADN et déterminer sa fonction. Mais l'ADN est le support physique d'algorithmes pour créer des protéines. Par conséquent, les algorithmes numériques permettent la modification des algorithmes génétiques. Les répercussions

pourraient être considérables dans l'avenir, car nous remplaçons l'ancienne évolution génétique, basée sur des mutations aléatoires, par des modifications ciblées du code génétique contrôlé par des algorithmes numériques. Ce n'est pas de la science-fiction : ces modifications ont déjà été appliquées sur des animaux, [pour traiter certaines maladies](#). Cela donne un certain poids à l'opinion des [transhumanistes](#), qui estiment que ces techniques pourraient, un jour, être aussi utilisées pour étendre les capacités biologiques des humains...

Le diagramme suivant illustre ces relations :



Support	ADN / Gènes	Cerveau / Mêmes	Stockage informatique / Code exécutable
Transmission	Réplication de l'ADN, reproduction biologique	Imitation, langage, livres, vidéo, ...	Copie digitale et réseaux informatiques
Mémorisation (exemples)	Fabrication de protéines et d'organismes biologiques, stratégie d'accouplement, habiletés de survie...	Pratiques culturelles et agricoles, recettes, lois, jeux musicaux, mythes, processus, opérations mathématiques.	[Sous-ensemble des colonnes précédentes].
Évolution	Lente	Rapide	Ultra-Rapide
Boucles de rétroaction			



Nous l'avons vu, les plates-formes numériques permettent l'acquisition de données et l'exécution d'algorithmes. Elles réduisent la distance et les incompatibilités entre les données et les algorithmes, et, ce faisant, elles agissent comme des catalyseurs, permettant plus facilement d'appliquer des algorithmes aux données, et de créer des algorithmes à partir de données. Les algorithmes aident les développeurs à créer des algorithmes, et aident les spécialistes de l'apprentissage automatique à transformer les données en algorithmes. Les algorithmes ainsi créés sont ensuite utilisés dans les applications et les services hébergés par les plateformes, qui permettront à leur tour d'acquérir plus de données et gagner de l'argent pour investir dans de nouvelles versions, etc. Les plus grandes sociétés d'IA comme Google, Facebook ou Baidu se battent d'ailleurs pour créer des écosystèmes de développeurs qui pourront créer et développer de nouvelles idées d'algorithmes, de nouvelles variantes, les combiner, etc. C'est probablement la principale raison pour laquelle elles ont, par exemple, toutes mis en open-source leurs logiciels sophistiqués d'apprentissage profond (Deep Learning), afin d'amener les développeurs à utiliser leurs plateformes pour partager, modifier ou combiner du code, héberger les résultats, et augmenter ainsi la rétention des algorithmes dans leur écosystème.

Dans un proche avenir, ces plateformes dans le Cloud permettront aussi le partage des connaissances entre robots. Supposons par exemple qu'une voiture autonome rencontre une situation inhabituelle. Celle-ci pourra être envoyée vers la plateforme du fabricant, qui pourra l'utiliser pour re-entraîner les algorithmes de détection, et envoyer la nouvelle version vers tous les véhicules. Le processus pourra être largement automatisé, permettant des adaptations rapides aux changements de l'environnement.

Cette automatisation peut aller plus loin. Comme nous l'avons vu au sujet du jeu de Go, les progrès de l'apprentissage automatique autonome (« [automated machine learning](#) ») permettent aux algorithmes de créer de nouveaux algorithmes avec peu, ou pas, d'assistance humaine. Les réactions entre algorithmes dans les plates-formes qui les hébergent peuvent de ce fait être considérées comme autocatalytiques : ce sont des méta-algorithmes qui utilisent la proximité entre données et algorithmes au sein des plateformes pour créer de



nouveaux algorithmes ou méta-algorithmes. Or selon certains scientifiques comme [Stuart Kauffman](#), l'autocatalyse joue un rôle majeur dans l'origine de la vie, et plus généralement dans [l'émergence](#) de systèmes auto-organisés. Nous pensons que l'autocatalyse des algorithmes numériques est la percée majeure de l'IA, qui explique le mieux les progrès exponentiels que nous constatons.

Tous les pays industrialisés sont maintenant conscients de cette percée, et luttent pour développer et contrôler les algorithmes d'IA, car ils deviennent une partie essentielle de leur souveraineté et de leur indépendance. Le contrôle du processus d'innovation et des plateformes deviennent d'une importance capitale. La Chine et les gouvernements américains, par exemple, en particulier les agences militaires et de renseignement comme le DARPA, investissent massivement dans la R&D autour du traitement des données et de l'intelligence artificielle, souvent en cofinancement avec les grands fournisseurs de plateformes commerciales. En résulte un flux de nouvelles idées et d'algorithmes, qui sont sélectionnés d'une manière ou d'une autre. Celles qui apportent un avantage sont incorporées dans ces plateformes pour les usages commerciaux, mais peuvent également être utilisées par les États pour contrôler l'information et améliorer leur défense (comme les armes basées sur l'IA, les systèmes de renseignement...). Nous voyons ici clairement un processus de Darwinisme Universel.

En réaction, on voit émerger des initiatives comme [OpenAI](#), financé par Elon Musk, dont le but est de *"contrecarrer les grandes entreprises qui peuvent gagner trop de pouvoir en possédant des systèmes de super-intelligence consacrés aux profits, ainsi que les gouvernements qui peuvent utiliser l'IA pour prendre le pouvoir et même opprimer leurs citoyens"*. Le premier projet est une plateforme open source pour aider à développer collaborativement des algorithmes d'apprentissage par renforcement, qui resteront libre de droits. Une telle plate-forme pourrait être un moyen d'éviter la rétention des algorithmes dans certains "organismes" comme les gouvernements et quelques grandes entreprises. Nous verrons si une telle initiative réussit, mais elle nous semble être une bonne illustration de la dynamique de l'évolution des algorithmes et ses conséquences que nous avons décrites.



5 - Conclusion

L'évolution des algorithmes est un sujet crucial pour comprendre l'évolution de notre civilisation. En regardant le passé, cette évolution peut à priori sembler imprévisible. Par exemple, personne, il y a 20, 10 ou même 5 ans, n'a été capable de prédire l'évolution des algorithmes d'IA ou la croissance de Facebook. Pourtant, cette évolution n'est pas aléatoire, car elle résulte de millions de décisions rationnelles, prises par une myriade d'acteurs comme les ingénieurs logiciels, les consommateurs, les entreprises et les agences gouvernementales.

Dans cet article, nous avons étudié l'idée que cette évolution a de fortes similitudes avec l'évolution biologique et culturelle, qui ont conduit à des ruptures telles que l'explosion cambrienne il y a 500 millions d'années, ou l'invention de l'agriculture il y a 12 000 ans. Notre thèse est que les mêmes principes du darwinisme universel s'appliquent, ce qui éclaire la manière dont les algorithmes évoluent au fil du temps, et ce dans des contextes aussi variés que le développement des logiciels dans les communautés open source, les plateformes numériques, la transformation numérique de notre société, les progrès de l'IA et l'impact de celle-ci sur notre société. Nous pourrions trouver d'autres cas qui cadrent bien avec ces principes, par exemple dans des domaines tels que la cybersécurité (comme la coévolution des logiciels malveillants et leur contre-mesure), l'apprentissage non supervisé (comme les très prometteurs [Generative Adversarial Networks](#)), ou l'optimisation des logiciels par l'introduction automatique de petites différences (« [Automatic A/B Testing](#) »).

Cette contribution ne prétend pas être un travail scientifique, ni conclure quoi que ce soit sur la pertinence du darwinisme dans d'autres contextes. Mais elle fournit un cadre explicatif de l'évolution des algorithmes indépendamment de leur support, qu'il soit génétique, culturel ou numérique. Nous sommes convaincus qu'il faut considérer les algorithmes de cette manière globale, pour comprendre la révolution numérique que nous vivons, déterminer quelles pourraient en être les prochaines étapes, et atténuer les risques pour nos démocraties. Pour aller plus loin, notre sentiment est qu'une théorie générale de l'évolution des algorithmes est nécessaire, qui soit bien intégrée dans des théories plus larges traitant de l'émergence de structures complexes dans la nature.



Les vues exprimées ici sont celles de l'auteur, et pas nécessairement celles de son employeur