

Comment l'IA impacte t'elle notre travail ?

Cédric Matthews CNRS , IBDM
Guillaume Gay , Centuri , AMU

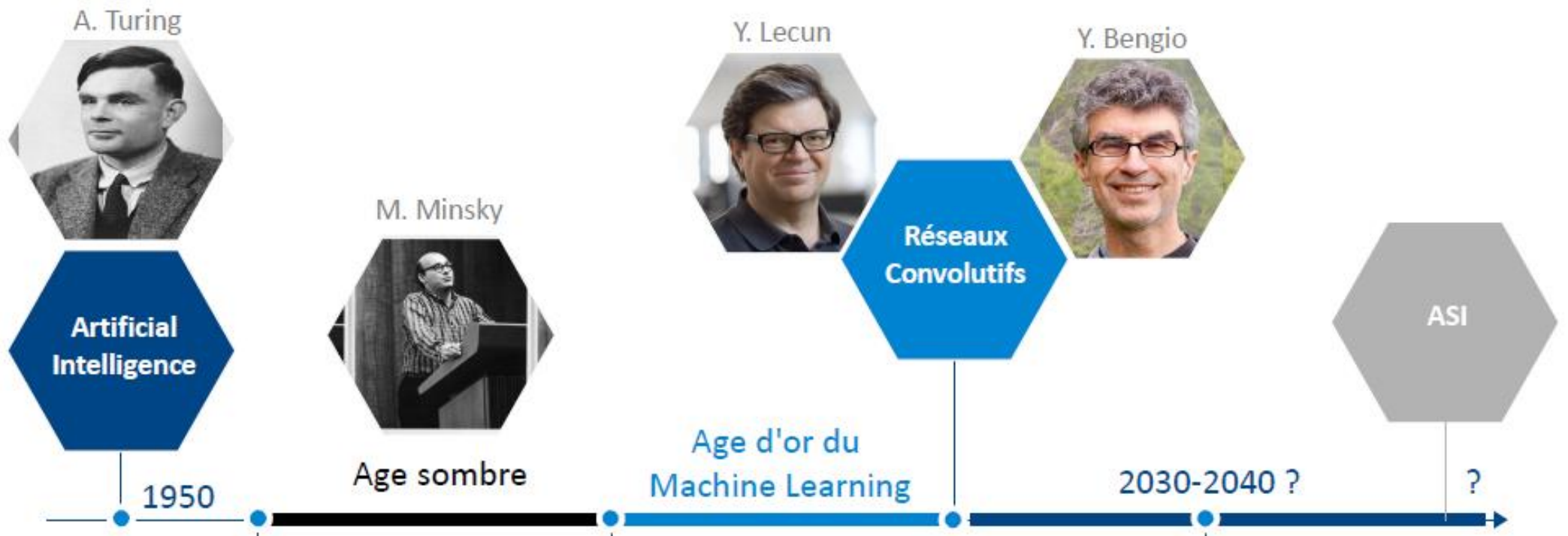
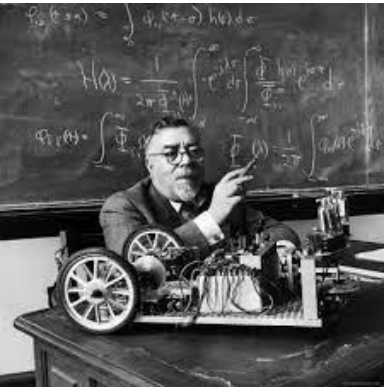
www.nature.com/nmeth / October 2021 Vol.18 No.10

nature methods

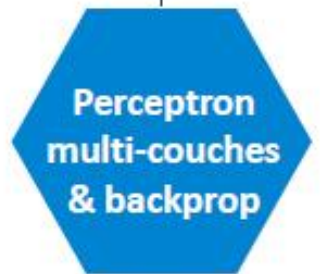


Adaptive optics for deep tissue imaging

Cybernétique
Norbert Wiener



F. Rosenblatt



G. Hinton



Rendre les machines capables d'effectuer n'importe quelle tâche qu'un humain peut réaliser.

Un constat d'aujourd'hui sur plateforme

Cédric :

« Par ailleurs, il apparait dans le sondage que nous avons réalisé en interne que 17 personnes vont avoir un projet associant du deep learning. Il serait peut intéressant de sourcer ce point de façon élargie pour que nous ayons une vision d'ensemble à ce sujet afin de regrouper les thèmes »

Léon :

« Je suis étonné que les utilisateurs aient des besoins en deep learning, avant ils n'avaient pas des besoins en "gradient morphologique" ou "top hat" par exemple. Je veux dire que le DL est une technique des analystes pour répondre a des besoins des utilisateurs... qui devraient s'exprimer en termes de "segmentation d'images difficiles" Ou "tracking de bcp de d'objets" "comptages fastidieux..." Etc non ? »

Cédric :

« Je pense que l'on passe dans l'effet de mode et de valorisation d'un projet car DL figure dedans. »

Pour exister faut il en être ?

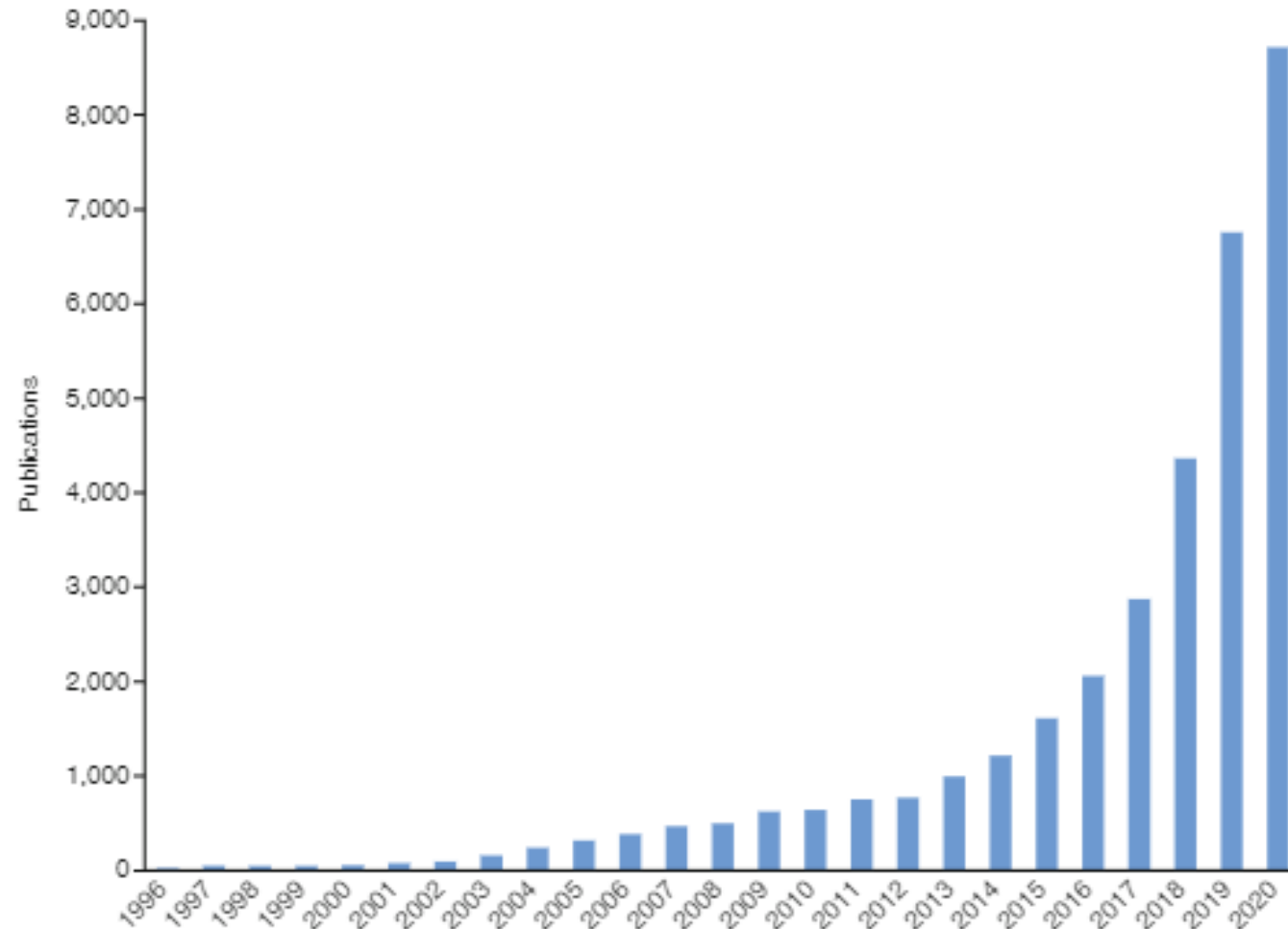


Fig. 1 | Exponential increase of ML publications in biology. The number of ML publications per year is based on Web of Science from 1996 onwards using the topic category for "machine learning" in combination with each of the following terms: "biolog*", "medicine", "genom*", "prote*", "cell*", "post translational", "metabolic" and "clinical".

Ce à quoi nous sommes techniquement confrontés

- Architecture complexe des modèles
- Grand nombre de paramètres à configurer
- Entraînement sur une grande quantité de données

- Difficultés pour obtenir de la reproductibilité dans les résultats mais aussi dans la réutilisation des méthodes
- Sur-confiance dans le résultat produit

Changement de méthode due à la complexité de l'approche

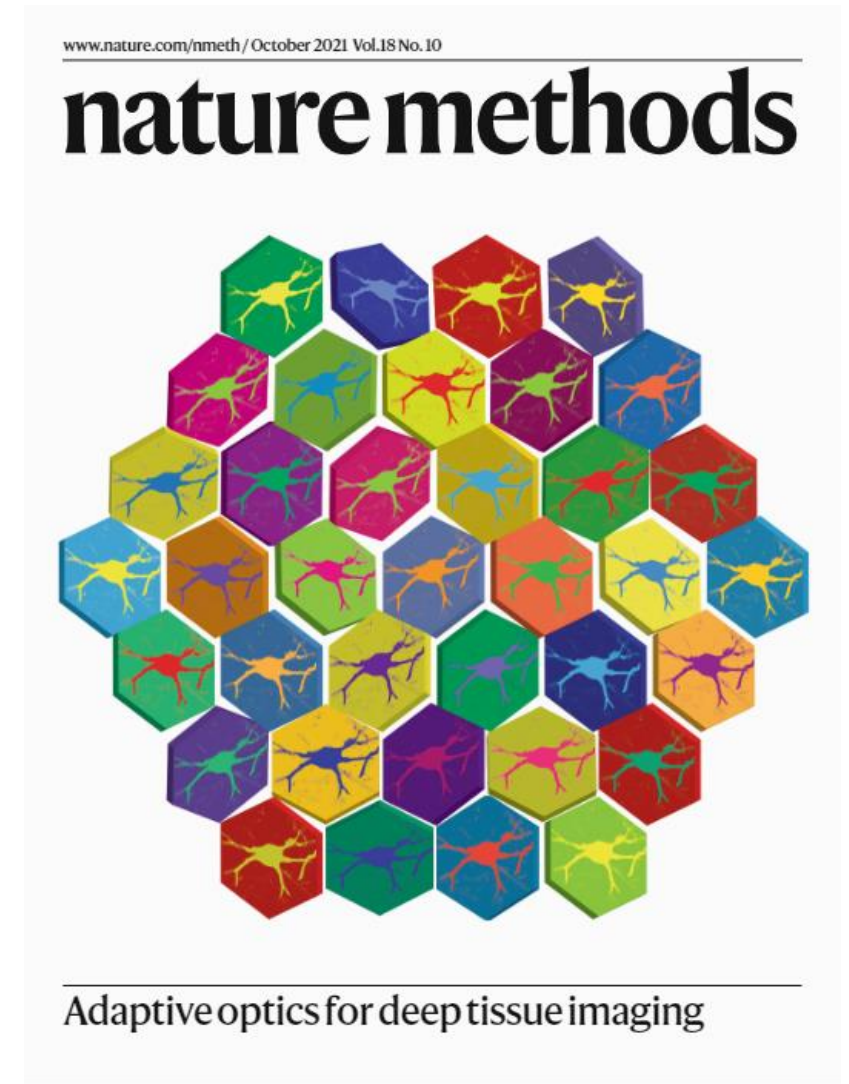
- On ne sait pas comment précisément les images sont manipulées
- L'imagerie quantitative dans un contexte produisant des effets non linéaires
- => reproductibilité et transparence de ce que l'on fait, sont encore plus importants que dans les méthodes classiques.

Des méthodologies apparaissent

- DOME, Data optimization Model Evaluation

Table 1 | Supervised ML in biology: concerns, the consequences they impart and recommendations

Broad topic	Be on the lookout for	Consequences	Recommendation(s)
Data	<ul style="list-style-type: none"> • Inadequate data size & quality • Inappropriate partitioning, dependence between train and test data • Class imbalance • No access to data 	<ul style="list-style-type: none"> • Data not representative of domain application • Unreliable or biased performance evaluation • Cannot check data credibility 	<ul style="list-style-type: none"> • Use independent optimization (training) and evaluation (testing) sets. This is especially important for meta algorithms, where independence of multiple training sets must be shown to be independent of the evaluation (testing) sets. • Release data, preferably using appropriate long-term repositories, and include exact splits. • Offer sufficient evidence of data size & distribution being representative of the domain.
Optimization	<ul style="list-style-type: none"> • Overfitting, underfitting and illegal parameter tuning • Imprecise parameters and protocols given 	<ul style="list-style-type: none"> • Reported performance is too optimistic or too pessimistic • The model models noise or misses relevant relationships • Results are not reproducible 	<ul style="list-style-type: none"> • Clarify that evaluation sets were not used for feature selection, preprocessing steps or parameter tuning. • Report indicators on training and testing data that can aid in assessing the possibility of under- or overfitting; for example, train vs. test error. • Release definitions of all algorithmic hyperparameters, regularization protocols, parameters and optimization protocol. • For neural networks, release definitions of training and learning curves. • Include explicit model validation techniques, such as <i>N</i>-fold cross-validation.
Model	<ul style="list-style-type: none"> • Unclear if black box or interpretable model • No access to resulting source code, trained models & data • Execution time impractical 	<ul style="list-style-type: none"> • An interpretable model shows no explainable behavior • Cannot cross compare methods & reproducibility, or check data credibility • Model takes too much time to produce results 	<ul style="list-style-type: none"> • Describe the choice of black box or interpretable model. If interpretable, show examples of interpretable output. • Release documented source code + models + executable + user interface/webserver + software containers. • Report execution time averaged across many repeats. If computationally tough, compare to similar methods.
Evaluation	<ul style="list-style-type: none"> • Performance measures inadequate • No comparisons to baselines or other methods • Highly variable performance 	<ul style="list-style-type: none"> • Biased performance measures reported • The method is falsely claimed as state-of-the-art • Unpredictable performance in production 	<ul style="list-style-type: none"> • Compare with public methods & simple models (baselines). • Adopt community-validated measures and benchmark datasets for evaluation. • Compare related methods and alternatives on the same dataset. • Evaluate performance on a final independent held-out set. • Use confidence intervals/error intervals and statistical tests to gauge prediction robustness.



Des méthodologies apparaissent

- AIMe2021, Standard de bonnes pratiques.



Reproducibility

R.1 Re-running

R.1.1 Do you provide all means (including dependencies) to easily re-run your AI? ☆

Yes No

R.2 Source code

R.2.1 Main method

R.2.1.1 Is the source code of your AI publicly available? ☆

Yes No

R.2.2 Data simulator

R.2.2.1 Is the source code of your data simulation publicly available? ☆

Yes No Doesn't apply

R.2.3 Pre-processing pipeline

R.2.3.1 Is the source code of your pre-processing pipeline publicly available? ☆

Yes No Doesn't apply

R.3 Pre-trained model

www.nature.com/nmeth / October 2021 Vol.18 No. 10

nature methods



Adaptive optics for deep tissue imaging

IA accélère t'elle la plateformeisation des services / usages ?

- Google colab
- Cell pose
- Zoo image avec modèles pré-entraînés disponibles
- Usage des méso-centres publics versus station de calculs local, ou travail graduel ?
- Omero nécessaire pour structurer les données (annotations) et les rendre accessibles, inter-opérables
- Nouveau cycle des données avec l'IA
- AI Me et certification / ranking de la méthodologie d'IA en biologie

Pose la problématique plus générale de Savoir créer des communs de la data

Risque des plateformes

CGU Google :

- **Chez Google :**

- En soumettant des contenus à nos Services, par importation ou par tout autre moyen, vous accordez à Google (et à toute personne travaillant avec Google) une licence, exclusive, entière, d'utilisation, d'hébergement, de stockage, de reproduction, de modification, de communication, de publication, de représentation ou de distribution public desdits contenus. Les droits que vous accordez dans cette licence sont limités à l'exploitation, la promotion ou à l'amélioration de nos Services et au développement de nouveaux Services. Cette autorisation demeure pour toute la durée de la protection de votre contenu, même si vous cessez d'utiliser nos Services (par exemple, la fiche d'entreprise que vous avez ajoutée à Google Maps).

- Lorsque vous utiliserez Google Drive vous accorderez une licence à Google. Cette licence vous donnera la liberté d'agir sur vos contenus. Même si vous supprimez vos contenus Google profitera toujours de cet accord. D'autres services de Google utiliseront vos contenus, ça peut aller très loin, il n'y a pas de limites. Ceux qui tiennent vraiment à préserver les droits de leurs propres créations vont fort probablement craindre ce service. Google devrait donner des précisions à ce sujet. Changera-t-il cette condition ? Voir les [Conditions d'utilisation](#) de Google.



Highly accurate protein structure prediction with AlphaFold

<https://doi.org/10.1038/s41586-021-03819-2>

Received: 11 May 2021

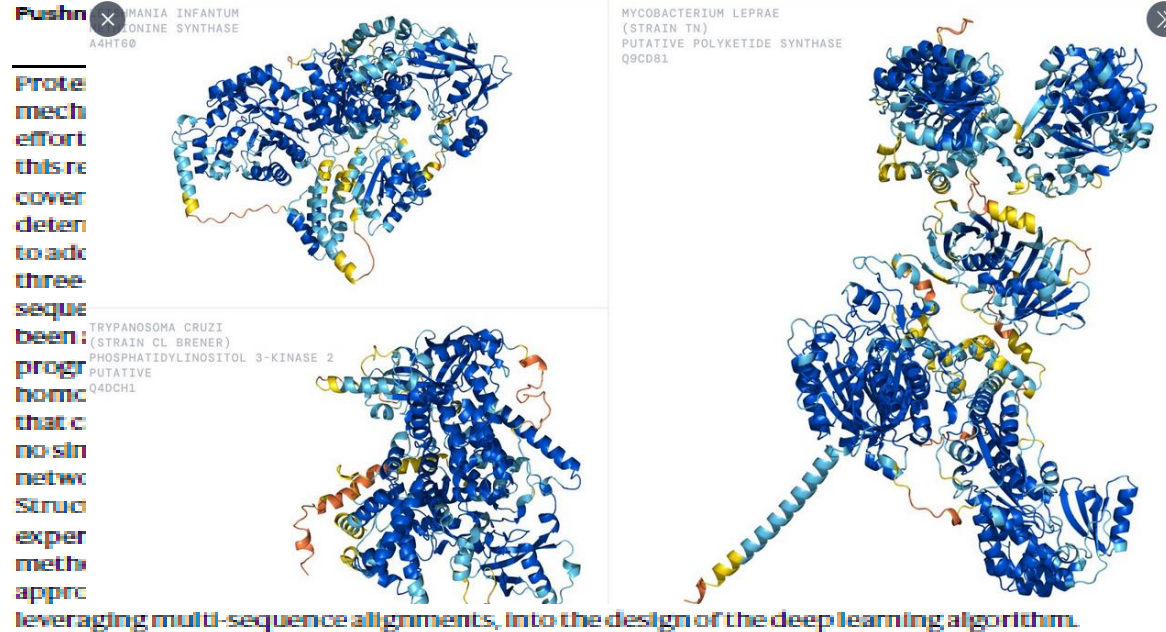
Accepted: 12 July 2021

Published online: 15 July 2021

Open access

 Check for updates

John Jumper^{1,2,3,4,5}, Richard Evans^{1,2}, Alexander Pritzel^{1,2}, Tim Green^{1,2}, Michael Figurnov^{1,2}, Olaf Ronneberger^{1,2}, Kathryn Tunyasuvunakool^{1,2}, Russ Bates^{1,2}, Augustin Zidek^{1,2}, Anna Potapenko^{1,2}, Alex Bridgland^{1,2}, Clemens Meyer^{1,2}, Simon A. A. Kohl^{1,2}, Andrew J. Ballard^{1,2}, Andrew Cowie^{1,2}, Bernardino Romero-Paredes^{1,2}, Stanislav Nikolov^{1,2}, Rishub Jain^{1,2}, Jonas Adler¹, Trevor Back¹, Stig Petersen¹, David Reiman¹, Ellen Clancy¹, Michal Zielinski¹, Martin Steinegger^{1,3}, Michalina Pacholska¹, Tamas Berghammer¹, Sebastian Bodenstein¹, David Ellrott¹, Orel Weisner¹, Andrew W. Senior¹, Koray Kavukcuoglu¹



 DeepMind
@DeepMind

Today we're adding 27 new proteomes (190k+ proteins) to the #AlphaFold Protein Structure Database w/ @emblebi. 17 of these represent Neglected Tropical Diseases, which continue to devastate the lives of more than 1 billion people globally: dpmc.ai/alphafolddb #WorldNTDDay 1/

12:14 PM · Jan 28, 2022 · Twitter Web App

313 Retweets 16 Quote Tweets

1,300 Likes

Quid de la brevetabilité des dispositifs créés avec des données ayant transitées par les centres Google ???

Question de l'extraterritorialité du droit américain à partir du moment où les données transitent par le sol américain ?

À l'ère de l'IA, causalité, corrélation deux mondes en coexistence pacifiques ?

Vu par Chris Anderson :

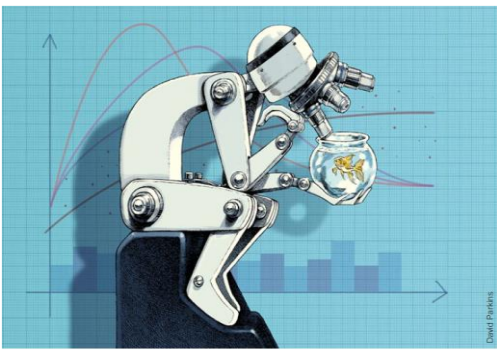
« Learning to use a "computer" of this scale may be challenging. But the opportunity is great: The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world.

Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all. »

Vu par Antoinette Rouvroy :

sous le terme de gouvernementalité algorithmique pour décrire ce que *Chris Anderson désigne la fin de la théorie* : « C'est-à-dire l'évolution vers un monde qui paraît de plus en plus fonctionner comme s'il était constitué lui-même de corrélations, comme si celles-ci étaient ce qu'il suffit d'établir pour en assurer le bon fonctionnement. »

- L'IA outil supplémentaire qui redynamise le système technique
- Faire repenser l'organologie en Biophotonique



Thinking Microscope: Deep Learning-Enabled Computational Microscopy and Sensing

From physics-inspired optical designs and devices, we are moving toward data-driven designs that will holistically change both optical hardware and software of next generation microscopy and sensing, blending the two in new ways. Today, we sample an image and then act on it using a computer. Powered by deep learning, next generation optical microscopes and sensors will **understand a scene or an object** and accordingly **decide** on how and what to sample based on a given task – this will require a perfect marriage of deep learning with new optical microscopy hardware that is designed based on data

ANI \longrightarrow AGI

La technique au service de l'image



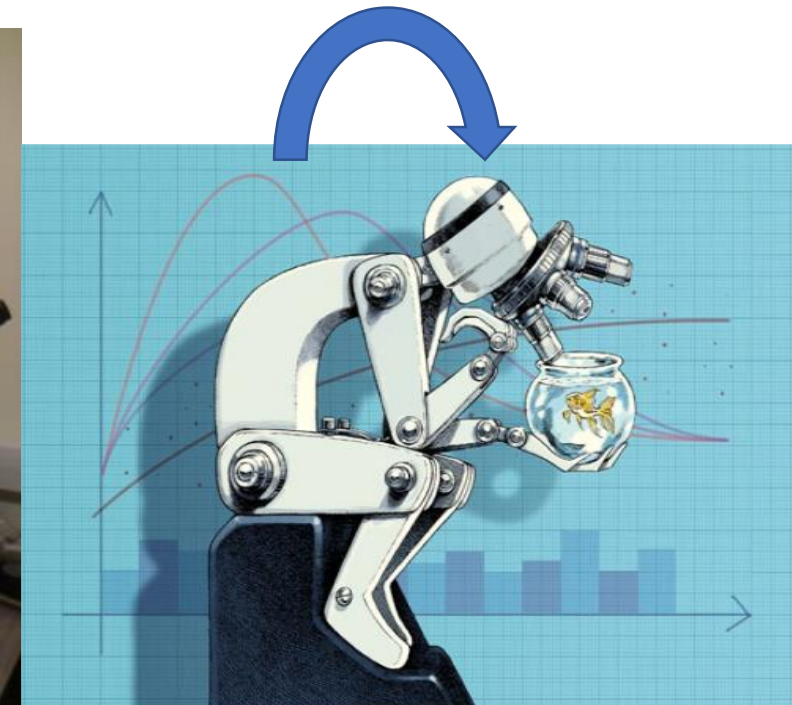
Microscope de van Leeuwenhoek, XVIIe



Microscope de marque Ernst Leitz Wetzlar Germany , 1962



Microscope de recherche, biphotonique de marque NIKON A1RS, 2012



Auto-diagnostic + correction par apprentissage

The thinking microscope

XVIIe

XXIe

Learning optimal wavefront shaping for multi-channel imaging

Elias Nehme[†], Boris Ferdman[†], Lucien E. Weiss, Tal Naor, Daniel Freedman, Tomer Michaeli, and Yoav Shechtman

> [Light Sci Appl.](#) 2021 Oct 7;10(1):210. doi: 10.1038/s41377-021-00649-9.

Couche
conception

An adaptive microscope for the imaging of biological surfaces

Faris Abouakil¹, Huicheng Meng¹, Marie-Anne Burcklen¹, Hervé Rigneault¹, Frédéric Galland², Loïc LeGoff³

Couche de
pilotage

Content-aware image restoration: pushing the limits of fluorescence microscopy

[Martin Weigert](#) , [Uwe Schmidt](#), [Tobias Boothe](#), [Andreas Müller](#), [Alexandr Dibrov](#), [Akanksha Jain](#), [Benjamin Wilhelm](#), [Deborah Schmidt](#), [Coleman Broaddus](#), [Siân Culley](#), [Mauricio Rocha-Martins](#), [Fabián Segovia-Miranda](#), [Caren Norden](#), [Ricardo Henriques](#), [Marino Zerial](#), [Michele Solimena](#), [Jochen Rink](#), [Pavel Tomancak](#), [Loïc Royer](#) , [Florian Jug](#)  & [Eugene W. Myers](#)

Couche
corrective

IA et softwares commerciaux

« Compute without meaning »

Nikon

non-invasive

accurate

time saving

high S/N

training

neural network

Convert.a

Enhance.ai

Segment.a

pattern recognition

low phototoxicity

quantitative

NIS.ai

Give your microscope the power of Artificial Intelligence

The new breakthrough NIS.ai software module uses data to apply a neural network to recognize easier faster image acquisition

Detailed description: The advertisement features a central graphic of a human brain with glowing nodes. A yellow circular arrow labeled 'neural network' surrounds the brain. Inside this circle are three smaller circles labeled 'Convert.a', 'Enhance.ai', and 'Segment.a'. The word 'training' is written along the top of the arrow, and 'pattern recognition' is written along the bottom. Six white circles with black text are connected to the brain by thin lines, representing benefits: 'non-invasive', 'accurate', 'time saving', 'high S/N', 'low phototoxicity', and 'quantitative'. The Nikon logo is in the top left. At the bottom, the text reads 'NIS.ai Give your microscope the power of Artificial Intelligence' followed by a smaller line of text: 'The new breakthrough NIS.ai software module uses data to apply a neural network to recognize easier faster image acquisition'.

L'IA-microscope comme Mégamachine* ?

- La machine a une dimension technique et matérielle mais elle a aussi une dimension sociale et symbolique
- Il y a un lien entre machine, organisation et dispositif de médiation
- Nous parlerons aussi de la machine comme machine invisible, c'est-à-dire comme dispositif organisationnel et institutionnel, comme une forme sociale qui permet à un projet de tenir.

*Lewis Mumford

Bruno Chaudet. Le Mythe de la machine dans les processus d'informations et de communications organisationnelles. Sciences de l'information et de la communication. Université Rennes 2 - COMUE UNIVERSITE BRETAGNE LOIRE, 2019

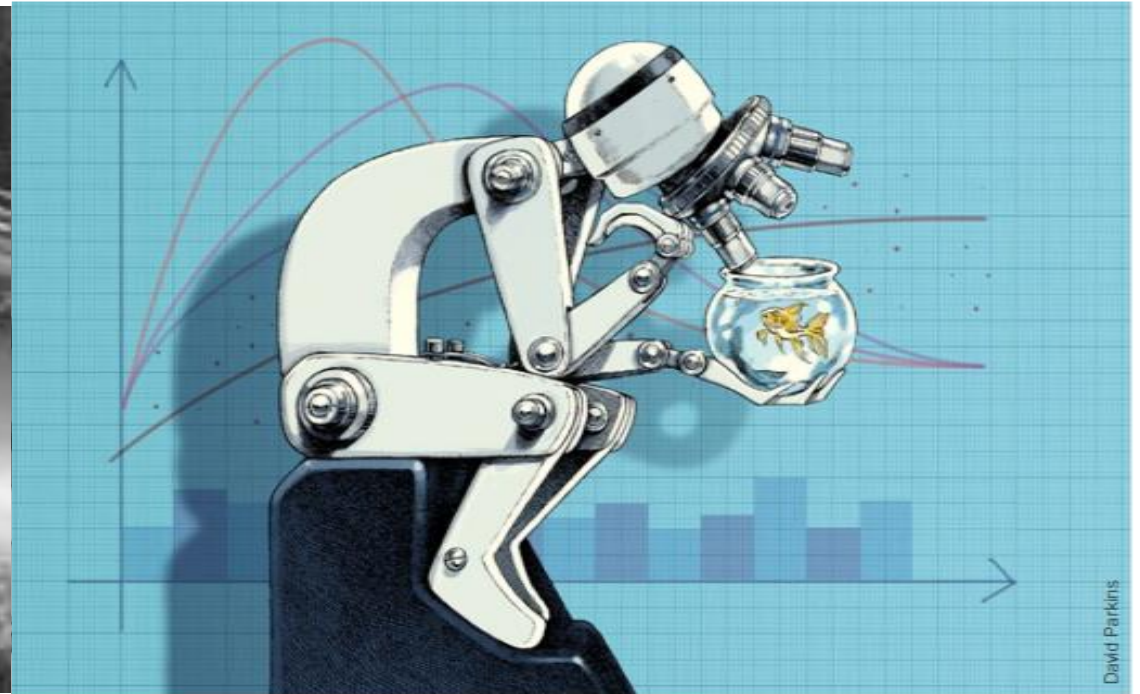
L'IA-microscope comme Mégamachine ?

- Si l'usage du DL-Microscope devenait une Mégamachine au sens de Lewis Mumford, nous dirions qu'elle est cette institution sociale (des scientifiques, Eubi, FBI...), technique (Microscopes, serveurs, algo..) fondée sur le mythe de la rationalité et qui prend forme dans les machines économiques (google, amazon), techniques (plateformes), productives (impact factors), politiques, matérielles.
- La Mégamachine est selon Lewis Mumford l'équivalent du groupement d'institutions qui émergeraient avec le mythe de la machine
- C'est-à-dire que nous sommes au début d'un changement de nos pratiques, de nos organisations qui précède toute nouvelles machines

Ne pas perdre la main sur son Intelligence

- D. Sage : « ne pas abandonner les méthodes classiques du traitement de l'image au détriment de l'IA, c'est juste un outil supplémentaire avec ses défauts et ses avantages »
- Ne pas utiliser le terme IA, qui est une fausse promesse dans les termes mais plutôt calcul intensif , massif
- Utiliser solutions libres de droits bien documentées
- Utiliser des langages (Julia, licence MIT) qui ne sont pas conditionnés à l'usage de plateforme comme google et former la génération montante à ces logiciels
- Charte ou recommandations de l'usage du calcul intensif par les plateformes (hébergement, méthodologie ...) , question des infrastructures nationales comme FBI... ou des réseaux de métiers (MITI)

« Pour voir, il ne faut pas avoir peur de perdre sa place »
Jean Luc Godard



Mais faut il la céder à un algorithme ?

Idées ouvertes

- Métrologie corrective utilisant l'IA , nouvel échelon dans l'intervention corrective d'un instrument dérivant ?
- Obtenir des images de qualité d'un 880 avec un 510 , falsification de performances ou aide à l'augmentation d'un instrument ?
- Utiliser des langages (Julia, licence MIT) qui ne sont pas conditionnés à l'usage de plateforme comme google et former la génération montante à ces logiciels
- Bonne pratique de l'IA sur plateformes par le Rtmfm, comme chaque technique, l'introduction à grande échelle, la diffusion doit être accompagnée

Une bouteille à la mer de l'IA

- Une petite réflexion dans le cycle rétention-protection que nous pratiquons entre humain et qui est court-circuité par la main invisible du numérique. On sait que la rétention tertiaire (le système numérique qui nous entoure) surdétermine R2 et R1.
- L'IA pratique de la même sorte par des cycles de rétention-protection. Les tagathons par exemple sont des cessions où l'humain transfère ces propres rétentions secondaires au système d'apprentissage artificiel ce qui représenterait la rétention primaire pour l'IA .
- Si on extrapole, la rétention secondaire pour l'IA serait la mise en interopérabilité des systèmes d'information (la conscience des systèmes globaux, qui n'existe que par le dialogue rendu possible entre eux) qui sont la résultante de nos apprentissages collectifs.
- Mais que serait alors la rétention tertiaire d'une IA , ce qui surdétermine les R2 et R1 de l'IA ?
- Ce qui pose la question de l'individuation collective des IA ?

Un point de vigilance supplémentaire

IA associée à l'efficiency économique va à l'encontre de la recherche académique



IA – Omero – cycle des données

l'IA à toutes les sauces, ne pas tomber pièges de

de la mode, échanges avec Leon sur constat
besoin interne

Letter to Alan Turing

Giuseppe Longo

<http://www.di.ens.fr/users/longo/>
CNRS et Ecole Normale Supérieure, Paris

January 16, 2018

Dear Alan,

It is with great joy that I have accepted the invitation to write you a personal letter. Even in your scientific writings your presence as a person is very strong, which is unusual for a mathematician. The traces of your personal life and your dramas extend beyond the limits of your own personal circumstances to concern us all – primarily because of the work that you did in World War II, but also because of the sufferings of a free young man, a homosexual, in a context that was totally hostile. I shall try to offer my own readings of some of your fundamental texts, placing them in relation to others that were equally revolutionary; and then I shall discuss with you some of the problems that we face, more than half a century later, including some that follow in the wake of your inventions. Science today is very different from the science that you knew, and it presents other challenges. Alongside the many possibilities that are offered to us – alongside the potentials (and actualities) for a global interaction of networks of what we could call modern *Turing machines* that connect us all, and for a debate of everyone with everyone on the face of this planet, and for a memory of humanity made available to humanity – I shall also tell you how new forms of techno-science are deforming and emptying out the “meaning” of the object of analysis, making it difficult to arrive at an inventiveness of scientific thought of the kind that you were so well able to express.

Points abordés

1. Apports / promesses / dangers des outils d'analyse en deep learning:
 - Constat de l'efficacité pour la segmentation, possibilité pour le contrôle qualité
 - Question de l'"augmentation" des instruments
 - Risque d'"oublier" les méthodes classiques et complémentaires
2. Ressources et données pour l'IA:
 - Délégation de l'architecture tech à google et al.
 - Propriété, licences, usages des données par la communauté (en particulier les ensembles d'apprentissage, patrimoine numérique)
 - Culture & bonnes pratiques