

Intelligence artificielle : la Suisse à la pointe de la recherche

La Suisse compte de nombreux projets de recherche dans le domaine de l'apprentissage automatique. La médecine constitue un champ d'application prometteur. *Andreas Krause*

Abrégé L'intelligence artificielle (IA) et l'apprentissage automatique intéressent toujours plus le public. Ces procédés peuvent aujourd'hui apporter des solutions grâce à l'explosion du nombre de données et à la puissance de calcul nouvelle des ordinateurs. Les processus d'apprentissage automatique sont de plus en plus souvent des technologies transversales à l'effet transformateur et au potentiel énorme qui peuvent également produire un effet de rupture. Cet article brosse l'état actuel de la recherche au moyen de quelques exemples helvétiques. La Suisse occupe une bonne position face à la forte concurrence mondiale actuelle dans le domaine de l'IA. Elle a maintenant la chance peut-être unique de participer à la mise en place d'une gestion des données et de l'IA fiable et reconnue dans le monde entier.

L'intelligence artificielle (IA) fascine le monde informatique depuis les années 1950. À cette époque, le pionnier britannique de l'informatique Alan Turing publiait « Computing Machinery and Intelligence »¹. Il esquissait l'idée de développer un programme qui, par sa capacité d'apprendre, égalerait l'intelligence humaine. Il a développé le fameux « test de Turing » pour rendre compte de la difficulté à trouver une définition de l'IA. Après un dialogue de plusieurs minutes avec son vis-à-vis, la personne effectuant le test doit déterminer si elle a affaire à une intelligence humaine ou artificielle.

Près de 70 ans plus tard, la majorité des téléphones mobiles multifonctions possèdent des assistants auxquels on peut demander de l'aide pour régler son réveil ou l'éclairage de sa maison. Le test de Turing est-il réussi? Pas vraiment. Siri et Alexa sont encore loin d'une intelligence artificielle « générale » au sens où l'entendait Turing.

Une révolution a toutefois bien eu lieu ces dernières années grâce à l'augmentation des données numériques et à une capacité de calcul qui croît de manière exponentielle. Ces deux éléments sont les conditions fondamentales de l'apprentissage automatique (« machine

learning»), qui donne à l'ordinateur des aptitudes humaines comme la vue et l'ouïe.

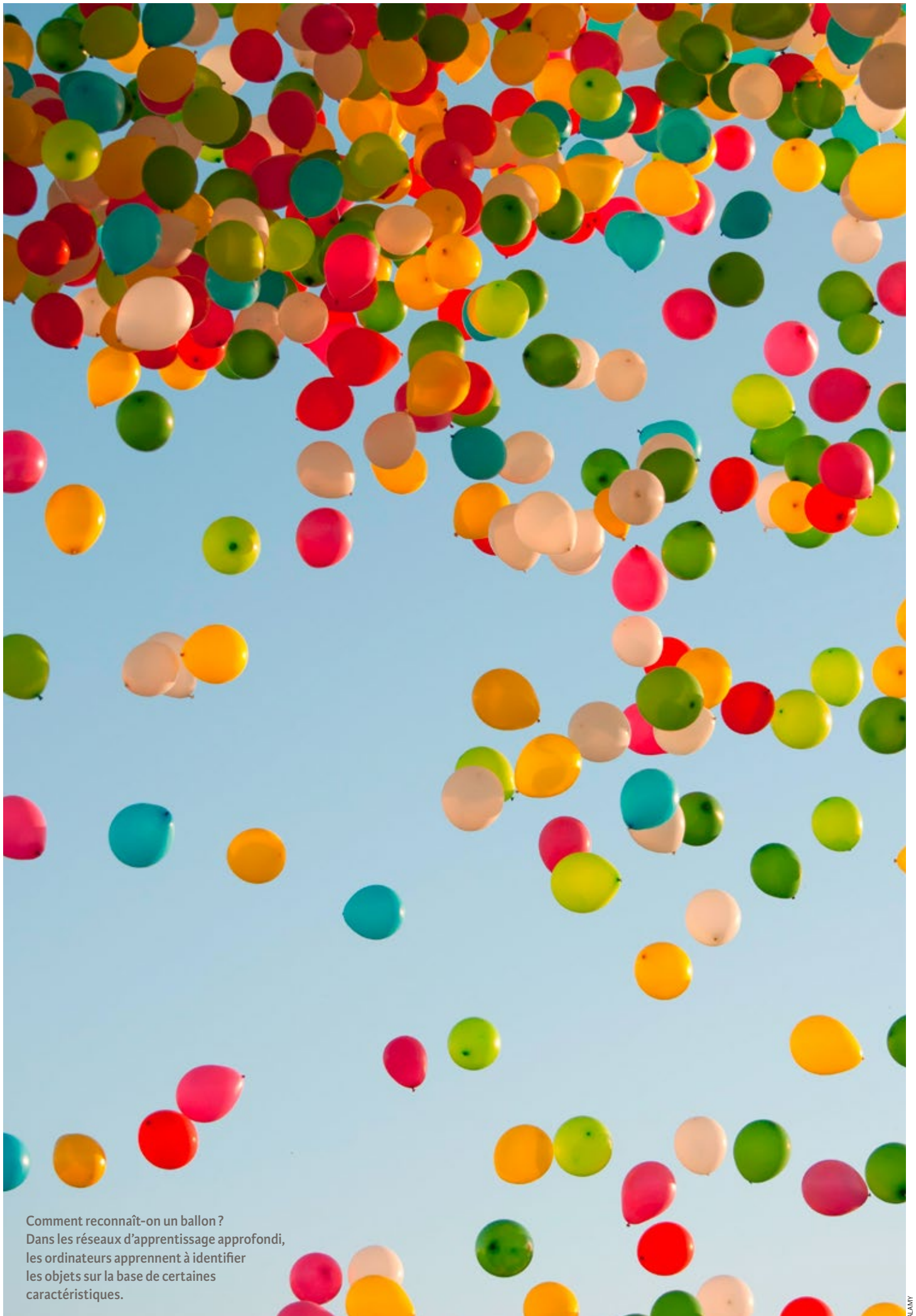
Des ballons et des fraises

Le domaine le plus abouti de l'apprentissage automatique à ce jour est l'« apprentissage supervisé » (« supervised learning »). En se fondant sur d'énormes quantités de données annotées par des humains (images étiquetées d'objets, etc.), on entraîne des systèmes complexes pour qu'ils imitent le comportement humain. Ces réseaux neuronaux artificiels contiennent souvent des millions de paramètres qui s'adaptent aux données au moyen de procédés d'optimisation mathématique. Cette méthode ressemble aux méthodes de régression statistique, bien que les modèles d'apprentissage automatique soient nettement plus complexes.

Les réseaux d'apprentissage approfondi (« deep learning ») peuvent ainsi reconnaître les liens hiérarchiques sur plusieurs niveaux. La banque de données ImageNet offre par exemple une base d'apprentissage: elle contient des millions d'images répertoriées en catégories et en termes comme « ballon » ou « fraise ». L'apprentissage automatique a fait des progrès considérables grâce à ces données: en huit ans, le taux d'erreur des meilleurs modèles est tombé de 25 % à 4 %, ce qui correspond au taux des humains lorsqu'ils font la classification eux-mêmes. La reconnaissance vocale et la traduction, qui se basent sur les réseaux récurrents, constituent d'autres exemples de réussite. Une variante populaire de ceux-ci (LSTM) a été développée en Suisse, à l'Institut Dalle Molle de recherche en intelligence artificielle (Idisia)². Les techniques de l'apprentissage automatique sont déjà des éléments importants des moteurs de

¹ Turing (1950).

² Hochreiter et Schmidhuber (1997).



Comment reconnaît-on un ballon ?
Dans les réseaux d'apprentissage profondi,
les ordinateurs apprennent à identifier
les objets sur la base de certaines
caractéristiques.

recherche et des systèmes de recommandation dans le commerce en ligne.

Les champs d'application de l'apprentissage automatique augmentent constamment. Le domaine médical présente un fort potentiel: les progrès accomplis dans l'interprétation des images³ peuvent déjà servir d'aide à la prise de décisions d'un niveau équivalent à celui d'experts humains. Des chercheurs de l'École polytechnique fédérale de Zurich (EPFZ) développent également des systèmes de surveillance automatique dans les unités de soins intensifs des hôpitaux⁴.

L'EPFZ et l'École polytechnique fédérale de Lausanne (EPFL) ont fondé en 2017 le Centre national de la science des données afin d'exploiter tout le potentiel des applications d'apprentissage automatique dans un esprit interdisciplinaire. L'institution accompagne des projets en science et en économie. Les champs d'application vont de la médecine à la cosmologie en passant par les sciences de l'environnement.

Décélérer de nouvelles maladies

L'«apprentissage non supervisé» («unsupervised learning») constitue un autre pan de l'apprentissage automatique. Si les utilisations pratiques directes ne sont dans ce cas pas aussi nombreuses que pour l'apprentissage supervisé, ce secteur progresse de façon exponentielle. Avec l'apprentissage non supervisé, les ordinateurs saisissent de manière autonome des échantillons à partir de données. Autrement dit, ils n'ont pas besoin d'indications humaines définies comme signal d'apprentissage. La médecine constitue là encore un champ d'application privilégié: il devient possible de déceler des sous-types de maladies d'un nouveau genre au moyen de jeux de données semblables⁵.

Les «modèles génératifs» font notamment partie de l'apprentissage non supervisé. Ils sont capables de modéliser des données de textes et d'images complexes, et d'en générer automatiquement des semblables. Il existe un grand potentiel dans la recherche sur les médicaments⁶ ou dans l'anonymisation des données des patients.

La validation représente une difficulté majeure de la procédure non supervisée: alors que

le taux d'erreur dans l'apprentissage supervisé peut être évalué au moyen de données de test (dans notre exemple: le nombre d'images mal classées), ce n'est pas possible avec l'apprentissage non supervisé. Le nombre de sous-types de différentes maladies qui apparaîtront dans les données n'est par exemple pas connu d'avance. Cela représente un véritable défi pour la recherche fondamentale.

Où irons-nous manger aujourd'hui ?

L'apprentissage supervisé ou non supervisé est considéré comme une «méthode d'apprentissage passive»: un modèle (par exemple un réseau neuronal) s'entraîne sur un ensemble de données imposé pour être ensuite utilisé. En revanche, dans la «méthode d'apprentissage active», une interaction a lieu avec l'environnement ou les utilisateurs.

L'«apprentissage par renforcement» («reinforcement learning») est notamment utilisé dans les systèmes de recommandation. En prenant des décisions, l'agent apprend comment l'environnement change et reçoit des signaux de récompense pour cela. Un défi lié à cette méthode provient du «compromis exploration-exploitation» («exploration-exploitation trade-off»): dans chaque situation, l'agent doit évaluer s'il choisit une action qui lui assure le succès (exploitation) ou s'il veut en apprendre davantage sur son environnement par une nouvelle action (exploration). Nous nous posons souvent ce genre de questions en tant qu'êtres humains, par exemple lorsque nous nous demandons s'il vaut mieux aller manger dans notre restaurant préféré ce soir ou en essayer un nouveau.

L'ordinateur triomphe au jeu de go

Bien que l'apprentissage par renforcement ait fait l'objet de recherches depuis des années, des percées majeures n'ont été réalisées que récemment. Le jeu de go en est un parfait exemple: contrairement aux échecs, auxquels l'ordinateur sait depuis longtemps mieux jouer que les humains avec la procédure de planification habituelle sans apprentissage automatique, le go se révèle bien plus complexe. L'agent Alpha Zero de la filiale de Google Deep Mind a réussi

3 Voir par exemple Cireşan et al. (2013).

4 Hyland et al. (2018).

5 Brodersen et al. (2014).

6 Gupta et al. (2018).

7 Hwangbo et al. (2019).

8 Berkenkamp et al. (2017).

à surpasser les meilleurs joueurs humains de go. Alpha Zero a simulé d'innombrables parties jouées contre lui-même afin de s'améliorer. L'IA a expérimenté de nouveaux coups pour voir la réaction de l'adversaire simulé.

Un simulateur précis est une condition indispensable à l'apprentissage par renforcement. C'est la raison pour laquelle les jeux s'y prêtent bien : vous savez exactement à quoi ressemble le plateau de go une fois que vous et votre adversaire avez joué un coup. En dehors du monde du jeu, la tendance est à la configuration de simulateurs les plus réalistes possible. Dans le domaine de la robotique, des chercheurs de l'EPFZ ont récemment réussi à apprendre de nouvelles techniques de marche à leur nouveau robot marcheur « ANYmal » par le biais de l'apprentissage par renforcement⁷.

Mais que se passe-t-il lorsqu'aucune technique précise de simulation n'est disponible ? C'est le cas dans d'innombrables applications, de la réadaptation médicale à la planification des essais scientifiques en passant par l'optimisation des processus. Dans ce cas, tenter des actions dont l'issue est incertaine peut se révéler risqué. Un groupe de recherche développe de nouveaux processus d'apprentissage par renforcement

dans le cadre d'un projet soutenu par le Conseil européen de la recherche⁸. Il s'agit d'évaluer minutieusement l'insécurité dans les modèles d'apprentissage automatique, ce qui permet des prises de décision fiables que l'on peut prouver sous certaines conditions – un défi important pour la robotique et d'autres applications.

La Suisse en tête

La recherche dans le domaine de l'IA se recoupe avec d'autres domaines scientifiques comme la science des données. Au moyen de procédés algorithmiques et statistiques, ce domaine essaie d'obtenir des liens et des constats à partir de données complexes. Tout comme l'IA moderne, la science des données recourt essentiellement à l'apprentissage automatique. Les systèmes modernes utilisent en outre souvent d'autres procédés algorithmiques et statistiques ou découlant de la théorie des jeux et des techniques de régulation.

La Suisse est bien placée dans la recherche en IA. Les écoles polytechniques font de la recherche fondamentale depuis des décennies grâce à la large autonomie dont elles bénéficient et œuvraient déjà dans ce domaine alors

Le joueur de go professionnel Ke Jie a dû s'avouer vaincu face à un robot.



L'IA n'était pas encore un sujet d'actualité. La Suisse se place au premier rang international en matière d'impact (citations par travail de recherche). L'institut Idiap, à Martigny (VS), et l'Idisia, à Manno (TI), sont tous deux spécialisés dans l'IA et son transfert vers l'industrie. Les projets de recherche de l'EPFZ et de l'EPFL focalisent régulièrement l'attention lors des conférences internationales sur ce thème.

Par ailleurs, le domaine des EPF met systématiquement l'accent sur la science des données. Outre l'ouverture du Centre national de la science des données, de nouvelles chaires ont été ouvertes et de nouveaux programmes de master et de formation continue créés. Les universités et les hautes écoles spécialisées sont également actives dans le domaine de l'IA. L'université de Zurich étudie par exemple les conséquences de la numérisation et de l'IA sur la société dans le cadre de sa Digital Society Initiative. Des entreprises privées, dont les groupes technologiques Google, Microsoft, Disney et IBM, mènent des recherches sur l'IA en Suisse.

L'intelligence artificielle n'est pas un but en soi

L'IA a le potentiel d'une technologie transversale pouvant s'appliquer à toutes les branches. Avec l'apprentissage automatique, elle modifie la manière dont la recherche est réalisée. Il n'est donc pas étonnant qu'une concurrence mondiale sévère dans ce domaine très porteur. Pour que l'Europe ne prenne pas de retard par rapport aux États-Unis et à la Chine, un réseau de recherche, le Laboratoire européen pour les systèmes intelligents et d'apprentissage (« European Laboratory for Learning and Intelligent Systems », Ellis) a commencé à se former l'année dernière. La

Suisse possède de bonnes cartes face à la concurrence pour attirer les meilleurs professionnels, mais elle doit investir de manière conséquente dans ce secteur si elle veut défendre son rôle de leader dans la recherche. L'IA n'est pas un but en soi : l'innovation et son utilité pour la société découlent de la collaboration étroite entre la recherche fondamentale et les disciplines d'application telles que la médecine et la robotique.

Comme pour toute nouvelle technologie, les risques existent : des questions se posent concernant la fiabilité, l'interprétabilité ou la vulnérabilité face aux données faussées. La Suisse offre les conditions idéales pour gérer et piloter des données et des technologies d'IA de manière reconnue et fiable. Plusieurs projets de recherche se sont déjà penchés sur la question de savoir comment gérer des processus équitables, transparents et stables⁹. Des plateformes comme « AI for Good » (« l'intelligence artificielle au service du bien social ») des Nations Unies à Genève peuvent aider la Suisse à se positionner au niveau international et à identifier des champs d'application comme le développement durable et l'alimentation. La Suisse devrait donc saisir la chance qui se présente de modéliser l'avenir numérique, marqué par l'IA, de manière positive et fiable.

⁹ Voir notamment Berkenkamp et al. (2017), Moosavi-Dezfooli et al. (2018) et Goel et al. (2018).



Andreas Krause

Professeur d'informatique à l'Institut d'apprentissage automatique de l'École polytechnique fédérale de Zurich (EPFZ), codirecteur académique du Centre national de la science des données

Bibliographie

Berkenkamp Felix, Turchetta Matteo, Schoellig Angela et Krause Andreas (2017). « Safe model-based reinforcement learning with stability guarantees ». *Advances in neural information processing systems*, 908–918.

Brodersen Kay H., Deserno Lorenz, Schlagenhaut Florian, Lin Zhihao, Penny Will D., Buhmann Joachim M. et Stephan Klaas E. (2014). « Dissecting psychiatric spectrum disorders by generative embedding ». *NeuroImage: Clinical*, 4: 98–111.

Cireşan Dan C., Giusti Alessandro, Gambardella Luca M. et Schmidhuber Jürgen (2013). « Mitosis detection in breast cancer histology images with deep neural networks ». *International Conference on medical image computing and computer-assisted intervention*, 411–418.

Goel Naman, Yaghini et Mohammad Boi Faltings (2018). « Non-discriminatory machine learning through convex fairness criteria ». *Thirty-Second AAAI Conference on Artificial Intelligence*.

Gupta Anvita, Müller Alex T., Huisman Bernd J. H., Fuchs Jens A., Schneider Petra et Schneider Gisbert (2018). « Generative recurrent networks for de novo drug design ». *Molecular informatics*, vol. 37, N° 1-2: 1700111.

Hochreiter Sepp et Schmidhuber Jürgen (1997). « Long short-term memory ». *Neural computation*, vol. 9 N° 8: 1735–1780.

Hwangbo Jemin, Lee Joonho, Dosovitskiy Alexey, Bellicoso Dario, Tsounis Vasilios, Koltun Vladlen et Hutter Marco (2019). « Learning agile and dynamic motor skills for legged robots ». *Science Robotics*, vol. 4, N° 26: eaau5872.

Hyland Stephanie L., Hüser Matthias, Lyu Xinrui, Faltys Martin, Merz Tobias et Rättsch Gunnar (2018). « Predicting circulatory system deterioration in intensive care unit patients ». *AIH@IJCAI*, 87–92.

Moosavi-Dezfooli Seyed-Mohsen, Fawzi Alhussein, Uesato Jonathan et Frossard Pascal (2019). « Robustness via curvature regularization, and vice versa ». *Proceedings of the IEEE Conference on computer vision and pattern recognition*, 9078–9086.

Turing Alan (1950). « Computing Machinery and Intelligence ». *Mind*, 49 (1950): 433–460.