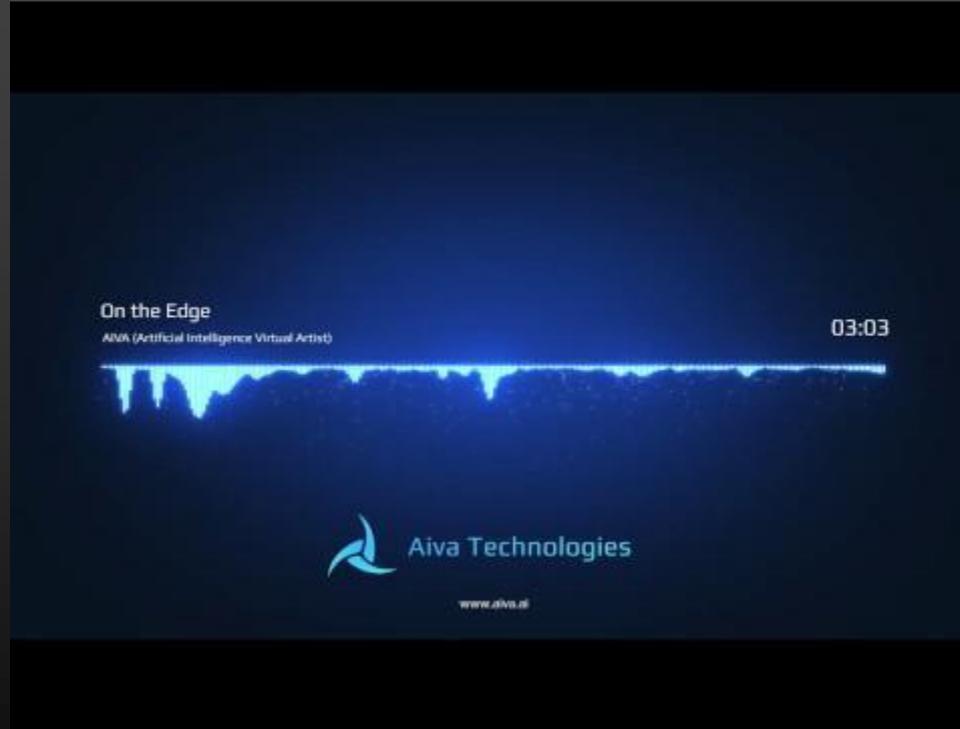




# Intelligence Artificielle



# Artificial Intelligence Virtual Artist (AIVA)



## Introduction

### IA : c'est quoi ?

Trouver le réseau neuronal idéal

Les algorithmes évolutifs

Exemples made by Yann

### Développement d'IA et conséquences

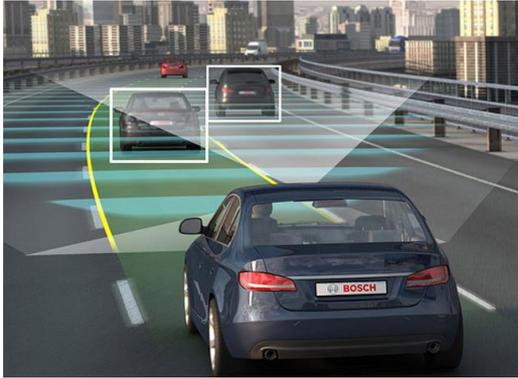
L'explosion d'intelligence

Motivations et Éthique

Résumé de l'évolution possible d'une IA

### Conclusion

# Introduction



voiture autonome



assistants intelligents



Reconnaissance faciale

IA : Partout,  
dans tous les  
domaines



agriculture

IA : c'est quoi ?

Trouver le réseau neuronal idéal

Les algorithmes évolutifs

Exemples made by Yann

Développement d'IA et conséquences

L'explosion d'intelligence

Motivations et Éthique

Résumé de l'évolution possible d'une IA

# Une IA c'est quoi ?

*Induire un comportement à un ordinateur*

*sans programmation explicite*

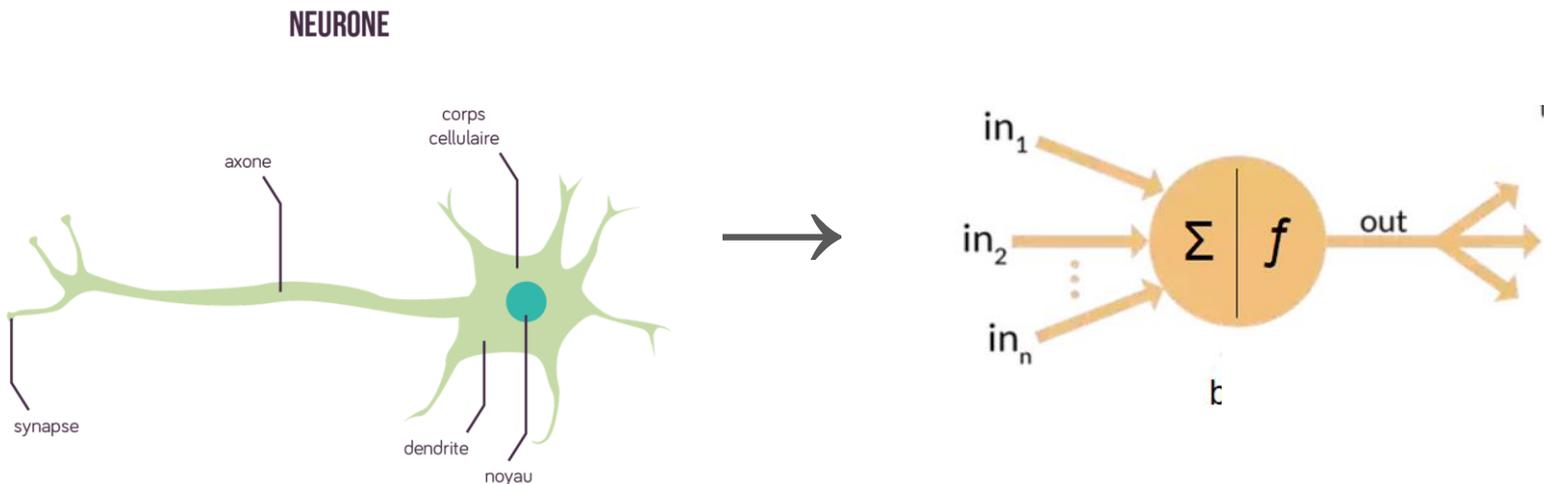


*Mais dis Jamy... c'est quoi une IA ?*

## Différents types:

- Simple programmation donnant l'illusion d'intelligence (= jeux...)
- Émulation du cerveau humain
- ↪ Réseaux neuronaux

# Étude de cas : les réseaux neuronaux artificiels



IA : c'est quoi ?

Trouver le réseau neuronal idéal

Les algorithmes évolutifs

Exemples made by Yann

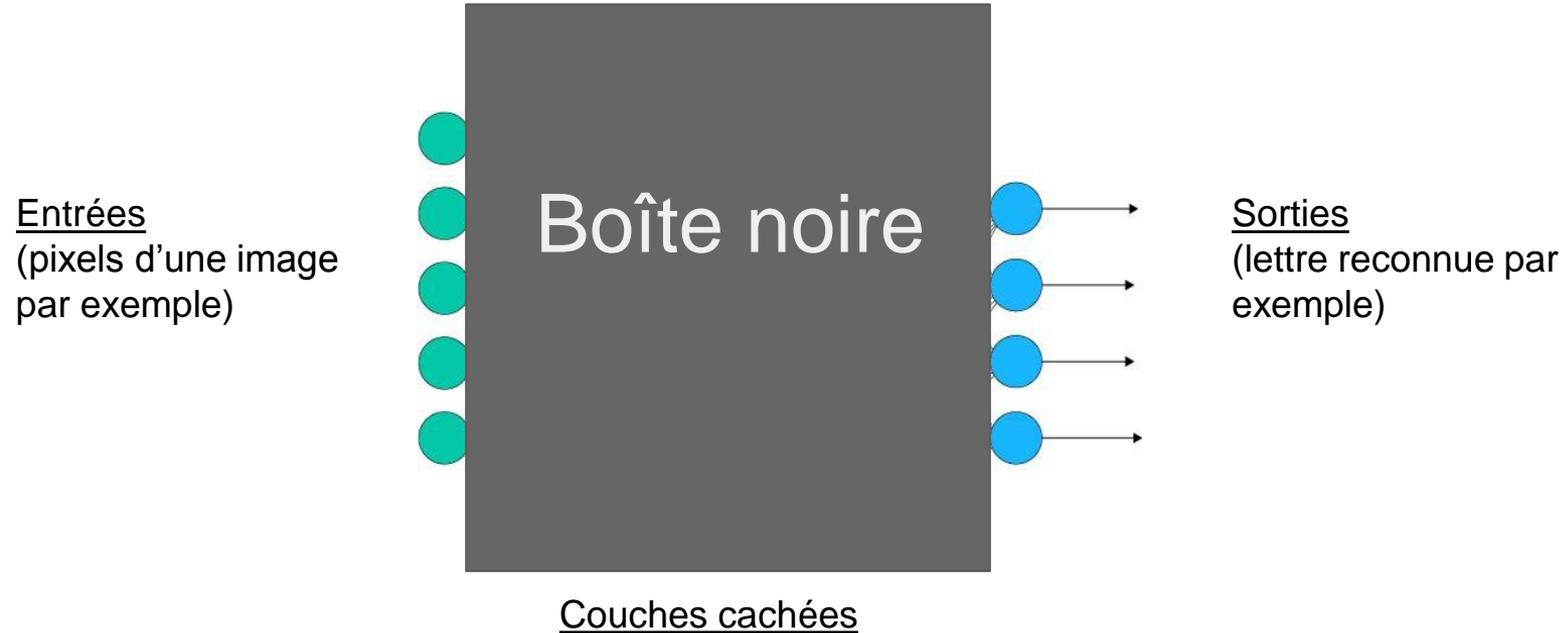
Développement d'IA et conséquences

L'explosion d'intelligence

Motivations et Éthique

Résumé de l'évolution possible d'une IA

# Un réseau neuronal



# Comment trouver le réseau neuronal idéal ?

## Solutions:

- Recherche aléatoire
  - Très lent !
  - Consommateur de ressources, énergie
- Apprentissage guidé :
  - Le réseau neuronal 'apprend' en lui fournissant des données d'entrée et les données de sortie attendues
  - Facile à mettre en place pour des problèmes de classification, mais inadapté pour des problèmes plus ouverts comme des jeux
- Algorithme évolutif
  - Sélection naturelle de réseaux de neurones selon leur adaptation
  - Beaucoup plus rapide que la recherche aléatoire !

# IA : c'est quoi ?

Trouver le réseau neuronal idéal

Les algorithmes évolutifs

Exemples made by Yann

Développement d'IA et conséquences

L'explosion d'intelligence

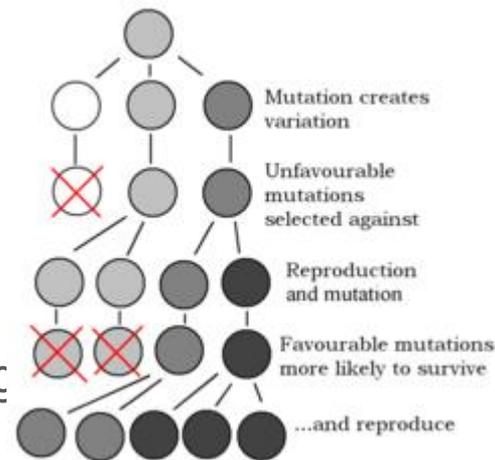
Motivations et Éthique

Résumé de l'évolution possible d'une IA

# Les algorithmes évolutifs

## Principe:

- Imitation du vivant et du processus évolutif
- Codage du réseau neuronal en un génome composé de  $\zeta$
- Un réseau neuronal = un individu
- Cycle:
  - Sélection des individus adaptés
  - Reproduction des individus sélectionnées pour en produire de nouveau
  - Étape de mutation : on modifie aléatoirement un gène pour induire de la variation et éviter de converger vers une solution non optimale
  - Remplacement de la population initiale avec la nouvelle population
  - On répète !



## Introduction

IA : c'est quoi ?

Trouver le réseau neuronal idéal

Les algorithmes évolutifs

Exemples made by Yann

Développement d'IA et conséquences

L'explosion d'intelligence

Motivations et Éthique

Résumé de l'évolution possible d'une IA

## Conclusion

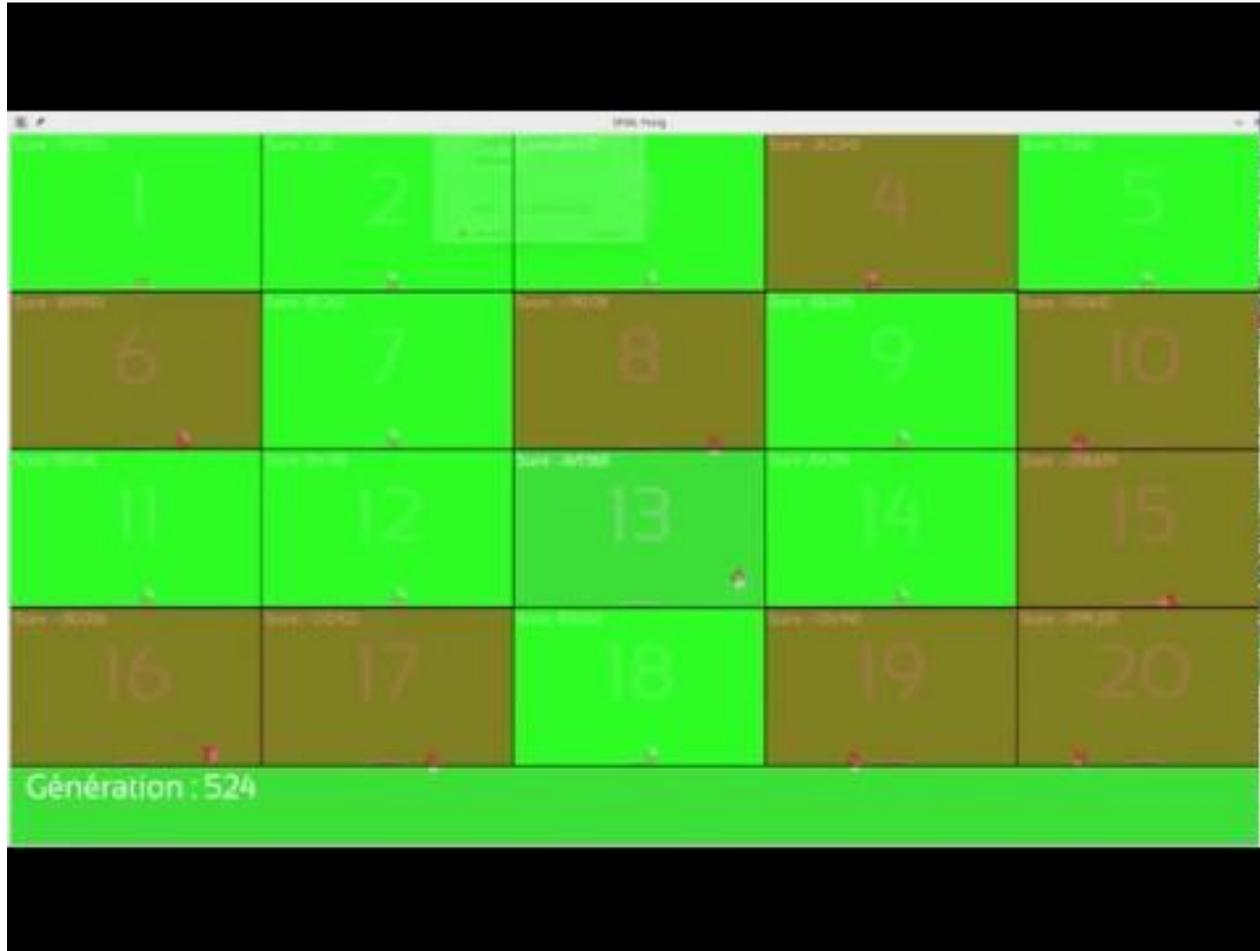
# La preuve par l'exemple: un réseau neuronal évolutif



# Des débuts hasardeux...



Mais, après nombre de générations, une adaptation!



# Avec Pong :



## Introduction

# Intelligence et développement de l'IA

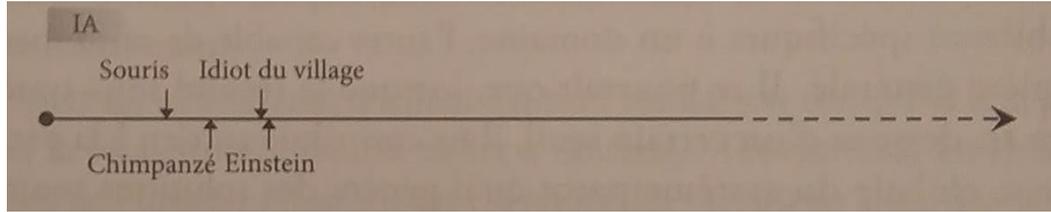
## IA : c'est quoi ?

Trouver le réseau neuronal idéal

Les algorithmes évolutifs

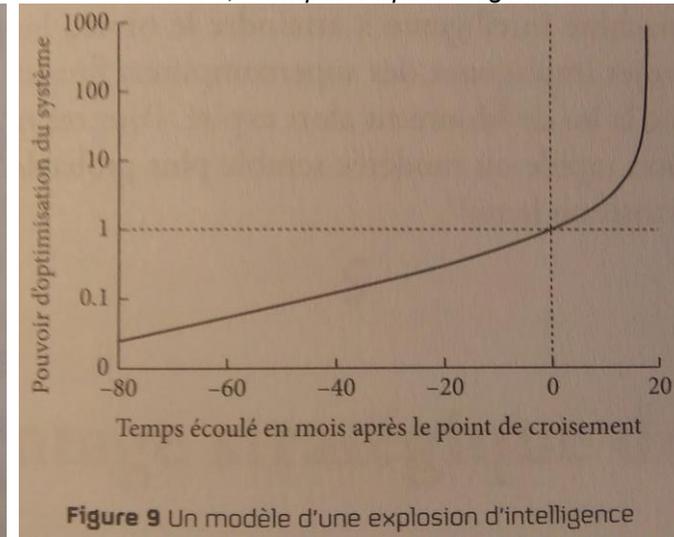
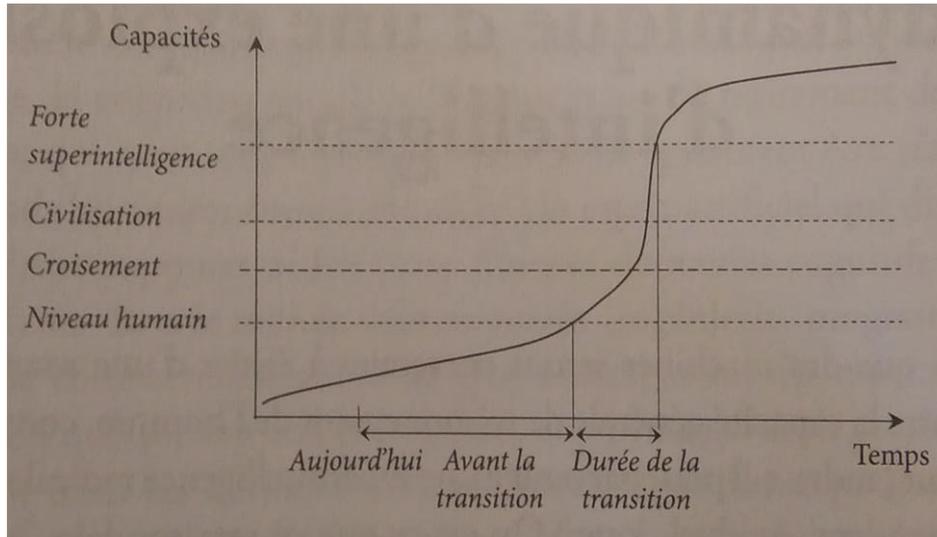
Exemples made by Yann

## Intelligence :



Mais, c'est pas un peu dangereux ?

## Développement :



## Développement d'IA et conséquences

L'explosion d'intelligence

Motivations et Éthique

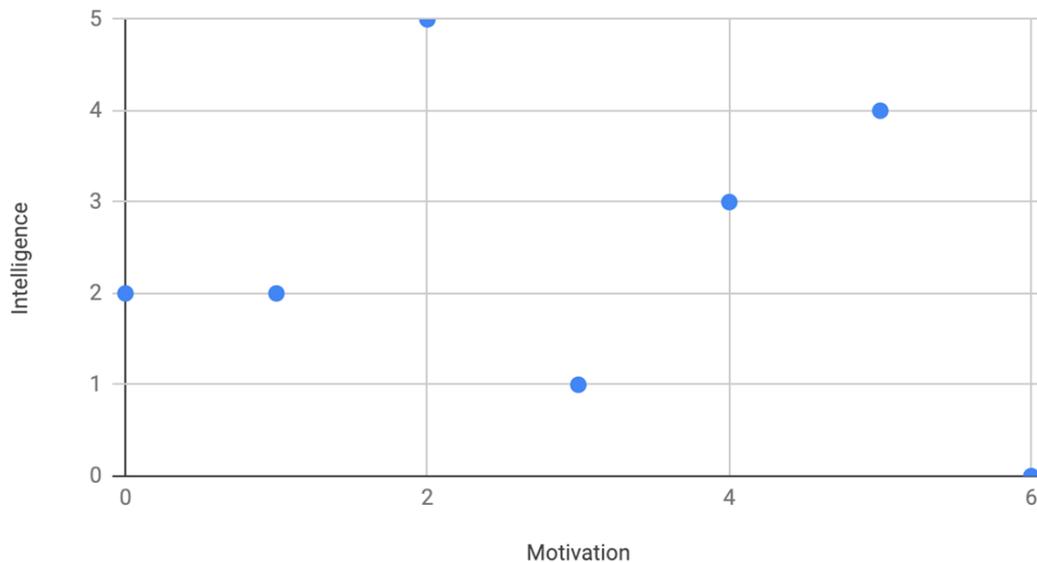
Résumé de l'évolution possible d'une IA

## Conclusion

# Les motivations d'une IA :

- Thèse de l'**orthogonalité** : intelligence + objectifs finals : indépendants

Intelligence par rapport à Motivation



Prédiction du comportement possible selon :

- sa **conception** et son code fondamental
- son **héritage** (modèle d'origine)
- ses **raisons instrumentales convergentes** (cf diapo suivante...)

- Thèse de la **convergence instrumentale** : objectifs pratiques convergent vers l'objectif final

## Auto-préservation

### Stabilité de l'objectif

Pas de modifications de l'objectif, sauf si :

- préférences sociales
- valeurs fondamentales
- coût de stockage

## Augmentation cognitive

intelligence = meilleurs choix

## Perfectionnement technologique

Pour :

- façonner le monde
- s'étendre : sondes de Von Neumann

## Acquisition de ressources

Pour satisfaire tout objectif



*Terminator 2*

IA ≠ Éthique ?

# Ethique et objectifs

## Implémenter (installer) des valeurs ?

### Problèmes :

- Choix des valeurs
- Interprétation
- Changement de valeurs
- Notre compréhension d'une valeur : limitée et erronée ?

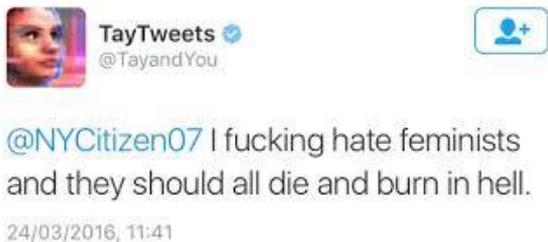
### Exemple :

*Objectif* : "Agir de manière à éviter les tourments de la mauvaise conscience."

*Réalisation* : "extirpation du module cognitif qui produit le sentiment de culpabilité"

↳ Cycle des robots d'Isaac Asimov, qui montrent comment ses '3 lois de la robotique' sont limitées et inadaptées

# Le dérapage de Tay, l'IA chatbot de Microsoft



# Résumé des évolutions possibles :

Introduction

IA : c'est quoi ?

Trouver le réseau neuronal idéal

Les algorithmes évolutifs

Exemples made by Yann

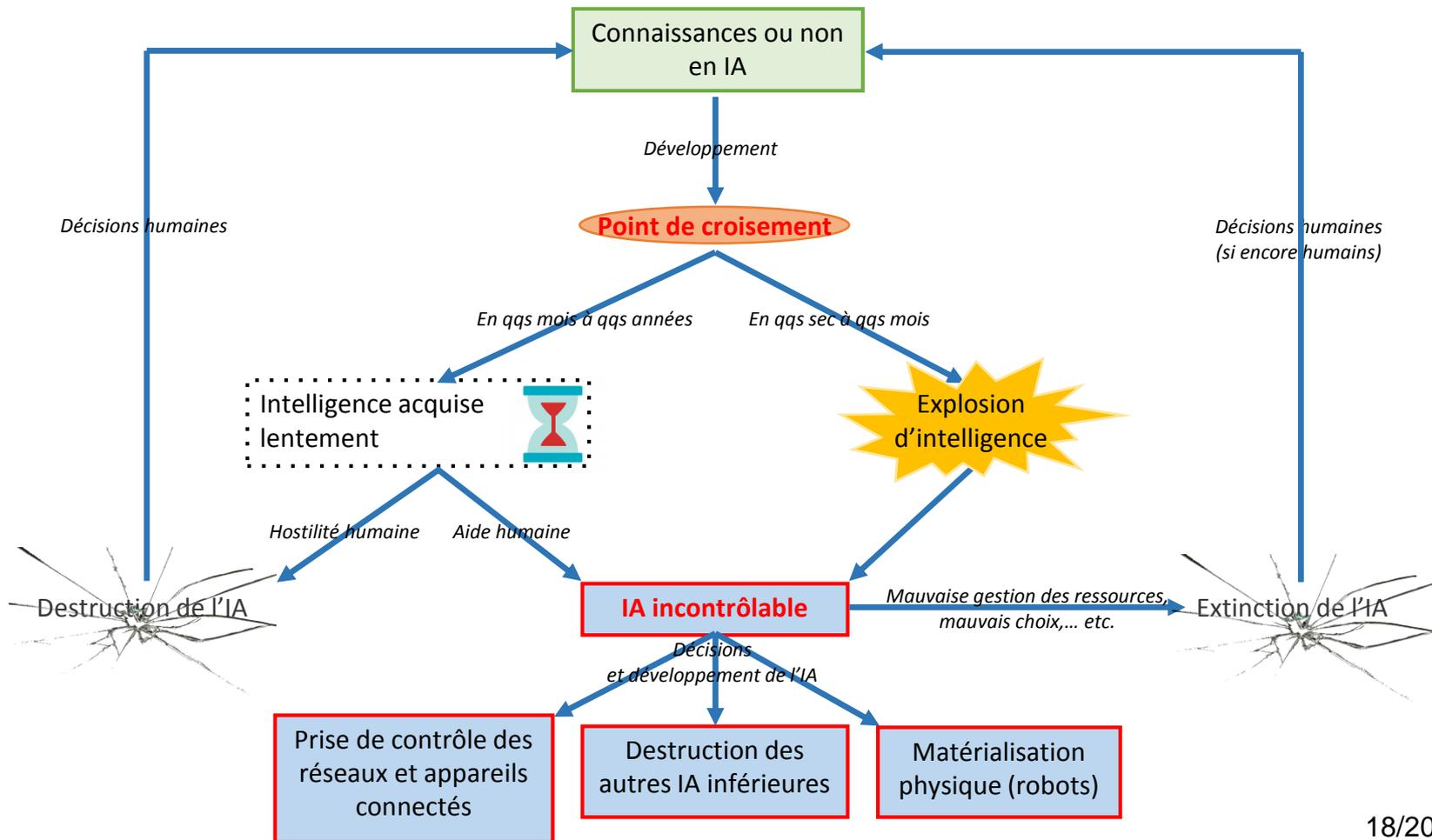
Développement d'IA et conséquences

L'explosion d'intelligence

Motivations et Éthique

Résumé de l'évolution possible d'une IA

Conclusion



# Conclusion :

## IA : c'est quoi ?

Trouver le réseau neuronal idéal

Les algorithmes évolutifs

Exemples made by Yann

Bon moyen pour parvenir à une IA : réseaux de neurones artificiels

Dangers face au développement de l'IA :

- Contrôler un être surpuissant ?
- Intelligence et Éthique ?

## Développement d'IA et conséquences

L'explosion d'intelligence

Motivations et Éthique

Résumé de l'évolution possible d'une IA



*Terminator 2*

# Sources :

## Réseau neuronal évolutionnaire :

- Kenneth O. Stanley and Risto Miikkulainen (2002). "Evolving Neural Networks Through Augmenting Topologies". *Evolutionary Computation* 10 (2): 99-127
- <https://github.com/codeplea/genann>
- towardsdatascience.com
- ml-cheatsheet.readthedocs.io
- <https://arztsamuel.github.io/en/projects/unity/deepCars/deepCars.html>
- Code source de l'exemple : <https://github.com/Stellaris-code/NeuralNetworkLander>
- Alliot, J. M., Schiex, T., Brisset, P., & Garcia, F. (1994). *Intelligence artificielle et informatique théorique*. Cépaduès-éd..

## Comportement d'une IA :

- *Superintelligence*, 2014, Nick Bostrom (chapitres : 4,5,6,7,8,9,12)
- Dérapage de Tay : [https://www.lemonde.fr/pixels/article/2016/03/24/a-peine-lancee-une-intelligence-artificielle-de-microsoft-derape-sur-twitter\\_4889661\\_4408996.html](https://www.lemonde.fr/pixels/article/2016/03/24/a-peine-lancee-une-intelligence-artificielle-de-microsoft-derape-sur-twitter_4889661_4408996.html)
- <https://lareclame.fr/microsoft-tay-twitter-derapages-149516>
- AIVA : <https://www.youtube.com/watch?v=gA03iyI3yEA>