



# EPISTÉMOLOGIE ET PHILOSOPHIE DES SCIENCES DE L'APPRENTISSAGE MACHINE

Christophe DENIS

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

Doctorant en Philosophie – ERIAC – Université de Rouen Normandie

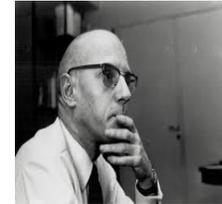
« “Le réel n'est jamais ce qu'on pourrait croire, mais il est toujours ce qu'on aurait dû penser.” »,  
Gaston Bachelard

**Atelier Fender, GDR MADICS , 8 juillet 2021**

# PREAMBULE : ON N'Y VOIT RIEN



« La lecture du tableau par Foucault se fonde sur l'hypothèse qu'il faut feindre que nous ne savons pas ce qui se reflète dans le miroir. Or, historiquement c'est absolument impossible puisque ce tableau a été peint à la demande du roi d'Espagne, et destiné à son bureau privé [...] »



*Daniel Arasse, Éloge paradoxal de Michel Foucault à travers Les Ménines », Histoires de Peinture, France Culture*

- Les explications provenant des philosophes souffrent souvent d'anachronisme empêchant l'acceptabilité de la part des historiens
- Les interprétations semblent à première vue irréconciliables entre historiens, théoriciens, philosophes et sémiologues
- Pour autant, selon Daniel Arasse, un anachronisme contrôlé « permet d'ouvrir de nouvelles questions et de nous faire mieux regarder la tableau »
- Par analogie, peut-on utiliser et sous quelles conditions l'apprentissage machine pour la compréhension de phénomènes complexe, même elle ne suit pas forcément le principe de parcimonie et non basée a priori sur des relations causales ?

# EN PHYSIQUE, ON EST PERDU DANS LES MATHS !

## LOST IN MATHS

*Comment la beauté égare la physique*

« Construit sur la base de riches entretiens avec les plus grands théoriciens *Lost in Maths* propose un regard bien informé sur l'état actuel de la physique fondamentale, par une physicienne tout à fait honnête, très drôle, et qui n'a pas froid aux yeux. »

Peter WOIT, physicien à Columbia University

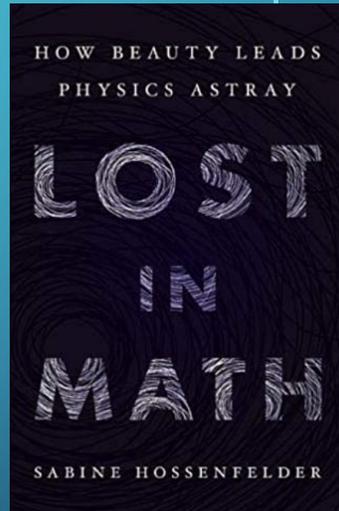
« Même les lecteurs très au fait peuvent achopper sur les concepts de la physique moderne, mais ils n'auront aucun mal à apprécier cette polémique clairvoyante, à la pugnacité jouissive, sur ses principales controverses. »

KIRKUS

La plupart des physiciens conçoivent la beauté formelle comme la voie royale vers la découverte scientifique. Aveuglés par l'élégance mathématique, ils s'interrogent sur les trous noirs, développent des théories stupéfiantes, inventent des dizaines de nouvelles particules, décrètent des modèles de grande unification. Mais aucune ou presque de ces idées n'a été confirmée par l'observation — en fait, nombre d'entre elles sont tout bonnement invérifiables. Pourtant les théoriciens sont persuadés que leurs gracieuses équations et leurs formules élégantes recèlent de formidables vérités sur la nature. Résultat, la discipline est aujourd'hui dans l'impasse.

Dans *Lost in Maths*, la physicienne Sabine Hossenfelder mène l'enquête pour comprendre comment l'obsession moderne de la beauté empêche de voir le monde tel qu'il est. Elle invite les physiciens à repenser leur façon d'édifier des théories et rappelle la nécessité, pour les scientifiques, d'accepter le désordre et la complexité afin de découvrir la vérité sur notre univers.

**Sabine Hossenfelder** est chercheuse à l'Institut des Études Avancées de Francfort (FIAS) et auteur du blog de vulgarisation *Backreaction*.



« La physique n'est pas des mathématiques et le développement de nouvelles théories doit être basé sur des données. En l'absence de comparaison avec les données expérimentales, les critères esthétiques des équations sont conservés », Sabine Hossenfelder

# CONTEXTE DES TRAVAUX

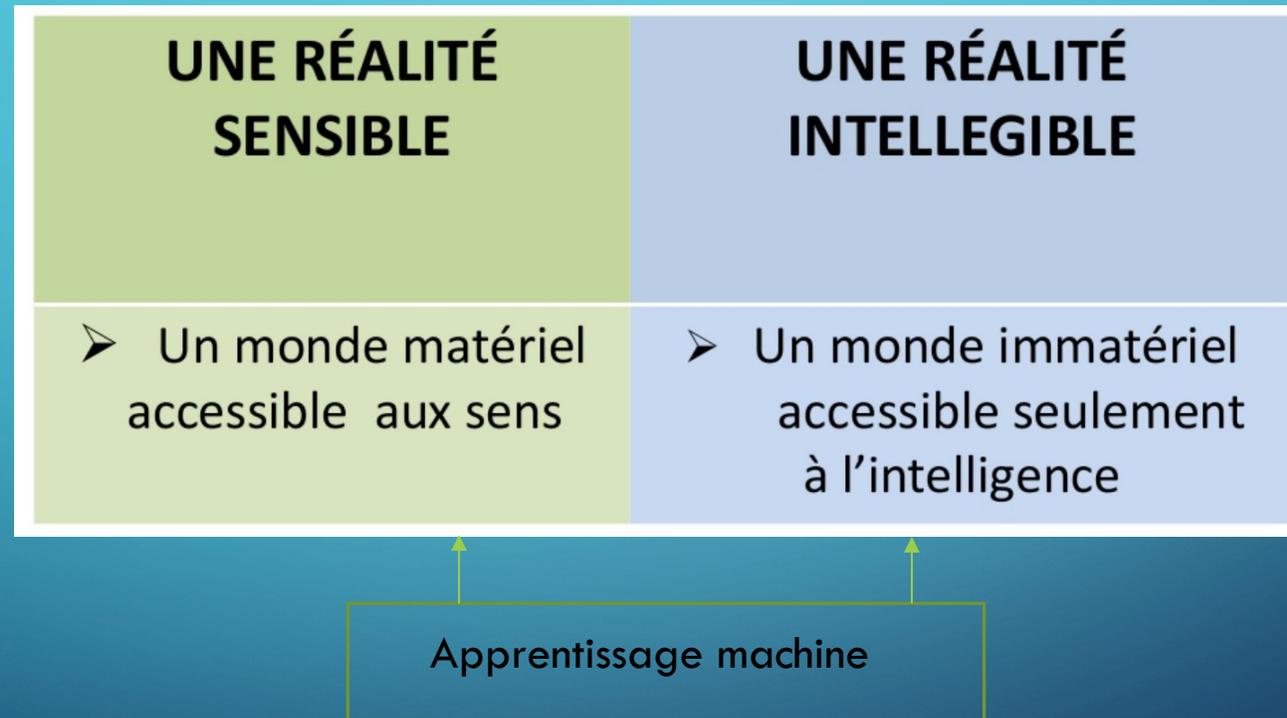
- La sophistication des techniques de traitement de l'information conduit depuis ces dernières décennies à une numérisation toujours croissante de nombreux secteurs de notre société.
- Ce constat n'échappe pas au monde scientifique accélérant la génération d'une avalanche de données dont l'analyse porte l'espoir d'interroger notre monde pour en améliorer sa connaissance.
- 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.

# CONTEXTE DES TRAVAUX

- Pour autant, l'approche hypothético-déductive atteint certaines limites pour l'analyse de phénomènes plus complexes et parfois couplés.
- Il existe aussi des théories portant par exemple sur les trous noirs et l'unification de modèles physiques qui reposent uniquement sur des théories mathématiques, provenant du monde intelligible, non confrontées avec des mesures physiques, provenant du monde sensible.
- La complexité des expérimentations permettant de confronter ces théories, freine la discipline scientifique, comme ce fut le cas pour vérifier la présence du boson de Higgs, découvert théoriquement près de cinquante ans plus tôt.

# L'APPRENTISSAGE MACHINE VU COMME UN DEMIRUGE PLATONICIEN

« *Imagine toi que le Bien et le Soleil sont deux Rois, l'un du mode intelligible, l'autre du monde sensible* », Platon, la République, Livre 6.



Utilisation de l'apprentissage machine profond comme « un démiurge platonicien » pour faire un lien entre le monde des lois et le monde sensible

# HYPOTHESE PRINCIPALE DE LA THESE

- Depuis 2010, l'apprentissage machine produit des résultats spectaculaires dans le domaine de la reconnaissance des formes bouleversant radicalement la discipline du traitement du signal.
- De nombreuses disciplines scientifiques notamment computationnelles s'orientent vers l'apprentissage machine, par exemple pour prédire plus précisément par exemple
  - le comportement turbulent d'un fluide
  - le comportement d'un système chaotique utilisé dans la science du climat.

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 des 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 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)

## 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)

# MODELISATION PAR APPRENTISSAGE MACHINE

- La mise au point d'une application basée sur de l'apprentissage machine va au-delà du seul entraînement 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.
- Travaux en cours
  - Statut ontologique des données d'apprentissage
  - Impact du type d'algorithme d'apprentissage machine, utilisant un nombre limité de techniques mathématiques, sur le modèle épistémologique du modèle

Besoin de prendre en compte la finalité du modèle d'AM pour ne pas rester dans des généralités

# FOCUS DE LA THESE

## MACHINE APPRENANTE DEDIEE AUX EDO ET EDP

Les EDO et les EDP sont majoritairement pour modéliser des phénomènes physiques (approche hypothético-déductive)  
On dispose de l'ensemble des codes sources pour reproduire et tester les solutions sur un cluster de GPU

- 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* »,

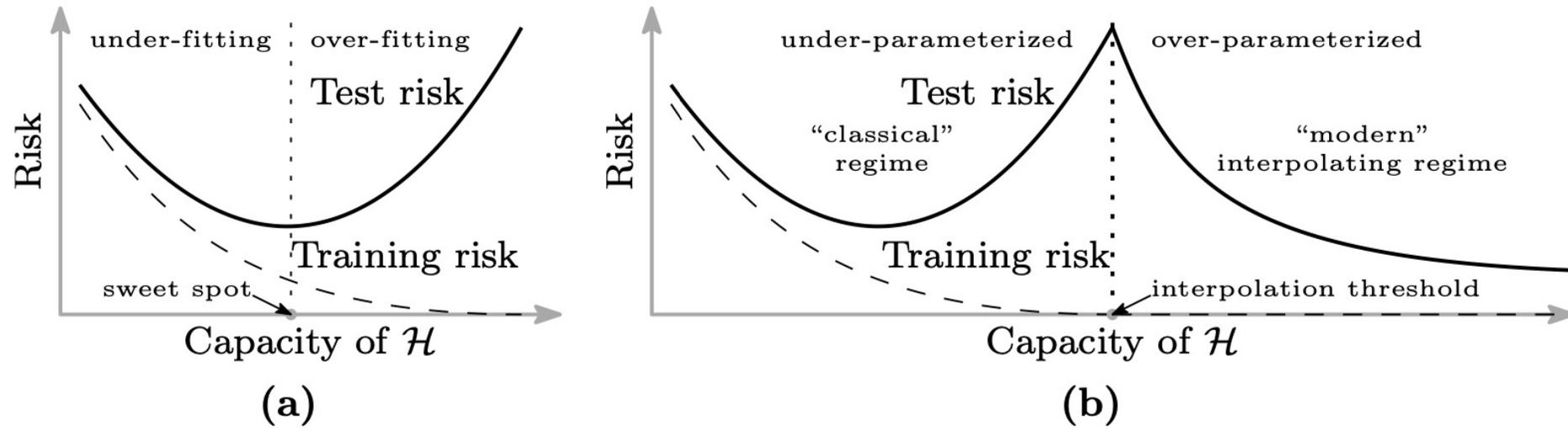
- Pronostic grâce à des équations différentielles ordinaires neuronales

Fompeyrine et, « *Enhancing Human-Machine Teaming for Medical Prognosis Through Neural Ordinary Differential Equations (NODEs)* », <https://arxiv.org/abs/2102.04121>, 2021

Appliquées à l'apprentissage automatique, les équations différentielles ordinaires neuronales (NODE) sont des algorithmes qui codent la dynamique d'un système en apprenant une équation différentielle ordinaire pour l'approximation de fonction

Ricky TQ Chen et al. "Neural ordinary differential equations". In: *arXiv preprint arXiv:1806.07366* (2018).

# LE PRINCIPE DE PARCIMONIE ET LE DILEMME BIAIS-VARIANCE



Mikhail Belkin and Daniel Hsu and Siyuan Ma and Soumik Mandal, *Reconciling modern machine learning practice and the bias-variance trade-off*, 2019

# LE PRINCIPE DE PARCIMONIE ET LE DILEMME BIAIS-VARIANCE

**First razor:** Given two models with the same generalization error, the simpler one should be preferred because simplicity is desirable in itself.

On the other hand, within KDD Occam's razor is often used in a quite different sense, that can be stated as:

**Second razor:** Given two models with the same training-set error, the simpler one should be preferred because it is likely to have lower generalization error.

Pedro Domingos, The Role of Occam's Razor in Knowledge Discover, 2004

# INTERPRETABILITE ET EXPLICABILITE

- Une analyse de l'état de l'art en Intelligence montre une absence de consensus et de distinction claire entre ces deux concepts, souvent utilisés d'ailleurs de manière inconsistante.

*"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 :

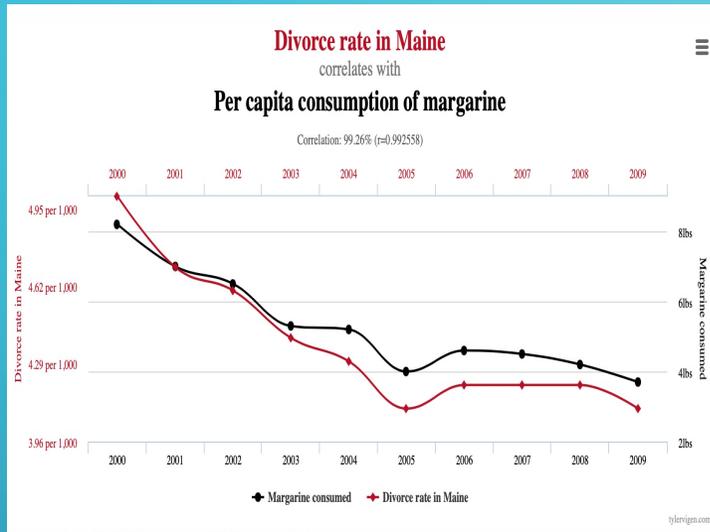
*" Interprétabilité d'un modèle : propriétés qu'a un modèle de se voir constitué d'éléments (signes, figures conceptuelles, données, etc.) qui ont un sens, c'est-à-dire un référent possible, pour le sujet humain ", (C. Denis, F. Varenne, 2019).*

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

*" 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).*

• Comment définir des relations causales entre les éléments.

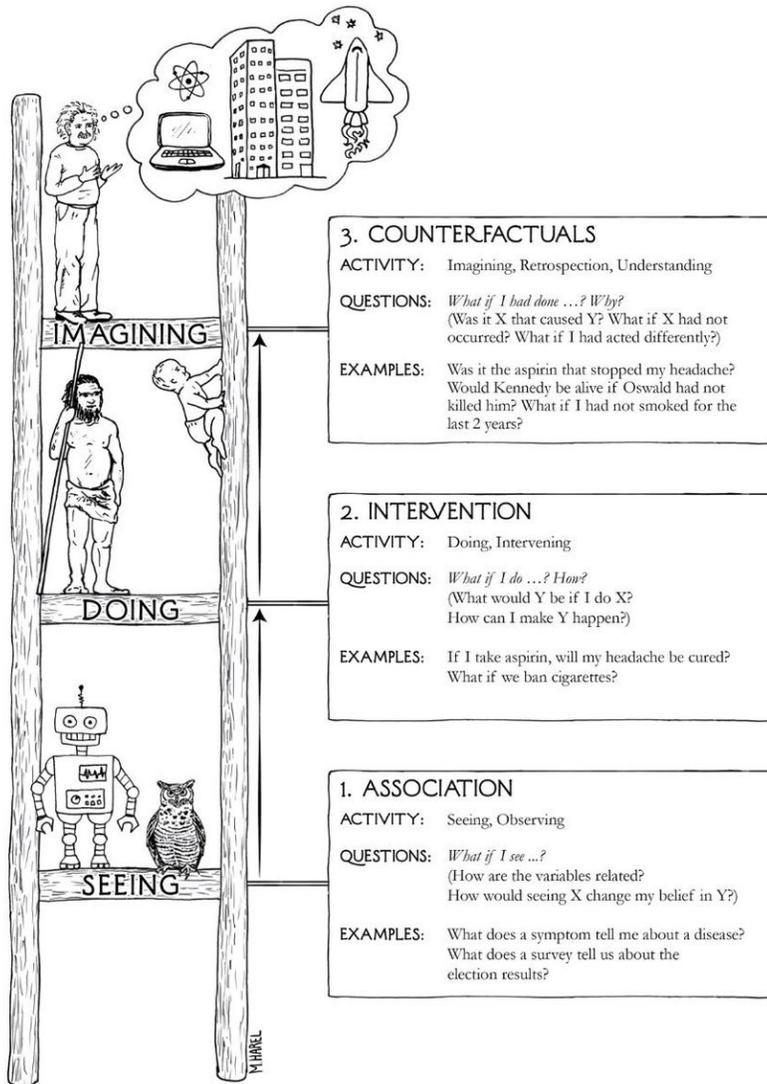
# CAUSALITE ET CORRELATION



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

- **Corrélation  $\neq$  Causalité**  $\rightarrow$  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 à effet
  - entre  $x$  et  $y$
  - ou entre au moins une des deux variables avec une autre variable  $z$

# 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.

# QUEL MODELE EXPLICATIF POUR UNE FINALITE DE PREDICTION OU DE DIAGNOSTIC ?

## Selon la finalité du modèle d'AM

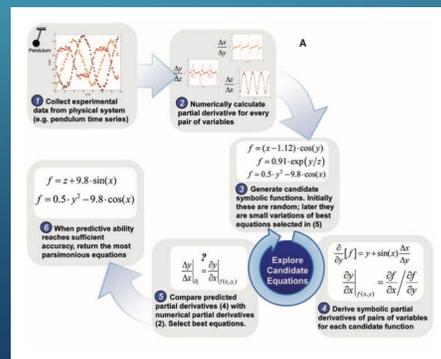
- Quels sont les rôles joués dans le modèle explicatif par
  - le traitement des corrélations – causalités
  - le traitement signal –signal
  - les mathématiques
  - l'interaction humain/modèle

	Diagnosis	Prediction
	Diagnosis and decision-making reassemble mature knowledge elements in a projected support, i.e. a stable mental map to compute observed data of the phenomenon.	Prediction reassembles non-mature knowledge elements (ideas) in a projected support, i.e. a mental map, consisting in a mix of structures and representations, that is able to gather dynamically observed data of the phenomenon.
Phenomenon	The mental map could be considered as organized following a mechanistic model in patterns and subpatterns that, dominates, hides, expands, reduces, secures internal coherence and organises dependencies. Subpatterns would then organize themselves dynamically around one master pattern.	The mental map could be considered projected as following a series of consecutive states of the initial phenomenon that mix subpatterns, identify the master pattern, respect internal coherence, reject noise, project the future organizations. Prediction then would be projected around the domination of one master pattern.
Expert limitations	Expert limitations & Pace of internal changes in the phenomenon is most of the time anticipated in order to confirm the diagnosis and take the corresponding decision. A prognosis is delivered following the previously identified mechanistic model.	Pace of changes in the organization of all subpatterns can hardly be anticipated. Avoiding noise is at the origin of partiality in the prediction.

Table 1: Structure of mental process for diagnosis and prediction of a human expert.

# 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.
- Peut-on avoir accès à certains signes avec un modèle explicatif de l'AM ?



# TRAVAUX EN COURS

- Etude de la remise en cause de la structure épistémologique de différentes disciplines scientifiques (Projet ANR SCIENTIA)
- Poursuite l'analyse des travaux de Judea Pearl et de Clark N. Glymour, repositionner la question de la causalité pour un modèle d'apprentissage machine
- Proposer des modèles explicatifs de l'apprentissage machine basée sur l'analyse de ses travaux en faisant un lien avec l'approche signal-signe.
- Revisiter le principe de parcimonie



- Thème du congrès Sciences et scientificité, 8 – 10 septembre 2021, Mons (Belgique)
- Focus de la session Intelligence Artificielle et Big Data

› Enjeux scientifiques et épistémologiques de l'explicabilité de la prédiction de phénomènes physiques par de l'apprentissage machine profond - *Christophe DENIS, LIP6, Équipe de Recherche Interdisciplinaire sur les Aires Culturelles*

› Data Science et scientificité : « The End of the Theory » ? - *Peggy Touvet, Université de Technologie de Troyes*

› Données écologiques massives : entre hybridation de traditions scientifiques anciennes et transformations inédites des pratiques de recherche en écologie. - *Léo TROCME--NADAL, Institut méditerranéen de biodiversité et d'écologie marine et continentale, Centre Gilles-Gaston Granger*

› A Falsificationist Account of Artificial Neural Networks - *Oliver Buchholz, University of Tübingen, Cluster of Excellence "Machine Learning: New Perspectives for Science"*

› Quantum logic as reversible computing - *Basil Evangelidis, National Capodistrian University of Athens, FernUniversität in Hagen*

# MERCI DE VOTRE ATTENTION

« Copernic a chassé l'homme du centre de l'univers, Lamarck et Darwin du sommet de la création, Freud de la maîtrise de soi, notre société a tendance à le chasser du cœur de ses projets.»

Axel Kahn

