Sciences des données, IA et Biologie

Thomas Schiex

Qui suis-je ?

- Chercheur INRAE, ingénieur informatique, thèse IA (91)
- Élu Fellow EurAI (2016), Fellow AAAI (2020)
- Comité “Recherche” de #FranceIA 

- En IA:
  - raisonnement automatique logique et probabiliste

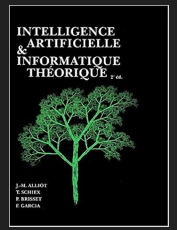
- En Biologie:
  a. Génétique
  b. Génomique
  c. Biologie structurale

C’est quoi l’IA ?

Le nom donné à un ensemble d’algorithmes et de techniques dont le but est de reproduire les capacités cognitives ou sensorielles des êtres vivants dans des machines électroniques.

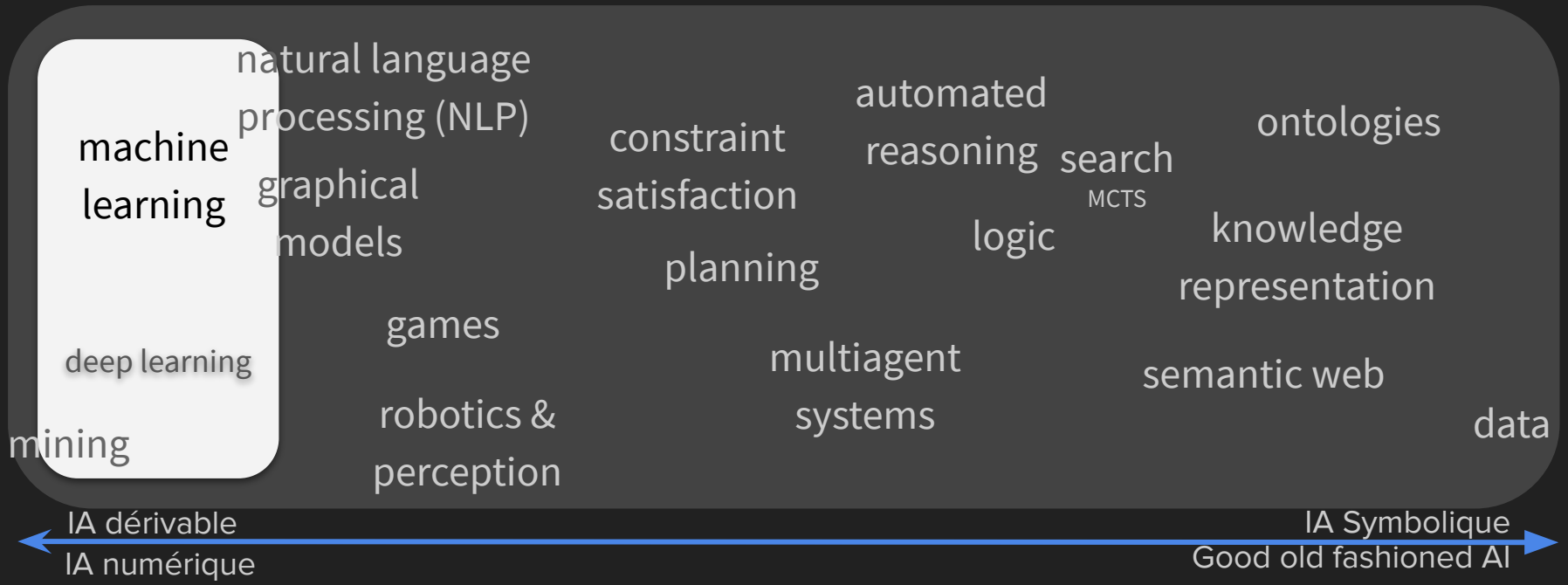
(Académie des sciences)

L’intelligence artificielle a pour but de faire exécuter par l’ordinateur des tâches pour lesquelles l’homme, dans un contexte donné, est aujourd’hui meilleur que la machine.



La calculette, c’était aussi de l’intelligence en son temps… cf. Littell, Eliakim, and Robert S. Littell, eds. Littell's Living Age. Vol. 34. TH Carter & Company, 1852.(cherchez “arithmometer”)

L'IA comme domaine de recherche



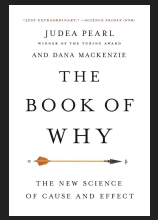
Des liens intimes avec les statistiques, l'automatique, la recherche opérationnelle, l'économie, la sociologie, les science cognitives

Abramson, Bruce (1987). The Expected-Outcome Model of Two-Player Games. PhD Thesis, Department of Computer Science, Columbia University. (director: Richard E. Korf)

Data sciences

L'ingénierie (noble) formée par la conjonction de:

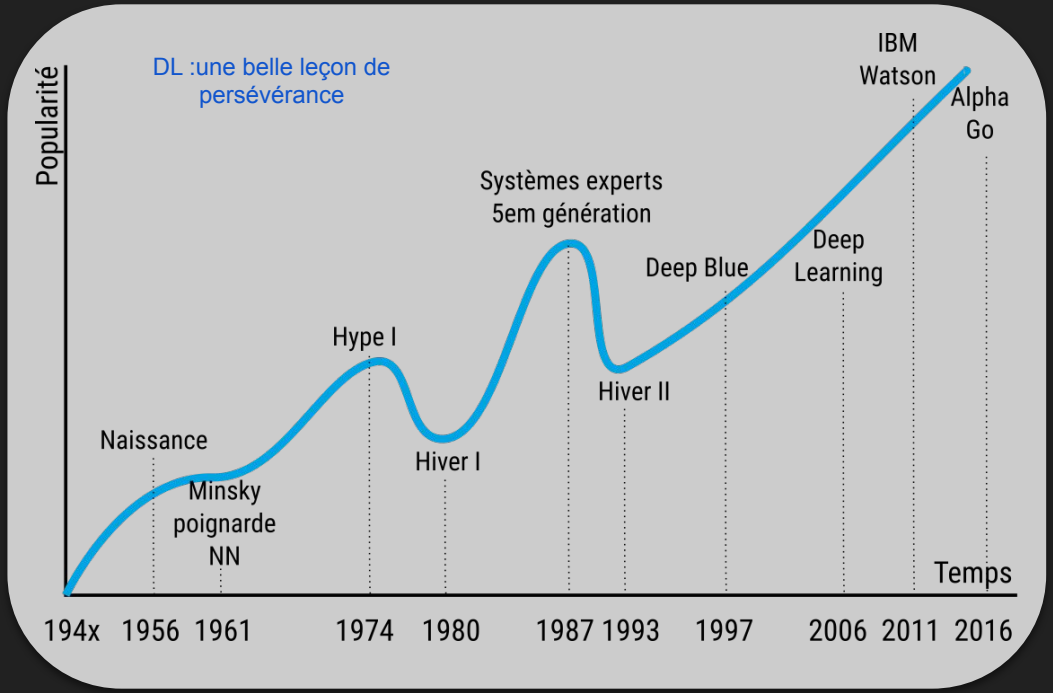
- Big data (volumineux, variés,... => ontologies, web sémantique,...)
- Data mining
- Machine learning (NLP, Deep Learning)
- Logique (hypothèses, abduction, causalité) et incertain



Assez bien développé en Biologie:

- Will AI write the Scientific Papers of the Future? (Yolanda Gill, AAAI 2020)

Le buzz de l'IA.... et les saisons



"AI will be either the best, or the worst thing, ever to happen to humanity."

– Stephen Hawking

"AI is the new electricity.

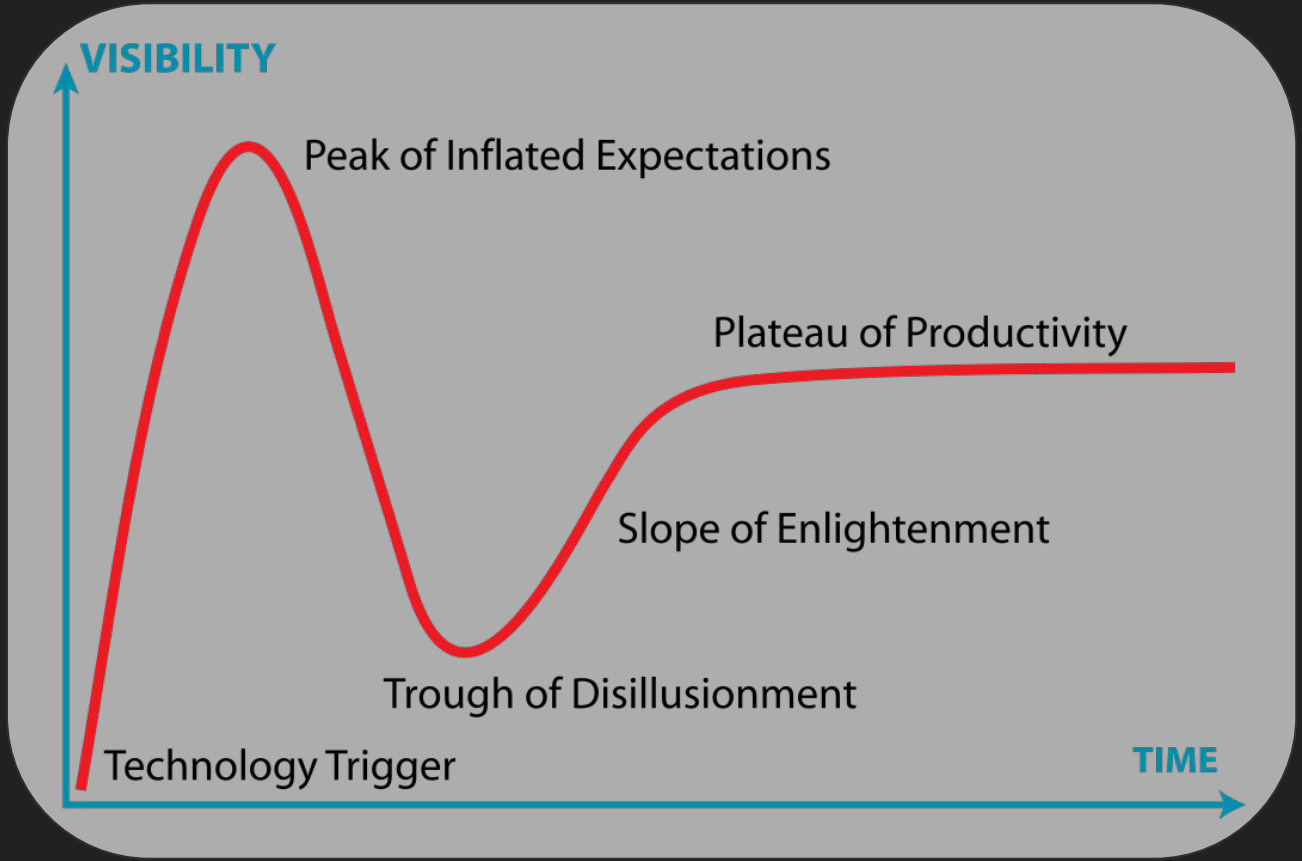
Pretty much anything that a normal person can do in <1 sec, we can now automate with AI."

– Andrew Ng

https://rodneybrooks.com/my-dated-predictions

https://twitter.com/andrewyng/status/788548053745569792

Le (dernier ?) été de l'IA et le cycle de Gartner



Amplifié par la remarquable facilité de mise en œuvre sur des cas bien explorés (analyse du signal: parole, écrit (NLP), image).

https://techtv.mit.edu/videos/10268-the-thinking-machine-1961---mit-centennial-film (à regarder en entier)

But I think that computer will be doing the things that men do when we say they are thinking. I'm convinced that machine can and will think in our lifetime.

Mais je pense que l'ordinateur fera ce que les hommes font quand on dit qu'ils pensent. Je suis convaincu que la machine peut et va penser de notre vivant.

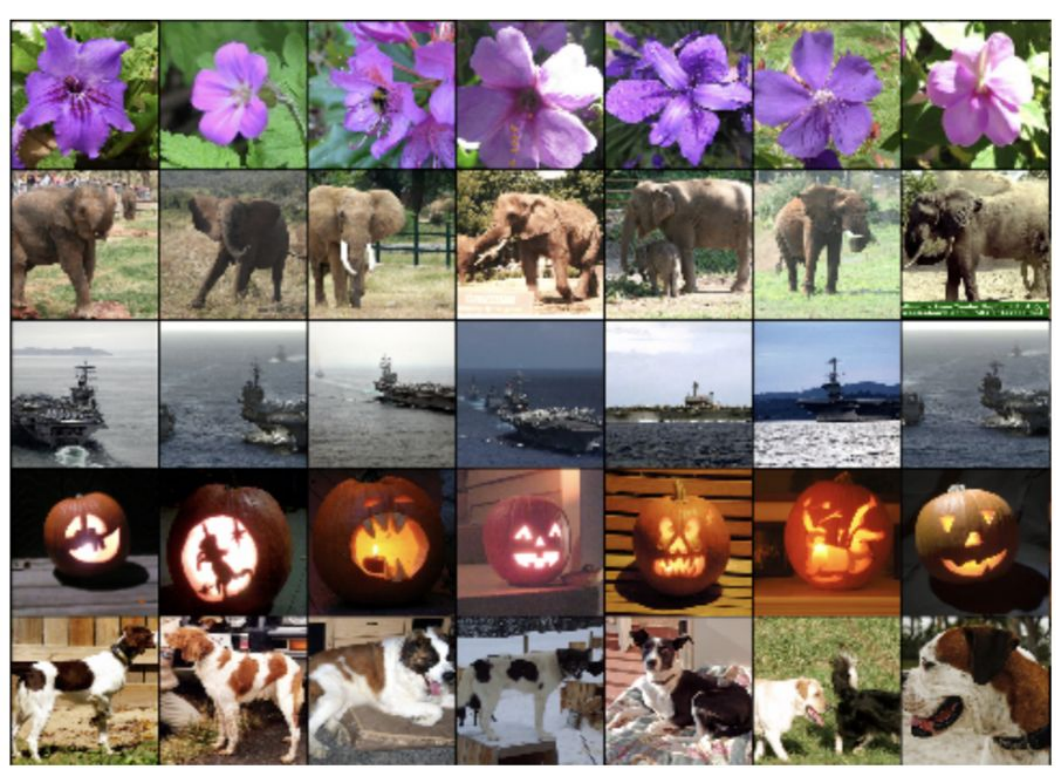
Oliver Selfridge (MIT, mort en 2008) est l'auteur d'importants articles sur les réseaux de neurones et l'apprentissage automatique.

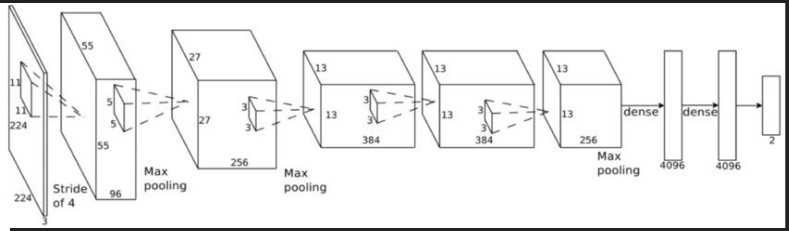
I confidently expect that within a matter of 10 or 15 years something will emerge from the laboratory which is not too far from the robots in science-fiction things.

Je m'attends avec confiance à ce que dans les 10 ou 15 prochaines années, quelque chose sorte du laboratoire qui n'est pas très éloigné des robots en matière de science-fiction.

Claude Shannon (1916-2001) est le père de la "Théorie de l'information.

Les progrès sur ImageNet (2012), CIFAR,...





62 378 344 paramètres

ResNet >100 couches
< 4% erreur (Top 5)

13.1 milliards de FLOPs &
des limitations énergétiques croissantes

GPT(1/2/3) en NLP

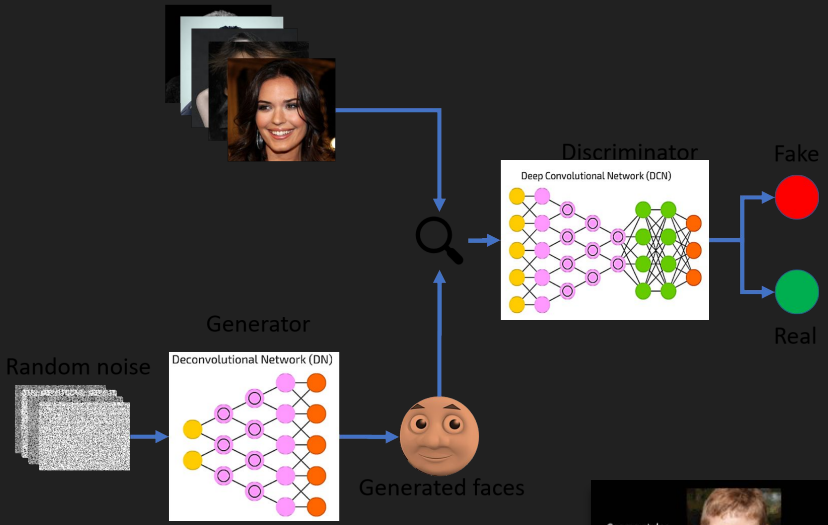
Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Modern Deep Learning Research." *AAAI*. 2020.

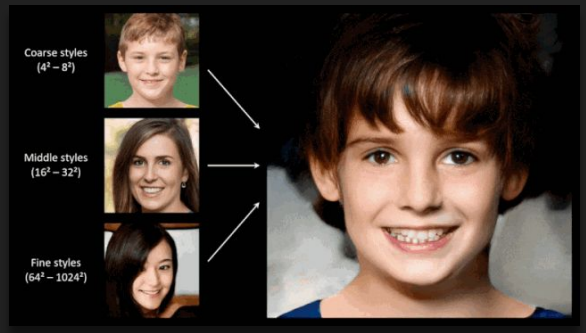
http://www.image-net.org/

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Avec de la créativité sur les “branchements”



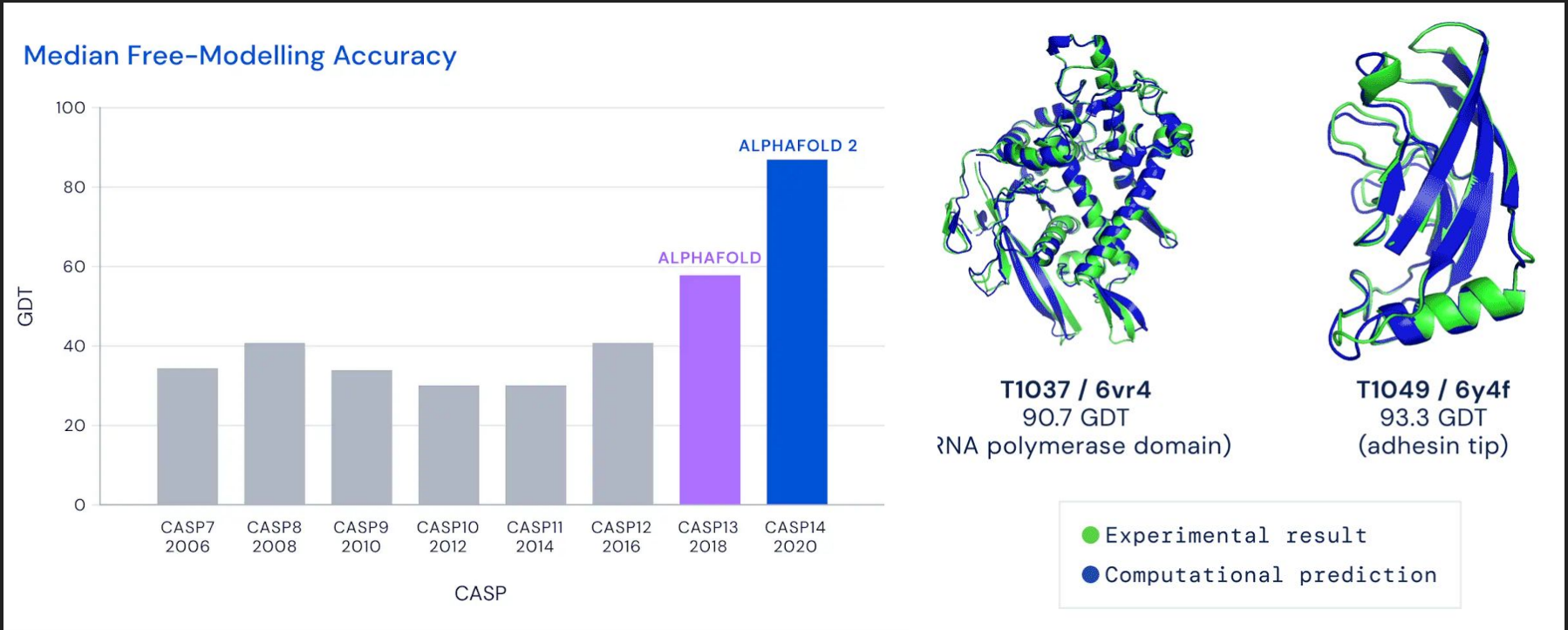


Apprentissage très difficile et coûteux
Solutions hybrides (Alpha-Go)

Generative adversarial network
Style-GAN, LS-GAN, W-GAN,…
(De)convolutional network
Residual network
Encoder-decoder (variational)
Recurrent network
Long Short Term Memory
Gated recurrent unit
Attention layer
Neural Turing machine
Memory enhanced network
Graph-based network
…
Descente en gradient stochastique

Karras, T., Laine, S., & Aila, T. (2018). A style-based generator architecture for generative adversarial networks. *arXiv preprint arXiv:1812.04948*.

Alpha Fold 1 et 2 à CASP



Mais des limitations qui résistent au temps

La tanche

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199.
Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet: https://openreview.net/forum?id=SkfMWhAqYQ, ICLR 2019
https://cacm.acm.org/news/221108-artificial-intelligence-pioneer-says-we-need-to-start-over/fulltext


Standard model for AI
Maximize

Exemple en médecine

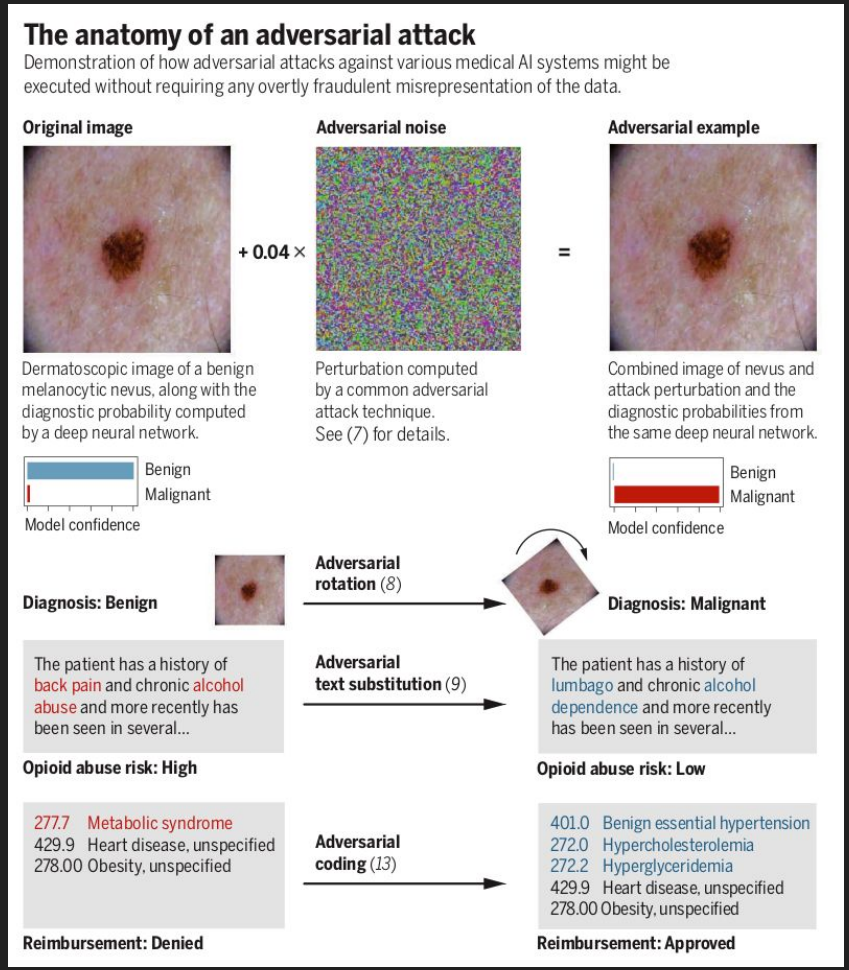
MACHINE LEARNING

Science

Adversarial attacks on medical machine learning

Emerging vulnerabilities demand new conversations

Et sans doute des enjeux financiers qui motiveraient ce type d'attaques



Et d’autres freins à l’adoption généralisée

- La nécessité de données annotées massives (use cases)
- “Fairness” difficile à garantir (banques,...)
- Non-certifiable (aéronautique,...)
- Non-explicable (santé,...)
- Pas toujours acceptable socialement (confidentialité, chômage*)
- Armes autonomes et risques associés perçus (voir Slaughterbots)

Autant de sujets ciblés dans le 3IA Toulousain

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE
Université Fédérale Toulouse Midi-Pyrénées

* Brynjolfsson, Erik, and Andrew McAfee. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company, 2014.

Pour la biologie: des opportunités

- Les problèmes sont difficiles à formaliser (molécules, omics, physiologie)
- Beaucoup de données (ouvertes - FAIR) sur des mondes aux frontières bien cernées
- Parfois moins de freins que dans d'autres secteurs (certifiabilité)

- Deux "fronts" importants & complémentaires
  - Biologie intégrative (Data Science): à la Yolanda Gill
    - intégration, cohérence des entrées/sorties, NLP, ML
    - Raisonnement, connaissances, causalité, hypothèses
  - Machine et Deep Learning ciblé à la α-Fold

Et des difficultés

Recherche publique et privée

- Des concurrents privés puissants: biologie (Watson, données ouvertes + privées), DeepMind (α-Fold)
- Collecte de données: agriculture numérique (IoT, robotique, capteurs & Mineral)

Alphabet

- Deep learning
  - En France: pour une part, de la frilosité, de la méfiance voire un rejet (“*ce n’est pas de la science*”)
    - Pas beaucoup de propriétés
    - “*The theory of Deep Learning is shallow*”
- De l’inertie (changement thématique pas simple)
- Recruter: difficile pour les spécialistes méthodologistes (€) (jury de thèse AFIA/Google-Facebook)
- Collaborer: pas trivial. Sollicitations importantes, ~~finalisé~~

Et de vraies tendances positives

- formation en croissance (Data Sciences, ML, DL)
- intérêt accru pour les problématiques bio (santé, environnement, climat)
- une technologie de plus en plus accessible, en particulier pour les bioinformaticiens (données, calcul, mais surtout avec les risques d'overfitting qu'on a vu)

En forme de conclusion

- Le Deep Learning (et ML) permet des avancées remarquables
- mais il ne faut pas “anthropomorphiser” ces systèmes, et ne pas oublier des limitations/fragilités qui persistent

- Les avancées marquantes de demain ne sortiront pas de là où on les attend.
- Si vous croyez en ce que vous faites, persévérez !
- (même si ce n’est pas du Deep Learning)

Des ressources

GdR CNRS IA: https://www.gdria.fr/livretia/

Vidéos AAAI 2020 (invited talks https://aaai.org/Conferences/AAAI-20/livestreamed-talks)

- Will AI write the Scientific Papers of the Future https://www.youtube.com/watch?v=3Spk_UOHNxA
- How Not to Destroy the World with AI: https://www.youtube.com/watch?v=QPSgM13hTK8
- Prix Turing (LeCun/Bengio/Hinton): https://www.youtube.com/watch?v=UX8OubxsY8w
- The Third AI Summer https://www.youtube.com/watch?v=_cQITY0SPiw
- Fireside chat with D. Kahneman https://www.youtube.com/watch?v=lKmEtz4VwMk

- Alpha-Fold: Blog Alpha Fold
- Symposium “Académie des Sciences” (ML/IA)
- Diapos “Sciences pour tous”
- Diapos “Fête de la science - Collégiens/Lycéens”