

UNIVERSITÉ
CÔTE D'AZUR



Psychologie des êtres artificiels
de l'Intelligence Artificielle
vers la Psychologie des êtres artificiels
#PsyArti

Journée d'étude IA et SHS, Université Côte d'Azur, 10 juin 2022

Pierre Crescenzo pierre.crescenzo@univ-cotedazur.fr <http://www.crescenzo.nom.fr/>

Comment les *Sciences Humaines et Sociales*
se saisissent-elles
de *l'Intelligence Artificielle* ?

Quelle part peut prendre l'informatique et
l'informaticien dans ce rapprochement ?

L'IA, qu'est-ce que c'est ? 1/2

Presque personne n'a la même définition de ce qu'est
l'Intelligence Artificielle.

*Et chacun trouve les définitions des autres un peu
inexactes, imprécises, maladroites, orientées, etc.*

Même entre tenants d'une même science. Alors, bien
sûr, en transversalité, c'est pire...

L'IA, qu'est-ce que c'est ? 2/2

- Définition du **Larousse** : *ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine*
- Mais c'est **restrictif**... Seuls les humains sont intelligents ? Seulement simuler ? ...
- **Intelligence**
 - fonctions mentales ayant pour objet la connaissance **conceptuelle** et **rationnelle**
 - aptitude à **s'adapter** à une situation, à **choisir** des moyens d'action en fonction des circonstances
- **Artificielle**
 - produit par le travail de l'homme, et non par la nature

Une définition : IA \approx une machine (ou une chose) non naturelle, qui paraît avoir (ou qui a) une intelligence (humaine ou comparable)

L'IA actuelle, à quoi ça sert ? 1/5

Ça peut lire et, en partie, comprendre.



En janvier 2018, deux IA ont battu des humains aux tests de compréhension de lecture de l'Université de Stanford.

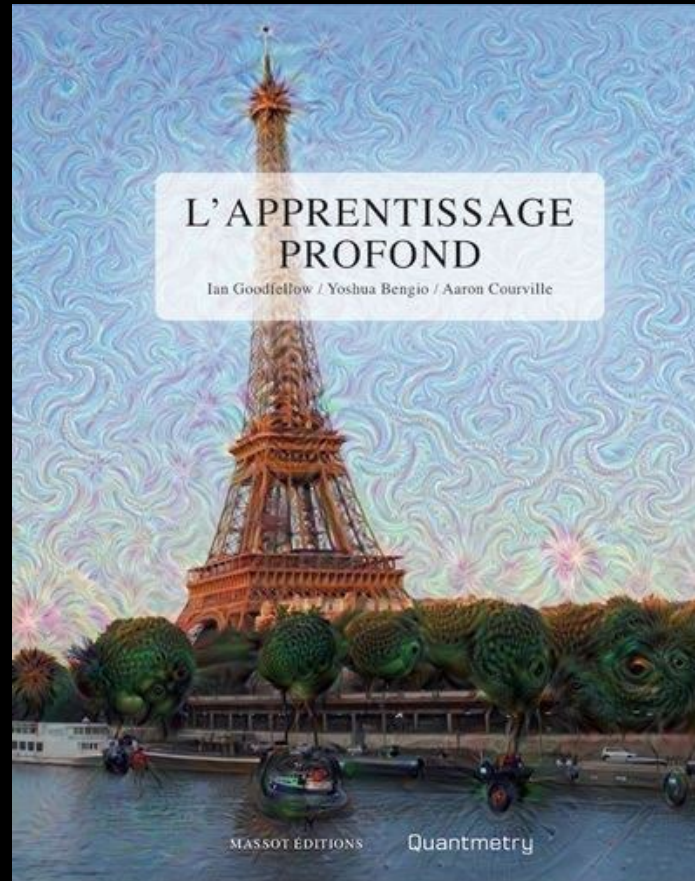
IA Microsoft : 82,65/100

IA Ali Baba : 82,44/100

humains : 82,30/100

L'IA actuelle, à quoi ça sert ? 2/5

Ça peut écrire, donc ça peut traduire.



En octobre 2018, un livre de 800 pages intégralement traduit en 12 h par une IA, de l'anglais vers le français, est sorti : *L'Apprentissage profond*.
La traduction a ensuite été relue et améliorée par des humains.

L'IA actuelle, à quoi ça sert ? 3/5

Ça peut reconnaître des choses sur des images.



Une IA peut faire la différence entre des chats et des chiens sur des images, après un apprentissage parfois long et difficile, supervisé ou non.

L'IA actuelle, à quoi ça sert ? 4/5

Ça peut reconnaître des choses sur des images, **ou pas**.



IA : Ceci est un **panda**,
j'en suis sûre à **57,7 %**.



modification
de la couleur
de quelques points



IA : Ceci est un **singe**,
j'en suis sûre à **99,3 %**.

Une IA peut se tromper, ou être trompée, là où un humain ne verrait aucune difficulté.

article en 2015 : *Explaining and Harnessing Adversarial Examples*
de Ian J. Goodfellow, Jonathon Shlens et Christian Szegedy

L'IA actuelle, à quoi ça sert ? 5/5

Ça peut reconnaître des choses sur des images, ou pas.



Stop



Stop



Stop



Limite
de
vitesse



Stop



Limite
de
vitesse

Une IA peut se tromper, ou être trompée, pas des défauts sur l'image.

article en 2018 : *Robust Physical-World Attacks on Deep Learning Visual Classification* de Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno et Dawn Song

L'IA, comment ça fonctionne ? 1/3

C'est, actuellement, principalement de l'Informatique et des Mathématiques.

Les systèmes experts :

- des faits
 - fait 1 : Le jour est clair.
 - fait 2 : La nuit est sombre.
 - fait 3 : Il fait soleil.
- des règles
 - règle 1 : S'il fait soleil alors il fait clair.
- un moteur qui agit
 - Je sais qu'il fait soleil (avec le fait 3).
 - Je peux donc en déduire qu'il fait clair (avec la règle 1).
 - Et je peux finalement en déduire qu'il fait jour (avec le fait 1).

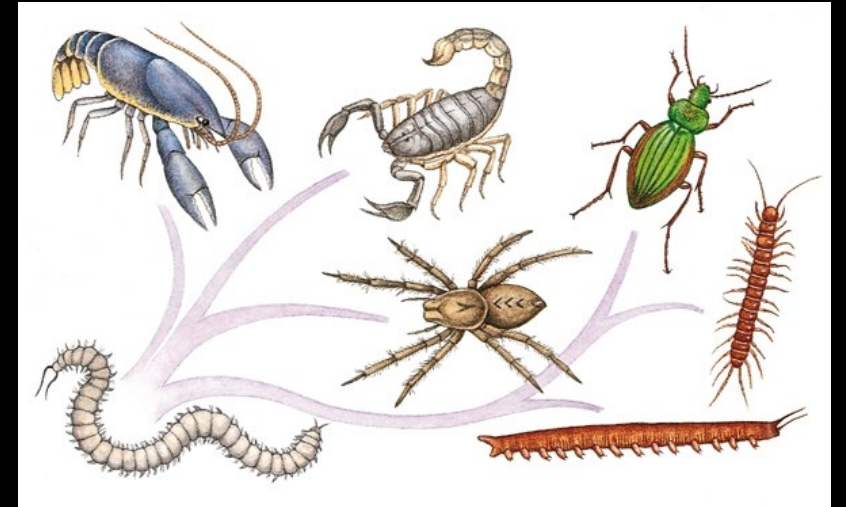
Une IA est ici un programme qui fait des raisonnements logiques.

L'IA, comment ça fonctionne ? 2/3

C'est, actuellement, principalement de l'Informatique et des Mathématiques.

Les algorithmes génétiques :

- des gènes binaires 0011010100111100111101111101
- un mécanisme de reproduction
- de la sélection et de la mutation



Une IA est ici un programme qui simule l'évolution génétique héréditaire d'individus dans une lutte pour survivre.

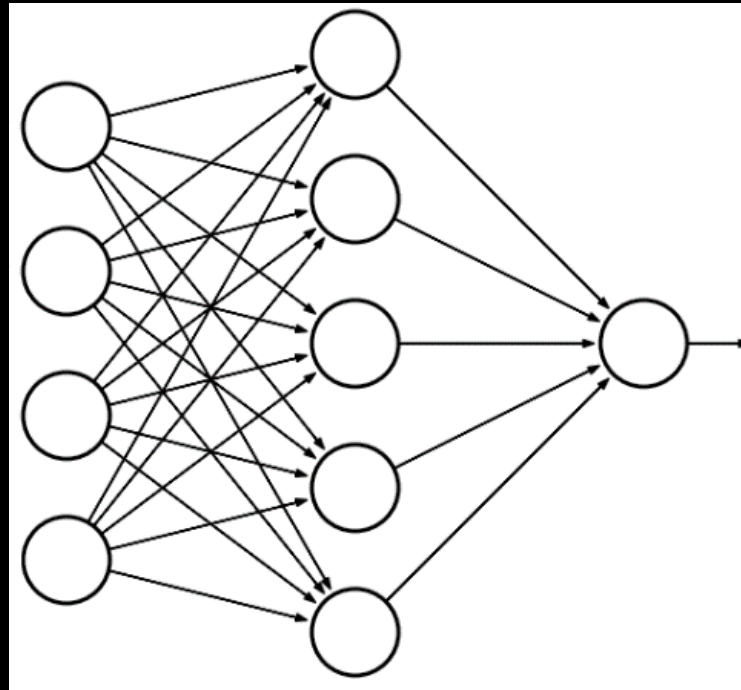
L'IA, comment ça fonctionne ? 3/3

C'est principalement de l'Informatique et des Mathématiques.

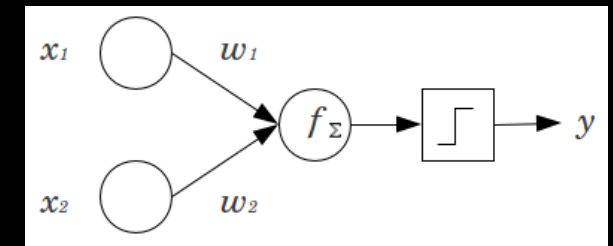
Les réseaux de neurones :

- une ou des entrées
- une ou des sorties
- de (nombreuses) couches de (nombreux) neurones

→ l'apprentissage profond
(*deep learning*)



exemple de neurone :



Une IA est donc ici un programme constitué d'un grand ensemble de fonctions mathématiques simples interconnectées, qui apprend et agit ensuite en fonction de ses apprentissages.

L'IA et ses métiers

C'est principalement de l'**Informatique** et des **Mathématiques**, mais pas seulement.

C'est aussi :

- du **Droit**
 - Les IA, les robots, etc. doivent-ils avoir des droits, des devoirs, des responsabilités ?
- de la **Psychologie**
 - Les IA, les robots, etc. sont-ils des machines préprogrammées ou ont-ils un comportement propre ?
 - S'ils sont agressifs ou idiots, comment améliorer cela ?
- de la **Médecine**
 - Comme pour beaucoup de métiers, les IA ou robots remplaceront-ils les médecins ?
 - Si une IA ou un robot a des bugs ou se comporte anormalement, peut-on le soigner ?
- de la **Sociologie**
 - Quelles sont les conséquences sur nos sociétés, dans nos entreprises, pour nos familles, de la présence de plus en plus forte des IA et des robots ?
- de la **Politique**
 - Les IA, les robots, etc. doivent-ils pouvoir se marier, voter, être traduits en justice... ?
- ...

L'Intelligence Artificielle, c'est beaucoup plus de questions que de réponses !

de l'Intelligence Artificielle
vers
les Êtres Artificiels

Que sont, ou qui sont, les êtres artificiels ?

Si définir l' **Intelligence Artificielle**
est **difficile** et sujet à questions,

définir les **Êtres Artificiels**
est **très complexe** et discutabile.

L'être artificiel, qu'est-ce que c'est ? 1/2

- Définitions du **Larousse** pour le mot **Être** :
 - tout ce qui vit et, spécialement, individu d'une espèce animale ; **créature**
 - personne, **individu**
 - **nature profonde** de quelqu'un
- Mais c'est **restrictif**... Seuls les animaux individuels sont-ils des êtres ?
- et **vague**... d'autres définitions font appel, outre aux concepts du vivant, à ceux d'intelligence, de conscience, de moral, de divin...
- **Artificielle** : produit par le travail de l'homme, et non par la nature

Une définition : Être Artificiel \approx création humaine sous forme d'individu ou groupe d'individus, disposant d'une certaine autonomie.

Exemples : intelligence artificielle, robot, objet connecté, logiciel, système, etc.

L'être artificiel, qu'est-ce que c'est ? 2/2

On ne cherche pas à savoir **a priori** si les êtres artificiels sont :

- **vivants** ou pas,
- **intelligents** ou pas,
- **conscients** ou pas.

En revanche, les thèmes du vivant, de l'intelligence, de la conscience, etc. sont **dans le champ de cette recherche**.

Nous considérons que **les êtres artificiels sont créés par les humains (ou par d'autres êtres artificiels)** hors des questions intéressantes de religion, divinité, spiritualité, métaphysique, philosophie, etc. que cette création pose.

Bien que pouvant soulever des questions parfois proches de celles abordées dans le transhumanisme, notre démarche est une étude scientifique a priori indépendante.

Être (artificiel et intelligent) ou ne pas être ?

Est-il nécessaire d'être intelligent pour être ?

La réponse est dans la question...

Tout être artificiel n'est pas intelligent.

Et ce sont les plus intelligents, ou les moins inintelligents, qui nous intéressent particulièrement.

Être (artificiel et vivant) ou ne pas être ?

Est-il nécessaire d'être vivant pour être ?

La réponse est là aussi dans la question...

Plus intéressant : est-il nécessaire d'être vivant pour être membre d'une, ou associé à, une société humaine ?

Actuellement, les êtres artificiels ne sont pas vivants.

Ils le seraient, au sens biologique du terme, que cela ne remettrait pas en cause ce thème de recherche, au contraire.

Être (artificiel et conscient) ou ne pas être ?

Est-il nécessaire d'être conscient pour être ?

*La réponse n'est pas vraiment dans la question...
car être un objet n'est pas vraiment être.*

Les êtres artificiels actuels ne sont pas conscients, même quand ils disposent d'éléments constitutifs d'une conscience.

Ceux qui seraient conscients, vivants ou pas, intelligents ou pas, posent des questions très intéressantes dans cette étude.

Les êtres artificiels et la conscience ^{1/2}

Le terme **conscience** a au moins deux acceptions :

- le fait pour un être de posséder une **représentation, même très simplifiée, de certaines activités qu'il rapporte à lui-même**. On parle alors plutôt de conscience de soi, de sa propre existence (*consciousness* en anglais).
- le fait pour un être de posséder une **représentation, même très simplifiée, du monde extérieur** et d'y réagir. On parle alors plutôt de conscience du monde, du contexte (*awareness* en anglais).

Factuellement, on peut dire que les êtres artificiels (*programmes informatiques*) ont :

- indiscutablement, une conscience du monde : prosaïquement, c'est la mémoire du programme informatique, modélisé grâce à ses structures de données,
- de façon moins systématique et évidente, une conscience de soi : certains programmes informatiques disposent d'un niveau méta incluant une représentation parfois assez poussée du programme lui-même, de ses objets en mémoire, de son fonctionnement.

Conscient ! Inconscient ! Quantique ?

Les êtres artificiels et la conscience 2/2

Des questions courantes concernant la **conscience** se posent aussi, voire plus, pour les êtres artificiels.

- Existe-t-elle vraiment ? N'est-elle pas seulement un concept humain ?
- Quelle est sa nature ? La même que pour les humains ?
- Quelle est son origine ? La même que pour les humains ?
- Comment se constitue-t-elle et comment se développe-t-elle ?
- Quel est son mode d'existence ?
- Comment peut-elle exister à partir d'entités non-conscientes ?
- Quelles sont ses propriétés ?
- Quelle est sa fonction ?
- A-t-elle une causalité propre et, si oui, de quelle nature ?
- Quelles relations la conscience a-t-elle avec les autres phénomènes de la réalité, physiques et mentaux ?
- ...

des Êtres Artificiels

vers

La Psychologie des êtres artificiels

La psychologie des êtres artificiels... 1/5

On a, plus ou moins, admis le concept des êtres artificiels...

Mais que peut-être une **psychologie** de ces êtres ?

- Ils sont. (*On n'y peut rien et on ne demande rien de plus.*)
- Ils **agissent et interagissent**
 - avec les humains,
 - avec les autres êtres naturels,
 - entre eux.
- Ils ont donc, de fait, des **caractéristiques**, des **comportements**, des **manières de faire**, des **postures**, des **traits apparents de personnalité**...

Bref... pas forcément un esprit ou un psychisme (*en tous cas aux sens où on l'entend traditionnellement*) mais des éléments constitutifs d'une **psychologie**, qu'elle soit **intrinsèque** ou simplement **ressentie**.

La psychologie des êtres artificiels... 2/5

Les êtres artificiels ont, ou pourraient avoir, une psychologie.

Et alors ?

On peut **étudier** leurs comportements, leurs traits de personnalité... *avec une bonne dose d'anthropomorphisme, forcément, mais on peut aussi essayer de s'en extraire.*

Rien que sur ces aspects, les chantiers sont innombrables !

Donc, la psychologie des êtres artificiels... 3/5

Les êtres artificiels ont, ou pourraient avoir, une psychologie.

Et alors ?

On peut **faire le lien entre leur psychologie et leur programmation** (*je suis informaticien* 😊) :

- ✓ Quelles méthodes, quelles techniques, quels algorithmes, quels modules, **quels objets informatiques ont amené ces comportements** (souhaités ou non) ?
- ✓ Quelles méthodes, quelles techniques, quels algorithmes, quels modules, **quels objets informatiques doivent être utilisés pour créer un comportement** particulier ?
- ✓ **Comment corriger un comportement anormal ou malvenu ?** Par (re)programmation ? En convainquant (nouvel apprentissage) l'être artificiel ? En considérant que l'on doit le soigner et non le modifier ? etc.

Donc, la psychologie des êtres artificiels... 4/5

Les êtres artificiels ont, ou pourraient avoir, une psychologie.

Et alors ? la suite...

On peut **s'interroger** sur leurs comportements et ce qu'ils impliquent ou déclenchent auprès des autres.

- Relèvent-ils du **choix**, et de la **responsabilité** de l'être artificiel ? *Avec les implications juridiques et sociétales qui en découlent.*

Certains pays permettent à des êtres artificiels (robots) d'être assermentés (avec pouvoirs de police ou de justice), voire d'acquérir la nationalité avec des droits (vote, démission) et devoirs (responsabilités, civisme).

- Doivent-ils être considérés comme **porteurs de sentiments** ou au moins **sujets de sentiments** par des êtres naturels, voire artificiels ?

Les robots ou objets sensoriels, réconfortants, sexuels, etc. sont communément acceptés. Certains pays envisagent la possibilité de se marier avec un robot ou une IA.

Donc, la psychologie des êtres artificiels... 5/5

Et alors ? et aussi...

Que l'on considère que ces objets sont des êtres ou pas, qu'ils ont une psychologie ou pas, on peut **s'émerveiller** ou **s'inquiéter** de certains de leurs comportements.

- Les êtres artificiels peuvent être, ou être perçus comme, désagréables, dérangeants, agressifs, etc. Ils peuvent l'être accidentellement, mais aussi avoir été conçus pour cela.
- Ils peuvent être spécifiquement conçus pour espionner, contraindre, attaquer, blesser, tuer... Certains sont explicitement des armes.
- Ils sont supposés être toujours sous un contrôle humain. S'ils ont été conçus ainsi... S'il n'y a pas d'erreur... Si tout se passe toujours comme prévu...
- Mais, dans les faits, ils sont très souvent présents de façon positive et aidante dans nos vies, parmi nous. **Et, surtout, la question ne se pose pas, ne se pose plus : ils sont...**

Les étudier sans préjugé, les comprendre sans fantasme ni aveuglement, c'est apprendre : à les concevoir, à s'en protéger, à vivre avec eux, à les intégrer, peut-être voir un jour émerger de nouveaux taxons...

Alors, les êtres artificiels ?

Définir l' **Intelligence Artificielle**
est **difficile** et sujet à questions et

étudier les **Êtres Artificiels**
est **très complexe** et passionnant !

UNIVERSITÉ
CÔTE D'AZUR



Psychologie des êtres artificiels
de l'Intelligence Artificielle
vers la Psychologie des êtres artificiels
#PsyArti

Journée d'étude IA et SHS, Université Côte d'Azur, 10 juin 2022

Pierre Crescenzo pierre.crescenzo@univ-cotedazur.fr <http://www.crescenzo.nom.fr/>