# EXPOSE INTRODUCTIF DE L'ATELIER « S'IL VOUS PLAIT, DESSINE MOI UNE EXPLICATION »

Organisé dans le cadre du programme de résidence datacraft

**Christophe DENIS**

Maitre de Conférences (HDR) au Laboratoire d'Informatique LIP6 – Sorbonne Université

Doctorant à l'ERIAC – Université de Rouen Normandie

christophe.denis@lip6.fr

Sorbonne Center for Artificial Intelligence, Paris, 15 novembre 2021




# CONTRIBUTIONS ÉPISTÉMOLOGIQUES À L'UTILISATION ÉTHIQUE ET CONVIVIALE D'UNE MACHINE APPRENANTE

LE CONTENU PRÉSENTÉ DANS CET COURS EST BASÉ SUR MES TRAVAUX DE THÈSE EN PHILOSOPHIE MENÉS AU SEIN DE ***L'EQUIPE DE RECHERCHE INTERDISCIPLINAIRE SUR LES AIRES CULTURELLES (ERIAC)*** À ***L'UNIVERSITÉ ROUEN NORMANDIE***, ET DIRIGÉ PAR ***FRANCK VARENNE***, SPÉCIALITÉ SUR L'EPISTÉMOLOGIE DES MODÈLES ET DES SIMULATIONS

AU LIP6, L'ENCADREMENT DE MES DOCTORANTS ***JULIO CARDENAS-CHAPELLIN***, ***THÉOPHILE BAYET*** ET ***DJES FREYS BILENGA*** DONT LEURS TRAVAUX PORTENT SUR L'APPLICATION DE L'APPRENTISSAGE MACHINE (GÉOPHYSIQUE, OBJECTIFS DE DÉVELOPPEMENT DURABLE AU SÉNÉGAL, MODÉLISATION DU CLIMAT AU GABON) ALIMENTENT GRANDEMENT MES RÉFLEXIONS PHILOSOPHIQUES

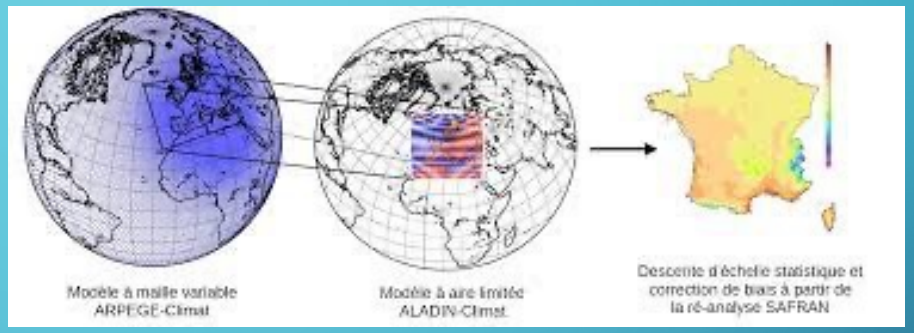
# PREMIERE PARTIE
# POSITIONNNEMENT DU PROBLÈME DE L'EXPLICABILITÉ DANS LA DÉMARCHE SCIENTIFIQUE

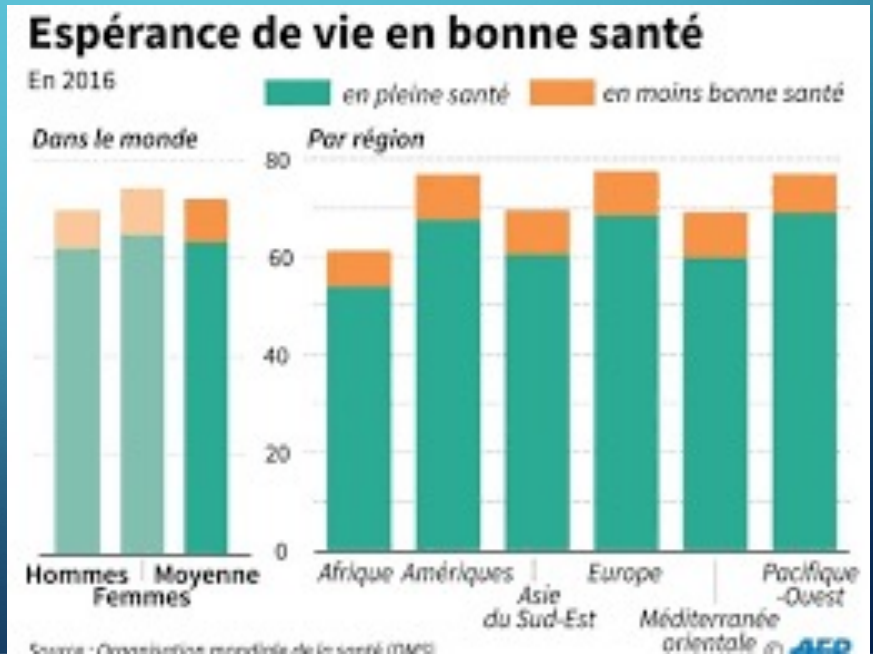
# RECONCILIER PROMOTHÉE ET EPIMÉTHÉE

« L'oppostion dressée entre la culture et la technique, entre l'homme et la machine, est fausse et sans fondement ; elle ne recouvre **qu'ignorance et ressentiment.** Elle masque derrière un facile humanisme une réalité riche en efforts humains »

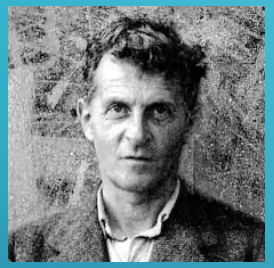
Gilbert Simondon, Du Monde d'Existence des Objets Techniques (MEOT), 1958

# BESOIN ACCRU DE MODELISATION POUR DEVELOPPER DES POLITIQUES PUBLIQUES



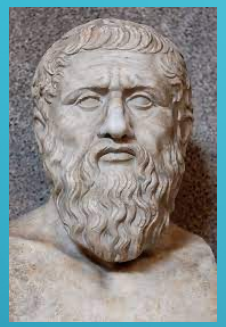
# QUI DIT VRAI ET FAUX ?



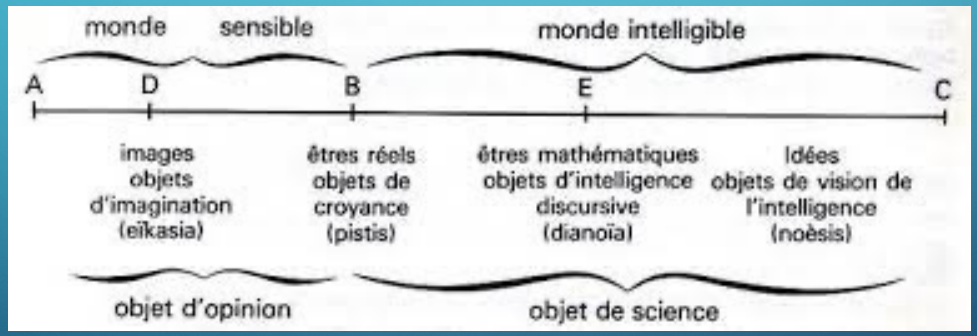
« **C'est** ce que les êtres **humains disent qui est vrai** et faux ; et ils s'accordent dans le langage qu'ils utilisent »

Ludwig Wittgenstein, Recherches Philosophiques, 1953

# SOUS QUELLES CONDITIONS LES OBSERVATIONS DU MONDE SENSIBLE PEUVENT IL DEVENIR CONNAISSANCE ?

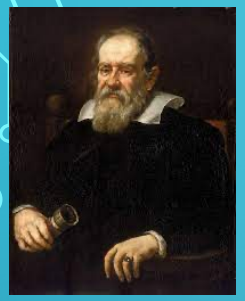


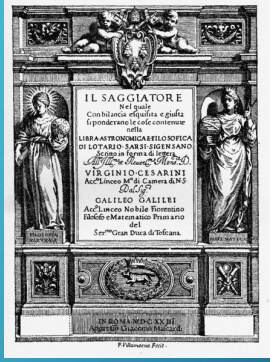
- *Analogie de la Ligne*
  - *- Socrate : Prends donc une ligne coupée en deux segments inégaux, l'un représentant le genre visible, l'autre le genre intelligible, et coupe de nouveau chaque segment suivant la même proportion ; tu classeras alors les divisions obtenues d'après leur degré relatif de clarté ou d'obscurité…..*
- *Platon, La République, Livre 6*




Comment décrire et comprendre les phénomènes naturels ?

# « LE » LANGAGE DE LA PHYSIQUE



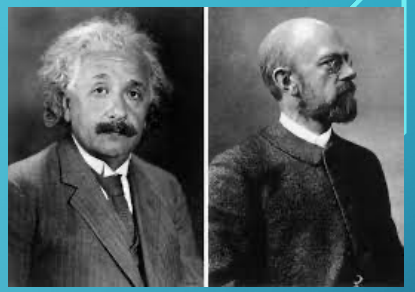



*La philosophie est écrite dans cet immense livre qui se tient toujours ouvert devant nos yeux, je veux dire l'Univers, mais on ne peut le comprendre si l'on ne s'applique d'abord à en comprendre la langue et à connaître les caractères avec lesquels il est écrit. Il est écrit dans la langue mathématique et ses caractères sont des triangles, des cercles et autres figures géométriques, sans le moyen desquels il est humainement impossible d'en comprendre un mot. Sans eux, c'est une errance vaine dans un labyrinthe obscur.*

*Galilée, L'essayeur, 1623*

Depuis les travaux fondateurs de Galilée portant sur la mathématisation du monde, la connaissance scientifique a été significativement améliorée grâce à l'approche hypothético-déductive qui décrit la physique effective d'un phénomène par un modèle utilisant le plus souvent des équations mathématiques.

# FORCE DU LANGAGE MATHEMATIQUE POUR FOURNIR DES EXPLICATIONS



- Exemple de la théorie de la relativité générale (Einstein et Hilbert 1915) : *théorie relativiste de la gravitation*

- Il s'agit d'une théorie mathématique non fondée sur des observations (les expériences de pensée d'Einstein)
- Elle a permis de fournir des explications concernant des écarts entre des observations par rapport à la théorie de la gravitation universelle proposée par Newton à la fin du 17$^{ième}$ siècle

# AVANCE DU PÉRIHÉLIE DE MERCURE

- Théorie de la gravitation universelle Newtonienne
  - La trajectoire d'une planète isolée autour du Soleil st une ellipse invariable.
- Ecart des observations avec la théorie
  - l'observation montre que le périhélie d'une planète se déplace lentement au cours des siècles (43 secondes d'arc par siècle, un degré = 3600 secondes d'arc)
- Quelle explication fournir à cet écart ?

# PLUSIEURS EXPLICATIONS POSSIBLES

- Vénus a en fait une masse 10% plus élevée expliquant des  cela aurait provoqué des irrégularités non observées dans l'orbite de la Terre
- Les perturbations sont dues à une hypothétique planète dont l'orbite est intérieure à celle de Mercure, nommée Vulcain  (planète observée)
- Modification d'une constante de la loi de gravitation de Newton $r^2$ à $r^{2,0000001574}$
- Les perturbations sont à l'origine de la masse du nuage zodiacal difficiles à estimer (explication davantage accepté avant l'explication d'Einstein grâce à ses travaux de la relativité générale en 1915.

Ceci  pose la question de l'acceptation d'une explication,  par exemple comment choisir plusieurs explications  concernant la prédiction d'une méthode d'apprentissage machine

# COMPREHENSION VS EXPLICATION

- En philosophie des sciences contemporaine, il n'existe pas de consensus sur la différence précise entre expliquer et comprendre.
- Cependant, une grande partie des auteurs (cf. Varenne, 2018, p. 18) s'accorde sur le fait d'associer :
  - l'explication à la causalité, plus précisément à des mécanismes
  - et la compréhension à l'unification d'une diversité de phénomènes sous un principe unique

# LA REVANCHE DES NEURONES ARTIFICIELS …

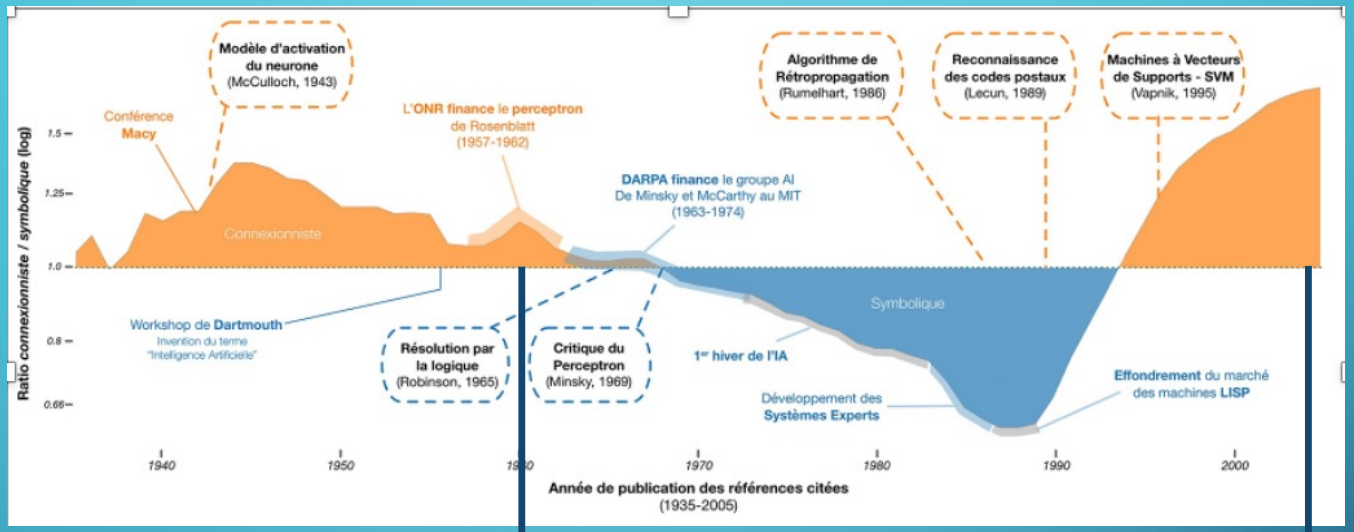
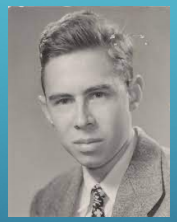
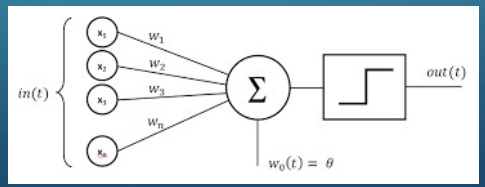
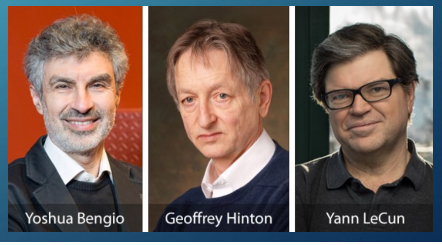
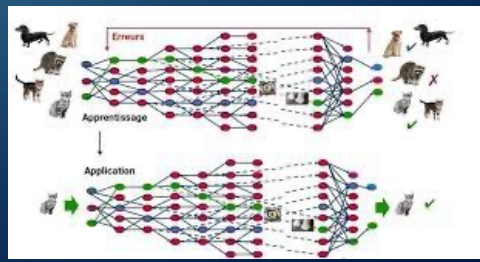



Figure extraite de l'article e
D. Cardon et al.
"La revanche des Neurones », Réseaux, 2018



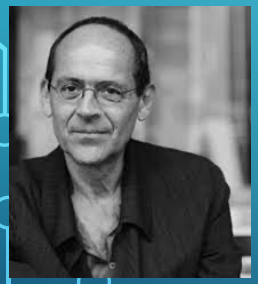



# LA REVANCHE DES NEURONES ARTIFICIELS

- Les résultats, souvent spectaculaires, de l'apprentissage machine (AM) suscitent à la fois de forts espoirs, des craintes légitimes, notamment en termes d'éthique et de transformation du travail et véhiculent un certain nombre de fantasmes.
- La conception de l'algorithme de rétro-propagation du gradient a permis de mettre à jour efficacement les poids des réseaux de neurones plus profonds.
- Les frontières de décision des réseaux de neurones sont devenues donc plus complexes, non linéaires, mettant fin à la critique émise en 1969 par M. Minsky concernant le perceptron, qui a freiné fortement l'activité de recherche sur les réseaux de neurones au profit de l'Intelligence Artificielle symbolique.

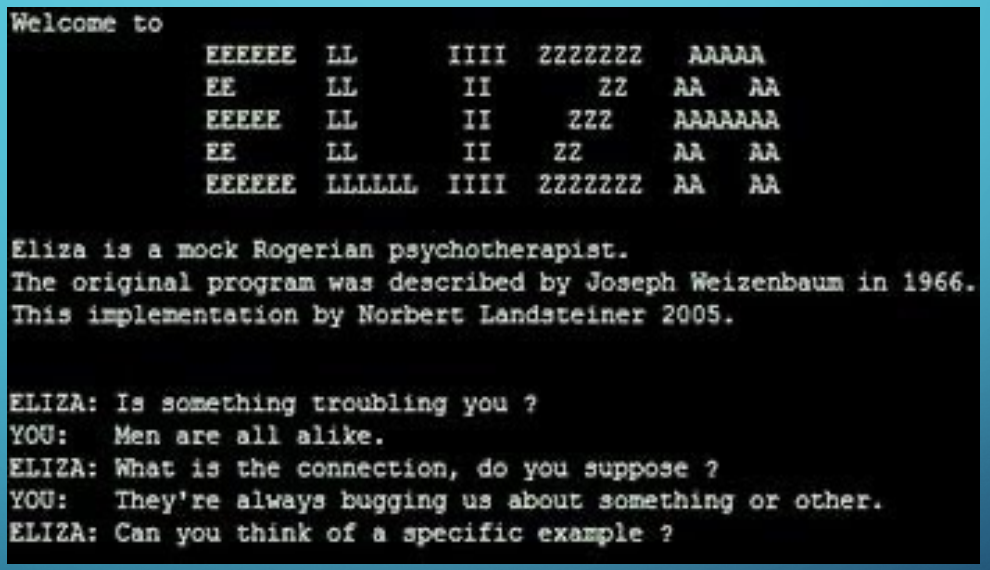
# RECONCILIER PROMOTHÉE ET EPIMÉTHÉE



« Nous n'arrivons plus à élaborer de savoir. Une technologie est un pharmakon : ce terme grec désigne ce qui est à la fois poison et remède mais il commence toujours par provoquer milles problèmes parce qu'ils commence par détruire les cadres constitués »

Bernard Stiegler, Libération, 2016.

# SE SOUVENIR D'ELIZA



# BESOIN D'INTÉGRER LA DEMANDE D'EXPLICATION DANS UN NOUVEAU MODE DE COMMUNICATION

- La célèbre école de Palo Alto (William Fry, John Weakland, Gregory Bateson, J… Haley, 1950) a défini une anthropologie de la communication en distinguant deux types de communication :

1. la communication *digitale* (analytique, logique, et précise). Elle explique et interprète : elle utilise les codes verbaux ;

2. la communication *analogique* c'est-à-dire affective, plus floue, utilise des symboles : Cette communication est essentiellement non verbale, comprise de tous.

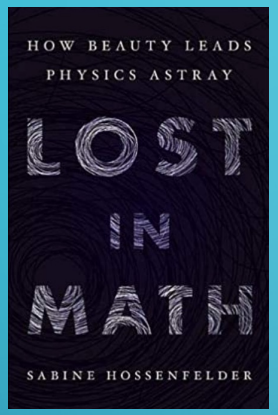
La situation linguistique d'un mode de communication comporte (i) le porteur de l'énoncé (énonciateur), (ii) le sujet de l'énoncé, (iii) l'énoncé et (iv) le(s) destinataire(s) du message

Dans une communication IA à destination d'un humain, le porteur de l'énoncé n'est jamais le sujet de l'énoncé, qui utilise pourtant (souvent) « je » (effet Eliza).

# BOULVERSEMENT DE DISCIPLINES SCIENTIFIQUES


- Depuis la conférence scientifique ECCV (European Conference on Computer Vision), en 2012, les capacités prédictives des réseaux de neurones profonds, en particulier les réseaux neuronaux convolutifs, ont bouleversé en profondeur la discipline du traitement et de la reconnaissance d'images.

# LA PHYSIQUE PERDUE DANS LA BEAUTE DES MATHÉMATIQUES




- *« Aveuglés par l'élégance mathématique, les physiciens ont développé des théories stupéfiantes, inventé des dizaines de nouvelles particules, décrété des modèles de grande unification. Mais aucune ou presque de ces idées n'a été confirmée par l'observation — en fait, beaucoup d'entre elles sont tout bonnement invérifiables. En dépit de ces contradictions, les théoriciens sont persuadés que leurs gracieuses équations et leurs formules élégantes recèlent de formidables vérités sur la nature. Et du fait de ces théories « trop belles pour ne pas être vraies », la discipline est aujourd'hui dans l'impasse. »*
- *Sabine Hossenfleder, Lost In Maths, 2019*

CHAOS THEORY

# Machine Learning's 'Amazing' Ability to Predict Chaos

17

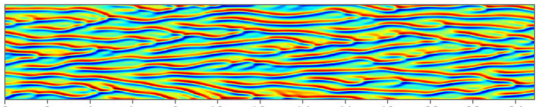
*In new computer experiments, artificial-intelligence algorithms can tell the future of chaotic systems.*

## Training Computers to Tame Chaos

A machine-learning algorithm has been shown to accurately predict a chaotic system far further into the future than previously possible.

**A Chaos Model**

Researchers started with the evolving solution to the Kuramoto-Sivashinsky equation, which models propagating flames:



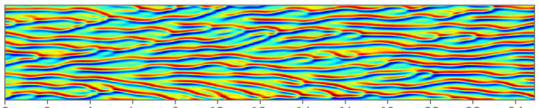
0 2 4 6 8 10 12 14 16 18 20 20 24

Time (in units of Lyapunov time*) ⟶

* Lyapunov time = Length of time before a small difference in the system's initial state begins to diverge exponentially. It typically sets the horizon of predictability, which varies from system to system.

**B Machine Learning**

After training itself on data from the past evolution of the Kuramoto-Sivashinsky system, the "reservoir computing" algorithm predicts its future evolution:

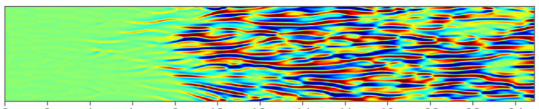


0 2 4 6 8 10 12 14 16 18 20 20 24

Time (in units of Lyapunov time) ⟶

**A – B Do They Match?**

Subtracting B from A shows that the algorithm accurately predicts the model out to an impressive 8 Lyapunov times, before chaos ultimately prevails:



0 2 4 6 8 10 12 14 16 18 20 20 24

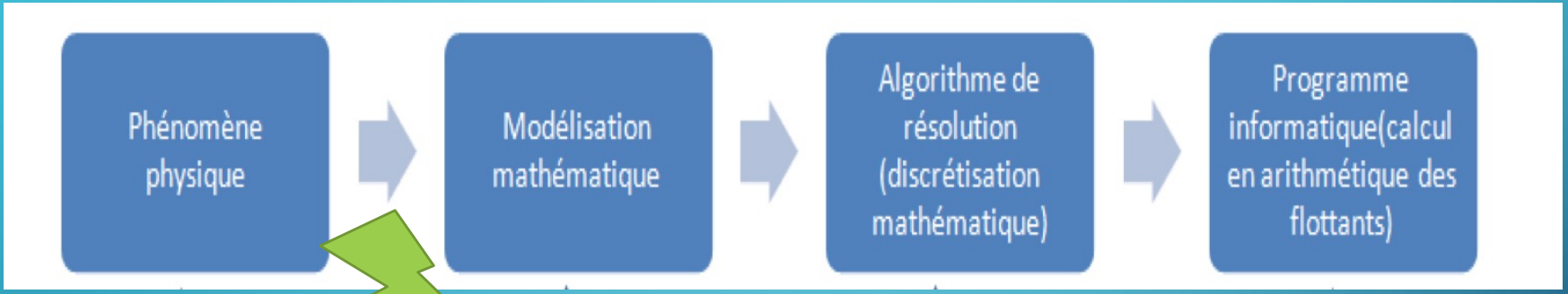
Time (in units of Lyapunov time) ⟶

# ATTRAIT DE L'APPRENTISSAGE MACHINE PAR DE NOMBREUSES DISCIPLINES SCIENTIFIQUES

- De nombreuses disciplines scientifiques notamment computationnelles développent des activités de recherche orientées vers l'apprentissage machine :

  - **diminuer la puissance de calcul nécessaire** pour mener des simulations numériques, en construisant un métamodèle
  - a**ugmenter les capacités prédictives d'un processus** déjà basé sur une approche statistique comme par exemple pour l'étude du climat ou en biologie moléculaire
  - **combler le manque de connaissance d'un phénomène modélisé** selon une approche hypothético-déductive, c'est-à-dire décrit par un modèle représentant d'une manière ou d'une autre la physique effective du phénomène (ses lois) et recourant souvent à des équations mathématiques tirées de ces mêmes lois.
  - Il s'agit donc, alternativement à cette approche modélisant les phénomènes effectifs supposés, de chercher par exemple
    - à prédire le comportement turbulent d'un fluide
    - la détection avec plus de précision d'un ensemble d'anomalies magnétiques en géosciences
    - le comportement d'un système chaotique avec des temps de prédiction dépassant ceux obtenus en résolvant les équations mathématiques

# L'APPROCHE HYPOTHETICO-INDUCTIVE EN PRATIQUE DANS LES CODES DE SIMULATION NUMERIQUE



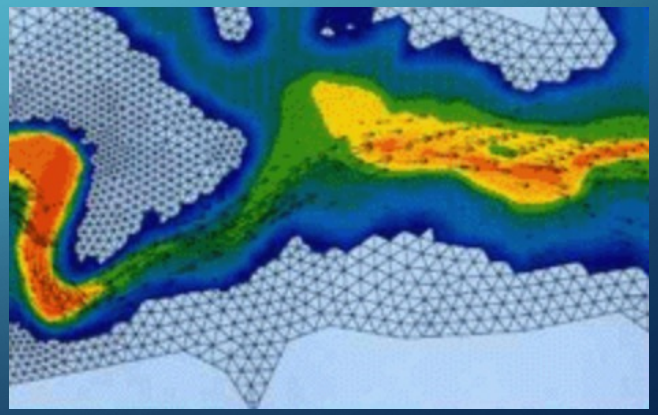

Pour certains phénomènes complexes

Incertitudes épistemiques

**Machine learning–accelerated computational fluid dynamics**

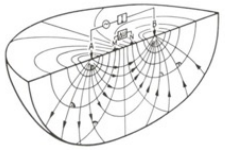
Dmitrii Kochkov, Jamie A. Smith, Ayya Alieva, Qing Wang, Michael P. Bre…

+ See all authors and affiliations



# « INVERSION DE SIGNAUX GEOPHYSIQUES PAR APPRENTISSAGE MACHINE PROFOND », JULIO CARDENAS-CHAPELIN

**Cas des méthodes électriques**



① **Acquisition : $\Delta V$, I**

↓

**Résistivité apparente**
(Valeurs observées)

② **Inversion :**

↓

**Résistivité réelle**
(Paramètres)

Exemple d'inversion en méthodes électriques

**Nouvelle approche :**
**Inversion non-classique**

**Collection de problèmes directs**

↓

**Valeurs observées**

- **Palier les faiblesses des algorithmes d'inversion actuels.**
- **Parvenir à un ajustement plus fin des paramètres d'amortissement.**

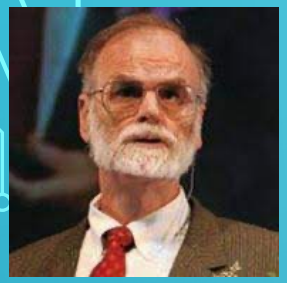
L' intérêt de l'application des algorithmes d'apprentissage profond pour l'inversion de données

# FIN DE LA THÉORIE ?

There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

Chris Anderson, Wired, 2008

# QUATRIÈME PARADIGME EN SCIENCE ?

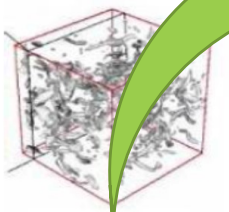


Jim Gray

## Science Paradigms

- Thousand years ago:
  science was **empirical**
  describing natural phenomena
- Last few hundred years:
  **theoretical** branch
  using models, generalizations
- Last few decades:
  a **computational** branch
  simulating complex phenomena
- Today:
  **data exploration** (eScience)
  unify theory, experiment, and simulation
  using data management and statistics
  - Data captured by instruments
    Or generated by simulator
  - Processed by software
  - Scientist analyzes database / files

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$



5

Cela peut être interprété comme une indépendance voire une discontinuité entre les « paradigmes » alors qu'il s'agit selon mon hypothèse d'une nouvelle fonction de connaissance ou une conjonction de fonctions de connaissance utilis précédents « paradigmes ».

# MODELISATION PAR APPRENTISSAGE MACHINE

- La mise au point d'une application basée sur de l'apprentissage machine va au-delà du seul entrainement de celle-ci
  - précédé toujours d'une étape de traitement des données
  - selon les applications
    - une phase de génération de données lorsque les mesures du phénomène ne sont pas en quantité suffisantes
    - une phase de transfert lorsque le domaine d'apprentissage de la méthode peut être différent du domaine spécifique d'utilisation.

Besoin de prendre en compte les fonctions de connaissance du modèle pour ne pas rester dans des généralités


# DEUXIÈME PARTIE : EPISTEMOLOGIE DES MODÈLES

CETTE PARTIE REPREND LE CONTENU D'UN EXPOSE EFFECTUE PAR FRANCK VARENNE SUR NOS TRAVAUX EN COMMUN LORS DE LA JOURNEE

'L'IA EST ELLE EXPLICABLE » LE JEUDI 17 OCTOTBRE 2019 À L'X

# MODÈLE COMME MEDIATEUR

**Caractérisation d'un modèle**

*« Pour un observateur B, un objet A* est un modèle d'un objet A dans la mesure où B peut utiliser A* pour répondre à des questions qui l'intéressent au sujet de A » (Minsky, 1965)*

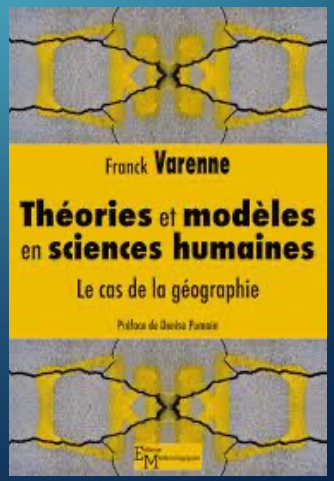
Cette caractérisation de modèle assure une médiation dans le cadre d'une questionnement

- Il existe un nombre assez important (mais limité) de fonctions de connaissance associées à une médiation

# RÉPERTOIRE DE 21 FONCTIONS DE CONNAISSANCE

**Tableau récapitulatif des fonctions des modèles**

| GRANDES FONCTIONS | FONCTIONS SPÉCIFIQUES | EXEMPLES[1] |
|---|---|---|
| I Faciliter l'appréhension sensible | 1. Rendre perceptibles certaines propriétés sur un substitut | Écorchés de cire, maquettes de molécules avec des billes… |
| | 2. Rendre perceptibles certains rapports sur un substitut | Diagrammes, cartes, organismes modèles, maquettes de bateaux… |
| | 3. Faciliter la mémorisation par une représentation ordonnée | Contines, images, théâtres mentaux, systèmes architecturaux, mémoire locale… |
| | 4. Condenser l'information pour faciliter l'accès et le rappel à volonté | Systèmes d'axes de symétrie, moments statistiques (moyenne, variance, etc.), paramètres de modèles statistiques analytiques… |
| II Faciliter la formulation intelligible | 5. Faciliter la compression de données pour préparer la conceptualisation | Modèles de données, modèles statistiques descriptifs ou synthétiques, enveloppe statistique… |
| | 6. Faciliter une sélection de types d'entités ou de propriétés | Modèles conceptuels, modèles de connaissance, classifications, hiérarchies, ontologies… |
| | 7. Faciliter la reproduction ou production de structures de données par des moyens intelligibles déductifs ou de calcul | Modèles phénoménologiques (à base de données), modèles descriptifs et/ou prédictifs, modèles de conception (ingénierie), modèles de synthèse par analyse spectrale de données… |
| | 8. Faciliter une explication | Modèle d'interaction, modèle de séquence finie d'interactions, modèle de mécanismes… |
| | 9. Faciliter une compréhension | Modèle d'optimisation, modèle à principe variationnel valant à échelle agrégée, modèle axiologique à l'échelle des individus (rationalité en valeur), geste mental, idéal-type… |
| III Faciliter la théorisation | 10. Faciliter une ébauche de théorie : modèle théorique | *Homo economicus*, modèles de la rationalité (utilitariste, limitée ou ordinaire), « théories de l'acteur rationnel »… |
| | 11. Faciliter une interprétation de théorie : modèle de théorie | Images mentales, modèles physiques de théories mathématiques… |
| | 12. Faciliter une illustration de théorie : modèle pour la théorie | Modèle des courants de fluide pour la circulation électrique, modèle d'oscillateurs électriques pour la théorie des dynamiques de populations… |
| | 13. Faciliter un test de cohérence interne de la théorie | Modèle sémantique, modèle concret (*i. e.* se référant à des objets) d'une théorie formelle, modèle des valeurs de vérité en théorie logique des propositions, modèle euclidien pour une géométrie… |
| | 14. Faciliter l'applicabilité de la théorie | Modèle sémantique approché ou incomplet, sous-structures empiriques[2]… |
| | 15. Faciliter la calculabilité d'une théorie | Modèle partiellement phénoménologique ou approché du fonctionnement de la théorie mathématique, modèle numérique, *computational template*[3] ou gabarit computationnel, modèle de simulation de type 2[4]… |
| | 16. Faciliter une hybridation et une cocalculabilité de plusieurs théories | Modèle mixte polyphase, modèle *ad hoc*, modèle asymptotique, modèle multi-échelles… |
| IV Faciliter la coconstruction des savoirs | 17. Faciliter une communication entre acteurs scientifiques | Base de données, ontologie explicite et ouverte, modèle de vulgarisation… |
| | 18. Faciliter la délibération et la concertation entre parties prenantes | Modèle multi-aspectuel pour la concertation, modèle d'exploration de scénarios concertés… |
| | 19. Faciliter la coconstruction de représentations et de modes de contrôle de système mixtes (humains/non-humains) | Modèle en recherche-action, modèle participatif, modélisation d'accompagnement… |
| V Faciliter la décision et l'action | 20. Faciliter l'intervention *sur* un système mixte et hétérogène | Modèle de décision, arbres de décision, modèle de crise, heuristique… |
| | 21. Faciliter une décision d'action *dans* un système principalement notionnel | Modèle d'anticipation de marché, modèle de produits dérivés en finance… |



# FONCTIONS DE CONNAISSANCE LES PLUS UTILISEES

**Tableau récapitulatif des fonctions des modèles**

| GRANDES FONCTIONS | FONCTIONS SPÉCIFIQUES | EXEMPLES[1] |
|---|---|---|
| I Faciliter l'appréhension sensible | 1. Rendre perceptibles certaines propriétés sur un substitut | Écorchés de cire, maquettes de molécules avec des billes… |
| | 2. Rendre perceptibles certains rapports sur un substitut | Diagrammes, cartes, organismes modèles, maquettes de bateaux… |
| | 3. Faciliter la mémorisation par une représentation ordonnée | Contines, images, théâtres mentaux, systèmes architecturaux, mémoire locale… |
| | 4. Condenser l'information pour faciliter l'accès et le rappel à volonté | Systèmes d'axes de symétrie, moments statistiques (moyenne, variance, etc.), paramètres de modèles statistiques analytiques… |
| II Faciliter la formulation intelligible | 5. Faciliter la compression de données pour préparer la conceptualisation | Modèles de données, modèles statistiques descriptifs ou synthétiques, enveloppe statistique… |
| | 6. Faciliter une sélection de types d'entités ou de propriétés | Modèles conceptuels, modèles de connaissance, classifications, hiérarchies, ontologies… |
| | 7. Faciliter la reproduction ou production de structures de données par des moyens intelligibles déductifs ou de calcul | Modèles phénoménologiques (à base de données), modèles descriptifs et/ou prédictifs, modèles de conception (ingénierie), modèles de synthèse par analyse spectrale de données… |
| | 8. Faciliter une explication | Modèle d'interaction, modèle de séquence finie d'interactions, modèle de mécanismes… |
| | 9. Faciliter une compréhension | Modèle d'optimisation, modèle à principe variationnel valant à échelle agrégée, modèle axiologique à l'échelle des individus (rationalité en valeur), geste mental, idéal-type… |

- l'analyse ou la réduction de données
- la description
- la prédiction
- l'explication

# MODÈLE À FONCTION D'ANALYSE DE DONNES

- Ce modèle s'applique sur la seule structure informationnelle des données du système cible, mais pas directement sur la structure des propriétés intrinsèques du système cible ni des relations mutuelles entre ces propriétés
- Ils sont faiblement prescriptifs ontologiquement.
- Il permet ensuite d'utiliser l'utilisation d'autres modèles : les modèles à fonction de description ou d'explication du système cible.
- C'est parce qu'ils traitent les données comme des signaux mais pas comme des signes.

# SIGNAL ET  SIGNE EN APPRENTISSAGE MACHINE

- L'une des motivations de la science est de tenter de parvenir aux signes (à l'ontologie, la réalité) par le biais des signaux dont l'on dispose :
  - Le signe désigne, qualifie ou quantifie une propriété
  - le signal indique, qualifie ou quantifie une interaction entre cet objet et un dispositif de mesure.
- Proposition [Denis & Varenne, 2019] :  les modèles d'AM « servent de représentation intermédiaire de la seule structure informationnelle du système cible » : L'AM ne permet de n'étudier que des signaux, sans fournir d'accès aux signes.

# MODÈLE À FONCTION PREDICTIVE

- Ce sont des cas particuliers de modèles descriptifs dynamiques (i.e. avec état initial et état final)
- Ils décrivent le système dynamiquement à travers au minimum deux types de données qui le représentent partiellement sans encore l'expliquer :
  - des données prédictives (nommées parfois « explicatives » de manière trompeuse en statistique inférentielle) utilisées par l'algorithme ou le modèle
  - des données comportementales ou prédites qui servent à évaluer la qualité de la prédiction, donc la qualité du modèle

# MODÈLE EXPLICATIF

- un modèle mathématique ou algorithmique est explicatif d'un système cible lorsque :
  - Il est au moins partiellement prédictif pour ce système
  - Il offre une représentation interprétable, c'est-à-dire signifiante et accessible à un esprit humain non aidé, à la fois des éléments dont il est composé et des processus élémentaires d'interaction qu'il met en œuvre (« sémantique cognitive »)

# TROISIÈME PARTIE : CONTRIBUTION EPISTÉMOLOGIQUE SUR L'APPRENTISSAGE MACHINE

# HYPOTHESE PRINCIPALE DES TRAVAUX

L'intelligence Artificielle, et plus précisément l'apprentissage machine, sous des conditions restant à déterminer, peut faciliter la compréhension de phénomènes à partir d'observations, pouvant ne pas être directement des mesures le concernant.

- Evolution du savoir et de l'intégrité scientifique en raison du caractère limité et contestable du « prédire sans comprendre » [Thom 2009].

# L'APPRENTISSAGE MACHINE COMME UN MODELE (OU UN MEDIUM)

**Caractérisation d'un medium dans le contexte de l'apprentissage machine**

"Pour un observateur B, une simulation par apprentissage machine A* est un medium d'un objet A dans la mesure où B peut utiliser A* à l'aide de signaux concernant et ne concernant pas A pour atteindre certaines propriétés de A (ses signes) ou au moins sa structure informationnelle » (Denis, 2021)

# INTERPRETABILITE

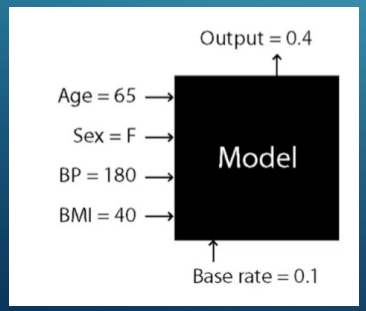
*"L'interprétabilité" fait référence au degré de* ***compréhensibilité*** *humaine d'un modèle ou d'une décision "boîte noire" donnée" (Lisboa, 2013 ; Miller, 2017).*

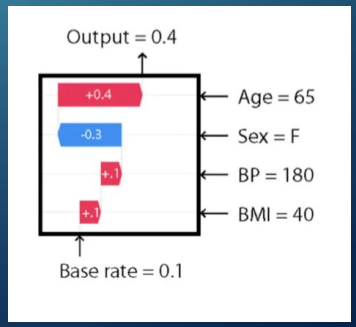
Définition alternative :

*" Interprétabilité d'un modèle : propriétés qu'a un modèle de se voir constitué d'éléments (signes, figures conceptuelles, données, etc.) qui ont un sens, c'est-à-dire un référent possible, pour le sujet humain ", (C. Denis, F. Varenne, 2019).*

→Définition sémantique liée à une ontologie non liée de à une compréhension.

→ Par exemple utilisation des indices de Shapley pour connnaître la contribution des données sur la prédiction (bibliothèque SHAP), que l'on verra dans le prochain atelier



# INTERPRETABILITE

"Extraire des information d'un modèle d'apprentissage machine n"est pas une condition suffisante pour le rendre interprétable !

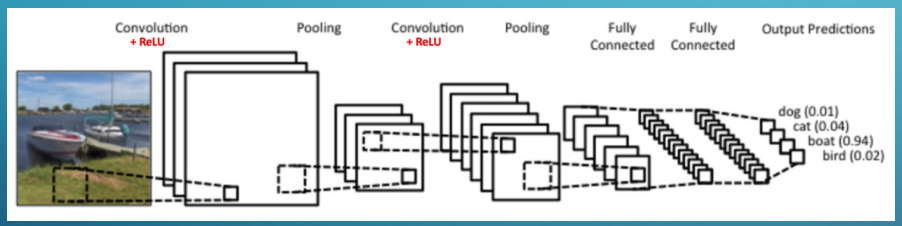
→ Encore faut-il que ces informations ont un sens, un référent possible pour un humain ou un groupe d'humain

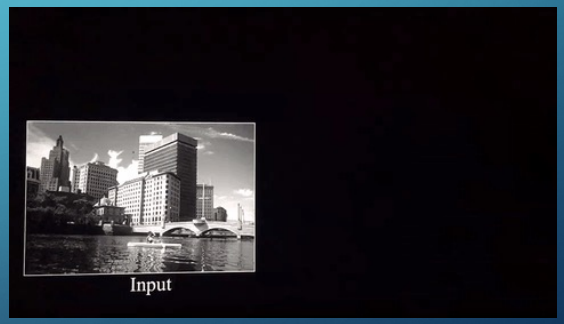
*Exemple : on imprime tous les poids des neurones d'un réseau de neurones profond. Il s'agit d'une information mais est ce que cela rend le modèle interprétable ?*

# EXPLICABILITE

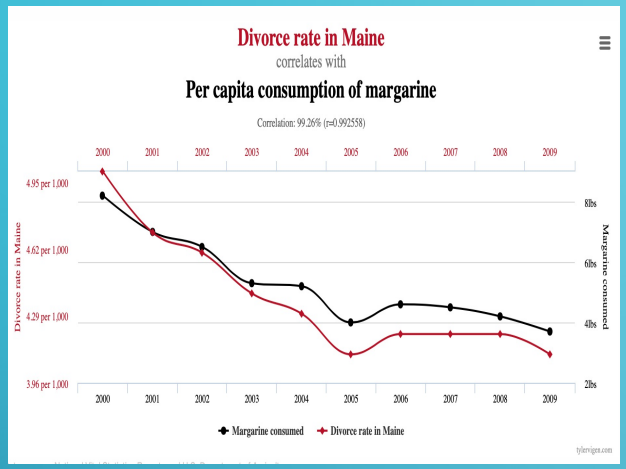
*" Explicabilité d'un modèle : capacité à exposer son effet sur les données d'entrée en plusieurs étapes connectées d'une manière que l'utilisateur peut interpréter de manière significative comme des causes ou des raisons. ", (C. Denis, F. Varenne, 2019).*

Illustration de l'explicabilité d'une réseau de neurones convolutif



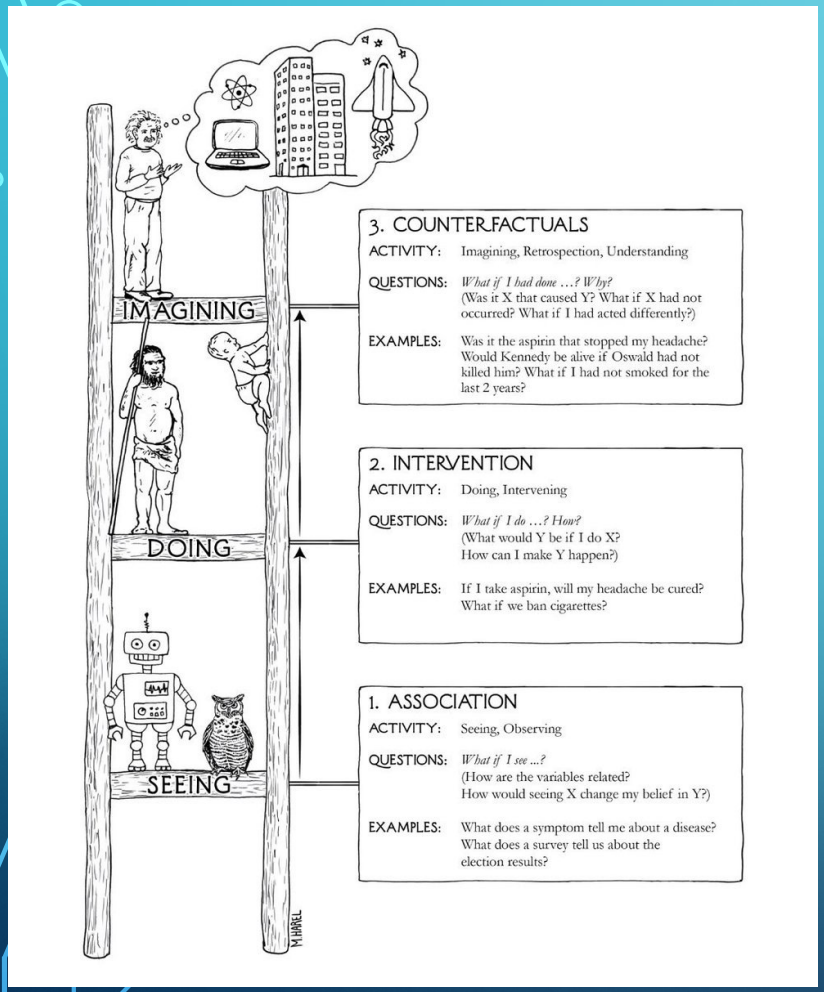
# CAUSALITE ET CORRELATION




- Depuis Kant, la détermination d'une chaine causale est la pire angulaire de la démarche scientifique moderne pour expliquer un phénomène
- Pour Hume, courant empiriste, une corrélation est une « habitude associative »

- Corrélation ≠ Causalité → Oui certes, mais besoin de formaliser davantage
- Une corrélation entre deux caractéristiques x et y peut être interprétée (ou non) comme une relation de cause connue à effet
  - entre x et y
  - ou entre au moins une des deux variables avec une autre variable z
- Ou le fruit d'un « hasard » qui peut générer une découverte (Eureka !)

# FORMALISATION DE LA CAUSALITE EN APPRENTISSAGE MACHINE




- J. Pearl « Causes of Effects and Effects of Causes », Sociological Methods & Research, 2015
- J. Pearl, « The book of Why »

- cadre conceptuel et méthodes mathématiques simples d'estimation de la probabilité qu'un événement soit une cause nécessaire d'un autre
- dilemme philosophique associé à la détermination de cas individuels à partir de données statistiques.
- → Stage en cours dans le cadre d'un projet CNRS

# CAS D'ETUDE
# MACHINE APPRENANTE DEDIEE AUX EDO ET EDP

Les EDO et les EDP sont majoritairement pour modéliser des phénomènes physiques (approche hypothético-déductive)

- Prédire les propriétés qualitatives des systèmes différentiels (stabilité, contrôle optimale)

Charton et al. « *Learning advanced computations from examples* », ICLR 2021 International Conference on Learning Representations

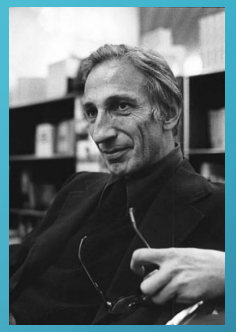
- Prédire un résultat de calcul symbolique (intégration)

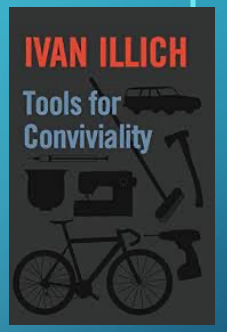
Charton et al. « Deep Learning for Symbolic Mathematics», ICLR 2020, International Conference on Learning Representations

- Découvrir des EDP (dépendant du temps) à partir de données dynamiques observées

*Long et al. «PDE-Net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network»,*

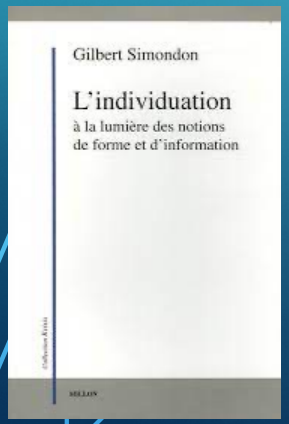
# EXAMEN EPISTÉMOLOGIQUE POUR DEFINIR DES CRITERES DE CONVIVALITE






- Illich définit trois critères indispensables pour qu'une instrumentation ou une institution soit considérée comme *juste* ou *conviviale* :
  - elle ne doit pas dégrader l'autonomie personnelle en se rendant indispensable ;
  - elle ne suscite ni esclave, ni maître ;
  - elle élargit le rayon d'action personnel.

Définition de  critères en effectuant un examen épistémologique des équations différentielles ordinaires neuronales (NODE) en rapport avec le principe d'individuation de Simondon



# MERCI DE VOTRE ATTENTION


- Ch. Denis, F. Varenne : "***Interprétabilité et explicabilité de phénomènes prédits par de l'apprentissage machine***", article accepté et en cours de publication dans ROIA, 2021
- Ch. Denis : "***Le périple de l'éthique de l'Intelligence Artificielle dans la révolution en cours des systèmes de soins***", Journal de Médecine Légale - Droit, Santé et Société, vol. 3 (3), Santé et intelligence artificielle Quelle(s) révolution(s) ?, pp. 17-21, (Eska) (2021)
- Th. Bayet, T. Brochier, Ch. Cambier, A. Bah, Ch. Denis, N. Thiam, J.-D. Zucker : "***A Machine Learning approach to improve the monitoring of Sustainable Development Goals : a case study in Senegalese artisanal fisheries***", CNIA 2021 : Conférence Nationale en Intelligence Artificielle, Bordeaux (virtuel), France, pp. 30-37 (2021)
- J. Cárdenas Chapellín, Ch. Denis, H. Mousannif, Ch. Camerlynck, N. Florsch : "*Réseaux de Neurones Convolutifs pour la Caractérisation d'Anomalies Magnétiques*", Actes CNIA 2021, Bordeaux (en ligne), France, pp. 84-90 (2021)
- Ch. Denis, J. Nicogossian : "***Du gène à l'octet : la communication phygitale pour une utilisation responsable de l'Intelligence Artificielle dans le domaine médical***", (2019)
- Ch. Denis, F. Varenne : "I***nterprétabilité et explicabilité pour l'apprentissage machine : entre modèles descriptifs, modèles prédictifs et modèles causaux. Une nécessaire clarification épistémologique***", Actes Conférence Nationale en Intelligence Artificielle (CNIA), PFIA 2019 (https://www.irit.fr/pfia2019/wp-content/uploads/2019/07/actes_CNIA_PFIA2019.pdf), Toulouse, France, pp. 60-68 (2019)
- Ch. Denis : "***Towards an explainable and convivial AI based tools: Illustration on medicine applications***", orateur invité, http://www.chistera.eu/christophe-denis, Tallinn, Estonia (2019)