THÈSE DE DOCTORAT

DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE-FRANCHE-COMTÉ

PRÉPARÉE À L'UNIVERSITÉ DE BOURGOGNE

École doctorale n°37

Sciences Physiques pour l'Ingénieur et Microtechniques

Doctorat d'Informatique

par

SARAH GHIDALIA

Étude sur les mesures d'évaluation de la cohérence entre connaissance et compréhension dans le domaine de l'intelligence artificielle

Thèse présentée et soutenue à Dijon, le 10 janvier 2024

Composition du Jury :

M. SÉBASTIEN FERRÉ	Professeur à l'Université de Rennes 1	Rapporteur
M. GILLES GESQUIÈRE	Professeur à l'Université Lumière Lyon 2	Rapporteur
M. BART LAMIROY	Professeur à l'Université de Reims	Examinateur
M. CHRISTOPHE NICOLLE	Professeur à l'Université de Bourgogne	Directeur de thèse
Mme AURÉLIE BERTAUX	Maître de conférences HDR à l'Université de Bourgogne	Codirectrice de thèse
Mme OUASSILA LABBANI-NARSIS	Maître de conférences HDR à l'Université de Bourgogne	Codirectrice de thèse

N° 1 2 7 7 7 0

Remerciements

Cette rédaction des remerciements marque la fin d'une aventure commencée à l'automne 2020 et je suis ravie que ce moment tant attendu arrive enfin.

Tout d'abord je tiens à remercier M. Sébastien Ferré, Professeur des Universités à l'Université de Rennes 1, ainsi que M. Gilles Gesquière, Professeur des Université à l'Université Lumière Lyon 2 pour avoir accepté de rapporter le présent manuscrit. Je souhaite leur exprimer ma gratitude pour les diverses suggestions et remarques qu'ils ont formulées dans le but d'enrichir et d'améliorer mon travail.

J'exprime également ma reconnaissance envers M. Bart Lamiroy, Professeur des Universités à l'Université de Reims Champagne-Ardenne, pour avoir accepté d'examiner cette thèse et de faire partie de mon jury.

Christophe, Ouassila et Aurélie merci de m'avoir accordé de manière rituelle une heure dans votre emploi du temps hebdomadaire, les discussions qui ont eu lieu à ces moments-là ont toujours été très enrichissantes. Pardon pour toutes les relectures d'articles avec des deadlines très (trop) courtes !

Christophe, je ne vous remercierai jamais assez d'avoir accepté d'être mon directeur de thèse. Devoir gérer mes états d'âme et mon caractère rebelle n'est pas une chose facile au quotidien pourtant vous y arrivez à merveille.

Ouassila, il est difficile d'exprimer à quel point je te suis reconnaissante pour toutes les heures que tu as passées à relire et me conseiller pour corriger ma prose. Tu as toujours été très présente durant ma thèse et je pense que peu de doctorants ont eu droit à un tel soutien de la part de leur directeur de thèse, pour cela je te remercie énormément. Merci également d'avoir pris du temps, en plus des réunions hebdomadaires pour discuter de mon travail voir simplement pour savoir comment j'allais d'un point de vue psychologique. Je sais que tu aimerais faire bien plus encore mais pour y arriver il faudrait que tes journées durent plus de 24h !

Aurélie, merci d'avoir répondu présente à chaque fois que j'ai eu besoin de tes conseils avisés. Tu as raison, l'heure du café (ou plutôt du thé pour moi) est le meilleur moyen pour échanger. Je te remercie d'avoir toujours réussi à trouver de la place dans ton emploi du temps chargé pour m'aider lorsque j'en avais besoin.

David, le seul membre de ma famille qui comprenne ce que je raconte quand je parle d'informatique. Merci de m'avoir fait à manger pendant plusieurs mois et de m'avoir aidé

à monter une cuisine l'année passée, cela m'a permis de consacrer plus de temps à mon travail de recherche.

Thierry et Isabelle, papa et maman, mes principaux sponsors depuis ma naissance. Merci pour votre soutien indéfectible, j'ai de la chance d'avoir une famille aussi aimante que celle-ci.

Simon, Nicolas, Cheikh, Davide, Pauline et tous mes autres collègues, merci pour votre soutien et nos discussions autour des repas de midi.

Arnaud, Corentin, Martha, Mylène, Samuel et Simon, onze ans d'amitié cette année, oui j'ai compté, je ne saurais comment vous remercier d'être vous-même.

Triton, merci de prendre de mes nouvelles régulièrement, et de me faire relativiser avec la rédaction de cette thèse.

Merci à ma famille au sens plus large, grands-parents, oncles, tantes, cousins et cousines, qui s'enquièrent toujours de comment je vais et comment mon travail avance. Je ne m'illusionne pas, je sais que vous faites cela pour être sûrs d'être invités au pot de thèse, surtout Olivier !

Sommaire

I Contexte et Problématiques 1

1 Introduction 3

1.1 Contexte . . . 3

1.2 Objectif de la thèse . . . 6

1.3 Plan de la thèse . . . 6

2 Contexte applicatif et orientation scientifique 9

2.1 Introduction . . . 10

2.2 H2020 RESPONSE : Les problématiques de la Smart City . . . 11

2.2.1 Les Districts et Bâtiments à Énergie Positive . . . 13

2.2.2 La place de l'usager dans la Smart City . . . 14

2.2.3 Les prédictions en Smart City . . . 17

2.3 Un problème de cohérence . . . 18

2.3.1 La cohérence en apprentissage automatique . . . 18

2.3.2 Pourquoi évaluer la cohérence ? . . . 19

2.3.3 La représentation de la connaissance . . . 23

2.4 La connaissance préalable ou *prior knowledge* . . . 23

2.4.1 Formalisation de la connaissance . . . 25

2.5 Conclusion . . . 29

3 État de l'art 31

3.1 Introduction . . . 33

3.2 Méthodologie de la SLR . . . 34

3.2.1 Planification de la revue . . . 34

3.2.2 Création de la revue . . . 38

3.3 Analyse statistique globale des études 40
3.3.1 Techniques d'hybridation utilisées pour combiner ontologie et apprentissage automatique (RQ1) 41
3.3.2 Identification des différents algorithmes d'apprentissage automatique utilisés (RQ2) 42
3.3.3 Usage du raisonnement déductif hors liens de subsomption (RQ3) . 44
3.3.4 Grandes thématiques de l'intelligence artificielle abordées (RQ4) . . 44
3.3.5 Domaines d'application des études (RQ5) 46
3.3.6 Analyse temporelle et géographique 48
3.4 Les trois catégories principales 50
3.4.1 Ontologie améliorée par l'apprentissage | Learning-Enhanced Ontology 50
3.4.2 Apprentissage automatique piloté par l'ontologie | Ontology-driven machine learning 58
3.4.3 Système d'apprentissage et de raisonnement | Learning and reasoning system 64
3.5 Positionnement 68
3.5.1 Les modèles de conception pour l'IA hybride 69
3.5.2 La taxonomie du neuro-symbolique 70
3.5.3 La combinaison de l'ontologie avec l'apprentissage automatique . . 71
3.6 Conclusion 71
3.6.1 Les trois défis de l'IA hybride 73
3.6.2 Bilan 74

II Contribution 75

4 Apprentissage automatique enrichi 77

4.1 Introduction 79
4.2 Enrichissement par les connaissances 80
4.2.1 Les données d'entraînement 81
4.2.2 L'architecture du modèle 85

4.2.3 La phase d'apprentissage . . . 87
4.2.4 Le modèle final . . . 88
4.2.5 Bilan sur l'ajout de connaissance en apprentissage automatique . . . 89
4.3 Évaluation de la qualité . . . 90
4.3.1 Évaluation de la qualité en apprentissage automatique . . . 91
4.3.2 Stratégies d'évaluation des modèles . . . 93
4.3.3 Limites dans l'évaluation des modèles sans cohérence . . . 97
4.3.4 Évaluation des différents types de cohérence . . . 99
4.4 Méthodologie d'évaluation de la cohérence . . . 109
4.4.1 Analyse de l'existant . . . 109
4.4.2 Approche d'évaluation des systèmes d'apprentissage automatique informés . . . 111
4.5 Conclusion . . . 115

5 Transformation des connaissances en contraintes 117

5.1 Introduction . . . 119
5.2 Exemple basé sur des facteurs prévisibles . . . 120
5.2.1 L'oscillateur harmonique amorti . . . 120
5.2.2 Prédire sans connaissance . . . 122
5.2.3 Prédire avec des connaissances . . . 124
5.3 Scénario Réel : Complexité et Chaos . . . 125
5.3.1 Modélisation de la durée de vie des matériaux . . . 126
5.3.2 Ajout de connaissances dans un réseau de neurones . . . 129
5.3.3 Bilan et limites . . . 131
5.4 L'apprentissage automatique informé par une ontologie . . . 133
5.4.1 Ontology-based Physics-Informed Machine Learning . . . 133
5.4.2 Mise en oeuvre de l'expérimentation . . . 138
5.4.3 Évaluation . . . 141
5.5 Conclusion . . . 147
5.5.1 Challenges . . . 147

III Conclusion **149**

6 Conclusion générale **151**

6.1 Travaux effectués 153

6.2 Perspectives 154

6.2.1 Vers une IA de confiance 155

6.2.2 Simulations de phénomènes physiques complexes 156

6.2.3 Modularisation des systèmes d'IA 157

Liste des Figures **187**

Liste des Tableaux **190**

Liste des publications **191**

I

Contexte et Problématiques

1

Introduction

1.1 Contexte 3
1.2 Objectif de la thèse 6
1.3 Plan de la thèse 6

1.1/ Contexte

L'intelligence artificielle (IA) est une des technologies les plus marquantes de ce début de 21ème siècle. Dans un monde de plus en plus interconnecté, son influence a radicalement transformé notre manière de vivre, de travailler et de communiquer. Les outils les plus récents comme ChatGPT [1], Midjourney [2], Dall-E [3] et consorts démocratisent l'usage de l'intelligence artificielle. Aujourd'hui, tout un chacun peut bénéficier d'un assistant virtuel capable de rédiger ses e-mails ou de créer une illustration à moindre coût.

Pourtant, il n'existe pas de consensus sur la définition du terme "intelligence artificielle" inventé il y a plus d'un demi-siècle [1, 2]. Marvin Minsky [3] suggère que l'IA se réfère à la capacité des machines à *résoudre des problèmes complexes*, tandis que Dimiter Dobrev [4] compare l'IA aux êtres humains, la définissant comme *"un programme qui, dans un monde arbitraire, ne se débrouillera pas plus mal qu'un humain"*, ce qui rappelle la définition originale de John McCarthy [5], autre pionnier de la discipline. Toutefois, comparer l'intelligence virtuelle à l'intelligence humaine pour tenter de trouver une définition claire pose un gros problème : il faudrait auparavant avoir une explication claire de ce qu'est l'intelligence humaine. Or, là encore, il n'existe pas de consensus à ce jour sur la définition formelle de l'intelligence humaine dans le monde des neurosciences.

1. https://chat.openai.com
2. https://www.midjourney.com
3. https://openai.com/dall-e-2

Certains auteurs ont proposé une définition plus pragmatique de l'IA en tentant de décrire son fonctionnement. Ainsi, Andreas Kaplan et Michael Haenlein la définissent comme *la capacité d'un système à interpréter correctement des données externes, à apprendre à partir de ces données et à utiliser ces apprentissages pour atteindre des objectifs et des tâches spécifiques grâce à une adaptation flexible* [6], mais cette définition réduit fortement l'IA au domaine de l'apprentissage automatique. Pei Wang met l'accent sur la capacité de l'IA à s'adapter à son environnement, même avec des connaissances et des ressources limitées [2], ce qui permet d'être moins restrictif quant aux techniques utilisées pour le faire.

Il est d'ailleurs dommageable d'avoir une définition trop limitée n'incluant pas les différentes facettes du domaine. Cette problématique ne date pas d'hier, l'IA est souvent cantonnée à la technologie du moment. Pei Wang faisait déjà ce reproche en 2007, époque à laquelle l'IA était vue principalement comme un système axiomatique, donc purement formel [7]. Aujourd'hui, c'est une vision diamétralement opposée qui prédomine. Pour beaucoup, il est inconcevable de parler d'IA sans implémenter un algorithme d'apprentissage automatique. Heureusement, Stuart Russel et Peter Norvig nous rappellent que ce domaine est bien pluridisciplinaire ; il inclut la robotique, le raisonnement logique, la prise de décision, le traitement du langage naturel, la vision par ordinateur, la résolution de problème et l'apprentissage automatique [8]. De là, une définition plus inclusive de l'intelligence artificielle pourrait être la suivante : l'IA est la science qui permet aux machines de percevoir, de comprendre et d'interagir avec le monde réel d'une manière proche de celle des êtres humains.

Pour comprendre cette perspective, nous pouvons faire appel à l'allégorie de la caverne exposée par Platon [9]. Comme les prisonniers enchaînés au fond de la grotte, qui ne perçoivent que des ombres et des échos du monde intelligible, les machines ont également une perception de notre monde limitée aux données qui leur sont fournies. Dès lors, comment permettre aux machines d'avoir un impact sur le monde tangible si elles ne peuvent pas comprendre les différents aspects, les complexités, et les nuances du monde réel ? Il faudrait progressivement sortir les machines de la caverne dans laquelle elles sont enchaînées. La première étape consiste à doter les machines de capacités de réflexion pour la résolution de problèmes. Pour ce faire, le processus cognitif humain s'appuie principalement sur deux formes de raisonnement : l'**induction** et la **déduction**. Le raisonnement inductif facilite la découverte de connaissances générales (lois, théorèmes, corrélations, etc.) à partir d'observations spécifiques. Le raisonnement déductif, quant à lui, permet d'appliquer des connaissances générales préexistantes à des cas spécifiques [10] [4].

4. Au XIXe siècle, Charles Sanders Peirce a identifié un troisième type de raisonnement, appelé **abduction**, qui est utilisé pour générer des hypothèses qui expliquent des observations spécifiques [10, 11]. L'abduction revêt une importance scientifique considérable, car elle concerne les questions de causalité, l'intelligence artificielle explicable (XAI) et, potentiellement même, la fiabilité. Cependant, notre travail ne se

Socrate, disciple de Platon, définit l'induction comme un mode de raisonnement qui consiste à tirer une conclusion générale à partir de plusieurs cas particuliers. Le raisonnement inductif est une forme de raisonnement ampliatif, c'est-à-dire que l'on tire des conclusions qui vont au-delà des informations contenues dans les prémisses [11]. Ce type de raisonnement est très proche des mécanismes de l'apprentissage automatique : établir un raisonnement (modèle) à partir de faits explicites (expériences). Ainsi, le modèle n'est pas explicitement écrit, au contraire, il est déduit des données d'entrée afin d'en extraire des informations (lois générales).

À l'inverse, le raisonnement déductif est basé sur le syllogisme défini par Aristote tel qu'une *"parole (logos) dans laquelle, certaines choses ayant été supposées, quelque chose de différent de ces suppositions résulte par nécessité de leur être ainsi"*[5]. En d'autres termes, le raisonnement déductif est la capacité de tirer des conclusions sur des faits individuels (expériences) à partir de connaissances génériques (loi générale). Lorsque Aristote écrit *choses supposées*, cela correspond à la prémisse de l'argument, et lorsqu'il écrit *résultats de la nécessité*, cela correspond à la conclusion de l'argument [12]. Dans le domaine de l'IA, le raisonnement déductif est principalement associé aux approches symboliques, communément appelées Good Old-Fashioned AI (GOFAI) [13].

La GOFAI englobe une gamme de techniques comprenant les systèmes basés sur la connaissance (par exemple, les systèmes experts), les systèmes multi-agents et les systèmes de raisonnement basés sur les contraintes. Ces approches s'appuient sur des outils symboliques tels que les graphes de connaissances, les règles logiques, les ontologies et le calcul algébrique pour faciliter le raisonnement déductif à l'aide de moteurs d'inférence. Ces moteurs de raisonnement déductif peuvent utiliser les axiomes décrivant les concepts dans la TBox (Terminology Box, i.e. les lois générales) pour déduire de nouvelles connaissances sur la partie ABox (Assertional Box, i.e. les faits spécifiques) de l'ontologie. Dans les systèmes d'information modernes, les ontologies, telles que les spécifications formelles et explicites des conceptualisations partagées [14], s'avèrent extrêmement précieuses. Elles répondent aux besoins d'interopérabilité des données et de formalisation des règles commerciales, qui sont des aspects essentiels des systèmes d'information contemporains.

Si l'apprentissage automatique est en plein essor au détriment des systèmes symboliques, c'est parce que notre société vit un accroissement de la donnée. Toutefois, cet accroissement des données va de pair avec une crise de la confiance. En effet, contrairement aux systèmes experts qui se basaient sur des connaissances reconnues dans lesquelles les êtres humains pouvaient se fier, aujourd'hui beaucoup de systèmes reposent uniquement sur des données qui sont parfois inexactes ou obsolètes. C'est pourquoi l'IA

concentrera pas sur l'abduction, mais plutôt sur la combinaison du raisonnement inductif et déductif à des fins de résolution de problèmes.

5. Analytiques antérieurs I.2, 24b18-20

hybride, qui a pour objectif de combiner les raisonnements, notamment inductif et déductif, est actuellement en plein essor. L'ajout de connaissance, par le biais du symbolique, a pour objectif de rendre les systèmes d'apprentissage automatique plus cohérents et plus fiables. Toutefois, une question importante demeure : comment évaluer la cohérence des systèmes d'intelligence artificielle mis au point ?

1.2/ Objectif de la thèse

L'objectif principal de cette thèse est de répondre à cette question. Ce travail de recherche est financé dans le cadre du projet H2020 RESPONSE (integ**R**at**E**d **S**olutions for **PO**sitive e**N**ergy and re**S**ilient citi**E**s) auquel le laboratoire CIAD est associé. Ce projet souhaite établir une vision stratégique pour la transition énergétique des villes intelligentes : des villes climatiquement neutres d'ici à 2050. RESPONSE vise à faire de la durabilité énergétique une vision réalisable en résolvant le trilemme énergétique (sécurité, équité/abordabilité, durabilité environnementale) au niveau des bâtiments, des îlots et des quartiers dans les villes intelligentes. Le projet s'appuie sur des systèmes énergétiques intelligents, intégrés et interconnectés, associés à des infrastructures urbaines axées sur la demande, à des modèles de gouvernance et à des services qui favorisent la durabilité énergétique. L'un des enjeux cruciaux de ce contexte est d'établir des prédictions sur les données de la SmartCity qui soient cohérentes avec la réalité physique des faits à venir pour éviter qu'une décision politique ne soit pas adaptée ou une action des services de la ville soit incomplète ou inappropriée. Le chapitre 2 présentera en détail le contexte applicatif du projet RESPONSE et le contexte scientifique en définissant l'élément clé de nos travaux : la **cohérence**. Nous verrons que cette cohérence des raisonnements nécessite de combiner différentes approches de raisonnement qui sont complémentaires, en particulier les raisonnements logiques via des ontologies et les approches d'apprentissages automatiques.

1.3/ Plan de la thèse

La thèse que nous présenterons ici s'inscrit dans le champ de l'informatique et plus particulièrement dans celui de l'apprentissage automatique, où la question de la notion de cohérence au sein des systèmes intelligents sert de fil conducteur. Notre objectif principal est d'analyser comment la cohérence, en tant que concept, peut être appréhendée et évaluée dans le domaine de l'intelligence artificielle, notamment en relation avec les connaissances préalables intégrées dans ces systèmes. Le fil rouge de notre travail réside dans l'étude approfondie de la cohérence, que nous aborderons sous différents angles, en explorant diverses méthodologies et applications.

Le premier chapitre pose les bases de notre étude en définissant les termes clés et le contexte de notre recherche. Nous nous intéresserons tout d'abord à la notion de Smart City et aux enjeux liés à l'utilisation de l'apprentissage automatique dans ce contexte où la cohérence entre les prédictions artificielles et les réalités de terrain reste une mesure primordiale avant de définir l'action politique. Ensuite nous nous pencherons sur la définition de la cohérence en rapport avec l'intelligence artificielle. Enfin, la notion de connaissance préalable, nécessaire à la définition de tout contexte de raisonnement, sera explorée de manière approfondie.

Le second chapitre présentera une revue systématique de la littérature qui nous permettra de cartographier l'état de l'art en matière de combinaison entre apprentissage automatique et modélisation de la connaissance préalable sous la forme d'ontologies. Notre analyse mettra en lumière les différentes façons dont ces deux domaines se croisent et interagissent, avec un focus particulier sur les techniques algorithmiques employées. Cette étude comparative nous permettra également de positionner notre recherche par rapport à d'autres travaux majeurs dans le domaine.

Le troisième chapitre abordera la question de la cohérence dans les systèmes d'apprentissage informés par les ontologies. Nous explorerons les différentes méthodes d'intégration des connaissances dans ces systèmes et analyserons comment la cohérence peut être évaluée en fonction des techniques utilisées. Nous nous intéresserons également à la qualité globale des systèmes d'intelligence artificielle, avec un accent particulier sur l'évaluation de la cohérence.

Enfin, le quatrième chapitre se concentrera sur l'application pratique de nos recherches, en examinant comment la cohérence d'un modèle peut être vérifiée par rapport à des lois physiques formulées sous forme d'équations différentielles partielles. Nous présenterons deux cas d'étude, l'un portant sur la prédiction des mouvements d'un oscillateur harmonique, l'autre sur l'estimation de la durée de vie d'un matériau, pour illustrer l'importance des connaissances préalables dans l'évaluation de la cohérence. Une nouvelle méthode de formalisation des connaissances dans une ontologie sera également proposée, avec une évaluation de son efficacité.

À travers ces quatre chapitres, nous espérons apporter un éclairage nouveau sur la notion de cohérence en apprentissage automatique, en explorant ses différentes facettes et en proposant une méthode innovante pour son évaluation. Cette thèse se veut être une contribution significative à la recherche dans le domaine de l'intelligence artificielle, en mettant en lumière l'importance de la cohérence dans la construction de systèmes intelligents fiables et pertinents.

2

Contexte applicatif et orientation scientifique

2.1 Introduction 10
2.2 H2020 RESPONSE : Les problématiques de la Smart City 11
2.2.1 Les Districts et Bâtiments à Énergie Positive 13
2.2.2 La place de l'usager dans la Smart City 14
2.2.3 Les prédictions en Smart City 17
2.3 Un problème de cohérence 18
2.3.1 La cohérence en apprentissage automatique 18
2.3.2 Pourquoi évaluer la cohérence ? 19
2.3.2.1 Le surapprentissage 20
2.3.2.2 La robustesse 20
2.3.2.3 Les biais 21
2.3.2.4 L'interprétabilité 22
2.3.3 La représentation de la connaissance 23
2.4 La connaissance préalable ou *prior knowledge* 23
2.4.1 Formalisation de la connaissance 25
2.5 Conclusion 29

Ce chapitre présente le contexte d'application de nos travaux de recherche : la ville intelligente et les deux défis scientifiques associés. Le premier défi concerne le domaine de l'ingénierie des connaissances et le second défi est lié au domaine de l'apprentissage automatique.

La résolution du premier défi, qui consiste à agréger des données hétérogènes, est largement étudiée dans la littérature scientifique. Cela nécessite l'utilisation de techniques d'alignement structurel et schématique des données. Un alignement sémantique est également requis, en particulier pour formaliser les connaissances du contexte de production des données. La formalisation de la connaissance permet de désambiguïser le sens réel des données au delà de leur simple description schématique. Dans le contexte de la ville intelligente, c'est un point important car l'analyse des données à pour objectif la construction d'une stratégie d'action politique ou sociale.

Or, cette stratégie se doit d'être cohérente avec la réalité de terrain, c'est là notre second défi. L'évaluation des prédictions réalisées grâce aux techniques d'apprentissage automatique doit permettre de vérifier cette cohérence. Finalement, tout l'enjeu est de concevoir des algorithmes d'apprentissages capables d'intégrer des connaissances préalables issues du contexte applicatif afin d'améliorer la cohérence de leur prédiction.

La présentation des notions de villes intelligente, de cohérence, de formalisation des connaissances préalables est le cœur de ce chapitre.

2.1/ Introduction

La ville intelligente, plus communément dénommée Smart City, est un concept urbain qui intègre différentes technologies dans le but d'améliorer la gestion et la qualité de vie de ses habitants. Un enjeu majeur pour la Smart City est le respect de la conformité juridique et politique. C'est un véritable défi lorsqu'il doit être appliqué au traitement de données massives comme celles remontées par des capteurs, qu'ils soient disséminés dans l'infrastructure de la ville ou au plus proche des usagers comme les smartphones.

L'emploi de techniques d'apprentissage automatique pour l'analyse et la prédiction sur les données de la Smart City s'est largement démocratisé depuis la dernière décennie. Capable de traiter massivement des données pour répondre à des questions telles que la prédiction de températures, de qualité de l'air ou d'améliorer le système de feux de circulation, ce type d'intelligence artificielle connaît un grand engouement. Pourtant, l'apprentissage automatique traditionnel ne prend pas explicitement en compte certaines contraintes de la Smart City qui ne sont pas représentées uniquement sous la forme de données. Cette lacune implique une plus grande divergence entre les prédictions des algorithmes et la réalité de terrain. Cet écart limite les possibles stratégies d'actions poli-

tiques pour répondre aux risques présentés dans les prédictions.

Dès lors, il paraît intéressant d'évaluer la cohérence entre les analyses et les prédictions réalisées avec les exigences légales ou les règles métier initiales. Or, l'évaluation de la cohérence n'est pas encore largement répandue dans les systèmes d'intelligence artificielle. Elle nécessite l'usage d'IA hybride capable de combiner des raisonnements inductifs sur des données et des raisonnements déductifs à partir de connaissances.

L'objectif de ce chapitre est de présenter le contexte applicatif et scientifique de la thèse en définissant les différents objets de l'étude. La première section s'intéresse à la définition de la Smart City et des différentes problématiques liées à la mise en place de prédictions sur des données de la ville [1]. La deuxième section présente la définition du terme cohérence lorsqu'il est appliqué à l'apprentissage automatique. La troisième et dernière section clarifie la notion de connaissance préalable utilisée dans la définition de la cohérence. Elle présente également en quoi une formalisation de cette connaissance peut faciliter l'évaluation de la cohérence.

2.2/ H2020 RESPONSE : Les problématiques de la Smart City

Les technologies de l'information et de la communication (TIC) sont au cœur des différents domaines de la Smart City. Toutefois, il n'est pas correct de résumer la Smart City à une ville faisant usage des TIC, le numérique n'étant pas la seule technologie impliquée dans ce type de projet. Il vaut mieux considérer la définition suivante : *"une ville intelligente est une zone géographique bien définie, dans laquelle des technologies de pointe telles que les TIC, la logistique, la production d'énergie, etc., coopèrent pour créer des avantages pour les citoyens en termes de bien-être, d'inclusion et de participation, de qualité de l'environnement et de développement intelligent ; elle est gouvernée par un groupe de personnes bien défini, capable d'énoncer les règles et la politique pour le gouvernement et le développement de la ville"* [2] [15]. Cette définition nous rappelle que l'amélioration de la qualité de vie de ses habitants est l'un des objectifs principaux d'une Smart City. Cette évolution de l'environnement urbain implique l'usage de différentes technologies au sens large.

La ville intelligente s'étend sur six domaines présentés sur la Figure 2.1 : l'environnement, la mobilité, l'économie, les citoyens, le Smart Living et la gouvernance [16, 17]. Les

1. Cette section a été rédigé en reprenant des éléments présents dans la publication suivante : S. Ghidalia, O. Labbani Narsis, A. Bertaux, and C. Nicolle, 'Ville intelligente : valeur et vérité des données numériques', in Smart City et prise de décision, Mare & Martin, 2023, pp. 47–62.

2. La citation originale est la suivante : "a smart city is a well defined geographical area, in which high technologies such as ICT, logistic, energy production, and so on, cooperate to create benefits for citizens in terms of well being, inclusion and participation, environmental quality, intelligent development ; it is governed by a well defined pool of subjects, able to state the rules and policy for the city government and development"

FIGURE 2.1 – Les six domaines de la Smart City [16]

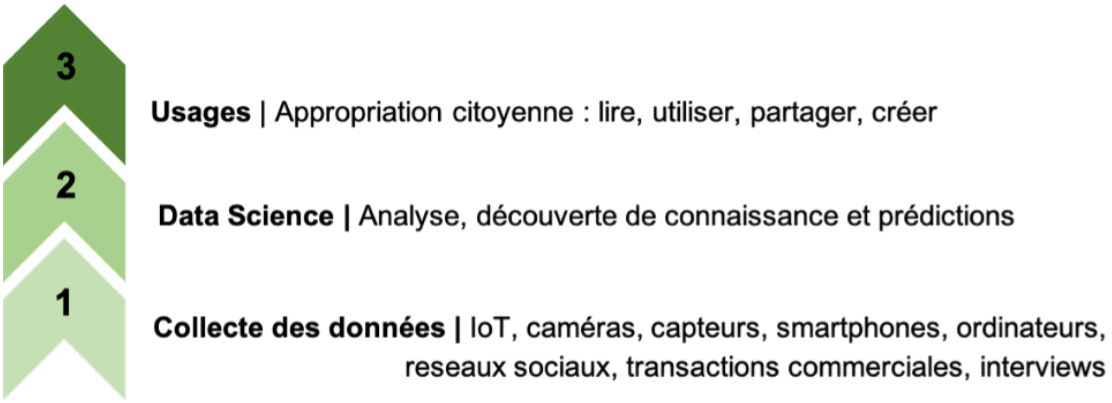

FIGURE 2.2 – Comment est utilisée la donnée dans une Smart City ?

technologies de l'information jouent un rôle essentiel dans chacun de ces six domaines de la Smart City, permettant le partage d'informations entre eux.

Les trois étapes principales du traitement de la donnée au sein d'une Smart City sont présentées par la Figure 2.2. D'abord la collecte, en général aux moyens de capteurs, de l'IoT, parfois d'applications numériques, d'interviews, ou même de smartphones. Ensuite, l'analyse des données avec pour objectif de mesurer, de découvrir de nouvelles connaissances et parfois de détecter des anomalies. Enfin, la communication des informations trouvées qui sont remontées au citoyen pour automatiser certains processus, améliorer sa qualité de vie ou bien l'informer et l'aider dans sa prise de décisions.

D'ici la fin de la décennie, nous aurons une plus forte concentration de la population mondiale au sein des villes. Or, les villes contribuent fortement à l'émission de gaz à effet de serre et consomment une grande majorité de l'énergie au niveau mondial [18]. Il est donc normal pour les différents acteurs mondiaux de s'intéresser de plus près aux Smart City s'ils veulent limiter l'augmentation de la température moyenne mondiale à 2°C

par rapport aux niveaux préindustriels comme ratifié dans les Accords de Paris par 189 États [3].

Pour l'Union Européenne (UE), ces problématiques sont d'autant plus fortes que d'ici 2050, c'est 80% de sa population qui habitera en ville [19]. Consciente des enjeux importants liés à l'urbanisation, l'UE a pris des engagements forts en terme d'environnement avec l'European Green Deal : un arrêt total des émissions de gaz à effet de serre (GHG) d'ici 2050, une croissance économique décorrélée de l'utilisation des ressources et la garantie qu'aucune personne ni aucun territoire ne seront laissés de côté. Cela explique pourquoi l'UE concentre ses efforts et octroie ses financements principalement si le projet de Smart City permet de répondre à des enjeux environnementaux [20].

2.2.1/ Les Districts et Bâtiments à Énergie Positive

L'habitat tient bien sûr une place très importante au sein des villes. Depuis l'antiquité, les êtres humains cherchent à construire des logements confortables en composant avec les éléments donnés par la nature. Ainsi, les habitations ont souvent été construites pour tirer avantage des rayons du soleil en hiver ou pour ne pas être trop exposé aux vents [21]. Ce concept d'habitation thermorégulé a d'ailleurs été étudié par Socrate qui avait adapté la forme et l'orientation de sa "Socratic House" pour créer une habitation durable et peu consommatrice d'énergie dès le V^{e} siècle av. J.-C [22].

C'est en se basant sur cette idée d'un habitat peu consommateur d'énergie qu'est né le concept de "Zero Energy Building" (ZEB), un bâtiment neutre en énergie. C'est-à-dire que l'énergie consommée dans le bâtiment doit être totalement générée au sein même de ce bâtiment [23]. Rapidement, l'idée encore plus ambitieuse d'un "Positive Energy Building" (PEB) a émergé. Cette fois-ci, le bâtiment doit produire à lui seul plus d'énergie qu'il n'en consomme. L'Europe a d'ailleurs décrit les capacités et contraintes des PEB dans son European Strategic Energy Technology (SET) Plan [24]. Un quartier composé d'un ensemble de bâtiments à énergie positive est appelé un "Positive Energy District" (PED), c'est ce genre de quartiers innovants que l'Europe veut mettre en place pour parvenir à son objectif de neutralité carbone d'ici l'horizon 2050. D'autant plus, qu'Europe, environ 40% de la consommation d'énergie est due aux immeubles [25].

Bâtir ou reconvertir d'anciens immeubles pour qu'ils produisent plus d'énergie que ce qu'ils en consomment est une stratégie qui demande des moyens techniques importants. Dans de nombreux cas, l'utilisation de meilleurs matériaux et l'installation de systèmes de création d'énergie renouvelable ne pourront pas suffire à générer plus d'énergie que le bâtiment en consomme. Pour limiter fortement la consommation d'énergie des bâtiments, il faut également établir des recommandations basées sur l'analyse d'une grande quantité

3. https ://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

de données.

Selon [23], il existe trois types de PED différents définis par l'European Energy Research Alliance (EERA) :

- PED autonomous : ils ne reçoivent jamais d'énergie de l'extérieur, mais ils peuvent envoyer au réseau leur surproduction d'énergie.
- PED dynamic : ils peuvent à la fois donner leur production, mais également recevoir de l'énergie d'autres sources (comme le réseau de distribution local) lorsqu'ils en ont besoin.
- PED virtual : ils permettent l'implémentation d'un système de stockage virtuel de l'énergie en dehors du site du PED.

L'analyse visant à réduire la consommation d'énergie doit considérer le mode de fonctionnement de chaque PED mais aussi sa manière de générer de l'énergie. En effet, il est important de ne pas étudier la consommation de l'énergie séparément de sa génération [26]. Ces deux aspects sont liés à bien des égards, c'est en optimisant les deux qu'on peut parvenir à une consommation énergétique optimisée.

Bien sûr, d'autres facteurs entrent en jeu lorsqu'il s'agit d'évaluer la consommation d'énergie. Tout d'abord, les conditions climatiques, extérieures aux bâtiments, elles influent fortement sur la consommation d'énergie à l'intérieur (chauffage, climatisation, etc.) [22,23]. Elles sont fortement liées aux comportements des usagers qui vont consommer de l'énergie en fonction de leurs besoins, de réguler la température ou d'augmenter la qualité de l'air à l'intérieur du bâtiment [25,27].

L'usager du bâtiment a un impact très important sur la consommation d'énergie, puisque cette consommation se base sur ses besoins au quotidien. Il faut donc considérer l'usager comme un facteur important dans la réduction de la consommation d'énergie [25–27].

2.2.2/ La place de l'usager dans la Smart City

Il existe beaucoup de définitions de la Smart City, mais la plupart d'entre elles partagent l'idée que la Smart City vise à améliorer la qualité de vie de ses citoyens, c'est-à-dire son confort et son bien-être.

Au début des années 2000, les Smart Cities étaient très orientées vers les TIC, non pas comme moyen, mais comme une fin en soi. Or, il faut prendre garde à ce genre d'organisation "top-down" pour ne pas créer une Smart City qui servirait plutôt l'État que les individus [28,29]. Un système qui repose uniquement sur les TIC (notamment les données et l'intelligence artificielle) sans prendre en compte les comportements et besoins humains ne pourra s'inscrire dans la durée à long terme. Suite aux critiques de l'organisation "top-down" initiale, il y a eu une évolution de l'image de la Smart City pour qu'elle soit plus orientée centrée sur le citoyen (citizen-centric) [28,30] au cours des années 2010.

Les TIC étaient toujours bien présentes, mais cette fois, elles devenaient un moyen de parvenir à une ville au service de ses habitants. Depuis quelques années, cependant, le terme "citizen-centric" devient un argument commercial [31], un moyen de promouvoir et de vendre la Smart City à travers le monde. Or, il est impératif de ne pas considérer l'usager comme un simple consommateur de la ville intelligente, mais comme un acteur principal.

Pour [29], il existe quatre modalités de citoyens principaux :

- les utilisateurs de service (service user) : ils consomment des services de la Smart City
- les entrepreneurs (entrepreuneurial) : ils participent activement à l'innovation et la co-création au sein de la Smart City
- les citoyens politiques (political) : ils jouent un rôle actif dans la gouvernance de la ville, notamment dans la prise de décision et les délibérations
- les participants civiques (civic) : ils participent à des activités communautaires qui ne sont pas directement liées à une activité économique

Pour être citizen-centric, la Smart City doit intégrer tout un panel de citoyens différents avec des engagements très disparates au quotidien. Cependant, il est important que l'amélioration de la qualité de vie concerne l'ensemble des citoyens de la Smart City [29]. Il faut veiller à ce que certains habitants ne soient pas favorisés au détriment d'autres.

Les citoyens de la Smart City sont eux-mêmes des générateurs de données. Que ce soit aux moyens de leurs interactions via les réseaux sociaux, par le biais de leurs objets connectés ou bien encore par au travers de diverses applications sur leurs smartphones (santé, mobilité, etc.). Il ne faut pas considérer les individus uniquement comme des consommateurs de services, ils peuvent également participer et apporter eux-mêmes de la valeur. Ces constatations ont permis la mise en œuvre de diverses applications basées sur le crowdsourcing, la célèbre application Waze est un bon exemple. Il permet notamment à ses utilisateurs d'indiquer les accidents et travaux sur la route en temps réel. Cela permet aux conducteurs d'éviter les embouteillages et de privilégier des routes secondaires, ce qui au final génère moins de pollution. Un autre exemple, encore plus proche de la Smart City, est celui de Cit-eazen : application française proposée par la société Engie pour permettre aux citoyens de remonter facilement des problèmes à la mairie [31]. Ces applications permettent d'impliquer l'utilisateur tout en lui rendant service au quotidien. On voit aussi que dans le cadre de certaines applications, tel que Waze, les individus peuvent complètement se passer des collectivités locales ou du gouvernement et s'organiser simplement au travers d'une application. Certaines organisations citoyennes se passent également du support de toute entreprise privée et préfèrent utiliser des outils open source pour se réunir et planifier leurs actions. On l'a notamment bien vu aux travers des contestations sociales qui ont eu lieu ces dernières années. Beaucoup ont démarré via les réseaux sociaux, l'organisation se faisant de manière totalement dématérialisée.

Aujourd'hui, les utilisateurs sont de plus en plus habitués aux services sur-mesure qui respectent leurs habitudes et leurs contraintes. Il faut donc éviter que la technologie explique aux gens comment ils devraient vivre, mais plutôt s'intéresser à leurs modes de vie pour que la technologie puisse s'y intégrer et l'améliorer. La difficulté c'est que ce service sur-mesure doit être industrialisé pour être proposé à très grande échelle, à l'ensemble des usagers de la Smart City.

L'humain est un facteur central dans la consommation d'énergie puisque c'est en fonction de son confort et ses besoins à lui que certaines actions (augmenter ou baisser le chauffage, allumer la climatisation, utiliser un lave-linge, etc.) consommatrices d'énergie vont avoir lieu. Il est capital de s'intéresser le plus possible aux habitudes de chacun pour que les recommandations faites soient les moins contraignantes possibles. Dans l'étude menée par [32], les résultats montrent que de manière générale, les occupants d'un bâtiment sont plus satisfaits de leur environnement lorsqu'ils ont la possibilité d'interagir avec lui. Créer un bâtiment entièrement autonome dans la gestion de la consommation d'énergie n'est donc pas forcément une bonne solution. D'un point de vue technologique, il est difficile de gérer totalement le chauffage, la climatisation et l'éclairage d'un bâtiment entier tout en satisfaisant l'ensemble de ses occupants. Il y aura toujours besoin d'interactions avec des êtres humains pour ouvrir des portes ou des fenêtres, gérer la luminosité d'une pièce grâce à l'éclairage électrique ou aux stores. La technologie ne peut pas contrôler ou s'ajuster parfaitement aux besoins de chaque utilisateur du bâtiment. Certains vont avoir besoin de fermer les stores à cause du soleil tandis que d'autre, au même moment et dans la même configuration, seront satisfaits sans toucher aux stores. Tout est une question de sensibilité. Or, dans cette même étude [32], il a été montré que les occupants qui avaient reçu une formation pour interagir avec leur bâtiment intelligent étaient en moyenne plus satisfaits que ceux qui n'avait eu une formation pour gérer les équipements. Cela nous apprend deux choses :

— Les êtres humains aiment interagir avec un environnement, surtout lorsqu'ils sont formés pour le faire.
— La technologie ne doit pas tout contrôler, mais plutôt de collaborer avec l'usager et de lui fournir les informations dont il a besoin pour éclairer sa prise de décision.

Ainsi, les recommandations faites grâce à la technologie doivent être expliquées en détail à l'usager [27] pour qu'il puisse les comprendre et les appliquer selon ses propres besoins du moment. En étant capable d'expliquer les recommandations faites par le système, l'usager sera plus enclin à les suivre et donc à être satisfait en termes de confort.

Il faut également considérer les différents profils d'usagers du bâtiment pour pouvoir leur recommander des actions au plus proche de leurs habitudes quotidiennes. Viser l'efficience énergétique demande de se baser sur les besoins et les demandes des utilisateurs [22]. Précisons également, que les conditions climatiques jouent également un rôle majeur dans la consommation d'énergie. Pour qu'un système de recommandation soit

performant, il faut qu'il intègre un maximum de paramètres provenant de son environnement.

2.2.3/ Les prédictions en Smart City

La Smart City est par définition un milieu hétérogène où se mêle des données issues de sources diverses (données démographiques, relevés météorologiques, données de capteurs, etc.) et des normes (lois juridiques, volonté politique, etc.) qu'il convient de faire respecter. Cela en fait un environnement complexe fortement contraint dans lequel il est parfois difficile de réaliser des prédictions. En effet, les techniques d'apprentissage automatique traditionnelles ne peuvent pas garantir le respect des contraintes durant l'apprentissage. On obtient alors des modèles capables de prédire à partir des données historiques, mais sans garantir le respect des normes, des lois et des règles métier qui régissent la Smart City.

Une autre problématique importante concerne la capacité d'une application à s'adapter au profil de chaque individu avec lequel elle interagit. En effet, un habitant de la Smart City n'aura pas besoin du même niveau d'informations qu'un technicien d'exploitation énergie, ni de celui d'un touriste, ni même de celui d'un responsable politique. De plus, il est intéressant de pouvoir qualifier la donnée en fonction de l'application que l'on veut réaliser par la suite. Une donnée issue d'un relevé de capteur n'aura pas le même intérêt en fonction de l'application finale dans laquelle elle sera utilisée.

Prenons comme exemple le cas d'un capteur de température qui se trouve à quatre mètres du sol dans une rue d'une Smart City française : le 5 décembre à deux heures du matin, ce capteur relève la température de 55℃ alors qu'une heure plutôt il avait relevé une température de -5℃ plus cohérente avec les normales saisonnières. En mobilisant nos connaissances et un peu de bon sens, on se rend vite compte que la température relevée à deux heures du matin est anormale. L'explication est simple : un incendie s'est déclaré quelques mètres en dessous du capteur, la température relevée par le capteur représente donc bien la "réalité" de ce qu'il se passe dans cette rue à cet instant-là. Si cette donnée est utilisée pour notifier les différents incidents survenus dans la ville, alors cette donnée brute est très utile, elle permet de voir qu'il s'est produit un incident majeur. En revanche, si l'application est une prédiction des îlots de chaleur, alors cette donnée est aberrante. Elle peut induire le modèle de prédiction en erreur s'il la prend en compte puisqu'elle indique une température bien trop élevée qui ne correspond pas à la température du quartier à ce moment-là.

Cet exemple met en évidence l'importance de contextualiser les données en fonction de l'application finale. Cela peut impliquer des étapes de prétraitement des données, de filtrage ou de pondération en fonction des objectifs spécifiques que l'on souhaite at-

teindre. En intégrant des connaissances propres à chaque application et en utilisant des approches adaptées, il est possible de fournir des informations ciblées et pertinentes aux utilisateurs de la Smart City, tout en évitant les erreurs ou les interprétations incohérentes.

Pour s'assurer d'avoir des prédictions qui reflètent plutôt bien la réalité de la Smart City, il est essentiel d'adopter une approche qui intègre explicitement les normes et contraintes propres à cet environnement complexe. Cette approche doit permettre de garantir que les prédictions sont conformes aux réglementations juridiques, aux politiques urbaines et aux autres normes spécifiques à la Smart City, favorisant ainsi le développement d'un environnement urbain intelligent et harmonieux. Elle doit garantir une utilisation efficace des données et une meilleure prise de décision dans divers domaines de la Smart City, en répondant aux besoins et aux objectifs spécifiques de chaque utilisateur. Pour atteindre cet objectif et fiabiliser la décision de l'action publique il est nécessaire de garantir la cohérence entre la prédiction numérique et la réalité de terrain.

2.3/ Un problème de cohérence

La cohérence désigne le fait que deux parties distinctes aient des rapports logiques entre elles. Par exemple un discours cohérent, est un discours au cours duquel les affirmations énoncées sont en adéquations logiques avec les paroles qui les ont précédées. La conclusion du discours peut être complètement fausse (si les prémisses sont fausses), néanmoins elle doit être valide d'un point de vue logique. Or pour être valide d'un point de vue logique il faut que les règles d'inférence utilisées garantissent qu'il est impossible de tirer une conclusion fausse à partir de prémisses vraies. Ainsi, il est contradictoire d'affirmer les prémisses et de nier la conclusion. Pour constituer un raisonnement valide l'inférence doit respecter les règles dictées par la logique [4].

2.3.1/ La cohérence en apprentissage automatique

Mitchell [33] définie la cohérence comme la capacité d'un modèle à produire des prédictions qui sont cohérentes avec les données d'entraînement [5] comme présenté sur la Figure 2.3. Une dizaine d'années plus tard, Yu et al. [34] étend cette définition à la capacité d'un modèle à produire des prédictions qui sont cohérentes avec les données d'entraînement ainsi qu'avec des connaissances préalables sur le domaine étudié. Le modèle est dès lors contraint d'être en cohérence à la fois avec les données d'entrée mais également avec des connaissances préalables (prior knowledge [6]) qui lui sont fournies comme

4. comme le principe du tiers exclu en logique classique
5. *training data*
6. ce terme est défini plus précisement en anglais : The prior knowledge comes from an independent source, is given by formal representations, and is explicitly integrated into the machine learning pipeline [35]

présenté sur la Figure 2.3. La cohérence n'est plus simplement une mesure de performance des prédictions par rapport à des données d'entraînement mais bien une mesure de la qualité du modèle produit par apprentissage vis-à-vis d'un ensemble de contraintes. Un des objectifs de cet ajout de contrainte est d'assurer une meilleure généralisation du modèle [36] sur des données d'entrée jusqu'alors inconnues et donc de produire des résultats plus fiables. Dans cette thèse, le terme cohérence se réfère à la définition de Yu et al., c'est-à-dire qu'un modèle pour être cohérent doit prendre en compte à la fois les données d'entraînement et la *prior knowledge*.

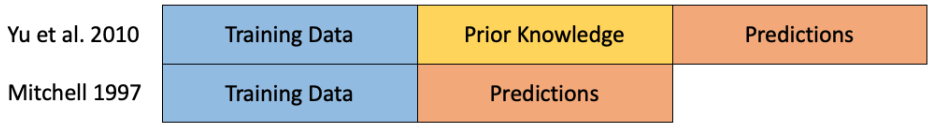

FIGURE 2.3 – Briques d'évaluation de la cohérence

À partir de la définition de la cohérence donnée par Yu et al. [34], nous proposons une définition mathématique de la cohérence, relative à la capacité d'un modèle à générer des prévisions cohérentes à la fois avec les données d'entraînement et les connaissances antérieures du domaine étudié. Dans cette définition, les données d'entraînement et les connaissances préalables sont considérées comme correctes et sans bruit.

Une hypothèse f est cohérente avec un ensemble d'exemples d'apprentissage S et un ensemble de connaissances antérieures liées au contexte K_D si et seulement si $f(x) = c(x)$ pour chaque exemple $\langle x, c(x)\rangle$ dans S et si $f(x)$ est cohérente avec n'importe quelle connaissance k dans K_D :

$$PK_{Consistency}(f, S, K_D) \equiv (\forall \langle x, c(x)\rangle \in S), f(x) \Vdash \{k \in K_D\} = c(x) \tag{2.1}$$

avec $c(x)$ la fonction cible sous-jacente. Il est important de noter que cette formule définit un idéal, les exemples d'apprentissage S et les connaissances préalables K_D sont parfaitement pris en compte par la fonction cible $c(x)$ sans aucun bruit. Cela implique aussi que l'ensemble des connaissances liées à la problématique étudiée sont bien présentes dans K_D. Toutes ces hypothèses sont bien souvent impossibles à vérifier en réalité, puisque les exemples d'entraînement et les connaissances préalables contiennent généralement des bruits ou des informations incorrectes. Il est du devoir du créateur du modèle de s'assurer que ces derniers soient les plus proches de la réalité possible.

2.3.2/ POURQUOI ÉVALUER LA COHÉRENCE ?

Un modèle d'apprentissage automatique idéal serait capable de donner des résultats conformes avec la réalité de notre monde physique. Pour ce faire, la qualité d'un mo-

dèle d'apprentissage est traditionnellement évalué avec des critères de performances [7] capables de mesurer la cohérence des résultats vis-à-vis des données d'entraînement mais pas vis-à-vis de connaissances préalables. L'intérêt du calcul de cohérence apparaît à plusieurs niveaux dans le processus d'apprentissage automatique. Ces éléments sont détaillés ci-dessous.

2.3.2.1/ Le surapprentissage

Le surapprentissage est un phénomène qui se produit lorsqu'un modèle est trop complexe par rapport aux données d'entraînement dont il dispose. Le modèle apprend les exemples d'entraînement "par cœur" sans comprendre les concepts sous-jacents qui se généraliseraient à de nouveaux exemples, ce qui peut conduire à des performances médiocres sur des données inconnues. Cela peut se produire, par exemple, lorsque le nombre de variables en entrée est trop important par rapport à la taille de l'ensemble de données d'entraînement (aussi appelée “malédiction de la dimension”). Dans certains cas, les techniques de réduction du suraprentissage comme la cross-validation et le réglage des hyperparamètres ne fonctionnent pas. En ajoutant une mesure de cohérence, le modèle peut être régularisé par rapport à des connaissances externes et donc éviter ce phénomène de surapprentissage [37].

Les connaissances préalables aident à régulariser le modèle et à éviter le surapprentissage, tout en limitant la complexité du modèle. Dans le cas où l'on dispose de peu de données d'entrée, le risque de surapprentissage est plus élevé car le modèle peut facilement sur-ajuster sur les données disponibles, en créant des hypothèses sur des relations qui n'existent pas réellement dans les données. En incorporant des connaissances qui guident le processus d'apprentissage et permettent de mieux généraliser à de nouvelles données il est possible d'éviter ce phénomène. Cela permet de pouvoir construire des meilleurs modèles même lorsqu'on a peu de données à disposition [37].

2.3.2.2/ La robustesse

Les modèles d'apprentissage automatique peuvent être sensibles aux perturbations dans les données d'entrée, ce qui peut entraîner des résultats imprévus. La robustesse des modèles en apprentissage automatique fait référence à la capacité d'un modèle à bien généraliser et à maintenir ses performances lorsqu'il est confronté à des données de test qui diffèrent de celles utilisées pour l'entraînement. En d'autres termes, un modèle est considéré comme robuste s'il est capable de faire des prédictions précises et fiables même avec des données qu'il n'a jamais vues auparavant. Cela rejoint le problème du

7. RMSE, MAE, accuracy, précision, etc.

suraprentissage vu au paragraphe précédent. Pour être robuste, un modèle doit être capable de détecter et de s'adapter aux variations et aux anomalies dans les données, sans surajuster ou sous-ajuster. Cela peut être réalisé en utilisant des techniques telles que la régularisation, la validation croisée et l'augmentation de données, qui aident à réduire les erreurs de généralisation et à améliorer les performances du modèle sur de nouvelles données.

Par exemple, le concept drift est un phénomène dans lequel les données d'entrée changent progressivement au fil du temps, ce qui peut rendre les modèles d'apprentissage automatique obsolètes ou moins performants. La robustesse d'un modèle d'apprentissage automatique se réfère à sa capacité à maintenir sa performance même lorsque les données d'entrée changent ou lorsqu'il est confronté à des scénarios imprévus. Pour faire face au concept drift, il est important de construire des modèles robustes qui peuvent s'adapter aux changements dans les données d'entrée et maintenir leur performance. Cela peut être fait en utilisant des techniques telles que le renforcement continu du modèle, l'apprentissage en ligne ou en mettant en place une surveillance continue pour détecter les changements dans les données et mettre à jour le modèle en conséquence. La mise en place de cette surveillance peut s'effectuer via la mesure de cohérence en permettant de détecter plus rapidement les erreurs et les incohérences dans les prédictions. En effet, la mesure de cohérence permet de comparer les prédictions d'un modèle avec les connaissances préalables et les sorties attendues, et ainsi d'identifier les cas où le modèle ne parvient pas à reproduire correctement les résultats attendus en théorie. En détectant ces erreurs plus rapidement, les développeurs peuvent ajuster le modèle ou les données d'entrée pour les corriger avant qu'ils ne se propagent et n'entraînent une dégradation de la performance du modèle. Cela peut contribuer à réduire la sensibilité du modèle au concept drift et à augmenter sa capacité à s'adapter aux changements dans les données d'entrée. L'ajout d'une mesure de cohérence peut accroître la robustesse du modèle face aux différentes perturbations que peuvent subir les données d'entrée.

2.3.2.3/ Les biais

Les modèles d'apprentissage automatique peuvent également être biaisés en raison de biais dans les données d'entraînement ou dans le processus d'apprentissage. Les biais en apprentissage automatique font référence à des préjugés ou des erreurs dans les données utilisées pour entraîner un algorithme d'apprentissage automatique. Ces biais peuvent provenir de plusieurs sources, notamment des biais humains dans les données d'entraînement ou des biais systémiques dans les modèles algorithmiques eux-mêmes. Par exemple, si un ensemble de données d'entraînement pour un modèle de recrutement ne contient que des candidats masculins, le modèle risque de favoriser les candidats masculins lorsqu'il sera utilisé pour évaluer des candidats réels, même s'ils sont moins

qualifiés que des candidates féminines. De même, si un modèle de reconnaissance faciale est entraîné sur des images de personnes blanches, il peut avoir des difficultés à reconnaître correctement des visages de personnes de couleur.

Les biais en apprentissage automatique peuvent avoir des conséquences graves, car ils peuvent entraîner des décisions injustes ou discriminatoires. C'est pourquoi il est important de les détecter et de les corriger dans les modèles d'apprentissage automatique pour garantir une utilisation juste et équitable de ces technologies. Pour ce faire, il existe des techniques telles que l'équilibrage des données d'entraînement, la régularisation, la sélection de caractéristiques non biaisées et la vérification de l'équité et de la non-discrimination dans les évaluations des modèles. L'ajout de prior knowledge au cours de ces différentes étapes peut être très bénéfique au modèle car la connaissance va lui ajouter des contraintes pour s'assurer qu'il reste conforme à l'éthique qu'on attend de lui.

En ajoutant une mesure de cohérence on s'assure que le modèle finalisé respecte bien les contraintes ajoutées, réduisant ainsi les biais et améliorant la justesse des résultats finaux. En effet, elle peut aider à détecter les incohérences dans les données d'entraînement, ce qui peut être un indicateur de biais. En identifiant ces incohérences, les modèles peuvent être ajustés pour tenir compte de ces différences, ce qui peut aider à réduire les biais. De plus, la mesure de cohérence peut être utilisée pour évaluer la qualité de la sortie du modèle d'apprentissage automatique. En comparant la sortie du modèle à la réalité connue, la mesure de cohérence peut identifier les situations où le modèle est biaisé.

2.3.2.4/ L'interprétabilité

Les modèles d'apprentissage automatique peuvent produire des résultats qui ne sont pas facilement compréhensibles pour les humains. L'interprétabilité des modèles en apprentissage automatique désigne justement la capacité à comprendre et à expliquer comment un modèle prend ses décisions. Cela implique de pouvoir décrire les relations entre les entrées et les sorties du modèle, ainsi que les facteurs qui ont influencé la prise de décision. L'interpretabilité est un critère important pour déterminer si les décisions prises par le modèle sont fiables et peuvent être comprises par les êtres humains. L'interprétabilité est donc essentielle pour évaluer la qualité des prédictions d'un modèle, mais aussi pour détecter les biais, les erreurs et les limites de ces prédictions. Ce qui rejoint les problématiques des paragraphes précédents. La mesure de cohérence peut aider à augmenter l'interprétabilité des algorithmes d'apprentissage automatique. En effet, si le modèle est cohérent, il sera plus facile de comprendre comment il prend ses décisions et donc de l'interpréter. En incluant des contraintes dans la fonction de perte ou en choisissant des architectures de modèles qui favorisent la cohérence, il est possible de développer des modèles qui sont intrinsèquement plus interprétables [38]. Cela peut être particulièrement

utile dans les domaines où l'interprétabilité est une exigence réglementaire ou éthique, tels que la médecine ou la finance.

2.3.3/ La représentation de la connaissance

Mobiliser des connaissances spécifiques, qu'elles soient linguistiques, physique, factuelle ou encore taxonomique, dans un système d'intelligence artificielle requiert de représenter cette connaissance dans un format adapté. Une représentation adaptée de ces connaissances permet de comparer les prédictions ou les résultats générés par un modèle avec les connaissances existantes et d'identifier d'éventuelles incohérences. Les prédictions d'un modèle de traitement du langage naturel peuvent être comparées à des règles grammaticales, à des structures syntaxiques ou sémantiques attendues pour détecter des erreurs grammaticales ou des ambiguïtés. Les résultats d'un modèle lié aux connaissances physiques peuvent être confrontés à des lois physiques, à des contraintes géométriques ou à des principes fondamentaux prédéfinis pour identifier toute violation de ces règles ou contraintes. De même, dans le cas d'un modèle basé sur des connaissances factuelles, les prédictions peuvent être évaluées en les comparant à des bases de données de faits établis, permettant ainsi la détection d'informations erronées ou contradictoires. Enfin, pour un modèle de classification ou de catégorisation fondé sur des connaissances taxonomiques, les prédictions peuvent être examinées en fonction de leur correspondance avec une structure hiérarchique de catégories prédéterminée.

La paragraphe suivant définit le terme de *connaissance préalable* utilisée dans l'évaluation de la cohérence et expose en détail les méthodes par lesquelles ces connaissances sont formalisées.

2.4/ La connaissance préalable ou *prior knowledge*

La définition de la connaissance en intelligence artificielle (IA) peut varier selon le contexte et les perspectives, mais généralement il s'agit d'un ensemble d'informations, de concepts et de règles stockées dans un système informatique, qui lui permettent de comprendre et d'interagir avec son environnement de manière intelligente. Ces informations, concepts et règles peuvent être acquises de différentes manières, comme l'apprentissage automatique ou la programmation manuelle.

La *prior knowledge*[8] est l'ensemble des connaissances provenant d'une source indépendante, représentées de manière formelle et intégrées explicitement dans le processus d'apprentissage automatique [35]. Il ne s'agit donc pas d'informations directement is-

8. aussi parfois dénommée background knowledge

sues de données brutes mais bien d'éléments existants indépendamment des données d'entraînement du modèle. La prior knowledge regroupe entre autres, différents types de connaissances comme les connaissances sur les lois physiques, les règles métier, les normes ou encore les lois juridiques d'un Etat.

Certaines connaissances, comme les normes ou les lois, sont en réalité des contraintes qui devront être impérativement respectées par les modèles d'apprentissage automatique. En effet, les véhicules autonomes sont obligés de respecter le code de la route pour avoir l'autorisation de circuler. Mais ce n'est pas le seul exemple, dans des domaines très encadrés juridiquement comme le secteur de la défense ou le secteur médical, de nombreuses contraintes doivent être respectées par les programmes informatiques au sens large. Les modèles issus de l'IA ne font pas exception et doivent également se conformer à certaines normes sous peine d'être inexploitables à cause de failles juridiques. Or, pour pouvoir respecter un ensemble de règles imposées, il faut que les modèles d'apprentissage automatique soient construits en ayant connaissance de ces lois [9] pour pouvoir les respecter. Présentes dans de nombreuses entreprises, essentielles à différents secteurs d'activité, les règles métier [10] sont des déclarations qui définissent ou contraignent certains aspects de l'activité d'une entreprise ou d'une organisation [39]. La conformité aux règles métier doit être vérifiée par les systèmes d'information car elles garantissent la sécurité, la qualité, le respect de la réglementation, la gestion des risques, le contrôle des coûts et bien d'autres aspects essentiels à la vie d'une entreprise. L'ajout de contraintes dans les algorithmes d'apprentissage automatique devient donc indispensable pour assurer la pérennité de leur exploitation au sein des entreprises et organisations. Sans contraintes, sans ajout de prior knowledge, on s'expose à des anomalies dans les modèles voir à des infractions commises par un système d'IA.

Depuis quelques années, l'inclusion de lois physique dans des algorithmes d'apprentissage automatique suscite de plus en plus d'intérêt et fait même l'objet d'un champ de recherche à part entière nommé Physics-Informed Machine Learning (PIML) [40]. Les sciences physiques trouvent des applications dans de nombreux domaines, par exemple la dynamique des fluides est utile en hydrogéologie ou en aéronautique, la physique des matériaux est utilisée en science de l'industrie, l'étude de l'optique et de l'acoustique peuvent être utile en santé, etc. Les équations de physiques rendent compte des principes généraux et des lois valables pour un ensemble d'objets. Toutefois, parfois elles ne suffisent pas à matérialiser des systèmes complexes avec exactitude. En effet, ces équations sont bien souvent déterministes, il faut donc connaître l'ensemble des paramètres initiaux pour parvenir à un résultat juste. C'est possible en théorie, c'est possible (parfois) dans le cadre d'une expérience contrôlée, mais en pratique c'est souvent complètement

9. on dit qu'ils sont informés par ces connaissances
10. aussi nommées règles d'entreprise ou règles de gestion

irréaliste. C'est pour cela que des algorithmes d'apprentissage automatique [11] capable de s'adapter à chaque objet de la réalité en s'appuyant sur le comportement passé d'autres objets similaires sont utilisés. Il est alors possible d'approximer le comportement que pourrait avoir chaque objet dans la réalité en prenant en compte des paramètres liés à l'objet lui-même sans pour autant être exhaustif dans les paramètres. Là encore, il reste quelques écueils : il faut savoir sélectionner les paramètres qui influencent le plus l'objet. Il faut créer un modèle capable de généraliser sur des données nouvelles donc inconnues. Il faut que ce modèle respecte les lois générales scientifiques qui font consensus. Ce dernier point, nécessaire à l'utilisation des modèles, est souvent négligé et peu respecté. Les algorithmes étant uniquement entraînés sur leur capacité à donner une réponse au plus proche de la réalité (d'après des données antérieures) ainsi que sur leur capacité à généraliser il arrive parfois que les modèles ne soient finalement pas cohérents avec les connaissances physiques éprouvées ce qui peut les rendre inutilisables en pratique. En s'assurant que les équations de physique sont respectées lors de l'entraînement d'un modèle d'apprentissage, les PIML font en sorte que les résultats soient plus cohérents avec les connaissances scientifiques.

Diverses formes de connaissances comme des lois ou normes juridiques, des règles métier ou encore des connaissances scientifiques doivent être traitées par les systèmes d'IA pour garantir leur cohérence. Il reste à savoir comment formaliser ces connaissances et par quels moyens les ajouter dans le processus d'apprentissage ?

2.4.1/ Formalisation de la connaissance

L'être humain s'est depuis longtemps attaché à transmettre ses connaissances à ses pairs ainsi qu'aux générations futures. La tradition orale à permis de transmettre des contes mythologiques, des lois, des religions et même des savoir-faire de manière informelle. Le compagnonnage ou le système d'apprentis permettaient à une personne d'apprendre un métier grâce à l'encadrement de "savants", la plupart du temps le savoir-faire se transmettait via des discussions et des travaux de mise en pratique. La formation des générations futures était ainsi assurée de manière informelle. Puis, avec l'invention de l'écriture, il a été rendu possible de commencer à formaliser la connaissance sous forme d'écrits afin de former plus de personnes dans un espace-temps beaucoup plus grand : le savant n'avait plus besoin d'être dans la même pièce que l'apprenti pour lui transmettre son savoir. À partir des années 1970, avec l'avènement des ordinateurs, les entreprises se sont intéressées à la manière d'intégrer l'ensemble de leurs connaissances dans les machines pour les mobiliser plus efficacement. Ce processus de gestion et d'exploitation des connaissances a forcé une réflexion sur la formalisation des connaissances puisqu'à cette époque, les machines ne pouvaient pas traiter des connaissances informelles. Au-

11. voir même de deep learning

jourd'hui, passer d'une connaissance informelle à une connaissance formelle est toujours un enjeu majeur car la connaissance formelle est plus facilement exploitable par les machines [12]. Ce qui change, c'est que désormais, les machines elles-mêmes sont capables de réaliser cette transformation si on les y entraîne.

Dans la définition de la prior knowledge, il est clairement exprimé que la connaissance doit être formalisée mais il n'est pas explicité sous quelle forme. Le champs de l'ingénierie des connaissances s'intéresse à la gestion, le stockage et l'utilisation de ces connaissances représentables identifiées comme faisant partie de l'univers du discours [42]. L'univers du discours U est un triplet qui comprend un ensemble fini d'objets noté $\mathbb{O}$, un ensemble fini d'attribut noté $\mathbb{A}$ et un ensemble fini de relations conceptuelles noté $\mathbb{R}$ [42], tel que :

$$U \triangleq (\mathbb{O}, \mathbb{A}, \mathbb{R}) \tag{2.2}$$

où l'ensemble des relations conceptuelles R est un produit cartésien qui peux prendre quatre formes différentes comme décrit ici :

$$\mathbb{R} \triangleq \mathbb{O} \times \mathbb{A} | \mathbb{A} \times \mathbb{O} | \mathbb{O} \times \mathbb{O} | \mathbb{A} \times \mathbb{A} \tag{2.3}$$

Quatre types de gestion des connaissances sont possibles en fonction des besoins en représentation des connaissances explicites et tacites comme le montre la Figure 2.4 [43]. Les connaissances tacites sont les plus difficiles à représenter, certaines d'entre elles ne font d'ailleurs pas partie de l'univers du discours, il s'agit de comportement, de perception, d'intuition propre à l'homme. L'observation d'un lion est une connaissance complexe à représenter de manière explicite, c'est pour cela qu'il a fallu attendre l'avènement des réseaux de neurones pour réaliser des traitements d'images plus efficaces : décrire formellement un lion n'explique en rien comment, en regardant une masse de pixels, nous pouvons en déduire qu'il y a un lion sur l'image. En revanche, les connaissances explicites comme le fait que le lion soit un félin, qu'il soit carnivore, qu'il vit dans la savane, qu'il a quatre pattes munies de griffes, une crinière et une queue sont beaucoup plus simple à représenter de manière formelle. Les connaissances tacites se transmettent facilement d'humain à humain, il suffit de montrer une fois un lion - sur une photo ou même sur un dessin - à un jeune enfant pour qu'il comprenne à quoi cet animal ressemble et qu'il sache le reconnaître par la suite. Cependant, cette connaissance n'est pas aussi facilement accessible à une machine qui doit être entraînée sur des milliers (ou parfois millions) d'images différentes pour apprendre à reconnaître un lion. Les connaissances

12. L'avènement des LLM pourraient nous faire penser que la connaissance informelle est parfaitement comprise par les machines mais ce n'est pas le cas à l'heure actuelle, le manque de cohérence de ces modèles en est la preuve [41].

explicites implémentées dans les machines pourront être utilisées par divers systèmes informatiques si elles ont été correctement formalisées en vue de cet usage. Dans cette optique, il faudra choisir une représentation qui permette à minima d'avoir une gestion de connaissance orientée système (system-oriented) voir une représentation dynamique plus complexe à réaliser mais souhaitable si l'on veut favoriser la collaboration homme-machine.

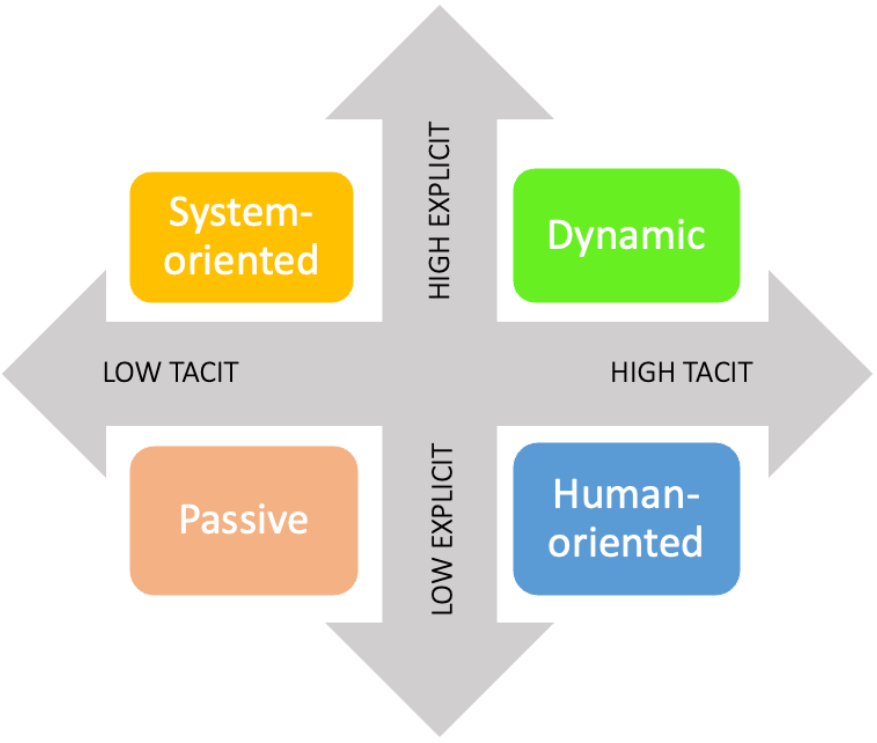

FIGURE 2.4 – Les quatre types de gestion des connaissances [43]

Pour représenter de manière la plus exhaustive possible les connaissances explicites, nous avons besoin de pouvoir définir des concepts et des termes au niveau sémantique, de décrire des individus à l'aide d'attributs associés, d'établir de relations entre ces éléments ainsi que de modéliser des règles utilisables dans une dynamique d'aide à la décision [44]. L'ensemble de ces exigences sont reflétées par le pouvoir d'expressivité offert par les différentes techniques de représentation des connaissances. Plus l'ensemble de ses éléments est respecté, plus le pouvoir d'expressivité de la base de connaissance est grand, plus il est possible pour les experts de représenter leurs connaissances de manière effective. L'expressivité est donc un critère de sélection important lorsqu'on choisit un modèle de représentation des connaissances au même titre que la compréhensibilité et l'accessibilité de la représentation [44]. La compréhensibilité exprime le fait que la connaissance soit décrite sous une forme compréhensible par un expert humain. Un langage comme Haskell [13] [45] fait partie des langages de programmation les plus expres-

13. https://www.haskell.org/

sifs, en revanche il n'est pas limpide pour un expert métier non familier avec le développement informatique. L'accessibilité du modèle s'assure que celui-ci puisse mobiliser l'ensemble des connaissances qui lui sont données à tout moment. Cette connaissance étant notamment utilisée lorsqu'on utilise un moteur d'inférence en vue de raisonner sur les connaissances. En général, plus la représentation à un grand niveau d'expressivité, plus l'inférence est complexe à réaliser ce qui ne garantit plus le critère d'accessibilité [46]. Il est souvent difficile de concilier ces trois critères de sélection que sont l'expressivité, la compréhensibilité et l'accessibilité dans un seul et même modèle de représentation des connaissances. Il est donc important de sélectionner son modèle en fonction de l'usage prévu.

Il existe quatre manières générales de représenter des connaissances : le schéma de représentation logique, le schéma de représentation par réseau [14], le schéma de représentation procédural et le schéma de représentation frame-based [47]. Les schémas de représentation logique permettent de représenter des connaissances sous forme de collection de formule logique en utilisant les notions de constante, de variable, de fonction, de prédicat, de connecteur logique et de quantificateur. La logique de premier ordre est la plus facilement représentée, mais les logiques multivariées, floues, modales, etc. peuvent aussi être modélisées grâce à ce type de schéma. Une description du monde utilisant une modélisation par objets (noeuds) et liens entre ces objets au moyen d'associations binaires (arc) peut être représentée dans un réseau sémantique. L'avantage de ce type de schéma est qu'il permet de représenter des relations de classification comme *est-membre-de* [15], *est-une-instance-de* [16], ou des relations d'agrégation comme *fait-partide* [17], ou encore des relations de généralisation comme *est-un* [18], *est-une-sous-classede* [19]. La connaissance peut en outre être présentée comme un ensemble de procédures au moyen d'un langage informatique comme le langage LISP [20] [48]. Les schémas de représentation procéduraux présentent l'avantage majeur de permettre la spécification des interactions entre les faits avec beaucoup de flexibilité (ce qui augmente l'accessibilité). Toutefois, ces modes de représentation sont plus difficiles à comprendre pour le commun des mortels [21] et bien sûr plus compliqué à modifier ce qui entraîne une maintenabilitée du système limitée. Enfin, le schéma de représentation frame-based, et plus spécifiquement le language OWL qui en fait partie, permet à la fois de représenter des objets associés à des attributs, de représenter les connexions entre ces objets (liens de subsomption ou non) et même de pouvoir y ajouter des règles logiques pour modéliser

14. aussi appelé réseau sémantique
15. *member-of*
16. *instance-of*
17. *part-of*
18. *is-a*
19. *subset-of*
20. mis au point par John McCarthy en 1960. Le LISP a été très utilisé en symbolique dans les années 1970, des machines dédiées à ce langage ont même été construites à cette époque.
21. sous-entendus les non-développeurs

les interactions plus complexes entre les entités. Il est également possible de raisonner sur les notions d'héritage ainsi que d'opérer des classifications. En résumé, si nous voulons pouvoir garantir une représentation d'une connaissance (par exemple pour exprimer la prior knowledge) qui soit compréhensible à la fois par les systèmes (haute expressivité) et par les êtres humains (compréhensibilité), il est préférable de s'orienter vers un schéma de représentation frame-based comme les ontologies avec le language OWL.

2.5/ Conclusion

L'utilisation de l'apprentissage automatique revêt une importance capitale dans l'analyse des données de la Smart City. Grâce aux algorithmes d'apprentissage automatique, nous sommes en mesure d'extraire des informations précieuses à partir de vastes ensembles de données recueillies, qu'il s'agisse des relevés de capteurs ou des données émises en temps réel par les citoyens. Ces algorithmes permettent de déceler des relations dissimulées, d'identifier des motifs significatifs et de prédire les schémas futurs. Ce processus s'apparente à un type de raisonnement inductif puisqu'il consiste à généraliser à partir des exemples observés pour établir des règles et des modèles prédictifs.

Néanmoins, il est important de souligner que les villes sont soumises à de multiples contraintes, notamment d'ordre juridique ou physique, qui doivent être respectées par l'algorithme de prédiction. Par conséquent, il devient essentiel d'évaluer le degré de conformité des prédictions aux regards des diverses règles qui s'appliquent. Pour ce faire, nous proposons d'évaluer la cohérence entre les prédictions générées et les connaissances préalables qui décrivent l'ensemble de ces contraintes de manière explicite.

Il est également primordial de formaliser ces connaissances au moyen d'un langage expressif, permettant ainsi de les représenter de manière précise et de les rendre utilisables par les machines. Pour représenter ces connaissances avec une grande précision, les rendre exploitables par les machines et plus facilement maintenables, il est recommandé de les formaliser sous la forme d'ontologies.

Les ontologies fournissent un cadre sémantique permettant de définir, de manière formelle, les connaissances et les contraintes inhérentes à un domaine spécifique. Elles permettent de décrire les concepts, les relations et les contraintes de manière structurée, facilitant ainsi la représentation et la gestion des connaissances dans un environnement complexe tel que la Smart City. L'intégration de l'ontologie au processus de prédiction permettait de mettre en œuvre un raisonnement déductif, afin d'appliquer les réglementations urbaines, les politiques environnementales ou encore des normes de sécurité définies dans l'ontologie pour orienter les prédictions.

Pour résumer, le raisonnement inductif permet d'exploiter les données en vue de déceler

des modèles, tandis que le raisonnement déductif facilite l'application des connaissances et des contraintes pour orienter les prédictions. En combinant l'apprentissage automatique (raisonnement inductif) avec les connaissances émanant de l'ontologie (raisonnement déductif), il pourrait être possible d'obtenir des résultats prédictifs plus fiables et cohérents avec les contraintes et les connaissances propres à la Smart City. Pour l'ensemble de ces raisons, la suite de notre travail porte sur l'exploration des travaux de recherche combinant des approches d'apprentissage automatique et des connaissances issues de l'ontologie. L'objectif final, dans le cadre de notre projet, est de créer des modèles qui reflètent encore mieux la réalité de la Smart City.

3

État de l'art sur la combinaison entre ontologies et apprentissage automatique

3.1 Introduction 33

3.2 Méthodologie de la SLR 34

3.2.1 Planification de la revue 34

3.2.1.1 Définition des questions de recherche 34

3.2.1.2 Sélection des moteurs de recherche scientifique et des mots-clés 35

3.2.1.3 Définition des critères d'inclusion et d'exclusion 37

3.2.1.4 Définition des critères de qualité 37

3.2.2 Création de la revue 38

3.2.2.1 Collecte des articles 38

3.2.2.2 Application des critères d'inclusion, d'exclusion et de qualité 38

3.2.2.3 Analyse 39

3.3 Analyse statistique globale des études 40

3.3.1 Techniques d'hybridation utilisées pour combiner ontologie et apprentissage automatique (RQ1) 41

3.3.2 Identification des différents algorithmes d'apprentissage automatique utilisés (RQ2) 42

3.3.3 Usage du raisonnement déductif hors liens de subsomption (RQ3) . 44

3.3.4 Grandes thématiques de l'intelligence artificielle abordées (RQ4) . . 44

3.3.5 Domaines d'application des études (RQ5) 46

3.3.6 Analyse temporelle et géographique 48
3.3.6.1 Analyse temporelle des études 48
3.3.6.2 Analyse géographique 48
3.4 Les trois catégories principales 50
3.4.1 Ontologie améliorée par l'apprentissage | Learning-Enhanced Ontology 50
3.4.1.1 Création ou modification d'ontologies par apprentissage | Ontology learning 51
3.4.1.2 Mappage d'ontologie | Ontology mapping 54
3.4.1.3 Raisonnement basé sur l'apprentissage | Learning-based reasoning 56
3.4.2 Apprentissage automatique piloté par l'ontologie | Ontology-driven machine learning 58
3.4.2.1 Apprentissage automatique informé | Informed machine learning 59
3.4.2.2 Explications de boite noire par usage d'ontologie | Ontologies explain black-box 62
3.4.3 Système d'apprentissage et de raisonnement | Learning and reasoning system 64
3.4.3.1 Système expert intégrant l'apprentissage | Expert system embedded learning 64
3.4.3.2 Application hybride | Hybrid application 66
3.5 Positionnement 68
3.5.1 Les modèles de conception pour l'IA hybride 69
3.5.2 La taxonomie du neuro-symbolique 70
3.5.3 La combinaison de l'ontologie avec l'apprentissage automatique . . 71
3.6 Conclusion 71
3.6.1 Les trois défis de l'IA hybride 73
3.6.2 Bilan 74

Ce chapitre concerne l'état de l'art. Pour réaliser ce travail nous avons utilisé une méthode appelée revue de littérature systémique qui permet de construire une analyse quantitative et qualitative des travaux de recherche étudiés. Cette méthode est reproductible. C'est-à-dire que la démarche d'étude est expliquée et elle peut être reproduite à l'identique sur le même corpus pour obtenir des résultats équivalents. Cette méthode peut ainsi être réutilisée au cours du temps pour identifier l'évolution des résultats de recherche sur le sujet. Notre état de l'art concerne les travaux de recherche qui combinent l'apprentissage automatique avec des ontologies.

3.1/ Introduction

La combinaison des raisonnements inductif et déductif est un champ d'étude très vaste en raison des multiples formes qu'ils peuvent prendre. Comme expliqué à la fin du chapitre précédent, nous nous limitons dans nos travaux à l'étude de la combinaison d'apprentissage automatique (pour le raisonnement inductif) avec les ontologies (pour le raisonnement déductif). L'objet est de connaître les différentes possibilités de combinaison qu'offrent ces deux paradigmes.

Pour ce faire, il est intéressant de faire appel à une étude quantitative et reproductive appelée revue systématique de la littérature (SLR [1]). En suivant la méthodologie spécifique aux SLR, on trouve plus de 400 papiers relatifs à la combinaison entre apprentissage automatique et ontologies. Après une sélection méticuleuse il devient possible de réaliser une cartographie précise de ce sujet [2].

La première section de ce chapitre explique en détail le protocole suivi pour réaliser cet état de l'art sous forme de revue systématique de la littérature. La deuxième section est une analyse globale des études permettant de répondre à l'ensemble des questions de recherche définies dans la partie précédente. La troisième section permet de détailler un peu plus les trois grandes catégories d'hybridation entre l'apprentissage automatique et les ontologies, notamment en mettant en avant les techniques algorithmiques utilisées. Enfin, la quatrième section positionne l'état de l'art réalisé par rapport à deux autres études majeures de l'IA Hybride.

1. en anglais *Systematic Litterature Review*
2. Cette SLR a été soumise au journal *Artificial Intelligence Review* en juillet 2022 sous l'intitulé "Combining Machine Learning and Ontology : A Systematic Literature Review", elle a été modifiée en juin 2023 suite au processus d'examen par les pairs.

3.2/ Méthodologie de la SLR

La méthodologie de cette SLR, présentée en Figure 3.1, s'inspire principalement de celle présentée par Kitchenham et al. [49]. Une fois le sujet d'étude bien identifié, la première étape consiste à planifier la revue en définissant les questions de recherche associées, en sélectionnant des moteurs de recherche d'articles scientifiques, en choisissant des mots-clés pertinents, en définissant des critères d'inclusion, d'exclusion et de qualité des articles. La deuxième étape est la réalisation de la SLR à proprement dit, avec la collecte de l'ensemble des articles scientifiques correspondants aux mots-clés établis au préalable ainsi que la sélection des articles pertinents grâce aux critères d'inclusion et d'exclusion. La qualification des articles et leur analyse font également partie de cette deuxième étape. La troisième et dernière étape consiste à créer un rapport synthétisant l'ensemble des informations contenues dans les articles sélectionnés ainsi que l'identification des défis futurs qui concerne le sujet d'étude.

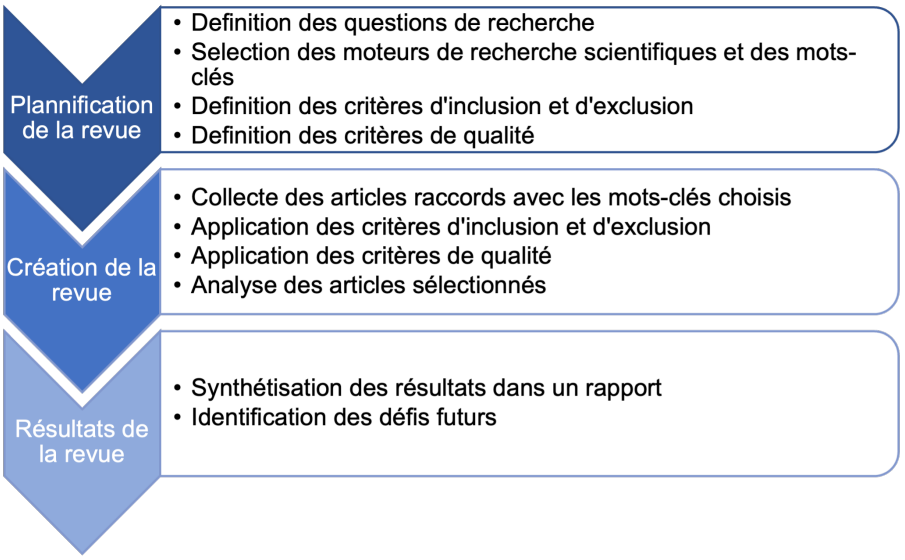

Figure 3.1 – Méthodologie de la SLR

3.2.1/ Planification de la revue

3.2.1.1/ Définition des questions de recherche

Les questions de recherche sont importantes dans une SLR car elles définissent le périmètre et l'objectif de la revue, aidant ainsi ses auteurs à concentrer leurs efforts et à obtenir des résultats pertinents. Une question de recherche est une question claire et spécifique formulée dans le contexte d'un sujet d'étude. Elle vise à guider la démarche

de recherche sur ce que la revue cherche à explorer, comprendre, analyser ou résoudre. L'objectif de cette SLR est de répondre aux cinq questions de recherche suivantes :

RQ1 : *Pourquoi et comment l'ontologie et l'apprentissage automatique sont combinés ?* Cette question à pour objectif de comprendre la motivation des travaux existants ainsi que les techniques d'hybridation utilisées

RQ2 : *Quels sont les algorithmes d'apprentissage automatique utilisés dans chaque travail ?* En effet, il existe de nombreux modèles d'apprentissage automatique, il est important de savoir lesquels sont les plus souvent utilisés dans ce sujet d'étude.

RQ3 : *Y a-t-il des travaux qui utilisent des règles autres que les liens de subsomption ?* Les ontologies sont généralement utilisées dans le processus de raisonnement pour garantir la cohérence hiérarchique d'un modèle et en déduire de nouvelles connaissances. Le plus souvent, les ontologies déduisent des liens de subsomption, mais certains travaux vont plus loin en ajoutant d'autres types de liens ou bien même des règles qui vont au-delà des simples relations hiérarchiques. Reconnaître les travaux utilisant des règles de raisonnement non hiérarchiques permet d'identifier d'autres types de cohérence appliquée à un modèle.

RQ4 : *Quels sont les différents types de thèmes principaux de l'IA présents dans les travaux combinant l'apprentissage automatique et l'ontologie ?* Plusieurs thèmes principaux, définis par le système de classification de l'ACM [3], sont récurrents dans le domaine de l'IA, comme la vision par ordinateur ou le traitement du langage naturel (NLP). Le traitement de chaque thème est particulier et fait souvent appel à des techniques différentes les unes des autres, il convient donc de s'intéresser aux particularités de traitement de ces différents thèmes.

RQ5 : *Quels sont les domaines d'application couverts ?* L'IA est utilisée dans différents domaines et joue un rôle important en aidant les humains à travailler avec de meilleures performances. Identifier le domaine d'application des travaux présents dans la revue pourra aider le lecteur qui s'intéresse à l'un d'eux en particulier.

3.2.1.2/ Sélection des moteurs de recherche scientifique et des mots-clés

Une fois les questions de recherche définies, il faut choisir les moteurs de recherche scientifique et les mots-clés qui seront utilisés pour sélectionner les études primaires. Afin de couvrir un large panel d'articles, cette SLR utilise trois moteurs de recherche scientifique différents : Web Of Science [4], ACM Digital Library [5] et Science Direct [6].

3. https://dl.acm.org/ccs
4. https://clarivate.com/webofsciencegroup/solutions/web-of-science/
5. https://dl.acm.org/
6. https://www.sciencedirect.com/

Ces trois moteurs de recherche sont recommandés par [50] car ils satisfont aux critères de reproductibilité des recherches, ainsi qu'à l'utilisation de termes booléens dans la requête. Google Scholar n'a pas été utilisé en moteur de recherche principal car il ne garantit pas assez la reproductibilité de la recherche [50] (pour que ce soit le cas il faut systématiquement passer sa navigation en privée ou vider ses cookies). Finalement, il y avait suffisant d'articles à traiter issus des trois moteurs de recherche principaux qu'une sélection secondaire n'a pas été nécessaire.

La recherche des mots-clés à été faite uniquement en anglais afin de travailler sur des études écrites exclusivement en langue anglaise. Une combinaison des deux principaux mots-clés, *ontologie* et *machine learning*[7] a été mise en place pour interroger les bases de données scientifiques sélectionnées. Cependant, le nombre d'articles obtenus était trop important et toutes les études n'étaient pas pertinentes pour la revue. La requête à ensuite été affinée en restreignant la recherche de ces mêmes mots-clés dans les partie titre, résumé et mots-clés, ce qui a permis d'obtenir un nombre plus restreint d'articles qui semblent pertinents.

Après une première analyse, il est constaté que le terme *deep learning*[8] est souvent utilisé à la place de *machine learning*, même s'il s'agit d'un sous-ensemble de cette technique. Le mot-clé *deep learning* à donc été ajouté à la requête, avec les mêmes restrictions que celles appliquées au terme *machine learning*. Par ailleurs, certains auteurs utilisent directement le terme *neural network*[9] (en particulier pour les articles les plus récents) sans mentionner explicitement les termes *machine learning* ou *deep learning*. Ce mot-clé à également été ajouté à la requête, même s'il ne concerne qu'une minorité d'articles (10 à 15 %) dans chaque requête.

Pour le mot-clé *ontology*[10] est utilisé seul, il n'est pas combiné avec d'autres mots-clés représentant différentes techniques sémantiques, telles que *taxonomy*[11], *knowledge modeling*[12], ou *knowledge graph*[13]. En effet, l'objet de cette revue étant l'utilisation des ontologies en particulier pour d'effectuer un raisonnement logique, il n'est pas nécessaire de s'intéresser à d'autres formes de représentation des connaissances. D'ailleurs une ontologie permet de pouvoir représenter à la fois des taxonomies et des graphes de connaissances.

Enfin, le mot-clé *artificial intelligence* a été ajouté à la requête car il permet de restreindre les résultats obtenus au domaine de recherche cible. Contrairement aux mots-clés précédents, la présence de ce terme est recherchée n'importe où dans l'article afin d'être le

7. en français apprentissage automatique
8. en français apprentissage profond
9. en français réseau de neurones
10. en français ontologie
11. en français taxonomie
12. en français modélisation des connaissances
13. en français graphes de connaissances

moins restrictif possible.

Sur la base des mots-clés sélectionnés et de différentes analyses des résultats obtenus la requête finale est la suivante :

“ontology" AND (“machine learning" OR
“deep learning" OR “neural network") AND “artificial intelligence"

Cette requête permet de cibler les études primaires concernées par les questions de recherche définies précédemment.

3.2.1.3/ Définition des critères d’inclusion et d’exclusion

Pour filtrer les articles issus de la recherche par mots-clés et conserver les documents pertinents qui seront utilisés pour répondre aux questions de recherche, une série de critères d’inclusion et d’exclusion a été établie.

Critères d’inclusion

InC1 : L’étude décrit une approche qui combine au moins une ontologie et au moins une technique d’apprentissage automatique.

InC2 : L’étude ne compare pas seulement les ontologies versus l’apprentissage automatique, elle les combine.

Critères d’exclusion

ExC1 : Les posters ou les démonstrations qui ne fournissent pas suffisamment de détails sur leur contribution.

ExC2 : Les articles en double renvoyés par divers moteurs de recherche.

ExC3 : Les articles qui ne sont pas rédigés en anglais.

ExC4 : Les articles non accessibles qui ne peuvent pas être récupérés en ligne.

ExC5 : Les livres (ou chapitres de livres) détaillant des articles collectés précédemment.

ExC6 : Les articles étendus des mêmes auteurs. Dans ce cas, l’article le plus récent est conservé.

ExC7 : Une revue existante ou une étude non primaire (il peut s’agir d’une étude secondaire ou tertiaire).

3.2.1.4/ Définition des critères de qualité

La qualité d’une SLR dépend de la qualité des articles examinés. Il est donc important d’évaluer rigoureusement les articles inclus dans notre SLR en tenant compte des cri-

tères de qualité suivants :

1. les études sont menées dans des instituts de recherche de haut niveau
2. les études sont publiées dans des revues et des conférences internationales de bonne qualité et sont référencées par des bibliothèques électroniques réputées
3. les motivations et les contributions sont clairement définies.

Pour évaluer la qualité de cette SLR, nous avons utilisé le Quality Assessment Instrument for Software Engineering systematic literature Reviews (QAISER) développé par [51].

3.2.2/ Création de la revue

Cette section décrit la manière dont la revue à été réalisé conformément au protocole défini en amont qui suit quatre étapes principales :

1. la collecte d'articles en fonction des mots-clés choisis,
2. appliquer les critères d'inclusion et d'exclusion,
3. appliquer les critères de qualité, et
4. l'analyse des articles sélectionnés.

3.2.2.1/ Collecte des articles

Cette première étape interroge les moteurs de recherche d'articles scientifiques sélectionnées avec l'ensemble des mots-clés définis en adaptant la requête de base à chaque base de données comme présenté dans le tableau 3.1. La première recherche a été effectuée à la fin du mois de mai 2021, et un total de 373 études a été collecté pour être analysé. Une deuxième recherche a été effectuée en février 2022 dans le but de mettre à jour le rapport d'analyse avec les nouvelles études publiées depuis la première recherche. En se limitant aux études publiées en 2021 et 2022 sur les bases de données scientifiques sélectionnées un lot important de 70 articles a pu abonder la collection initiale.

3.2.2.2/ Application des critères d'inclusion, d'exclusion et de qualité

Après avoir recueilli l'ensemble des 443 articles, les quatre premiers critères d'exclusion ont été appliqués à l'aide de Zotero [14], un outil de gestion des références, afin de supprimer les posters, les tutoriaux ainsi que les articles dupliqués et inaccessibles, comme résumé sur la Figure 3.2. Les 351 articles restants étant tous rédigés en anglais, seul les trois derniers critères d'exclusion restent à vérifier. Pour ce faire, il a fallut lire les titres

14. https://www.zotero.org/

TABLE 3.1 – Requête finale utilisée pour chaque moteur de recherche d'articles scientifiques

Base d'articles scientifiques	**Requête**	**Mai 2021**	**Fev. 2022**
ACM	[Abstract : ontology] AND [[Abstract : "machine learning"] OR [Abstract : "deep learning"] OR [Abstract : "neural network"]] AND [All : artificial intelligence]	101	9
Science Direct	artificial intelligence AND Title, abstract, keywords : "ontology" AND ("machine learning" OR "deep learning" OR "neural network")	106	23
Web of Science	TS=(ontology AND ("machine learning" OR "deep learning" OR "neural network")) AND WC="Artificial Intelligence"	166	38

et les résumés de chaque étude et éliminer celles qui ne respectaient pas ces critères. Dans le même temps, les études qui ne correspondaient pas aux critères d'inclusion, sur la base de leur titre et résumé, ont également été éliminé de la sélection. Seul 153 études ont été conservées et étudiées, 25 d'entre elles ne correspondant finalement pas aux deux critères d'inclusion, il a été décidé de les supprimer de la collection finale suite à leur lecture. En conséquence, 128 articles ont fait l'objet d'une analyse approfondie présentée dans cette SLR.

Nous avons également appliqué les critères de qualité définis précédemment pour évaluer les études primaires sélectionnées. La qualité des bases de données scientifiques utilisées a permit de garantir les trois critères établis pour l'ensemble des études sélectionnées.

3.2.2.3/ ANALYSE

Afin de pouvoir répondre à l'ensemble des questions de recherche, différents attributs (décrit dans le Tableau 3.2) ont été extraits de chacune des études primaires analysé. Une analyse statistique, présentée à la section suivante, à ainsi pu être réalisée à partir des données récoltées correspondants à ces variables.

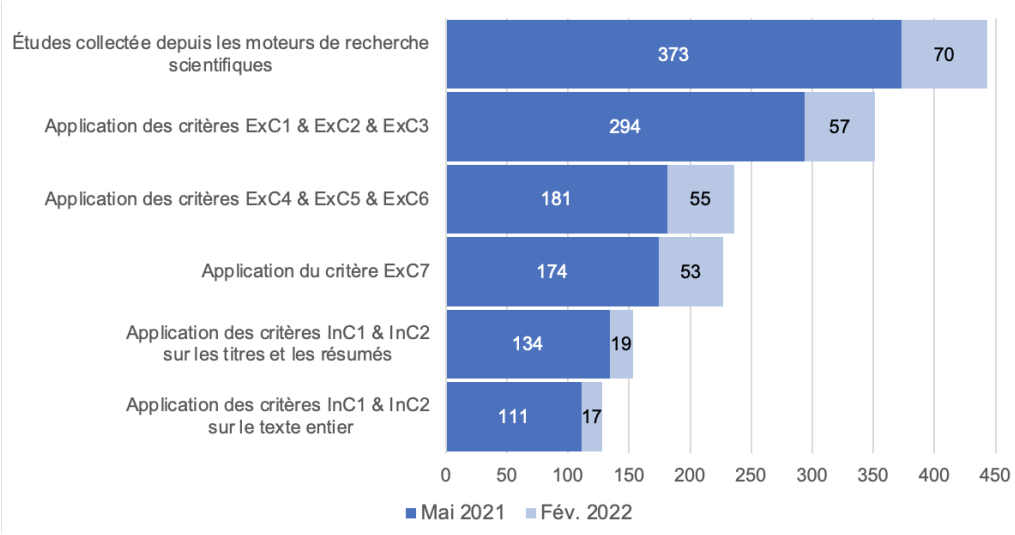

FIGURE 3.2 – Selection des articles

TABLE 3.2 – Données extraites des études sélectionnées

Attribut	**Description**
Année	Année de publication
Pays	Pays où le premier auteur est basé
Algorithme d'apprentissage automatique	Le (ou les) algorithme(s) d'apprentissage automatique utilisé(s) dans l'étude
Raisonnement via l'ontologie	Présence d'un raisonnement déductif, ou au moins de règles formelles permettant un raisonnement déductif
Grand thématique de l'IA	Thème connu de l'intelligence artificielle tel que décrit par le système de classification informatique ACM s'il en est question dans l'étude
Catégorie	Catégorie de l'étude, classé selon la classification présenté dans cette SLR

3.3/ ANALYSE STATISTIQUE GLOBALE DES ÉTUDES

Une analyse statistique globale des études combinant l'utilisation de l'apprentissage automatique et des ontologies offre un aperçu précieux des tendances et des résultats de ce domaine interdisciplinaire. Cette analyse a permit de réaliser une première classification des différents travaux tout en révélant les algorithmes prédominants, les domaines d'application courants et les avantages obtenus en fusionnant ces deux approches. Le détails des trois catégories principales est donné dans le paragraphe 3.4.

3.3.1/ Techniques d'hybridation utilisées pour combiner ontologie et apprentissage automatique (RQ1)

Après avoir lu les articles sélectionnés, trois groupes de combinaisons d'ontologie et d'apprentissage automatique se sont distingués : Learning-Enhanced Ontology, Ontology-driven machine learning, and Learning and reasoning system [15]. Ces trois groupes principaux et leurs sous-groupes (tous présents sur la Figure 3.3) ont été partiellement nommés grâce à des travaux récents qui se concentrent sur différentes formes de combinaison entre l'apprentissage inductif et le raisonnement déductif. Cela explique pourquoi nous trouvons parfois le terme sémantique au lieu de ontologie, alors que cette SLR traite seulement des articles qui utilisent une ontologie pour la partie symbolique. Le lecteur pourra ainsi faire plus facilement le lien avec d'autres articles traitant de la combinaison entre apprentissage automatique et raisonnement symbolique (comme le neuro-symbolique).

Ces trois groupes principaux et leurs sous-groupes sont détaillés dans le paragraphe 3.4 afin de pouvoir répondre plus en détail aux questions de recherche RQ1, RQ2, et RQ3.

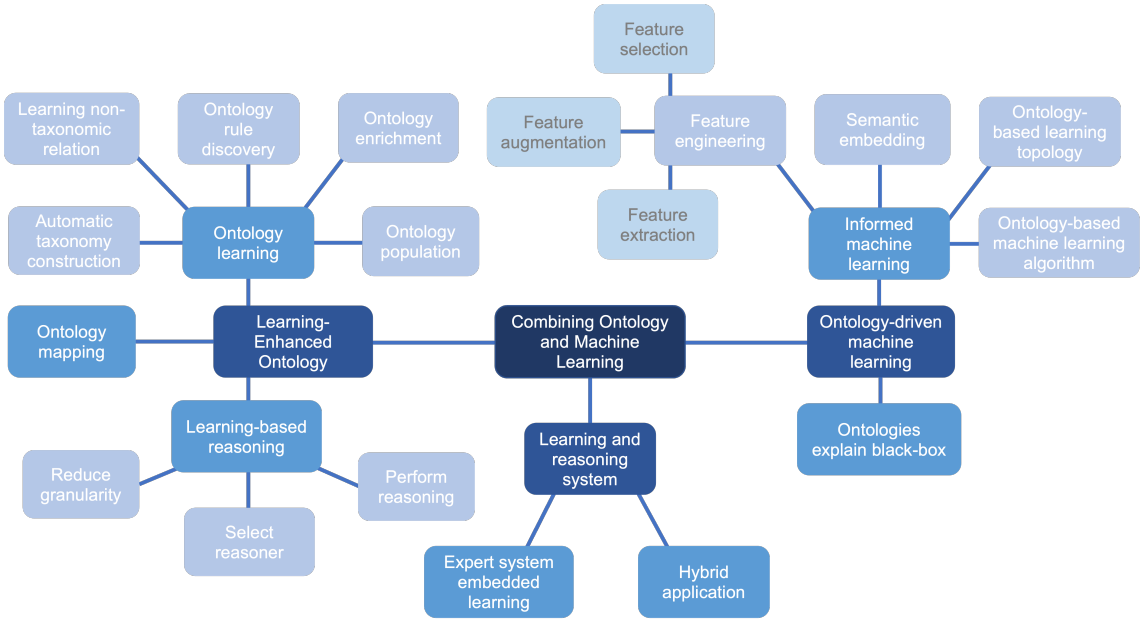

Figure 3.3 – Catégories de combinaisons d'ontologie et d'apprentissage automatique présentées dans cette SLR

15. en français respectivement : Ontologie améliorée par l'apprentissage, Apprentissage automatique piloté par l'ontologie et Système d'apprentissage et de raisonnement. Les noms anglais sont utilisés pour faciliter au lecteur leur utilisation dans les moteurs de recherche internationaux.

3.3.2/ Identification des différents algorithmes d'apprentissage automatique utilisés (RQ2)

Les choix d'algorithmes d'apprentissage automatique pour chaque cas d'application sont influencés par des facteurs tels que la nature spécifique de la tâche, la disponibilité des données et les objectifs propres à chaque cas. La sélection adéquate d'un algorithme d'apprentissage automatique est un facteur décisif pour la réussite d'un projet, car il a un impact majeur sur les résultats obtenus.

C'est pourquoi la question de recherche RQ2, qui vise à identifier les algorithmes d'apprentissage automatique utilisés dans chaque cas d'application étudié, revêt une importance cruciale. Cette investigation permet de mettre en lumière les tendances actuelle et les meilleures pratiques en matière d'utilisation d'algorithmes pour des projets spécifiques.

Pour ce faire, une analyse statistique des algorithmes les plus couramment utilisés pour l'hybridation des techniques d'apprentissage avec les ontologies a été réalisée sur l'ensemble des travaux sélectionnés dans cette SLR. La Figure 3.4 et le Tableau 3.3 présentent les résultats globaux de cette analyse toutefois il est à noter que de plus amples détails sont donnés dans la section 3.4.

Trois des quatre principales approches d'apprentissage automatiques sont retrouvées au sein des articles étudiés :

- l'apprentissage supervisé : 110 articles
- l'apprentissage non-supervisé : 37 articles dont 12 sont de l'apprentissage auto-supervisé
- l'apprentissage semi-supervisé : 1
- l'apprentissage par renforcement : aucun article de cette étude ne fait partie de cette catégorie

Les réseaux neuronaux sont très répandus dans les études sélectionnées, comme le montrent la Figure 3.8 et le Tableau 3.3. Ils sont en particuliers présents dans les catégories supervisée et auto-supervisée mais un peu moins dans la catégorie non supervisée car leur application aux problèmes de *clustering*[16] est plus récente. Dans ce dernier cas, les auteurs préfèrent souvent d'autres algorithmes de regroupement plus classiques tels que le k-means, la classification ascendante hiérarchique (CAH), l'analyse en composantes principales (ACP) ou l'allocation latente de Dirichlet (LDA).

16. en français classification non-supervisé

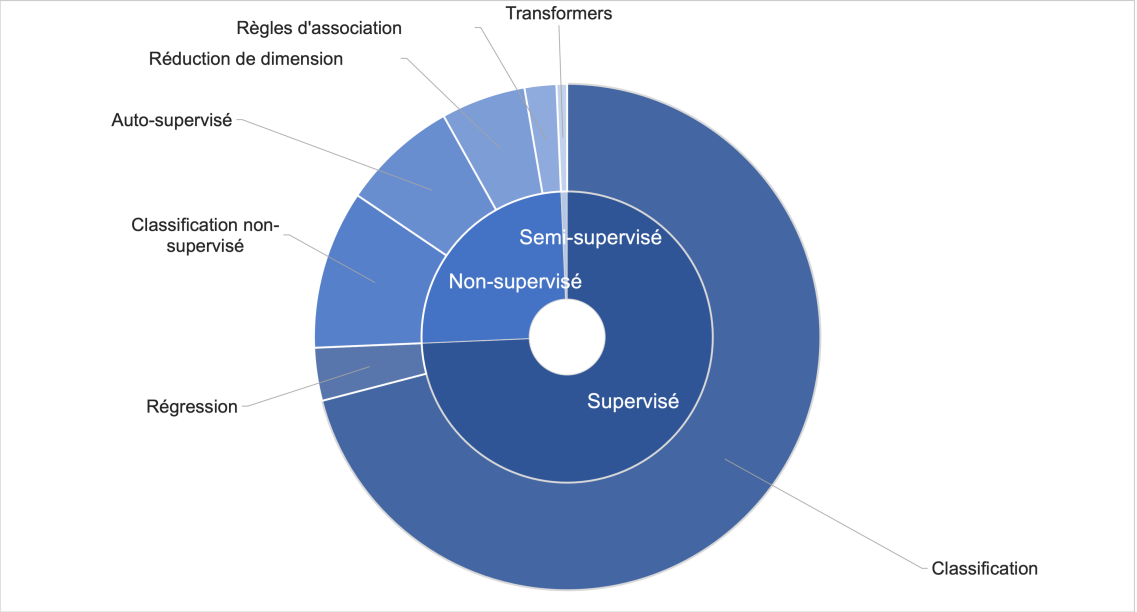

FIGURE 3.4 – Types d'algorithmes d'apprentissage automatique présents

TABLE 3.3 – Utilisation des algorithmes d'apprentissage automatique

Apprentissage supervisé	
Réseau de neurones	[52–109]
Systèmes probabilistes	[110–129]
SVM	[130–140]
Méthodes ensemblistes	[141–150]
Combinaison de plusieurs algorithmes	[151–158]
Modèle linéaire	[159, 160]
Modèle vectoriel	[161]
Shapelet	[162]
Apprentissage non-supervisé	
Clustering	[66, 90, 95, 104, 120, 126, 139, 146, 163–169]
Réduction de dimension	[117, 131, 149, 153, 159, 170, 171]
Association de règles	[106, 172]
Réseau de neurones	[173]
Systèmes probabilistes	[174]
Apprentissage auto-supervisé	
Réseau de neurones	[60, 69, 74, 81, 84, 99, 109, 166, 175–178]
Apprentissage semi-supervisé	
Transformers	[179]

3.3.3/ Usage du raisonnement déductif hors liens de subsomption (RQ3)

Il est intéressant de noter qu'une majorité d'articles emploient le raisonnement déductif seulement à des fins de raisonnement hiérarchique régit par des règles de subsomption (cf. Figure 3.5). Par raisonnement déductif, nous entendons ici le raisonnement ontologique, c'est-à-dire la déduction de nouveaux faits à partir de règles générales. Dans cette SLR, la plupart des articles n'utilisent que la contribution sémantique, en particulier les concepts et les relations hiérarchiques des ontologies, ils ne se concentrent pas sur la découverte de nouveaux faits basés sur le raisonnement ontologique. Ainsi, seuls 37% des articles examinés, répertoriés dans le Tableau 3.4, décrivent une forme de combinaison entre l'apprentissage inductif et le raisonnement déductif avec des règles non hiérarchiques. Il semble que de nombreux auteurs utilisent les ontologies comme de simples taxonomies améliorées (avec des relations non heuristiques entre les concepts) mais n'utilisent pas de règles plus complexes pour l'inférence.

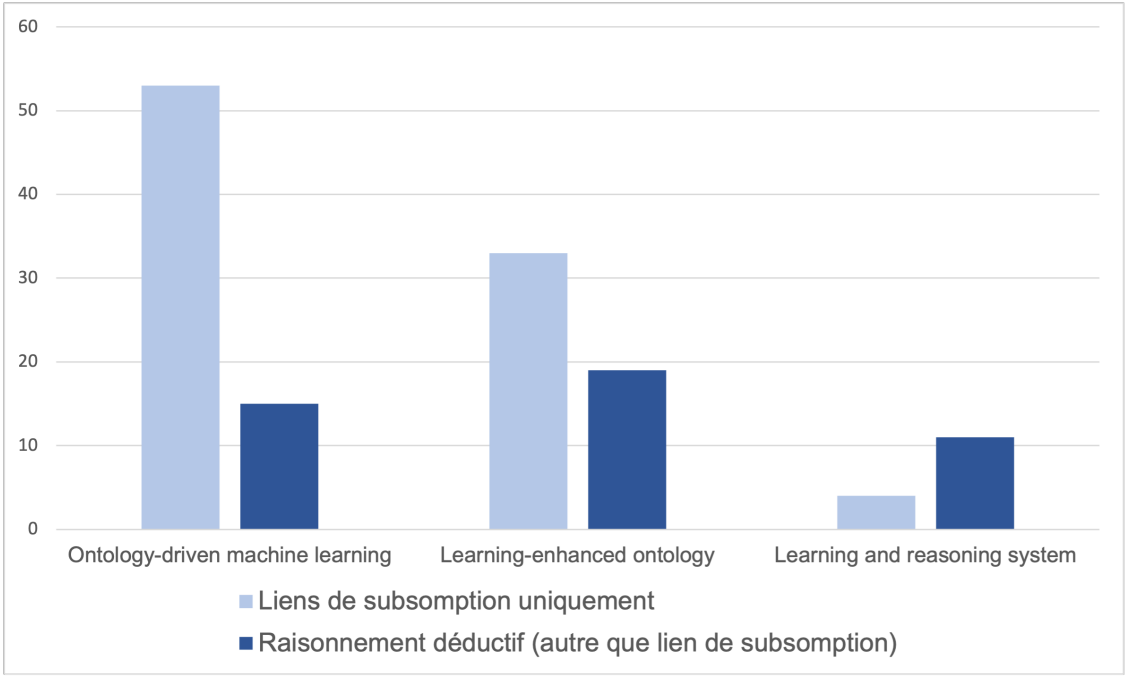

Figure 3.5 – Proportion de raisonnement déductif dans les études par catégorie

3.3.4/ Grandes thématiques de l'intelligence artificielle abordées (RQ4)

Le système de classification ACM[17] définit plusieurs thèmes principaux impliqués dans l'intelligence artificielle comme le traitement du langage naturel (NLP), la vision par ordinateur, les systèmes multi-agents, les séries temporelles et la planification.

17. https://dl.acm.org/ccs

TABLE 3.4 – Liste des articles proposant un raisonnement déductif non limité aux liens de subsomption par grande catégorie

Ontology-driven machine learning	[61, 65, 71, 72, 84, 85, 100, 105, 106, 108, 115, 124, 151, 152, 164]
Learning-enhanced ontology	[55, 61, 103, 106, 107, 113, 114, 117, 118, 121, 122, 124, 127, 131, 138, 142, 146, 156, 172]
Learning and reasoning system	[56, 57, 66, 73, 86, 92, 111, 139, 145, 159, 162]

L'analyse des articles se rapportant à ces différentes thématiques a révélé que la majorité d'entre eux se penchaient sur des problématiques liées au NLP. Cela est particulièrement vrai pour les études classées dans les catégories Ontology-driven machine learning et Learning-enhanced ontology, comme illustré dans la Figure 3.6.

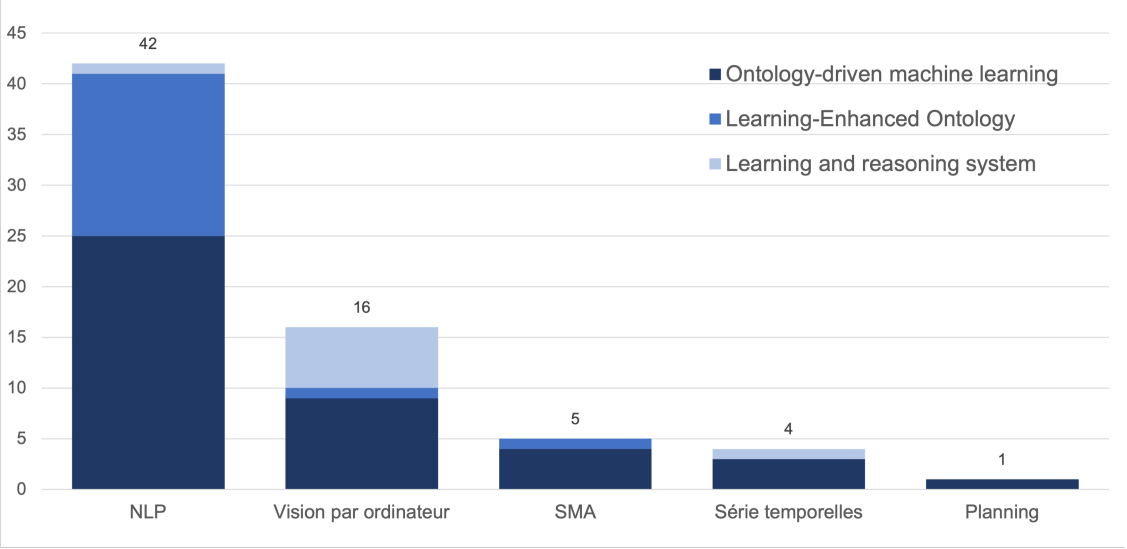

FIGURE 3.6 – Thème de l'intelligence artificiel par catégorie

Dans une moindre mesure, certains articles abordent le domaine de la vision par ordinateur, en particulier en ce qui concerne la reconnaissance d'images. Par contre, peu d'articles se consacrent aux systèmes multi-agents et aux séries temporelles. En outre, un seul article aborde la thématique de la planification. Un récapitulatif de l'ensemble des études abordant ces différentes thématiques est donné dans le Tableau 3.5.

Ces observations soulignent la prédominance des problématiques liées au traitement du langage naturel dans l'hybridation de l'apprentissage automatique avec les ontologies. Cette tendance peut être expliquée par l'utilisation fréquente des ontologies dans les travaux qui portent sur la sémantique. En effet, les ontologies fournissent une struc-

TABLE 3.5 – Liste des articles associés à leur thématique d'IA

NLP	[55,60,61,69,74–76,83,84,88,89,96,97,99,101,102,104,108–110, 112, 116, 119, 128–130, 132, 135–138, 157, 161, 164, 169–171,175,179]
Vision par ordinateur	[52,53,56,59,63,66,73,80,83,90,91,94,98,100,139,178]
Systèmes multi-agents	[65,68,71,120,146]
Séries temporelles	[78,85,151,162]
Planification	[120]

ture formelle qui permet de représenter les connaissances et les relations sémantiques entre les concepts. Dans le domaine du NLP, où la compréhension et la représentation sémantique des données textuelles ont une place prépondérante, les ontologies offrent une solution adaptée pour ajouter une couche de sens et de contexte aux informations textuelles.

Bien que moins présentes, les autres thématiques abordées montrent qu'il est possible de tirer profit des ontologies dans des applications variées allant du traitement d'images à l'exploitation de systèmes multi-agents en passant par la prédiction de séries temporelles.

3.3.5/ DOMAINES D'APPLICATION DES ÉTUDES (RQ5)

Une grande majorité des études, comme le montre la Figure 3.7, ne se concentrent pas sur un seul domaine d'application. En effet, leurs auteurs ont choisi de résoudre un problème particulier en basant leur travail sur des ensembles de données généralistes ou interchangeables afin de permettre la réutilisation de leur travail dans divers domaines d'application.

Cependant, le domaine d'application le plus rencontré est celui de la santé (24% des études), notamment parce que les ontologies médicales telles que SNOMED [18] ou les ontologies biologiques telles que GeneOnto [19] sont souvent utilisées dans le contexte médical. La santé présente également des contraintes particulièrement fortes en matière d'explicabilité des résultats et de normes à respecter, l'usage de données sémantiques et de raisonnement ontologique est donc tout à fait approprié pour ce type de problématique [180].

18. https://bioportal.bioontology.org/ontologies/SNOMEDCT
19. http://geneontology.org/docs/download-ontology/

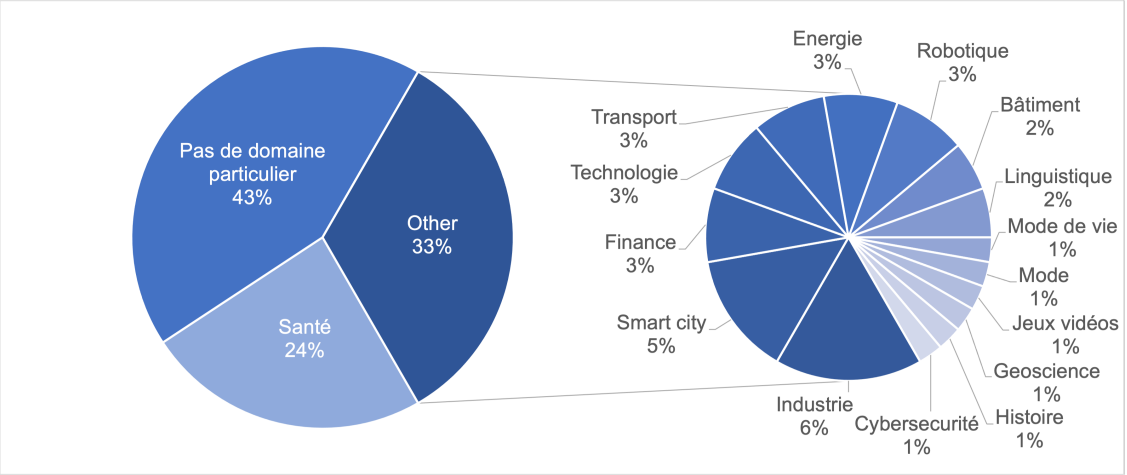

FIGURE 3.7 – Domaines d'application présents dans cette SLR

TABLE 3.6 – Liste des articles associés à leur domaine d'application

Santé	[55, 57, 58, 74–77, 81, 82, 88, 91, 93, 97, 102, 106, 114, 117, 124, 126, 132, 134, 140, 145, 147, 152, 154, 157, 160, 177]
Industrie	[78, 94, 105, 111, 122, 123]
Smart city	[56, 86, 151, 159, 162]
Finance	[60, 110, 172]
Technologie	[113, 129, 161]
Transport	[92, 151, 153]
Energie	[85, 156, 162]
Robotique	[59, 137, 173]
Bâtiment	[108, 176]
Linguistique	[99, 101]
Mode de vie	[167]
Mode	[53]
Jeu vidéo	[68]
Géoscience	[175]
Histoire	[174]
Cybersécurité	[121]

Les autres domaines d'application relevés dans cette SLR sont plus marginaux, comme le démontre la Figure 3.7, avec moins de 10 articles chacun. Cependant, le Tableau 3.6

dresse une liste exhaustive des études rattachées à chacun de ces domaines.

3.3.6/ Analyse temporelle et géographique

En plus de l'analyse principale destinée à répondre aux questions de recherche, une analyse temporelle et géographique des études a également été entreprise pour identifier les pays qui se sont montrés les plus investis dans la question de l'hybridation entre l'apprentissage automatique et les ontologies, ainsi que depuis combien de temps cette thématique suscite de l'intérêt.

3.3.6.1/ Analyse temporelle des études

La Figure 3.8 illustre l'évolution du nombre d'articles publiés traitant de ce domaine d'étude. Le premier article recensé dans cette revue systématique de la littérature remonte à l'année 2000 (cf. 3.2.1.2). Par conséquent, cette SLR offre un panorama détaillé de l'état de l'art concernant la combinaison des ontologies avec l'apprentissage automatique sur une période de plus de 20 ans.

Depuis 2010, une augmentation notable du nombre d'études portant sur l'hybridation entre l'apprentissage automatique et les ontologies a été constatée, avec une accélération particulièrement marquée ces dernières années. En effet, 57% des études analysées ont été publiées après 2018.

Au fil des années, une tendance se dessine également dans l'utilisation croissante des réseaux neuronaux. Le groupe des réseaux neuronaux englobe une gamme d'algorithmes allant du simple perceptron aux techniques de pointe telles que les *transformers*. Comme déjà présenté ci-dessus et détaillé plus amplement dans la section 3.3.2, il est intéressant de noter que les réseaux neuronaux sont présents dans la majorité des articles étudiés.

3.3.6.2/ Analyse géographique

Les diagrammes de la Figure 3.9 et de la Figure 3.10 mettent en évidence la distribution géographique des contributeurs en tenant compte de la localisation du premier auteur de chaque article. Les continents les plus fortement représentés sont l'Europe, l'Asie et l'Amérique du Nord. Parmi les pays européens, l'Italie et l'Espagne se démarquent avec environ une douzaine d'articles chacun inclus dans cette SLR. Toutefois, c'est avant tout la diversité des pays contributeurs (tels que le Royaume-Uni, la France, l'Allemagne, l'Autriche, la Pologne, la Grèce, la Bulgarie, la Roumanie, la Belgique, la Lituanie et la Serbie) qui propulse l'Europe en tête des continents les plus actifs. En Asie, c'est plus spécifiquement la Chine qui contribue de manière significative à la deuxième place de ce

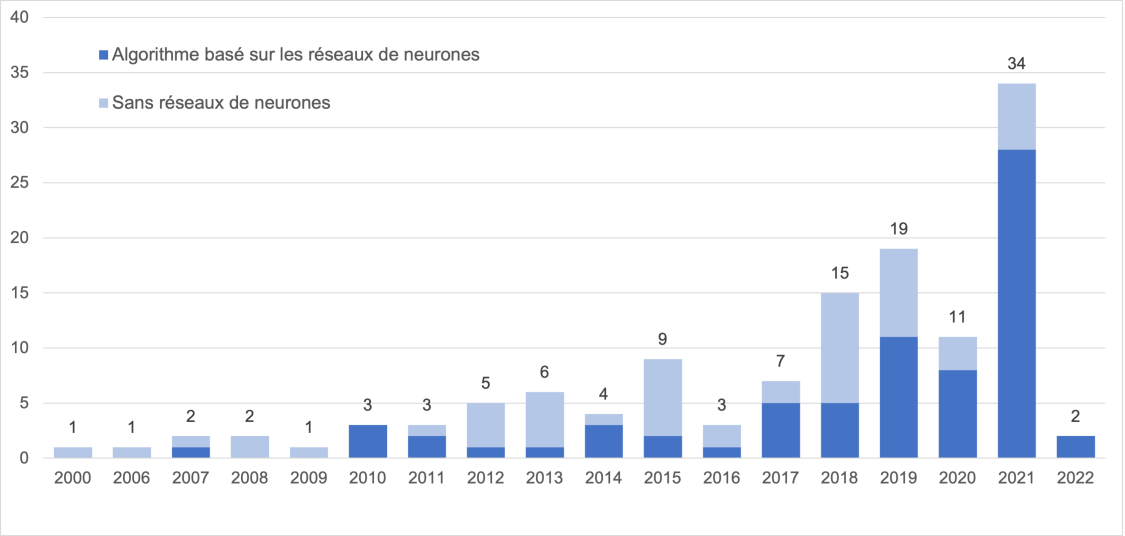

FIGURE 3.8 – Évolution du nombre de publications sur la combinaison des ontologies et de l'apprentissage automatique

classement. En Amérique du Nord, les États-Unis jouent un rôle majeur en fournissant une part substantielle des études analysées.

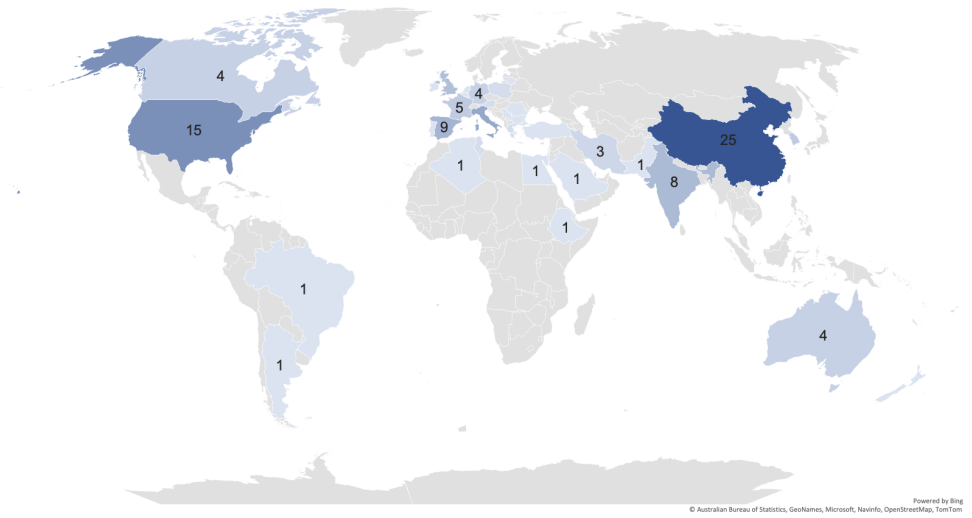

FIGURE 3.9 – Pays où est basé le premier auteur

Ces données prennent une signification accrue lorsqu'elles sont croisées avec les informations relatives aux budgets mondiaux consacrés à la recherche et au développement (R&D) [20]. La prééminence de la Chine et des États-Unis dans notre classement est cohérente avec les investissements annuels conséquents alloués à la R&D dans ces pays. En revanche, les contributions substantielles de l'Italie et de l'Espagne sont plus complexes à expliquer en termes de dépenses de R&D même s'il est notable que ces deux pays font

20. http://uis.unesco.org/apps/visualisations/research-and-development-spending/

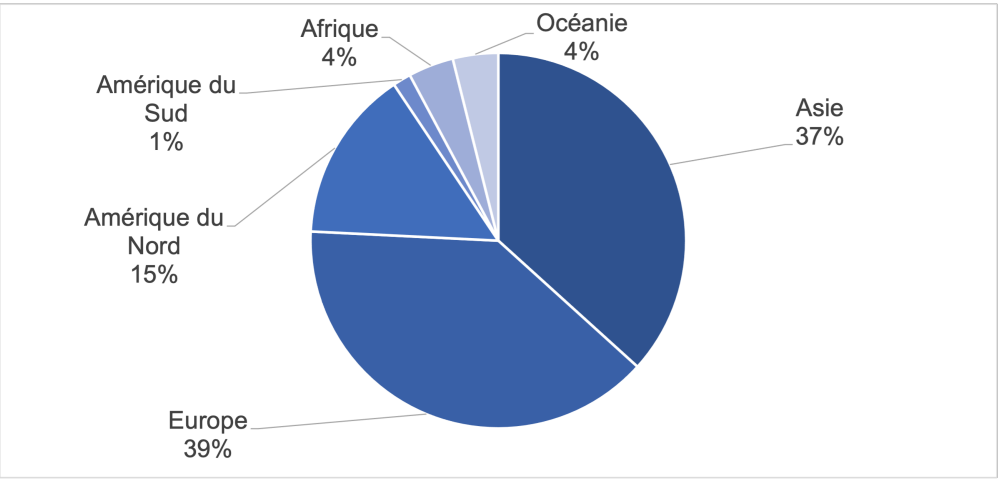

FIGURE 3.10 – Continent où est basé le premier auteur

partie des 40 premiers du classement de l'*Investment Monitor*[21]. En Italie, il est intéressant de noter que plusieurs études émanent de l'Université de Trente, qui se consacre spécifiquement à la recherche en neuro-symbolique.

Cette analyse démographique met en lumière des équipes universitaires spécifiques qui s'engagent activement dans la recherche sur l'hybridation des techniques d'apprentissage avec les ontologies. Elle illustre également comment certains pays et universités, malgré des variations dans les dépenses de R&D, se démarquent dans ce domaine d'étude par leur engagement et leur contribution significative.

3.4/ LES TROIS CATÉGORIES PRINCIPALES

3.4.1/ ONTOLOGIE AMÉLIORÉE PAR L'APPRENTISSAGE | LEARNING-ENHANCED ONTOLOGY

Il existe trois stratégies principales pour optimiser l'exploitation d'ontologies grâce à l'apprentissage automatique, comme illustré dans la Figure 3.3. En premier lieu, l'ontology learning englobe un éventail de techniques dédiées à la création et à la maintenance automatisées[22] des ontologies à l'aide d'algorithmes d'apprentissage [181]. Ensuite, l'ontology mapping, aussi appelé alignement d'ontologies, dont le but est d'assurer l'interopérabilité des ontologies en identifiant les correspondances existantes entre celles-ci. Pour finir, le raisonnement basé sur l'apprentissage[23] englobe un ensemble de techniques conçues pour simplifier le raisonnement déductif dans le contexte ontologique en recourant à l'apprentissage automatique.

21. https://www.investmentmonitor.ai/ai/ai-index-us-china-artificial-intelligence
22. ou semi-automatisé
23. en anglais *learning-based reasoning*

3.4.1.1/ Création ou modification d'ontologies par apprentissage | Ontology learning

L'*ontology learning* est le processus par lequel des ontologies sont générées ou enrichies automatiquement à partir de diverses sources de données et de connaissances. Le processus d'*ontology learning* implique la collecte et parfois l'analyse de données provenant de sources variées, telles que des textes, des bases de données, des documents Web, et même des ontologies déjà existantes. À l'aide d'algorithmes d'apprentissage automatique, les informations issues de ces données sont ensuite extraites pour être utilisées dans l'identification de concepts, de relations et de propriétés qui pourraient potentiellement être intégrés dans une ontologie. La Figure 3.11 illustre ce mécanisme en montrant que les données sont traitées par un algorithme d'apprentissage automatique, symbolisé par un entonnoir, avant d'être transformées en éléments de la TBox ou de la ABox d'une ontologie, symbolisé par le polygone annoté "A/T Box" associée à une roue cranté. Cette dernière permet de symboliser l'ontologie finale créé sur laquelle il est possible désormais possible d'inférer. L'*ontology learning* a fait l'objet d'un examen approfondi dans diverses revues de la littérature récentes, en raison de son potentiel à fournir une assistance précieuse à la création d'ontologies, tâche traditionnellement gourmande en terme de temps et de ressources [182–184].

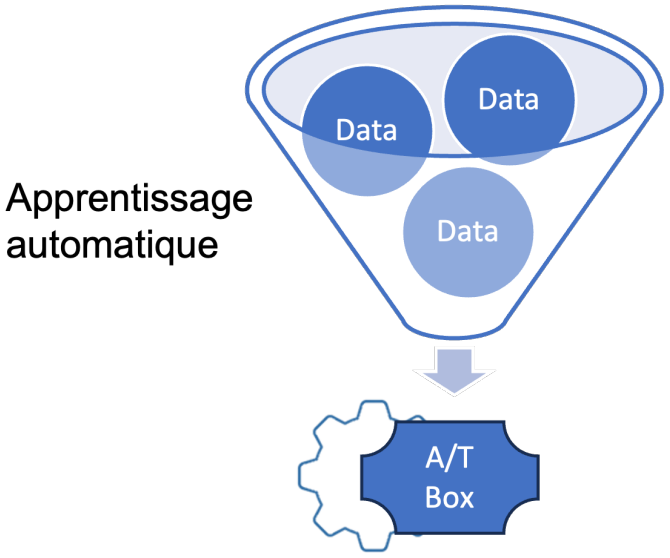

Figure 3.11 – Mécanisme de l'*Ontology learning*

L'objectif de cette revue n'est pas de fournir une exploration exhaustive du domaine, mais plutôt de présenter une vue d'ensemble des principales techniques existantes en matière d'*ontology learning*. Nous avons identifié 40 études dédiées à cette catégorie, ce qui signifie que ces articles abordent la création ou l'enrichissement automatique (ou semi-automatique) d'une ontologie à l'aide de divers algorithmes d'apprentissage automatique [185]. L'objectif est de réaliser cette création sans nécessiter l'intervention d'un expert humain. Étant donné que ce recours est souvent coûteux, parfois sujet à des

erreurs et moins rapide qu'une machine, il devient impératif de s'en passer autant que faire ce peut, en particulier dans un contexte d'application temps réel. L'*ontology learning* a pour but d'optimiser à la fois le processus de génération d'ontologie en l'accélérant, tout en renforçant sa fiabilité grâce à la réduction de l'intervention humaine propice aux erreurs [24].

Au sein des articles examinés dans cette revue, plusieurs se penchent sur l'élaboration automatisée de taxonomies [55, 62, 99, 104, 121, 128, 132, 150, 157, 165, 166, 171]. Cette démarche englobe trois domaines distincts : l'extraction de la terminologie propre au domaine, la découverte de nouveaux concepts, ainsi que l'établissement des relations hiérarchiques entre les concepts au sein de la taxonomie [186]. Afin de constituer une taxonomie complète de manière automatisée (comprendre une série de concepts organisés et liés entre eux de manière hiérarchique), ces articles ont fréquemment recours à une combinaison d'au moins deux des trois domaines évoqués précédemment. Les algorithmes utilisés dans ces travaux sont souvent liés au traitement NLP à l'instar des réseaux de neurones récurrents (RNN) qui sont idéaux pour traiter des données séquentielles comme le texte [55, 62, 157] ou bien SVM [132]. En prétraitement des données, d'autres algorithmes liés à la manipulation de données textuelles tels que le récent Word2Vec [99, 166] ou des algorithmes non supervisés permettant de faire de la modélisation thématique et/ou de réduire la dimension comme LDA [171] ou Truncated Singular Value Decomposition (TSVD) [165] (cf. Tableau 3.7).

D'autres articles vont plus loin que la construction automatique des taxonomies en réalisant l'apprentissage de relations non hiérarchiques entre les concepts [55,99,150]. L'ajout de ces relations non taxonomiques permet d'accéder au statut d'ontologie tel que définit au Chapitre 2. Nous notons que dans ces deux articles, des technologies récentes impliquant des réseaux neuronaux (tels que LSTM ou la technique de *word embedding* Word2vec comme montré dans le Tableau 3.7) ont été utilisées pour récupérer les liens non hiérarchiques entre les concepts d'une ontologie.

Certains articles explorent également la découverte de règles au sein des données brutes, qui sont ensuite incorporées dans l'ontologie [79,104,118,122,156,172,173]. Pour intégrer de nouvelles règles au sein d'une ontologie, on peut recourir à des mécanismes de règles d'association en utilisant des algorithmes tels que FCA [104] ou APRIORI [172]. Mais là encore, les réseaux de neurones sont mobilisés en majorité [79, 156, 173] (cf. Tableau 3.7).

Comme évoqué dans le Chapitre 2, le processus de population de l'ontologie consiste à compléter la TBox avec une base de faits, c'est-à-dire des instances concrètes qui représentent des individus, des objets ou des faits du domaine d'intérêt et qui constituent

24. ce qui n'empêche pas de mobiliser ensuite les experts humains pour évaluer l'ontologie ainsi conçue et rectifier à la marge ses éventuels défauts

la partie ABox de l'ontologie [63, 74, 110, 114, 119, 129, 137, 138, 166, 174]. Ce processus permet l'ajout automatisé de nouveaux individus aux concepts préexistants de l'ontologie [187]. La majorité des algorithmes utilisés pour la population d'ontologies sont basés sur des techniques probabilistes (cf. Tableau 3.7), telles que les modèles de Markov cachés (HMM) [174] ou les champs aléatoires conditionnels (CRF) [114] qui sont couramment utilisées pour extraire et classifier des entités et des relations à partir de textes non structurés. Les machines à vecteurs de support (SVM) sont également fréquemment employées, notamment pour leur capacité à gérer efficacement la classification et l'extraction d'entités dans des documents textuels [129, 137, 138]. Les SVM faisaient partie des algorithmes privilégiés en matière de traitement du langage naturel avant l'essor des réseaux de neurones, qui sont désormais utilisés de manière prédominante dans les travaux les plus récents [63, 74]. Quant aux algorithmes de réduction de dimension spécifiques au NLP, ils sont essentiels pour gérer la complexité des données textuelles. Des techniques telles que l'allocation de Dirichlet latente (LDA) et l'indexation sémantique latente (LSI) sont précieuses pour extraire des informations sémantiques pertinentes dans de grands corpus textuels [110].

Enfin, l'enrichissement d'ontologie fait référence à la procédure de mise à jour et d'amélioration d'une ontologie existante, principalement de sa TBox, en y ajoutant de nouveaux concepts, propriétés, relations ou règles [113, 124, 126, 146, 148, 158]. Cette démarche vise à étendre la représentation des connaissances d'une ontologie pour maintenir sa pertinence et son utilité d'une ontologie au fil du temps. L'enrichissement permet de mieux aligner l'ontologie avec la réalité du domaine qu'elle représente et de garantir qu'elle reste une ressource précieuse pour la prise de décision, la recherche d'informations et le raisonnement. Il est intéressant de noter que l'utilisation prédominante d'algorithmes probabilistes [113, 124, 126] et de méthodes ensemblistes [146, 148] est observée dans ce contexte.

TABLE 3.7 – Algorithmes d'apprentissage automatique utilisés pour l'*ontology learning*

Ontology Learning		
Construction automatique de taxonomies	Réseaux de neurones	RNN : [62, 157], [55] (LSTM), Word2vec : [99, 166]
	Système probabiliste	CRF : [55, 121], MLN : [128], Naïve Bayes : [157]
	SVM	[132]
	Réduction de dimension	LDA : [171], TSVD : [165]
	Clustering	HAC : [166]
	Algorithmes multiples	[150]
Apprentissage de relations non hiérarchique	Réseaux de neurones	LSTM : [55] Word2vec : [99]

Suite à la page suivante

Table 3.7 – *Suite de la page précédente*

Ontology Learning		
	Algorithmes multiples	[150]
Découvertes de règles	Réseaux de neurones	ANN : [156], BPNN : [79], CNN : [173]
	Système probabiliste	Arbre de décision : [156](M5), Réseau bayésien : [122]
	Règles d'association	FCA : [104], APRIORI : [172]
	Clustering	SOM : [173]
Population d'ontologie	Réseaux de neurones	[74], CNN : [63]
	Système probabiliste	HMM : [174], CRF : [114], Naïve Bayes : [119], Arbre de décision : [63]
	SVM	[129, 137, 138]
	Réduction de dimension	LDA : [110], LSI : [110]
Enrichissement d'ontologie	Méthodes ensemblistes	Random Forest : [146, 148]
	Système probabiliste	HMM : [126], CRF : [113], Réseau bayésien : [124]
	Algorithmes multiples	[158]
	Clustering	COCLU : [126]

Les différentes catégories de l'*ontology learning*, telles que la construction automatique des taxonomies, l'apprentissage de relations non hiérarchiques, la découverte de règles, la population de l'ontologie et l'enrichissement de l'ontologie, offrent des méthodes variées pour développer et maintenir des ontologies. Cette approche joue un rôle essentiel dans la gestion des connaissances, la recherche d'informations et la prise de décision dans des domaines complexes car elle permet de se passer d'experts humains (du moins en partie) tout en continuant d'assurer la création et la maintenance d'ontologies alignées avec la réalité.

3.4.1.2/ Mappage d'ontologie | Ontology mapping

L'ontologie mapping, également connu sous le nom d'alignement d'ontologies, a pour objectif de découvrir des correspondances entre des termes aux sens similaires dans deux ontologies distinctes, tout en veillant à maintenir la cohérence de la structure globale de l'ontologie [188]. L'objectif principal de l'*ontology mapping* est de créer une correspondance sémantique entre les éléments des ontologies afin de faciliter l'interopérabilité entre ces ontologies et étendre leur portée terminologique en alignant à la fois leurs concepts, leurs propriétés mais également leurs instances [64, 70, 75, 87, 103, 141, 149, 155, 165, 168, 176, 179]. L'alignement de concepts, consiste à trouver des équivalences entre des concepts similaires, comme "logement" et "habitation" ; l'alignement de

propriétés, met en correspondance les relations entre les concepts, telles que "a pour propriétaire" et "possédé par" ; et enfin, l'alignement d'instances, associe des individus spécifiques d'ontologies différentes représentant la même réalité, par exemple, liant "lampadaire UX345TV" dans une ontologie à "UX345TV lampadaire" dans une autre.

L'utilisation de l'apprentissage automatique dans le mapping d'ontologies permet d'automatiser le processus de découverte de correspondances sémantiques, ce qui est particulièrement utile dans des environnements où des ontologies hétérogènes doivent être intégrées.

Ce processus, illustré dans la Figure 3.12, commence par la préparation des ontologies sources, modélisé par les polygones annotés "A/T Box", notamment en sélectionnant des caractéristiques pertinentes sur la TBox et la ABox. Ensuite, un modèle d'apprentissage automatique, représenté là encore par une forme d'entonnoir, soit supervisé (avec des correspondances connues) soit non supervisé (découvrant automatiquement des correspondances potentielles) est entraîné. Ce modèle est ensuite utilisé pour prédire des correspondances entre les éléments communs aux deux ontologies, comme l'illustre le polygone "A/T Box" final situé entre les deux ontologies de départ. Ces correspondances sont bien souvent soumises à une évaluation de qualité, notamment de cohérence, puis intégrées pour améliorer l'interopérabilité entre les ontologies source. Pour finir, un post-traitement peut être appliqué pour affiner les résultats.

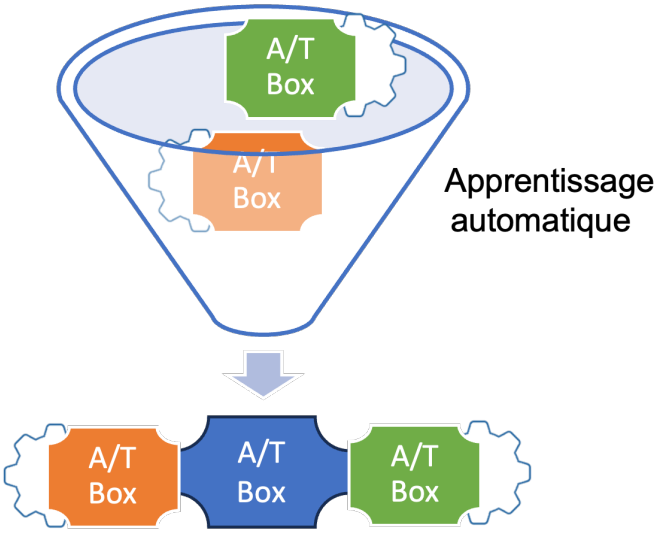

FIGURE 3.12 – Mécanisme de l'*Ontology mapping*

Les algorithmes d'apprentissage automatique utilisés pour réaliser l'alignement des ontologies sont détaillés dans le tableau 3.8. Dans cette catégorie, on observe que les réseaux neuronaux sont largement prédominants [64, 70, 75, 87, 103, 165, 168, 176, 179]. On trouve également dans cette catégorie l'utilisation de méthodes ensemblistes, notamment Random Forest (bagging) [141, 149] ou une combinaison de plusieurs classifieurs [155].

En outre, des techniques de réduction de dimension telles que l'Analyse en Composantes Principales (ACP) sont également explorées [149].

TABLE 3.8 – Algorithmes d'apprentissage automatique utilisés pour l'alignement d'ontologies

Ontology mapping	
Méthodes ensemblistes	Random Forest : [141, 149], Ensemble classifier : [155]
Réseaux de neurones	[64,168], MLP : [70,75], LSTM : [103], IAC : [87] Word2vec : [176], Fuzzy ART : [165] Transformers : [179](BERT)
Réduction de dimension	ACP : [149]

L'*ontology mapping* est un processus essentiel pour harmoniser des ontologies distinctes et favoriser l'interopérabilité des données et des connaissances. Grâce à l'utilisation de techniques d'apprentissage automatique, notamment les réseaux neuronaux et les méthodes d'ensemble, ce processus d'alignement des ontologies est facilité, rendant possible son application à de vastes ensembles de données. En fin de compte, cela contribue à favoriser l'échange efficace d'informations dans un environnement numérique en constante évolution.

3.4.1.3/ RAISONNEMENT BASÉ SUR L'APPRENTISSAGE | LEARNING-BASED REASONING

Les articles de la catégorie raisonnement basé sur l'apprentissage traitent de la prise en charge du raisonnement ontologique par des techniques d'apprentissage automatique [61, 106, 107, 117, 127, 131, 142]. Le problème majeur du raisonnement ontologique réside principalement dans sa lenteur d'exécution lors de sa mise en oeuvre dans des contextes réels. Dans notre société axée sur la rapidité, en particulier pour les systèmes temps réel, les algorithmes d'apprentissage automatique se présentent comme une alternative nettement plus rapide au moteur d'inférence classique comme Pellet [189] ou HermiT [190].

L'apprentissage automatique est exploité de différentes manières dans le raisonnement ontologique. Il peut être employé pour affiner la granularité du raisonnement, réduisant ainsi le champ d'action du raisonneur en adoptant une approche locale plus ciblée [117,131] ou pour prédire le temps de raisonnement d'un moteur d'inférence sur une ABox particulièrement volumineuse [142]. Une autre application consiste à sélectionner le moteur d'inférence le plus adapté à un cas d'application spécifique [127]. En outre, le raisonnement déductif lui-même peut être effectué par un algorithme d'apprentissage automatique qui en utilisant les ontologies fournies en entrée acquiert le savoir associé aux associations entre les concepts, aux règles implicites et aux modèles de relations entre les entités [61, 106, 107]. Cette approche peut parfois introduire un certain degré

d'incertitude ou d'imprécision, ce qui peut potentiellement compromettre la cohérence parfaite garantie par un moteur d'inférence déductif classique, car ceux-ci se conforment aux règles de la logique formelle et n'acceptent que des conclusions strictement déduites des axiomes et des faits présents dans l'ontologie. Cependant, il est important de souligner que le degré de cohérence et de précision dans le raisonnement ontologique basé sur l'apprentissage automatique peut varier en fonction des approches spécifiques utilisées et de la manière dont les modèles sont configurés et entraînés. Certaines méthodes d'apprentissage automatique peuvent être calibrées pour fournir des réponses avec un degré élevé de cohérence et de précision, tandis que d'autres peuvent être orientées vers des réponses plus probabilistes.

Ces diverses méthodes d'exploitation de l'apprentissage automatique dans le raisonnement ontologique s'appuient sur la capacité des algorithmes d'apprentissage automatique à extraire des modèles, des associations et des relations complexes à partir des ontologies en entrée. Cela permet d'enrichir l'ontologie par le biais d'un raisonnement déductif plus rapide, aboutissant ainsi à une ontologie plus complète et informée. La figure 3.13, illustre ce processus où l'apprentissage automatique (symbolisé par un entonnoir) permet d'apprendre à réaliser un raisonnement déductif sur des ontologies similaires.

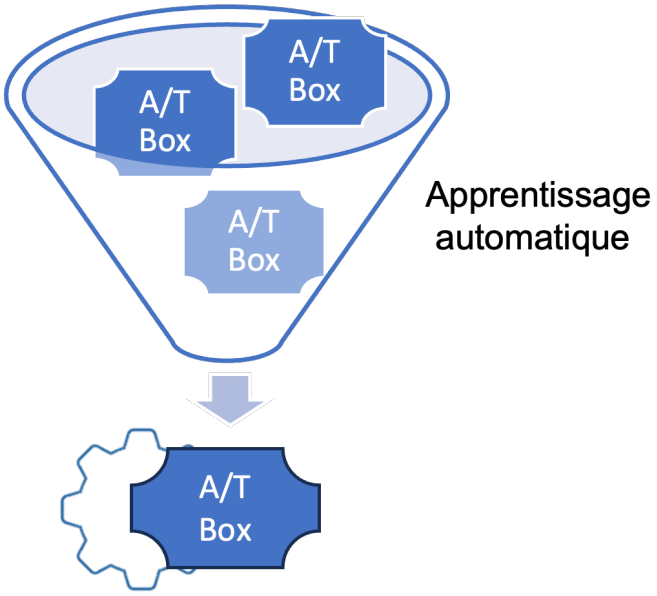

FIGURE 3.13 – Mécanisme du raisonnement déductif basé sur l'apprentissage

Les algorithmes d'apprentissage automatique utilisés pour le raisonnement déductif dans le contexte des ontologies peuvent varier en fonction des tâches spécifiques et des approches adoptées (cf. Tableau 3.9). Les réseaux de neurones, y compris les réseaux de neurones profonds, sont souvent utilisés pour capturer des relations complexes entre les concepts et les entités dans les ontologies. Ils sont donc adaptés aux tâches de clas-

sification, de prédiction et de génération de relations, toutes nécessaires pour effectuer un raisonnement ontologique via un algorithme d'apprentissage automatique [61, 107]. L'algorithme APRIORI, qui repose sur les règles d'association, est aussi employé dans ce contexte [106]. De plus, les arbres de décision et les méthodes de réduction de dimension sont utilisés pour réduire la granularité [117, 131, 142], et les arbres de décision peuvent également aider à choisir le raisonneur approprié [127].

TABLE 3.9 – Algorithmes d'apprentissage automatique utilisés pour le raisonnement déductif

Raisonnement basé sur l'apprentissage		
Réduction de la granularité	Système probabiliste SVM Réduction de dimension	Arbre de décision : [117](C5.0) [131] ACP : [117, 131], Algorithme Boruta : [142]
Selection d'un raisonneur	Système probabiliste	Arbre de décision : [127]
Effectue le raisonnement	Réseaux de neurones Règles d'association	RNN : [61, 107] APRIORI : [106]

Ces multiples approches illustrent les opportunités d'optimisation du raisonnement déductif grâce à l'application d'algorithmes d'apprentissage automatique, principalement en vue de significativement raccourcir les délais d'exécution de ce processus. Bien qu'elles puissent parfois présenter moins de cohérence par rapport à un raisonnement classique, elles se révèlent plus bénéfiques dans des scénarios réels où un important volume de données doit être soumis à ce type de d'inférence.

3.4.2/ APPRENTISSAGE AUTOMATIQUE PILOTÉ PAR L'ONTOLOGIE | ONTOLOGY-DRIVEN MACHINE LEARNING

L'approche de l'ontology-driven machine learning constitue une variante spécifique du *semantic data mining*, qui intègre des connaissances, ici sous forme d'ontologie spécifiquement, au sein d'une étape du processus de data mining, comme mentionné dans l'étude menée par Dou et al. en 2015 [191].

Les deux catégories trouvées dans cette SLR, à savoir l'apprentissage automatique informé [25] et l'utilisation des ontologies pour éclaircir la notion de "boîte noire" [26], sont présentées dans la Figure 3.3. Les détails concernant ces deux catégories sont fournis dans

25. en anglais *Informed machine learning*
26. en anglais *Ontologies explain black-box*

les paragraphes suivants.

3.4.2.1/ Apprentissage automatique informé | Informed machine learning

L'apprentissage automatique informé est un concept défini dans l'article de Von Rueden et al. [35] comme l'utilisation par un algorithme d'apprentissage d'une source d'information hybride composée de données et de connaissances préalables. Dans le cadre de cette SLR, les études classées dans cette catégorie utilisent comme source hybride des données et une (ou plusieurs) ontologie(s) -symbolisé par la polygone annoté "A/T Box"- pour produire des résultats finaux sous forme de donnée, comme illustrée dans la Figure 3.14. L'apprentissage automatique informé est particulièrement répandu dans le domaine de la physique, où il facilite l'intégration des principes physiques au sein des algorithmes d'apprentissage automatique. Cette intégration vise à améliorer les performances et à mieux s'adapter aux modèles modélisations du domaine de la physique.

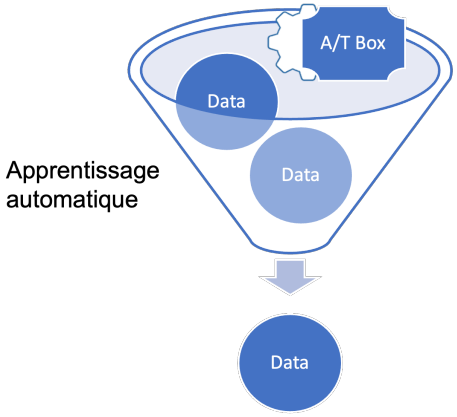

Figure 3.14 – Mécanisme de l'*Informed machine learning*

L'apport de connaissances préalables peut se produire à différentes étapes du pipeline d'apprentissage automatique, telles que lors de la préparation des données de formation (par le biais de l'ingénierie des caractéristiques ou l'intégration des connaissances), lors de la création de l'architecture du modèle (en utilisant une topologie basée sur des informations issues l'ontologie par exemple) ou pendant la phase d'apprentissage de l'algorithme. Les différents points d'insertion des connaissances sont expliqués plus en détail dans le Chapitre 4.

L'ingénierie des caractéristiques désigne le processus de sélection, d'extraction et d'augmentation de variables (caractéristiques) à partir des données brutes dans le but d'améliorer les performances des modèles d'apprentissage automatique [192]. Cette catégorie regroupe trois paradigmes différents : l'extraction, l'augmentation et la sélection de caractéristiques.

L'extraction de caractéristiques a pour objectif de réduire le nombre de variables d'entrée

d'un modèle tout en les modifiant pour en optimiser l'utilisation. Elle altère les variables initiales en se basant sur la connaissance préalable de l'ontologie afin d'obtenir des caractéristiques pertinentes [65, 83, 85, 88, 90, 91, 94–96, 105, 109, 110, 120, 125, 135, 136, 140, 154, 160, 164, 169, 170]. L'augmentation des caractéristiques implique l'ajout de nouvelles variables à l'ensemble de données, en se basant sur la connaissance préalable de l'ontologie [67, 79, 82, 112, 116, 129, 133, 134, 144, 147, 152, 163]. Cette augmentation est souvent effectuée dans le but d'apporter des informations complémentaires à l'ensemble de données initial. Au contraire, la sélection de caractéristiques à pour objectif de réduire le nombre de variables en ne conservant que celles qui sont les plus pertinentes, tout en préservant les variables initiales sans les modifier [58]. La sélection de variables simplifie le modèle, favorisant ainsi son interprétation, accélérant son apprentissage, améliorant sa capacité de généralisation, et supprimant les caractéristiques inutiles ou peu informatives. Une majorité des articles étudiés utilisent ces caractéristiques ainsi modifiées dans des réseaux de neurones, cependant on remarque que cette catégorie est une des plus variées en terme d'algorithme d'apprentissage utilisés (cf. Tableau 3.10). En effet, l'ontologie apporte de nouvelles connaissances aux données d'entrées, son usage ne dépend donc pas d'un type d'algorithme en particulier. Ce dernier est par conséquent choisi en fonction de chaque cas d'application et ne subit pas d'autre contrainte particulière.

L'intégration des connaissances [27] est similaire à la catégorie précédente, car elle touche également à la modification des données d'entrées mais d'une façon très spécifique. Cette technique est utilisée pour convertir des variables en vecteurs numériques dans un espace multidimensionnel. Cette représentation vectorielle facilite principalement l'utilisation des données dans les réseaux de neurones [60, 61, 69, 76, 81, 84, 89, 93, 101, 102, 108, 151, 153, 175, 177], mais elle s'avère également utile pour d'autres types d'algorithmes (notamment SVM et XGBoost comme on peut le voir dans le Tableau 3.10). L'article le plus ancien dans cette revue systématique qui emploie cette technique remonte à 2018, suggérant que la recherche dans ce domaine est relativement récente. Les articles sélectionnés font usage du *Knowledge Graph Embedding* (KGE) ou de l'Ontology Embedding, qui représente une évolution particulière du KGE. Le KGE est une approche qui repose sur l'utilisation de modèles spécifiques tels que TransE, TransR ou DistMult. Son objectif principal est de convertir les connaissances contenues dans un graphe de connaissances en vecteurs numériques en attribuant à chaque entité et à chaque relation du graphe des vecteurs numériques qui les représentent (c'est une approche similaire à celle utilisée par Word2Vec).

La topologie d'apprentissage basée sur l'ontologie simplifie la sélection d'une structure de modèle appropriée en tirant parti de la connaissance préalable [68]. Au lieu de recourir à des méthodes de recherche d'hyperparamètres, parfois longues et complexes, comme la recherche par grille (*GridSearch*), cette approche permet d'approximer la topologie

27. en anglais *knowledge embedding*, dérivée du terme *feature embedding*

d'un modèle (par exemple, le nombre de couches cachées et le nombre de neurones dans ces couches pour un réseau neuronal) en utilisant des informations préalables. L'exemple unique répertorié dans cette SLR s'intéresse à la structure d'un modèle de réseau de neurones.

La phase d'apprentissage basé sur l'ontologie permet d'intégrer directement les connaissances dans le processus d'entraînement du modèle, habituellement en utilisant une fonction de coût spécifique [28]. Contrairement à une fonction de perte classique qui mesure simplement la différence entre les prédictions du modèle et les valeurs réelles, une fonction de coût intégrant des connaissances prend en compte des informations externes sous forme de contraintes ou de régularisations pour guider l'apprentissage. Elle s'assure ainsi, durant la phase d'apprentissage, qu'une certaine cohérence avec les connaissances préalables est bien respectée en minimisant la fonction de coût lorsque c'est bien le cas (et en l'augmentant dans le cas contraire). Cette méthode est principalement basée sur les réseaux neuronaux [53, 71, 78, 100, 106, 178] mais d'autres approches exploitant des techniques probabilistes, telles que les réseaux bayésiens, les modèles de Markov ou les arbres de décision [52, 59, 115, 124, 143] sont également capables de mettre à profit une fonction de perte incluant des connaissances.

TABLE 3.10 – Algorithmes d'apprentissage automatique utilisés pour l'*informed machine learning*

Informed Machine Learning		
Extraction de caractéristiques	Réseaux de neurones	[90,94,96], ANN : [65], CNN : [91,109], RNN : [88, 105](LSTM), [85](GRU), Word2Vec : [109] VGG16 : [83]
	SVM	[135, 136, 140]
	Système probabiliste	Bayesian network : [125], Naïve Bayes : [110], Decision tree : [120](C4.5), EM : [120]
	Méthodes ensemblistes	Random forest : [170]
	Modèle linéaire	Regression logistic : [160]
	Algorithmes multiples	[95, 154]
	Clustering	[120], k-means : [169], HAC : [164]
Augmentation de caractéristiques	Réseaux de neurones	[67], ANN : [79, 82], MLP : [152], RNN, CNN et HAN : [97]
	Système probabiliste	HMM : [112], Bayesian network : [116], Decision tree : [129, 147](C4.5)
	Méthodes ensemblistes	Random forest : [144]
	SVM	[133, 134, 152]
	Clustering	k-means : [163], [167](Farthest First)

Suite à la page suivante

28. en anglais *loss function*

Table 3.10 – *Suite de la page précédente*

Informed Machine Learning		
	Règles d'association	FCA : [147]
Sélection de caractéristiques	Réseaux de neurones	[58]
	SVM	[58]
Intégration des connaissances	Réseaux de neurones	[151, 153], MLP : [76], CNN : [102], RNN : [61, 108], [93](GRU), [60, 69, 81, 89, 101](LSTM), Autoencoders : [177], Word2Vec : [60, 69, 81, 84]
	SVM	[130]
	Méthodes ensemblistes	XGBoost : [150]
	Algorithmes multiples	[151]
	Réduction de dimension	LDA : [81, 153]
Architecture basée sur l'ontologie	Réseaux de neurones	[68]
Apprentissage basé sur l'ontologie	Réseaux de neurones	[71, 106], LSTM : [78], CNN : [53], CAE : [178], [100](LTN)
	Système probabiliste	Bayesian networks : [52], [124](CBN), Markov : [59] (CRF), Decision tree : [115](C4.5)
	Méthodes ensemblistes	Random Forest : [143]

L'apprentissage automatique informé vise à optimiser l'utilisation des connaissances préalables tout en capitalisant sur les performances exceptionnelles des algorithmes d'apprentissage automatique. Cette synergie entre la puissance de calcul des algorithmes d'apprentissage automatique et la sagesse des connaissances humaines offre un potentiel considérable pour résoudre des problèmes complexes et favoriser l'innovation dans divers domaines, de la médecine à la finance en passant par l'industrie.

3.4.2.2/ Explications de boite noire par usage d'ontologie | Ontologies explain black-box

Dans le domaine de la recherche en intelligence artificielle, les réseaux neuronaux sont souvent perçus comme des "boîtes noires" car bien que leur comportement en entrée et en sortie soit observable, le mécanisme de raisonnement sous-jacent reste obscur. Par conséquent, le développement de l'intelligence artificielle explicative (XAI) est devenu essentiel. L'explicabilité d'un modèle algorithmique se réfère à sa capacité à présenter une séquence cohérente d'étapes interconnectées pouvant être interprétées par les humains comme des causes ou des raisons derrière le processus de prise de décision [193].

Cette capacité permet de clarifier l'algorithme ou ses résultats, ce qui facilite la compréhension des raisons pour lesquelles certaines décisions sont prises. Outre la question de la confiance, le manque d'explicabilité des modèles d'IA a également soulevé des défis juridiques dans divers domaines tels que la défense militaire, les soins de santé, l'assurance et les véhicules autonomes. L'incapacité à fournir des explications claires et compréhensibles sur les décisions prises par l'IA pose des problèmes juridiques dans ces domaines.

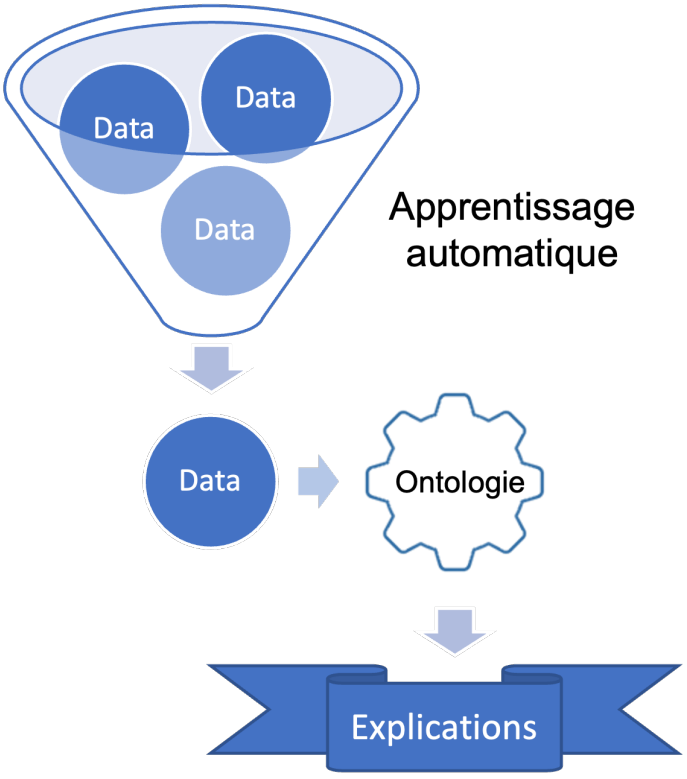

FIGURE 3.15 – Mécanisme de l'*Ontology explain black-box*

Lorsque les connaissances préalables ne sont pas intégrées, les explications se basent principalement sur des corrélations mathématiques entre les données et les résultats, ce qui ne garantit pas toujours la solidité et la fiabilité des indicateurs. L'explication de boîte noire par une ontologie repose sur l'idée d'utiliser une structure de connaissances formelle et explicite pour éclaircir *a posteriori* le fonctionnement interne d'un modèle d'intelligence artificielle (IA) considéré comme une "boîte noire". C'est-à-dire que lorsqu'une prédiction est générée par un modèle d'apprentissage automatique classique, comme représenté par un entonnoir sur la Figure 3.15, l'ontologie est utilisée pour fournir des explications sur cette prédiction. Pour une explication globale du modèle, l'ontologie peut être utilisée pour montrer comment les concepts et les entités de l'ontologie sont liés aux caractéristiques ou aux données d'entrée du modèle, décrivant ainsi le processus de raisonnement global du modèle. Pour une explication locale, l'ontologie peut être utilisée pour mettre en évidence comment des concepts ou des entités spécifiques de l'onto-

logie ont contribué à la prédiction particulière pour un individu donné. Dans cette SLR deux études utilisent des ontologies pour améliorer l'explicabilité des modèles : une de ces études visait à fournir une explication globale du modèle [72], tandis que l'autre se concentrait sur la fourniture d'explications locales [77].

Les réseaux de neurones étant un type d'algorithmes particulièrement sujet à ce phénomène de "boite noire", il est normal de les trouver comme sujet d'étude principal des deux articles susmentionnées, comme montré dans le Tableau 3.11.

TABLE 3.11 – Algorithmes d'apprentissage automatique utilisés pour l'*ontology explain black-box*

Explication des boites noires	
Réseaux de neurones	[72], GRU : [77]

Les explications des "boîtes noires" basées sur une ontologie offrent un cadre logique et cohérent pour les explications, alignant ainsi les décisions du modèle sur les connaissances du domaine. Cette approche facilite la transparence des modèles d'apprentissage automatique et contribue à l'établissement de la confiance des utilisateurs dans ces systèmes, ouvrant ainsi la voie à des applications peut-être plus responsables et plus éthiques.

3.4.3/ Système d'apprentissage et de raisonnement | Learning and reasoning system

Cette catégorie englobe toutes les applications intégrales qui recourent à l'apprentissage automatique et aux ontologies pour leur fonctionnement. Une application se réfère à un logiciel capable d'exécuter une ou plusieurs tâches spécifiques au sein du même domaine, tel qu'un système d'aide à la décision pour la gestion des maladies cardiaques [57]. Les études incluses dans cette catégorie et ses sous-catégories, montrées dans la Figure 3.3, portent sur des systèmes d'application complets, et pas seulement des mécanismes spécifiques comme l'apprentissage d'ontologies ou l'ingénierie des caractéristiques, par exemple.

3.4.3.1/ Système expert intégrant l'apprentissage | Expert system embedded learning

Un système expert est une application informatique capable de reproduire le raisonnement déductif humain et prendre des décisions ou résoudre des problèmes dans un domaine spécifique. Il se compose de divers éléments, notamment une base de connaissances qui contient des informations spécifiques au domaine, des règles de décision et des heuristiques, un moteur d'inférence qui effectue des déductions logiques en utilisant

les informations disponibles et bien souvent une interface utilisateur [194].

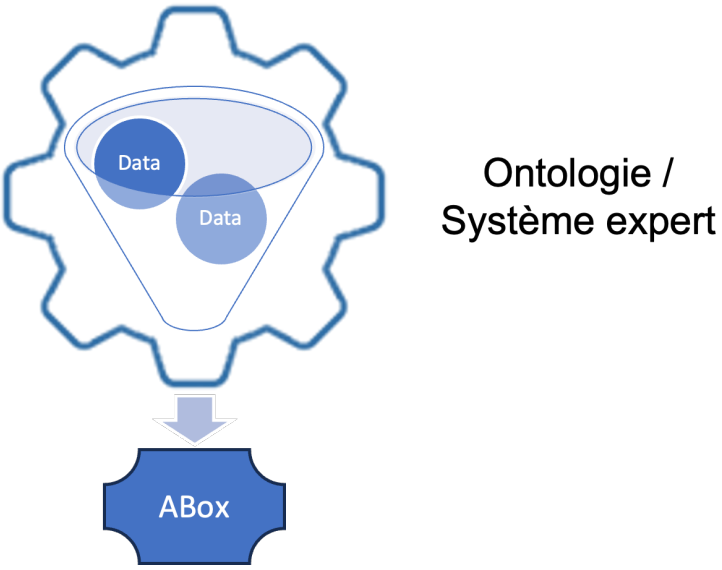

FIGURE 3.16 – Mécanisme d'un système expert intégrant de l'apprentissage automatique

Étant fondés sur un raisonnement exclusivement déductif, les systèmes experts ne sont pas appropriés pour exploiter efficacement certaines sources de données, comme des images, ni pour accomplir des tâches de classification sophistiquée lorsque l'ensemble complet des règles n'est pas explicitement formulé. Lorsque de telles capacités sont requises, le système expert peut faire appel à l'apprentissage automatique, intégrant ainsi les résultats obtenus. Cette approche peut également être pertinente pour effectuer des tâches telles que l'imputation des valeurs manquantes, essentielles au raisonnement déductif, dans le cas où le système expert manquerait de données pour effectuer un raisonnement précis [145, 159].

Dans le cadre de l'apprentissage intégré au sein du système expert, la composante d'apprentissage automatique, représenté sur la Figure 3.16 par un entonnoir, est intégrée au sein d'un système expert, ici représenté par une roue crantée symbolisant une ontologie. Le ou les modèles d'apprentissage intégrés peuvent être considérés comme des sous-modules du système expert. Comme on le voit également sur la Figure 3.16, les résultats produits sont principalement de nouveaux faits inférés par le système expert représenté par un polygone annoté "ABox".

Le système expert qui combine une ontologie avec un moteur d'inférence intègre un ou plusieurs modèles d'apprentissage qui utilisent un raisonnement inductif pour mieux percevoir certains éléments, par exemple en classifiant du texte ou des images. Tâches pour lesquelles l'usage d'un raisonnement purement déductif est sous-optimal. Cette approche vise à enrichir la base de connaissances du système expert, en particulier sa ABox, pour améliorer ses propres capacités de raisonnement déductif.

Dans cette SLR, les études utilisent l'apprentissage pour combler les valeurs manquantes dans leur système expert, pour ce faire elles font appel à des algorithmes de régression linéaire multiple et l'algorithme de boosting Adaboost comme présentés dans le Tableau 3.12.

TABLE 3.12 – Algorithmes d'apprentissage automatique utilisés pour les systèmes experts intégrant l'apprentissage

Système expert intégrant l'apprentissage	
Modèle linéaire	Régression linéaire : [159]
Méthodes ensemblistes	Adaboost : [145]

3.4.3.2/ APPLICATION HYBRIDE | HYBRID APPLICATION

Une application hybride est un système qui exploite de manière élaborée à la fois le raisonnement inductif et déductif dans des applications qui ont pour objectif de se rapprocher au maximum de la réalité. Ces systèmes d'intelligence artificielle hybrides sont constitués de plusieurs modules interconnectés qui collaborent entre eux [54, 56, 57, 66, 73, 80, 86, 92, 98, 111, 139, 161, 162]. La Figure 3.17, illustre bien ce mécanise d'hybridation en symbolisant la fusion entre l'apprentissage automatique (entonnoir) et l'ontologie (roue crantée) qui permet de prédire des résultats. Grâce à l'intégration de multiples modules, ces systèmes peuvent capitaliser sur les avantages des approches basées sur l'apprentissage, telles que l'apprentissage automatique, ainsi que sur les techniques de raisonnement symbolique, telles que les ontologies.

Les études présentes dans cette catégorie mobilisent plusieurs techniques présentées précédemment dans cette SLR. En combinant plusieurs approches, des applications complexes comme un système de contrôle des feux de circulation intelligent [86], la reconnaissance d'événements à partir de remontée de capteurs [162], l'aide au tri de déchets électroniques [73], la prédiction de la trajectoire d'un avion en se basant sur les instructions de contrôle du trafic aérien [54] ou encore détection automatique d'événements suspects via une caméra de surveillance [56] peuvent être mises en place.

La plupart des systèmes d'application hybrides privilégient l'utilisation de réseaux neuronaux ou de méthodes probabilistes, comme illustré dans le Tableau 3.13. L'identification d'événements à partir des données fournies par des capteurs utilise l'algorithme de classification Shapelet conçu pour reconnaître des motifs discriminants dans des séries temporelles. Cette approche est particulièrement pertinente dans le contexte de la Smart City.

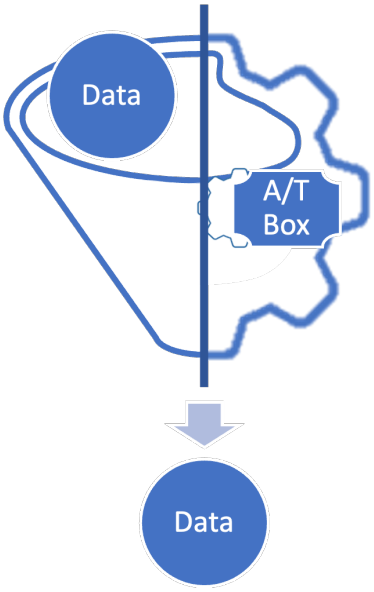

FIGURE 3.17 – Mécanisme de l'application hybride

TABLE 3.13 – Algorithmes d'apprentissage automatique utilisés pour les applications hybrides

Application hybride	
Neural networks	[57], R-CNN : [66, 73], CNN : [98] ANN : [80], RBFNN : [86], LSTM : [54], GRU : [92], YOLO : [56]
Probabilistic system	Markov : [111](HMM), LogitBoost : [57]
SVM	[139]
Shapelet	[162]
Vectorial model	Rocchio algorithm : [161]
Clustering	k-means : [139]

En tirant parti de la puissance de l'apprentissage automatique, et en exploitant la richesse des connaissances offerte par la structure formelle des ontologies, les applications hybrides continuent de repousser les limites actuelles de l'IA. Cette hybridation crée des perspectives d'innovation dans des secteurs tels que la Smart City ou la santé car elle permet de mieux prendre en compte les contraintes du monde réel.

3.5/ Positionnement de la revue dans le domaine de l'intelligence artificielle hybride

Il est essentiel de reconnaître que la combinaison entre ontologie et apprentissage automatique s'inscrit dans un paradigme plus large appelé hybridation de l'intelligence artificielle, qui vise à combiner différents types de raisonnement. Van Bekkum et al. [195] décrivent plusieurs modèles de conception pour l'IA hybride avec sept modèles élémentaires qui caractérisent les types de données d'entrée et de sortie, ainsi que les mécanismes employés pour le traitement des données (prédiction, déduction, formation, etc.). Ces modèles élémentaires sont ensuite combinés pour former des modèles de conception plus complexes qui délimitent divers scénarios d'hybridation. Afin de faciliter l'utilisation de cette classification, nous avons associé chaque catégorie décrite dans la section précédente au modèle de conception correspondant. Les résultats de cette mise en correspondance sont présentés dans le Tableau 3.14.

Table 3.14 – Alignement de nos catégories d'hybridation avec les modèles de conception de Van Bekkum et al. [195] et la taxonomie de Kautz [196]

Catégorie	**Modèles de conception [195]**	**Taxonomie [196]**
Learning-enhanced ontology		
Ontology Learning	Design Pattern 4	Neuro|Symbolic
Ontology Mapping	Design Pattern 4	Neuro|Symbolic
Raisonnement basé sur l'apprentissage	Design Pattern 10	Neuro :Symbolic→Neuro
Semantic data mining		
Informed machine learning	Design Pattern 7	Symbolic Neuro symbolic Neuro_{Symbolic}
Ontology explain black-box	Design Pattern 5	Neuro|Symbolic
Learning and Reasoning system		
Expert System Embedded Learning	Design Pattern 12	Symbolic[Neuro]
Hybrid application	Design Pattern 12	Neuro|Symbolic

3.5.1/ Les modèles de conception pour l'IA hybride

Les modèles de conceptions créés par [195] ont pour objectif d'être les plus généraux possibles en matière d'IA hybride. Ils décrivent donc l'ensemble des systèmes utilisant une combinaison de techniques connexionistes avec des techniques symboliques.

Le modèle de conception numéro 4 apparaît comme le plus approprié pour décrire *Ontology Learning* et *Ontology Mapping*. Il est spécifiquement utilisé pour l'apprentissage avec une sortie symbolique. Ce modèle de conception présente un bloc primaire qui apprend à partir de données textuelles et un bloc secondaire capable de déduire des informations à partir d'un nouveau modèle sémantique. La partie symbolique est dès lors construite ou enrichie grâce à l'apprentissage automatique.

Le modèle de conception numéro 5 est consacré au phénomène largement reconnu de "boîte noire" inhérent à certains algorithmes d'apprentissage automatique, en particulier les réseaux neuronaux profonds. Dans ce modèle de conception, un modèle symbolique est utilisé après l'apprentissage d'un modèle d'apprentissage pour interpréter les résultats obtenus en tirant parti des connaissances antérieures. Cela correspond étroitement à notre catégorie *Ontology explain black-box* où l'accent est mis sur l'utilisation de ressources ontologiques pour fournir des explications compréhensibles aux sorties d'un modèle d'apprentissage.

Le modèle de conception numéro 7 traite spécifiquement de l'apprentissage éclairé par des connaissances préalables, s'alignant parfaitement sur notre catégorie *Informed Machine Learning*. Le principe fondamental qui sous-tend ce modèle de conception est l'inclusion des connaissances antérieures dans le pipeline du modèle d'apprentissage automatique. En incorporant des connaissances pertinentes, l'objectif est d'améliorer à la fois les performances mais également les capacités de généralisation du modèle. Comme nous l'avons observé précédemment, l'intégration des connaissances peut se produire à différents stades du processus d'apprentissage. Il s'agit notamment de l'intégration des connaissances dans les données d'apprentissage, de l'intégration dans l'architecture du modèle, de l'intégration pendant le processus d'apprentissage du modèle et même de l'intégration post-hoc après la phase d'apprentissage [29].

Le modèle de conception numéro 10 est consacré à l'exploitation de la puissance de l'apprentissage automatique, en particulier des réseaux de neurones, pour permettre le raisonnement logique. Dans ce modèle de conception, un réseau neuronal est formé pour effectuer des tâches de raisonnement logique, ce qui correspond étroitement à notre catégorie *raisonnement basé sur l'apprentissage*. Cette approche présente des avantages notables, notamment une meilleure capacité de mise à l'échelle du raisonnement logique par rapport aux méthodes traditionnelles qui peuvent se heurter à des goulets d'étran-

29. ces différentes possibilités sont détaillées plus avant dans la section 4.2

glement lorsqu'elles traitent de grandes ontologies. En outre, le raisonnement basé sur l'apprentissage présente une plus grande résistance aux données bruyantes ou manquantes, améliorant ainsi la robustesse globale du processus de raisonnement.

Enfin, le modèle de conception numéro 12 est axé sur la conception de systèmes d'IA hybrides plus complexes qui peuvent être qualifiées d'applications réelles. Contrairement à un composant monolithique unique, les systèmes d'IA hybrides sont composés de multiples modules interconnectés qui communiquent entre eux. Ce modèle de conception correspond à la catégorie *Learning and Reasoning system* identifiée dans notre analyse systématique de la littérature. L'objectif de ces systèmes d'IA hybrides est d'exploiter les synergies entre les modules d'apprentissage et les modules symboliques, afin de produire des modèles plus fiables avec une transparence et une reproductibilité accrues. En intégrant de multiples modules, ces systèmes peuvent bénéficier à la fois des avantages des approches basées sur l'apprentissage et des avantages des techniques de raisonnement symbolique.

3.5.2/ La taxonomie du neuro-symbolique

Un sous-groupe bien reconnu au sein de l'IA hybride est celui du neuro-symbolique, qui se concentre sur l'intégration des méthodes symboliques avec les réseaux neuronaux, en particulier avec les réseaux de neurones profonds. [196] a introduit une taxonomie complète qui classe les diverses approches neuro-symboliques, fournissant un cadre structuré pour les comprendre et les catégoriser. En effet, de manière similaire aux modèles de conception de [195], les groupes décrits dans la taxonomie de [196] s'alignent également sur les catégories identifiées dans notre SLR, en particulier lorsque l'algorithme d'apprentissage employé est un réseau neuronal (cf. Tableau 3.14).

La catégorie *Informed machine learning* englobe deux types d'approches neuro-symboliques. La première est la "Symbolic Neuro symbolic", qui consiste à transformer des données brutes à l'aide d'une intégration symbolique. Cette technique est couramment employée dans les tâches de NLP, où les données sont converties en vecteurs à l'aide de méthodes telles que Word2vec et GloVe. La seconde approche consiste à utiliser l'architecture "Neuro_{Symbolic}" pour les systèmes d'apprentissage automatique plus complexes. Cette architecture facilite la conversion des règles symboliques en modèles de réseaux neuronaux, comme le font les réseaux de tenseurs logiques [100].

L'architecture "Symbolic[Neuro]" associe la reconnaissance de formes à un cadre symbolique de résolution de problèmes, ce qui permet d'améliorer les capacités de résolution de problèmes. Cette architecture est spécifiquement appliquée dans la catégorie *Expert System Embedded Learning* où l'apprentissage permet de peupler un système expert qui pourra ensuite prendre des décisions conformément à un ensemble de règles établies.

L'architecture "Neuro|Symbolic" occupe une place importante dans notre étude, englobant les catégories suivantes : *ontology learning*, *ontology mapping*, *ontology explain black-box*, et *hybrid application*. Cette architecture ressemble beaucoup à l'architecture "Symbolic[Neuro]" mais utilise des coroutines au lieu de sous-programmes. Elle met l'accent sur la communication entre un système symbolique et un système neuronal, ce qui est particulièrement pertinent pour notre catégorie *hybrid application*. Nous avons choisi de placer les trois autres catégories dans cette architecture parce qu'elle correspond le mieux à leurs fonctionnalités respectives, même si la communication bi-directionnelles entre les deux systèmes peut être plus limitée que dans le cas des applications hybrides.

Enfin, l'architecture "Neuro :Symbolic→Neuro" est analogue à la catégorie *Learning-based reasoning*, qui permet d'opérer un raisonnement symbolique grâce à un réseau de neurones. Dans le cas du raisonnement basé sur l'apprentissage, le réseau apprend des règles logiques pour effectuer un raisonnement déductif sur de nouvelles entrées. L'avantage de cette approche est que le réseau de neurones n'effectue pas de raisonnement en suivant explicitement des règles étape par étape ; au lieu de cela, il fait des prédictions basées sur le résultat attendu du raisonnement déductif. Comme indiqué précédemment, cette approche réduit considérablement le temps de calcul, en particulier lorsqu'il s'agit d'ontologies de grande taille. En revanche, il ne garantit pas toujours la cohérence des résultats obtenus.

3.5.3/ La combinaison de l'ontologie avec l'apprentissage automatique

Les travaux d'études menés dans notre SLR se sont concentrés sur l'hybridation de l'apprentissage automatique avec les ontologies. Bien qu'ayant certains points communs avec les modèles de conception de Van Bekkum et al. et la taxonomie du neuro-symbolique faite par Kautz, ils présentent une approche globalement différentes comme l'illustre la Figure 3.18.

En s'intéressant plus spécifiquement aux ontologies, dans le contexte de l'IA hybride, le concepteur offre un cadre plus formel aux connaissances qu'il souhaite mobiliser dans son système. C'est un grand avantage notamment lorsqu'il s'agit de systèmes complexes ayant besoin d'une grande interopérabilité associée à une grande capacité d'évolution comme on peut en trouver au sein de la Smart City.

3.6/ Conclusion

Notre étude quantitative, au cours de laquelle nous avons intégralement examiné et classé 128 articles, nous a permis de dresser une cartographie complète, à date, de l'hybridation entre ontologie et apprentissage automatique.

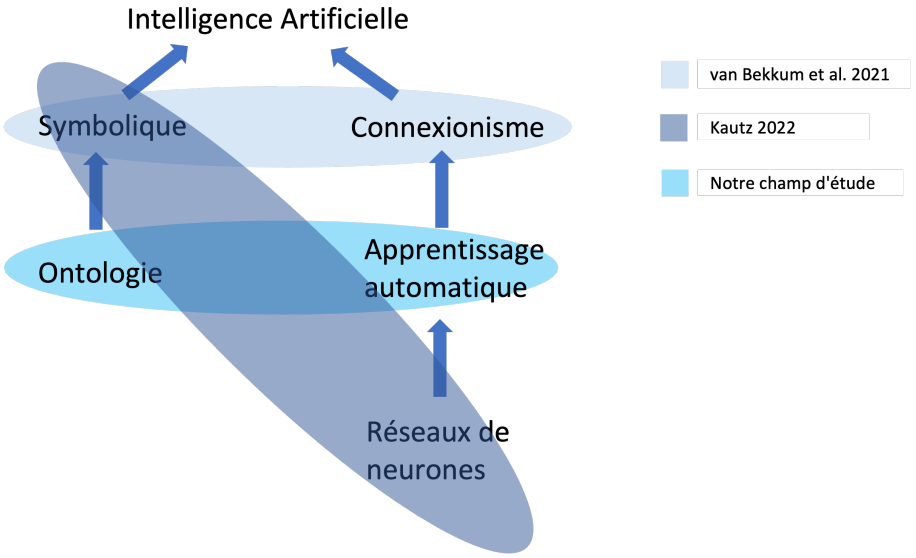

FIGURE 3.18 – Différences entre les trois paradigmes d'étude de l'IA Hybride

L'une de nos principales conclusions est que ce domaine ne se limite pas exclusivement à l'apprentissage automatique informé. Il intègre également la création d'ontologies automatisée par l'apprentissage ainsi que des systèmes plus complexes imbriquant différentes techniques de raisonnement inductif et déductif.

Les différentes questions de recherche auxquelles nous avons répondu nous ont permis d'identifier plusieurs choses. Premièrement que l'hybridation de raisonnement est un domaine en plein essor, avec un nombre de publications de plus en plus important dans les trois dernières années. Deuxièmement, il est évident que cette discipline trouve une forte application dans le domaine de la santé, en raison des nombreuses contraintes spécifiques qui y sont présentes. De plus, nous avons relevé d'autres cas d'utilisation explicitement liée à la Smart City ou approchant, comme la gestion de l'énergie, des bâtiments et des transports, démontrant ainsi que l'IA hybride est une solution pertinente pour le développement de modèles de prédiction dans ce contexte. Troisièmement, nous avons remarqué que ces techniques sont utilisées en majorité pour informer l'apprentissage automatique, en particulier pour le traitement du langage naturel et la vision par ordinateur. Néanmoins, quelques travaux se concentrent sur les séries temporelles, ce qui peut s'avérer utile pour le traitement des données provenant de capteurs.

De plus, pour chacune des trois grandes catégories d'hybridation identifiées, nous avons répertorié les divers algorithmes d'apprentissage automatique employés.

3.6.1/ Les trois défis de l'IA hybride

Comme l'illustre cette étude, la combinaison de l'ontologie et de l'apprentissage automatique reste un grand défi. Le terme général pour définir cette association est l'hybridation des intelligences artificielles. Dans cet état de l'art nous avons identifié trois principaux défis liés à l'utilisation d'ontologies combinées à l'apprentissage automatique.

— Le premier défi est la preuve formelle de l'expressivité et de la décidabilité de l'ontologie. Entre taxonomie et ontologie formelle, la représentation sémantique des connaissances est un équilibre entre l'expressivité et la décidabilité. Les logiques de description sont utilisées pour formaliser l'ontologie et déterminer ce niveau d'expressivité/décidabilité. Une grande majorité des articles étudiés ne mentionne pas le raisonnement déductif au-delà des liens de subsomption rendu possible par les ontologies et les moteurs d'inférence. Beaucoup d'entre eux utilisent une ontologie pour sa contribution au niveau sémantique. L'ontologie est utilisée comme une taxonomie améliorée puisqu'elle a l'avantage de représenter également les relations non hiérarchiques entre les différents éléments de l'information.

— Le deuxième défi est la capacité à expliquer les résultats d'un algorithme d'apprentissage automatique. Depuis une vingtaine d'années, l'intelligence artificielle explicable (XAI) est devenue un domaine de recherche intensif. Cette caractéristique est un enjeu crucial pour certains secteurs industriels sensibles tels que la santé, la banque, l'assurance ou encore la défense. Cette explicabilité peut être obtenue par des techniques telles que les explications locales interprétables (Local Interpretable modèle-agnostique (LIME) [197], SHapley Additive exPlanations (SHAP) [198], ou par le raisonnement symbolique. Des articles récents dans le domaine neuro-symbolique traitent de l'explicabilité [199, 200], mais nous n'avons trouvé que deux articles qui traitent de l'explicabilité dans l'apprentissage profond, en utilisant l'ontologie [72, 77]. Il s'agit d'une une méthode intéressante qui peut être utilisée lorsqu'une explication globale [72] ou locale [77] est nécessaire. Dans ces travaux, une ontologie de domaine spécifique au domaine à expliquer est utilisée pour augmenter la qualité de l'explication.

— Le troisième défi concerne la gestion de la cohérence pendant l'apprentissage de l'ontologie. La construction d'une ontologie de domaine par des experts humains reste très coûteuse et sujette aux erreurs, c'est pourquoi la construction automatique d'ontologies continue d'être développée. La gestion des changements dans l'ontologie au cours du processus d'apprentissage de l'ontologie ou de la cartographie de l'ontologie nécessite un contrôle de cohérence. La gestion de la cohérence permet de garantir la validité de raisonnement des différentes versions de l'ontologie. Cette étude de la cohérence est bien menée par [146] qui s'intéresse à l'enrichissement d'une ontologie, ainsi que par [139] et [66], classés dans

la catégorie population d'ontologies.

3.6.2/ Bilan

Un aspect particulièrement surprenant à l'issue de cette étude est le constat que, à l'exception du domaine de l'apprentissage d'ontologie, seuls deux articles dans les autres domaines identifiés abordent la question de l'évaluation de la cohérence des prédictions par rapport aux réalités physiques ou aux connaissances préalables. Le manque de considération envers ce problème, bien qu'il soit central pour notre domaine d'application, a été le moteur de notre étude visant à formaliser une approche d'évaluation générique pour les systèmes d'apprentissage automatique enrichis par des connaissances préalables. L'absence de réponse à cette question a orienté la suite de ce travail de recherche. Comment formaliser une approche d'évaluation générique de systèmes d'apprentissage automatique enrichie par des connaissances préalables ? Notre objectif principal étant de développer une mesure de cohérence entre les prédictions numériques et la réalité sur le terrain.

II

Contribution

4

Apprentissage automatique enrichi par la connaissance

4.1 Introduction 79
4.2 Enrichissement par les connaissances 80
4.2.1 Les données d'entraînement 81
4.2.1.1 L'ingénierie des caractéristiques 82
4.2.1.2 L'intégration des connaissances 84
4.2.1.3 Les simulations 85
4.2.2 L'architecture du modèle 85
4.2.3 La phase d'apprentissage 87
4.2.4 Le modèle final 88
4.2.5 Bilan sur l'ajout de connaissance en apprentissage automatique . . 89
4.3 Évaluation de la qualité 90
4.3.1 Évaluation de la qualité en apprentissage automatique 91
4.3.2 Stratégies d'évaluation des modèles 93
4.3.2.1 Le processus de conception d'un modèle 94
4.3.2.2 Les étapes d'évaluation d'un modèle 95
4.3.3 Limites dans l'évaluation des modèles sans cohérence 97
4.3.4 Évaluation des différents types de cohérence 99
4.3.4.1 La cohérence avec les données d'entraînement 99
4.3.4.2 La cohérence avec les connaissances 104
4.4 Méthodologie d'évaluation de la cohérence 109

4.4.1 Analyse de l'existant 109
4.4.2 Approche d'évaluation des systèmes d'apprentissage automatique informés 111
4.5 Conclusion 115

Ce chapitre vise à comprendre comment et à quel moment il est important d'évaluer la cohérence d'un système d'apprentissage automatique supervisé. Pour cela, il détaillera à quels endroits, dans un système d'apprentissage, il est possible d'ajouter de la connaissance. Cette notion est importante, car selon l'endroit, l'évaluation de la cohérence sera traitée différemment. Ces différentes possibilités et leurs impacts seront expliqués. Ce chapitre permettra aussi de démontrer que la cohérence s'inscrit dans une démarche plus globale d'évaluation de la qualité d'un système d'intelligence artificielle. Pour cela, ce chapitre présentera une proposition d'un protocole d'évaluation incluant l'examen de la cohérence.

4.1/ Introduction

L'apprentissage automatique a facilité la manière dont nous résolvons des problèmes complexes et variés, des tâches de traitement de langage naturel à la vision par ordinateur en passant par la prédiction de données financières. Cependant, les limites de cette méthode se font sentir lorsque le contexte d'étude impose des contraintes et des normes spécifiques qui doivent être rigoureusement respectées. Les modèles n'arrivent pas toujours à capter ces informations dans les données d'entraînement et fournissent alors un modèle incohérent par rapport à la réalité.

L'intelligence artificielle symbolique, d'un autre côté, permet de représenter et d'exploiter ces contraintes et ces normes sous forme d'un ensemble de connaissances sur lequel une forme de raisonnement déductif est opérée. Une combinaison de ces deux paradigmes, comme discuté dans le précédent, est peut-être la solution pour obtenir des modèles suffisamment flexibles, reposant sur des données empiriques tout en restant cohérents avec la réalité.

Dans nos travaux, les ontologies ont été choisies pour représenter formellement les connaissances, car elles présentent de nombreux avantages intéressant pour la Smart City. Elles facilitent la capture et la gestion de la connaissance sur les infrastructures, les services, les politiques, les réglementations et d'autres aspects complexes inhérents à une ville intelligente. Elles jouent un rôle important dans l'interopérabilité des systèmes et des données provenant de différentes sources, favorisant ainsi une meilleure collaboration et une intégration plus fluide des informations. Leurs capacités de raisonnement permettent d'inférer de nouvelles informations à partir des données existantes, ce qui peut être précieux pour la prise de décision, la détection de problèmes et l'optimisation des opérations urbaines. De plus, les ontologies étant conçues pour être aisément extensibles, i.e. elles peuvent inclure de nouveaux concepts et relations au fil du temps, elles s'adapteront sans peine aux évolutions technologiques et aux besoins changeants de la population.

Ce chapitre traite de l'étude des systèmes d'apprentissages informés par les ontologies. L'objectif principal est de comprendre comment assurer la cohérence de tels systèmes en s'assurant que les prédictions qu'ils font sont bien en adéquation avec les connaissances du domaine d'application.

La première section de chapitre explore en détail les diverses manières d'incorporer de la connaissance au sein d'un système d'apprentissage. En effet, cette base théorique est importante, car l'évaluation de la cohérence peut dépendre de la technique d'ajout des connaissances utilisées. Dans la deuxième section, nous expliquons en quoi la question de la cohérence s'inscrit dans une démarche plus globale d'évaluation de la qualité d'un système d'intelligence artificielle. Enfin, la troisième section présente un état de l'art sur l'évaluation de la cohérence en apprentissage automatique informé. Cette analyse plus approfondie du sujet est suivie par la présentation d'un protocole d'évaluation incluant l'examen de la cohérence.

4.2/ Méthodes d'enrichissement par les connaissances

Pour concevoir un algorithme d'apprentissage automatique, il faut avoir trois éléments de base que sont des données d'entraînement, une architecture appropriée au problème à résoudre ainsi qu'une phase d'apprentissage pensée pour optimiser les résultats [35, 201]. Ces trois éléments doivent être optimisés au mieux pour obtenir un modèle final fiable. En toute logique, il est possible d'ajouter de la connaissance dans l'ensemble des quatre phases constituant un algorithme d'apprentissage automatique présenté en Figure 4.1.

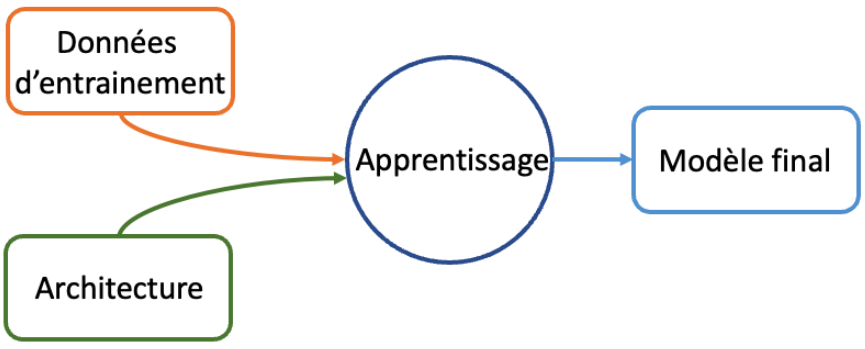

Figure 4.1 – Organisation d'un modèle d'apprentissage automatique d'après [201]

Les réflexions suivantes se basent à la fois sur cette observation et sur la SLR présentée au chapitre 3, dans laquelle une analyse approfondie des méthodes de combinaison de l'apprentissage automatique avec les ontologies a été réalisée. Cette SLR a permis de mettre en évidence les différentes manières d'intégrer de la connaissance au sein d'un

modèle d'apprentissage automatique. La section qui suit présente un état de l'art succinct de ces techniques.

4.2.1/ Les données d'entraînement

Les données d'entraînement font quasiment systématiquement l'objet de manipulation en vue de leur utilisation dans un processus d'apprentissage automatique. Une première étape de nettoyage des données brutes est bien souvent nécessaire pour supprimer les valeurs incohérentes (souvent appelées valeurs aberrantes) ou bruitées, traiter les valeurs manquantes de manière adéquate ainsi que supprimer les variables redondantes ou trop corrélées. L'apport de connaissance, qui plus est formalisée, est utile pour faciliter le nettoyage des données puisqu'elle permet d'automatiser cette étape fastidieuse [202]. Les valeurs aberrantes peuvent par exemple être plus facilement reconnues à l'aide de règles métiers. C'est le cas pour les données d'OpenFoodFacts [1], issues d'une production participative, ces données sont nombreuses, mais souvent erronées, l'application de règles métiers simples et reconnues (ex : "Pour 100g de produit, il ne peut y avoir plus de 100g de sucre") permet de repérer et de traiter une grande partie des incohérences de la base de données. De même, une connaissance sur les interactions entre variables permet également de pouvoir trouver les valeurs de certaines données manquantes (ex : Dans OpenFoodFacts le Nutri-Score est parfois manquant alors que les variables nécessaires à son calcul sont bien renseignées, connaître l'équation du Nutri-Score permet de le reconstituer).

D'autres étapes de préparation des données sont également indispensables pour améliorer les résultats du modèle. Elles sont réalisées en s'adaptant à la fois au type de données d'entrée et à la problématique que le modèle doit résoudre. Là encore, ces transformations mobilisent parfois les connaissances du concepteur de l'algorithme pour savoir lesquelles utiliser et dans quel ordre. Ainsi, dans les problèmes de classification si les classes à prédire ne sont pas équilibrées, des techniques de sur-échantillonnage ou de sous-échantillonnage peuvent être appliquées pour équilibrer les données et éviter tout biais résultant des classes dominantes. Lorsque les variables d'entrée sont dans des unités différentes, une étape de normalisation des données est vivement conseillée. De même, lorsque l'amplitude des données est importante et/ou que la distribution des données d'entraînement est loin de suivre une loi Normale, il est prudent de passer les données au logarithme.

En ce qui concerne les données textuelles ou les données de langage naturel, des étapes telles que la tokenisation (division en mots ou en sous-unités de sens), la suppression de la ponctuation, la suppression des mots vides (mots courants mais peu informatifs),

1. https://fr.openfoodfacts.org/data

la lemmatisation (réduction des mots à leur forme de base) et la vectorisation (conversion des mots en vecteurs numériques) sont couramment effectuées pour préparer les données avant l'entraînement du modèle. Pour le traitement d'images, le redimensionnement (modifier la taille de l'image), la normalisation (mettre à l'échelle), l'augmentation des données (en ajoutant toutes les rotations d'une même image) et le recadrage sont également des processus de préparations des données couramment effectués.

Bien que ces diverses techniques de préparation des données reposent essentiellement sur des connaissances, elles sont rarement formalisées, ce qui ne facilite pas leur utilisation dans des travaux futurs. En effet, l'ajout de ces connaissances ne se fait pas de manière automatisée. Elles sont souvent implémentées lors de la préparation des données selon la volonté du développeur en charge ce qui augmente la taille du code et réduit la maintenabilité de celui-ci. Cela ne permet pas de capitaliser sur les connaissances déjà acquises, et réduit la possibilité de les partager simplement.

Cependant, pour pallier à cette absence d'automatisation certains travaux mobilisent des ontologies formalisées dans leur processus de transformation des données brutes comme cela est montré dans la section 3.4.2 de la SLR. Les trois principales possibilités de mobiliser des connaissances formelles pour préparer les données d'entraînement du modèle sont l'ingénierie des caractéristiques (en anglais *feature engineering*), l'intégration des connaissances (en anglais *knowledge embedding*) et la simulation.

4.2.1.1/ L'ingénierie des caractéristiques

L'ingénierie des caractéristiques désigne le processus de sélection, d'extraction et d'augmentation de variables (caractéristiques) à partir des données brutes dans le but d'améliorer les performances des modèles d'apprentissage automatique.

La sélection permet de conserver seulement les variables les plus intéressantes pour le modèle ce qui réduit sa complexité (parfois en facilitant son interprétabilité), améliore sa vitesse d'apprentissage et sa généralisation tout en éliminant les caractéristiques redondantes ou peu pertinentes. Le processus de sélection des données peut-être réalisé de multiple façon sans impliquer l'ajout systématique de connaissance préalable. En particulier en utilisant des méthodes statistiques basées sur l'analyse des corrélations tels que le coefficient de Pearson [203], le coefficient de corrélation de Spearman [204], le Tau de Kendall [205], l'analyse de la variance (ANOVA) [206], le test du khi-deux [207] ou l'information mutuelle [208]. Une analyse du modèle a posteriori peut également être réalisée afin d'identifier quelles sont les variables les plus explicatives de celui-ci, en particulier sur les arbres de décision. D'autres techniques modernes comme LIME [197] et SHAP [198] peuvent également être utilisées. Connaître les caractéristiques les plus discriminatoires du modèle permet d'éliminer celles qui au contraire apportent peu d'infor-

mation du modèle d'apprentissage automatique. Les connaissances du domaine peuvent également aider à la sélection des données lorsque les liens et les règles qui s'opèrent entre les différentes variables d'entrée sont connus. L'ontologie, qui permet de formaliser ces relations entre les caractéristiques et leurs propriétés, peut être utilisée pour repérer les variables les plus pertinentes et les moins corrélées entre elles [58]. Les règles de sélection des données peuvent être établies à partir des relations entre les propriétés contenues dans l'ontologie.

L'extraction de caractéristiques poursuit un objectif semblable à celui de la sélection, à savoir réduire le nombre des variables d'entrée d'un modèle, tout en altérant ces mêmes variables afin d'en optimiser leur usage. En effet, plutôt que de simplement piocher parmi les variables existantes, l'extraction va les modifier pour en obtenir la quintessence, i.e. des variables contenant le plus d'informations possibles et ne ressemblant plus à celles de départ. C'est le rôle d'un algorithme de réduction des dimensions comme l'Analyse en Composante Principale (ACP) [209, 210] qui transforme un ensemble de caractéristiques initiales en un nombre plus réduit de composantes (des axes) de variance maximale. C'est également le cas d'autres techniques comme t-SNE ou LDA utilisées pour l'analyse de texte. La structure hiérarchique de l'ontologie peut être mise à profit pour réduire les variables et les transformer en retrouvant les concepts principaux qui peuvent correspondre à plusieurs variables initiales [160]. L'ontologie peut également être mise à profit dans des travaux liés au NLP, en particulier pour identifier les topics principaux présents dans un texte [170]. La réduction d'un texte en topic est souvent réalisée par des méthodes statistiques comme LDA, mais l'ontologie apporte davantage de cohérence sémantique lorsqu'elle est utilisée à cette fin. Une autre façon de garantir ce type de cohérence est d'utiliser l'ontologie comme interface de correspondance entre des termes [105], ce qui est particulièrement utile lorsqu'on manipule des sources de données hétérogènes.

A contrario, les variables peuvent également avoir parfois besoin d'être augmentées, c'est-à-dire que de nouvelles caractéristiques vont être ajoutées aux variables initiales, souvent dans le but d'apporter un complément d'information au jeu de données initial [211]. Bien que des méthodes impliquant des statistiques puissent être utilisées, cet ajout se fait principalement par l'apport de connaissances déjà existantes qui peuvent être mobilisées directement depuis une ontologie. C'est le cas pour un projet de reconnaissante d'activités humaines dans un habitat intelligent, où les remontées de capteurs sont augmentées grâce à une ontologie des activités humaines qui décrit des règles inhérentes à chaque activité (ex : "Si une personne est dans la chambre, elle semble allongée, les lumières de la pièce sont éteintes et la porte de la chambre est fermée, alors l'utilisateur est en train de dormir") avant d'être passées à un algorithme k-means [163]. Dans le domaine de la bio-informatique, l'ontologie Gene Ontology [2] est souvent utili-

2. http://geneontology.org/

sée pour annoter des données brutes afin d'apporter plus d'informations pertinentes en entrée, comme cela peut être le cas pour la prédiction de protéines synthétisées en réponse au stress oxytatif [147] ou pour la prédiction de la localisation subcellulaire de protéines [134]. Là encore, dans le domaine du NLP, l'ontologie peut apporter de nouveaux éléments aux données textuelles en entrée, par exemple en ajoutant des précisions à des notes médicales où certains termes sont enrichis par des précisions importantes [97].

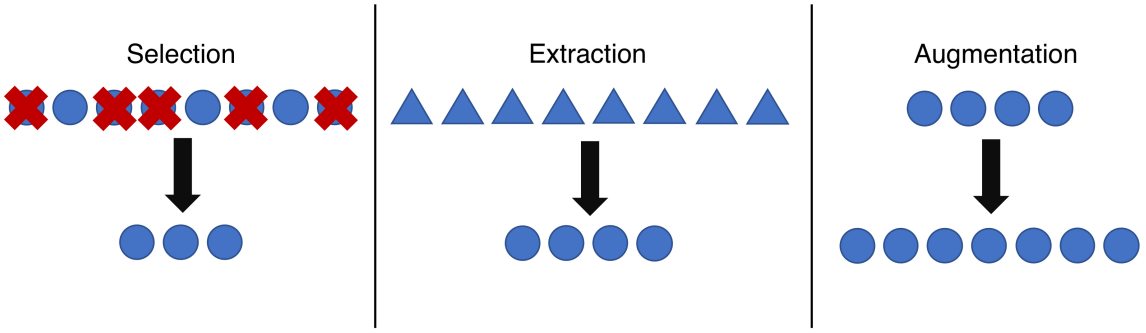

FIGURE 4.2 – Les trois techniques utilisées en ingénierie des caractéristiques : la sélection, l'extraction et l'augmentation

4.2.1.2/ L'INTÉGRATION DES CONNAISSANCES

L'intégration des connaissances, dérivée de l'intégration des caractéristiques [3], peut-être vue comme un type particulier d'extraction de caractéristiques. En effet, l'intégration des caractéristiques transforme les variables et les données brutes en vecteurs afin de pouvoir faciliter leur utilisation par un réseau de neurones. L'objectif est d'obtenir un espace vectoriel de dimensions réduites qui représente le mieux les informations contenues dans les données. La décomposition en valeurs singulières [4] (SVD) est une méthode de réduction de dimension, à l'instar de la ACP, qui permet de factoriser une matrice [212]. En traitement du langage, l'algorithme Word2Vec permet de générer des vecteurs à partir des mots en mettant en avant les mots qui sont en relation les un avec les autres (ex : le mot "roi" est proche du mot "homme", tout comme le mot "reine" est proche du mot "femme") [213].

L'intégration des connaissances propose d'étendre ce type de transformation en ajoutant des connaissances dans les vecteurs de données. C'est notamment possible grâce à l'*ontology embedding*, qui permet de transformer une ontologie en vecteurs [5] en sélectionnant les informations pertinentes (exemples : concepts, relations, propriétés, etc.) [108, 153, 177]. Le processus d'*ontology embedding* est souvent précédé par un

3. en anglais *feature embedding*
4. en anglais *Singular Value Decomposition*
5. voir aussi *graph embedding* terme plus usité qui permet de transformer un graphe de connaissance en vecteurs [214]

traitement des données avec Word2Vec, l'ontologie permettant alors d'ajouter de nouvelles connaissances aux données d'entrée. D'après les résultats fournis par [153] les modèles ajoutant des connaissances via une ontologie obtiennent de meilleurs scores de performances que les autres.

4.2.1.3/ Les simulations

Une simulation consiste à représenter le fonctionnement d'un processus, qu'il soit physique, industriel, biologique, économique ou militaire, en utilisant un modèle concret où les paramètres et les variables reflètent ceux du processus en question. L'objectif principal d'une simulation est de comprendre comment le système fonctionne dans des conditions spécifiques et d'explorer différentes situations sans avoir à manipuler le système réel directement. Cela permet d'obtenir des informations précieuses sans les risques ou les coûts associés à la manipulation d'un système réel.

Si l'apprentissage automatique est parfois mis à profit dans la création de simulations, elles nécessitent souvent une compréhension approfondie du domaine et du problème étudié ce qui implique l'usage de connaissances dans leur élaboration. Les simulations réalisées à partir de connaissance préalables peuvent également servir à simuler des données de sortie qui sont ensuite consommées par des algorithmes d'apprentissage automatique [215–217]. C'est un bon moyen de créer des données d'entraînement cohérentes avec les connaissances issues d'un domaine en particulier. Là encore, on peut imaginer l'ontologie comme une source de connaissance fiable permettant de créer des simulations cohérentes. L'avantage étant qu'une même ontologie est capable d'être exploitée pour différentes simulations dont les cas d'usage ne sont pas toujours rigoureusement identiques.

4.2.2/ L'architecture du modèle

L'architecture d'un modèle d'apprentissage automatique fait référence à la structure d'organisation des différents éléments qui compose l'algorithme d'apprentissage final, qu'il est possible de désigner sous le nom de pattern final, ainsi que les hyperparamètres qui lui sont associés. Hyperparamètre est le nom choisi pour désigner l'ensemble des paramètres qui doivent être fixés avant l'entraînement effectif du modèle. La structure d'un modèle peut être formée d'un algorithme seul (ex : régression linéaire ou *Random Forest*) ou bien composée de plusieurs algorithmes différents (ex : *Random Forest* suivi d'un perceptron multi-couches (MLP)). Ces différences d'architecture sont particulièrement bien illustrées par les réseaux de neurones qui présentent de nombreuses topologies comme celle d'un réseau entièrement récurrent (en anglais *fully recurrent network*), ou d'une carte auto-organisée (en anglais *self-organizing map*), ou encore d'un réseau de neu-

rones à convolution (en anglais *convolutional neural network*) par exemple. L'architecture finale retenue dépend de chaque cas d'application spécifique. En effet, elle doit prendre en compte à la fois la nature des données (une série temporelle n'aura pas exactement les mêmes besoins de traitement qu'un ensemble d'images) ainsi que du type de résultat souhaité (classification, régression, analyse non supervisée, etc.).

L'architecture doit également se conformer dans la mesure du possible à un certain nombre de contraintes, éventuellement fournies par un apport de connaissances préalables provenant d'une ontologie, que ce soit au niveau de la sélection de la structure du modèle ou au niveau du réglage des hyperparamètres [218–220].

Le choix de structure d'un modèle d'apprentissage est souvent décidé par le concepteur du modèle en fonction de ses connaissances vis-à-vis du cas d'usage. Dans certains cas, l'existence d'une ontologie peut contribuer à la mise au point d'une architecture capable de prendre en compte les spécifications de la problématique étudiée [220]. La sélection d'une topologie de réseau de neurones ou de réseau bayesien peut être guidée par une ontologie [68, 221]. La base de connaissance formelle étant mis à profit pour créer un ensemble de structures cohérentes avec le contexte du cas d'usage et les résultats qui sont attendus. L'ontologie permet également de construire des structures originales déterminées par un ensemble de contraintes définies pour une problématique particulière. Ainsi, la classification automatique des fonctions des protéines doit autant que faire se peut obtenir des résultats cohérents avec le graphe acyclique dirigé (DAG) représenté par Gene Ontology, l'usage d'une architecture qui respecte la contrainte hiérarchique du DAG est indispensable pour espérer obtenir des résultats proches de la réalité [222,223]. Les réseaux de neurones en graphes [6] (GNN) utilise une autre structure particulière de réseau de neurones qui permet de traiter les graphes de connaissance grâce à la forme de son architecture [224,225]. En effet, le réseau est constitué d'un MLP (Multi Layer Perceptron) distinct pour chaque composant du graphe [7] (ce qui correspond à une couche du réseau). Le graphe de connaissance issu de l'ontologie peut tout à fait servir de base à l'élaboration d'un GNN, néanmoins l'exploitation d'un raisonnement déductif au sein d'un GNN est une réflexion, qui à notre connaissance, n'a pas encore été menée.

La nature des hyperparamètres dépend entièrement de l'algorithme d'apprentissage automatique utilisé pour construire le modèle. Les réseaux de neurones ont comme hyperparamètres le nombre de couches et de nœuds [8], les fonctions d'activations utilisées par chacun des couches, le taux d'apprentissage, la taille des batch ou le nombre d'epochs [9] tandis que l'algorithme *Random Forest* a besoin de connaître le nombre maximum d'arbres dans la forêt où le nombre maximum de niveau de décision pour chaque

6. en anglais *Graph Neural Network*
7. i.e. chaque nœud et chaque arrête
8. ces deux paramètres ont un impact sur la structure du modèle en lui même
9. Un epoch représente une itération complète à travers l'ensemble de données d'entraînement.

arbre par exemple. En général, ces paramètres sont choisit sur la base de règles heuristiques et ils sont ajustés manuellement, ce qui peut être très long, car évaluer les performances d'une seule configuration des paramètres du réseau neuronal peut nécessiter plusieurs heures [219]. Les règles heuristiques sont souvent utilisées pour guider la prise de décision et la résolution de problèmes dans des situations où les solutions exactes ne sont pas facilement accessibles où lorsque le processus de résolution nécessite trop de temps ou de ressources (comme c'est le cas pour les hyperparamètres). Ces règles fournissent des lignes directrices approximatives qui aident à trouver des solutions acceptables dans des circonstances où l'application de méthodes rigoureuses mathématiques est difficile. Ce type de règles peut être représenté de manière formelle dans une ontologie qui pourra ensuite être utilisée lors de l'étape de détermination des hyperparamètres du modèle [218, 220].

4.2.3/ La phase d'apprentissage

L'étape d'apprentissage correspond à la phase au cours de laquelle le modèle apprend à partir des données d'entraînement fournies : le modèle ajuste ses paramètres internes pour être en mesure de faire des prédictions ou des classifications qui semblent correctes en fonction des métriques d'évaluation définies. Certaines fonctions permettent d'ajuster les biais du modèle au cours de l'apprentissage, elles sont vues comme des hyperparamètres étant donné qu'elles sont définies avant l'entraînement de celui-ci. Toutefois, c'est bien au cours du processus d'apprentissage que ces paramètres sont mobilisés.

Une de ces métriques, spécifique aux réseaux de neurones, est la fonction de coût (ou fonction de perte [10]) qui quantifie la différence entre les prédictions faites par un modèle d'apprentissage automatique et les valeurs réelles associées aux données d'entraînement. La fonction de coût est en général déterminée par la nature de la tâche que le modèle d'apprentissage automatique doit accomplir : une classification binaire et une régression n'utiliseront pas la même fonction de coût [226]. Comme pour les autres hyperparamètres, cette fonction peut donc être déterminée grâce à des règles heuristiques connues d'une ontologie. Cependant, on peut aller plus loin que cela en intégrant des connaissances préalables au sein même de cette fonction de coût comme le font certains réseaux de neurones informés par la physique (PINN) [37, 40]. Ces connaissances peuvent même être exprimées sous formes de règles logiques avant d'être transformées en contraintes exploitables dans la fonction de coût [227]. L'ajout de connaissances préalables au sein de cette fonction de perte peut améliorer les performances de généralisation du modèle [37].

La fonction d'activation joue également un rôle très important dans la conception d'un

10. en anglais *loss function*

réseau de neurones. Elle agit comme une porte qui détermine si un neurone transmettra ou non son signal à la couche suivante en fonction d'une certaine condition. Elle introduit une non-linéarité essentielle aux calculs du réseau, permettant au réseau de modéliser des relations complexes en autorisant ou non à ses neurones de propager de l'information en fonction d'un seuil prédéfini. Elle reproduit ainsi le comportement biologique du seuil de stimulation qui représente le niveau minimum de stimulation nécessaire pour qu'un neurone du cerveau humain génère un signal électrique qui se propage ensuite le long de son axone. Les fonctions d'activation sont souvent des fonctions purement mathématiques choisies en tenant compte de la nature des données d'entrée ainsi que de la problématique visée (classification binaire, multi-label ou régression par exemple). Toutefois, ces fonctions d'activation peuvent également être basées sur des connaissances issues notamment d'une ontologie [228].

Concernant les réseaux de neurones le paramètre optimiseur (en anglais *optimizer*), qui tient un rôle important dans l'ajustement des poids et des biais permettant de minimiser la fonction de coût, est encore à ce jour un sujet de recherche important. Toutefois, il n'existe pas à notre connaissance d'étude mêlant de la connaissance préalable à la définition d'une fonction d'optimisation.

4.2.4/ Le modèle final

L'évaluation *a posteriori* des modèles d'apprentissage automatique est une pratique essentielle pour garantir la performance, la fiabilité et la pertinence de ces derniers dans des applications réelles, comme discuté dans la section 4.3.4. Avant de commencer l'évaluation, il est important de définir ce qui est considéré comme une réussite pour le modèle : cela peut être une bonne performance sur certaines métriques et/ou le respect des normes établies dans le domaine. En effet, l'évaluation des modèles d'apprentissage automatique n'est pas seulement une question de performance pure, mais aussi de cohérence et d'interprétabilité. Il est possible d'obtenir de bons scores de performance tout en ayant un modèle plutôt incohérent [229] ce qui posera problème lors du déploiement à cause d'une mauvaise généralisation de celui-ci.

L'ajout d'une connaissance préalable dans cette évaluation est un atout puisqu'il permet de contextualiser les prédictions et de valider leur pertinence. Les connaissances peuvent servir à identifier et à corriger les biais indésirables introduits par le modèle, en particulier lorsque ces biais ne sont perceptibles que par des experts du domaine. Elles peuvent également aider à valider le modèle en comparant ses prédictions avec des vérités de terrain ou des données externes, ce qui peut être vu comme une évaluation de la cohérence avec des connaissances préalables. Là encore, l'ensemble des types de cohérence (décrites dans la section 4.3.4) à évaluer pour un cas d'usage particulier doit

être prédéfini bien en amont, lors de la conception du modèle. Cela permettra de mettre en place un processus d'évaluation adapté, capable de rendre compte des incohérences éventuelles du modèle.

Si l'ontologie peut être mise à profit dans l'évaluation du respect d'un ensemble de contraintes, comme décrit ci-dessus, elle peut également jouer un rôle dans la recherche d'explicabilité du modèle construit. L'utilisation des ontologies dans l'explicabilité de l'intelligence artificielle (XAI) est une approche à considérer pour rendre les modèles d'apprentissage automatique plus compréhensibles et interprétables car elles fournissent un contexte sémantique aux variables et aux prédictions du modèle [72, 77]. Cela permet d'expliquer les décisions prises par le modèle en les reliant à des connaissances du domaine renforçant ainsi la compréhension des utilisateurs.

En fonction des résultats de l'évaluation, des ajustements peuvent être apportés au modèle, aux hyperparamètres ou au pré-traitement des données pour améliorer les performances ou rendre le modèle plus interprétable. La connaissance formalisée à une fois de plus un rôle à jouer dans cette étape pour favoriser la confiance et l'adoption des modèles d'apprentissage automatique.

4.2.5/ Bilan sur l'ajout de connaissance en apprentissage automatique

L'ajout de connaissance dans les différentes étapes, de la collecte des données à l'évaluation du modèle, en passant par son entraînement permet de l'aider à être plus cohérent avec le monde réel. En effet, sans apport de connaissance préalable, les modèles ne peuvent prendre en compte l'ensemble des contraintes liées à la réalité physique des êtres humains.

En entrée, les concepteurs de modèles d'apprentissage peuvent intégrer des informations préexistantes et des connaissances expertes issues d'une ontologie dans les données d'entraînement pour les aider à guider le pré-traitement des données. En plus des concepts, relations et règles, l'ontologie peut également mettre à profit ses capacités de raisonnement déductif pour aider au nettoyage des données d'entrée. De même, en sortie lors de l'étape d'évaluation, l'interprétation des résultats du modèle peut également faire appel à ce type de raisonnement pour donner du sens aux prédictions et aux décisions prises par le modèle tout en veillant à ce qu'elles soient conformes à des réglementations ou à des contraintes spécifiques.

Dans la phase de conception de l'architecture, les règles associées à un raisonnement déductif peuvent guider le choix des structures du modèle, des couches et des connexions en fonction des contraintes du problème. En revanche, ce type de raisonnement est plus complexe à mobiliser dans la phase d'apprentissage. Toutefois, des règles transformées en contraintes peuvent être incorporées pour réguler l'entraînement du mo-

dèle afin d'éviter le sur-apprentissage.

En conclusion, l'ajout de connaissances à chaque étape d'un algorithme d'apprentissage automatique peut renforcer la robustesse, les performances et l'interprétabilité du modèle final. Cependant, il est essentiel de trouver le bon équilibre entre l'utilisation de connaissances externes et la capacité du modèle à tirer profit des données d'entraînement. En effet, si les connaissances sont prépondérantes par rapport aux données, les modèles ne pourront plus les exploiter avec efficacité car ils seront trop contraints. Cela équivaudrait à utiliser un procédé purement symbolique comme un système expert.

L'ajout de connaissance dans les modèles ne nécessite-t-il pas également de revoir la méthode d'évaluation de ceux-ci ? En effet, jusqu'ici les modèles d'apprentissage supervisés étaient jugés principalement au regard de leurs résultats par rapport aux données d'entrée. N'est-il pas temps de les éprouver également sur des connaissances ?

4.3/ Évaluation d'un système d'intelligence artificielle

L'évaluation des algorithmes d'apprentissage automatique s'intègre dans le contexte plus global du contrôle de la qualité des systèmes d'intelligence artificielle. A l'instar de l'ingénierie de logiciel, ces systèmes doivent subir des processus de contrôle qualité avant leur déploiement dans le monde réel, étant donné leur influence significative sur ce dernier [230].

En ingénierie logicielle, les critères de qualité sont généralement l'aptitude à répondre au besoin de l'utilisateur, la conformité aux différentes exigences spécifiques, le fait d'être exempt de défaut ou d'imperfection ainsi que la satisfaction du client [230]. Pour réaliser cela, les concepteurs peuvent très souvent s'appuyer sur un cahier des charges rédigé par le client qui annonce les spécifications fonctionnelles du logiciel et parfois énumère les différentes normes ou contraintes devant impérativement être respectées. L'ISO 8402 définit la qualité de manière formelle par la formule suivante : $Q = P/E$, où P représente la performance du logiciel et E les attentes du client. Lorsque la qualité Q est égale à 1, alors les attentes du consommateur sont totalement comblées ce qui est optimal [230].

Parmi les différentes propriétés susceptibles d'améliorer la qualité il y a des exigences fonctionnelles, comme l'exactitude, et des exigences non fonctionnelles comme la fiabilité, l'efficacité, la robustesse, la facilité d'utilisation, l'interopérabilité, l'équité, la maintenabilité, la réutilisabilité et l'interprétabilité [230, 231]. Cependant, il est observé que certains critères de qualité en apprentissage automatique sont largement privilégiés au détriment d'autres qui sont parfois complètement ignorés [231].

4.3.1/ Évaluation de la qualité en apprentissage automatique

L'exigence fonctionnelle la plus largement étudiée en apprentissage automatique supervisé est l'exactitude [231], qui mesure la capacité d'un modèle à réaliser des prédictions conformes par rapport aux données avec lesquelles il a été entraîné. [231] définit l'exactitude, en particulier dans le cadre d'une classification, de la manière suivante :

$$E(h) = Pr_{x \in \mathcal{D}}[h(x) = c(x)] \tag{4.1}$$

Avec $\mathcal{D}$ étant la distribution des futures données inconnues, x un élément de l'ensemble des données appartenant à $\mathcal{D}$ et h le modèle d'apprentissage automatique qui est testé. $E(h)$ est la probabilité que $h(x)$, i.e. le label prédit par le modèle h pour l'entrée x, et $c(x)$, i.e. le vrai label, soit identiques. Cette formule peut bien sûr être adaptée dans le cadre d'une régression en remplaçant le terme label par celui de "prédiction numérique".

Atteindre une précision élevée avec des données du passé peut fournir des indications sur les performances futures avec des données à venir. Toutefois, les résultats antérieurs ne peuvent présager entièrement les résultats futurs, les performances réelles du modèle ne peuvent être évaluées qu'à l'aune de ces données futures. Or, elles sont par définition souvent indisponibles lors de la conception de l'algorithme. Pour simuler ces données inconnues, et afin de vérifier que le modèle n'est pas trop sujet au sur-apprentissage, les concepteurs divisent souvent les données antérieures en un jeu d'entraînement et un jeu de test. Le jeu de test reste inconnu du modèle durant tout son apprentissage et ne sert qu'à des fins d'évaluation de celui-ci. Des pratiques similaires bien que plus sophistiquées comme la validation croisée sont très souvent mises en place pour s'assurer d'une meilleure exactitude sur des données inconnues.

La seconde exigence, cette fois-ci non fonctionnelle, la plus étudiée, est la robustesse du système d'apprentissage automatique qui décrit la capacité d'un modèle à ne pas être trop impacté des éléments perturbateurs. Le terme de robustesse étant formellement définit pour l'ensemble des logiciels par le glossaire standard de l'IEEE (IEEE Std 610.12-1990) comme “The degree to which a system or component can function correctly in the presence of invalid inputs or stressful environmental conditions” [232]. Dans le cas de l'apprentissage automatique des conditions de perturbations peuvent être engendrées par des données incorrectes, bruitées, aberrantes, un changement dans l'utilisation d'un framework ou dans son processus d'apprentissage. Cette mesure de la robustesse d'un système d'apprentissage automatique est également définit par [231] avec la formule suivante :

$$r = E(S) - E(\delta(S)) \tag{4.2}$$

Avec S un système d'apprentissage automatique exempt de perturbations, tandis que $\delta(S)$ est un système d'apprentissage automatique perturbé sur au moins l'un de ses composants comme ses données, son processus d'apprentissage ou son architecture. $E(S)$ étant une mesure d'exactitude du modèle non perturbé, la robustesse est la différence entre $E(S)$ et $E(\delta(S))$. Un modèle robuste n'a donc pas une différence de performance significative en présence de données bruitées.

Cet aspect est étudié par [233], qui montre comment un classifieur identifie correctement un panda dans une image au départ, puis identifie à la place un gibbon car un bruit imperceptible pour l'œil a été ajouté à l'image. Un tel comportement est extrêmement problématique, car un acteur malveillant peut venir perturber les données d'entrée dans le but de modifier consciemment les prédictions réalisées par le modèle [234]. La bonne circulation à bord d'un véhicule autonome peut-être grandement perturbée par l'ajout d'un simple autocollant sur un panneau stop comme le montre [235], le panneau étant dès lors reconnu comme une limite de vitesse et non plus une obligation de s'arrêter à l'intersection. Dans une Smart City, ce type de comportement inapproprié peut conduire à de nombreux accidents puisque les interactions entre les usagers et les éléments mobilier de la ville sont multiples : travaux, affichages occasionnels, évolution de la végétation, etc.

Le besoin d'équité, bien que moins prépondérant, est également un domaine étudié par les concepteurs de modèles d'apprentissage automatique [231]. Les êtres humains étant susceptibles d'avoir des préjugés (parfois des biais inconscients), il faut s'assurer que ceux-ci n'impactent en rien les algorithmes d'apprentissage automatique. L'exemple très connu est un système capable de sélectionner des CV intéressants pour une entreprise qui finit par sélectionner uniquement des hommes blancs issues d'écoles prestigieuses en ignorant complètement les autres candidatures [236]. Toutes les caractéristiques sensibles comme le genre, l'âge, l'origine, la religion, la couleur de peau, le fait d'être enceinte ou non, le statut marital ou le handicap ne doivent pas avoir d'impacts négatifs sur résultats d'un modèle d'apprentissage qui puisse conduire à une injustice.

Toutefois, il n'existe pas à l'heure actuelle de définition qui fasse consensus concernant cette notion d'équité. De fait, l'équité est souvent spécifique à un domaine en particulier, et les règles qui doivent être scrupuleusement respectées pour l'un ne s'appliquent pas toujours à l'autre. Par exemple, un prêt bancaire ne saurait être octroyé en fonction des caractéristiques génétiques d'une personne, ces informations ne doivent à aucun moment être prises en compte par la banque, tandis que le montant de son salaire annuel est lui une information capitale pour ce type d'application. Pour le choix d'un traitement médical approprié c'est l'inverse, les informations génétiques d'une personne peuvent être primordiales dans ce contexte tandis que le montant total de son salaire annuel n'est pas une information que le modèle doit prendre en compte. Le respect de l'éthique lié a l'équité dépend donc de différents facteurs comme le contexte d'application et le carac-

tère néfaste qu'un préjugé peut apporter aux résultats.

Cependant, en dehors de l'exactitude, de la robustesse et de l'équité, l'étude menée par [231] montre que moins de 3% des articles modélisant un système d'apprentissage automatique s'intéressent aux exigences d'interprétabilité et de pertinence du modèle.

L'interprétabilité des modèles possède deux aspects distincts, le premier est la transparence (comment l'algorithme fonctionne) et le second est l'explicabilité des résultats à posteriori (pourquoi le modèle trouve-t-il tel résultat en fonction de telle donnée d'entrée) [231]. La demande d'interprétabilité des modèles est en plein essor, notamment dû aux besoins légaux imposés dans certains secteurs comme la défense, la santé ou la finance [237]. En Smart City, le sujet de l'interprétabilité a également son importance pour les citoyens qui demande plus de transparence vis-a-vis des actions menées par le gouvernement de la ville. Le fait d'être capable d'expliquer le raisonnement qui a mené à une décision permet aussi une meilleure confiance dans les systèmes virtuels de la part des êtres humains, ce qui peut faciliter l'adoption d'une technologie par la population. Toutefois, l'interprétabilité des modèles n'est pas encore un sujet tout à fait résolu par les procédés existants comme le montre [237].

Enfin, la pertinence du modèle est une exigence fonctionnelle également peu vérifiée par les concepteurs de modèles d'apprentissage automatique [231]. Elle mesure l'adéquation entre la complexité d'un modèle et ses données d'entraînement afin de s'assurer que le modèle produit n'est pas trop complexe par rapport à celles-ci. Faute de quoi, le système d'apprentissage final serait plus susceptible d'être sujet au sur-apprentissage car il ne parviendrait pas à généraliser sur des données futures et pourrait manquer de robustesse. La question du sous-apprentissage et du sur-apprentissage est un point central en apprentissage automatique car il survient rapidement, surtout lorsque le nombre de données en entrée est limité, c'est pourquoi plusieurs mesures permettent de s'assurer qu'un modèle en est exempt. Il existe des mesures spécifiques qui s'assurent de la pertinence d'un modèle comme le *Perturbed Model Validation* (PMV) ou diverses méthodes pour détecter un sur-apprentissage ou un manque de robustesse puisque ces deux problématiques sont sous-jacentes à la pertinence du modèle [231].

4.3.2/ Stratégies d'évaluation des modèles

Des parallèles peuvent être établis entre le génie logiciel et la conception d'un système d'apprentissage automatique, en particulier sur l'étape d'évaluation de ceux-ci [238, 239]. Bien que leur conception diffère, les logiciels étant entièrement programmés pour une tâche spécifique, tandis qu'en apprentissage automatique, l'objectif est justement que les machines apprennent sans être explicitement programmées.

Les systèmes d'apprentissage automatique sont généralement construits autour de

quatre phases principales : la compréhension métier, l'acquisition des données et leur pré-traitement, la modélisation, l'évaluation et le déploiement du modèle. En règle générale c'est l'étape d'évaluation qui conditionne le déploiement final d'un système d'apprentissage automatique.

4.3.2.1/ Le processus de conception d'un modèle

Avant de mettre en place un processus d'évaluation, il faut comprendre le rôle qu'il joue dans le cycle d'élaboration globale d'un système d'apprentissage automatique. Il existe plusieurs méthodologies de conception d'un tel système dont les premières, à l'instar de *Knowledge Discovery Databases* (KDD) ont été mis en point dans les années 1990 [240,241]. Dans cette approche, les étapes ce suivent les unes à la suite des autres d'une manière linéaire, similaire à celle du modèle Waterfall [242], développé pour le génie logiciel dans les années 1970. Toutefois, contrairement à Waterfall, le concepteur peut être amené à opérer des modifications sur les étapes précédentes en fonction des résultats obtenus à l'étape finale. Le cadre proposé par KDD positionne l'étape d'évaluation à la fin du processus avec celle d'interprétation des résultats, lui conférant ainsi le rôle de contrôler du modèle. L'ensemble des étapes précédentes peut ainsi être ajusté en se basant sur l'évaluation réalisée, qui révèle les limites du modèle et fournit des indications pour d'éventuelles améliorations. Par la suite, des entreprises (comme le consortium formé par NCR, SPSS et Daimler-Benz ou Microsoft) ont fournit d'autres cadres de conception, respectivement nommés *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [243] et *Team Data Science Process* (TDSP) [244], qui mettent en avant les interdépendances entre chaque phase du cycle, autorisant des ajustements du modèle plus fréquents. En effet, il n'est pas nécessaire d'attendre la phase finale, intégrant l'évaluation, pour mettre à jour les étapes précédentes.

La figure 4.3 représente le cadre de conception le plus usité à l'heure actuelle, CRISP-DM, dans les projets d'apprentissage des données [245]. Cette méthode favorise une approche itérative, ce qui signifie que les étapes peuvent être révisées au besoin pour améliorer les résultats. Dans CRISP-DM, l'étape d'évaluation joue un rôle prépondérant puisque c'est elle qui conditionne la phase finale, celle de déploiement du modèle. C'est grâce au déploiement que le modèle pourra être utilisé par le client final, il faut donc que l'évaluation s'assure que le modèle soit en capacité de répondre correctement aux attentes du client, ce qui est un critère de qualité important comme vu précédemment. Il convient également de noter que la compréhension métier du cas d'application est une étape clé dans le processus, elle est présente au démarrage du projet mais également si l'évaluation effectuée montre des faiblesses dans le modèle. La bonne prise en compte du contexte est un élément important de la réussite d'un projet d'apprentissage automatique.

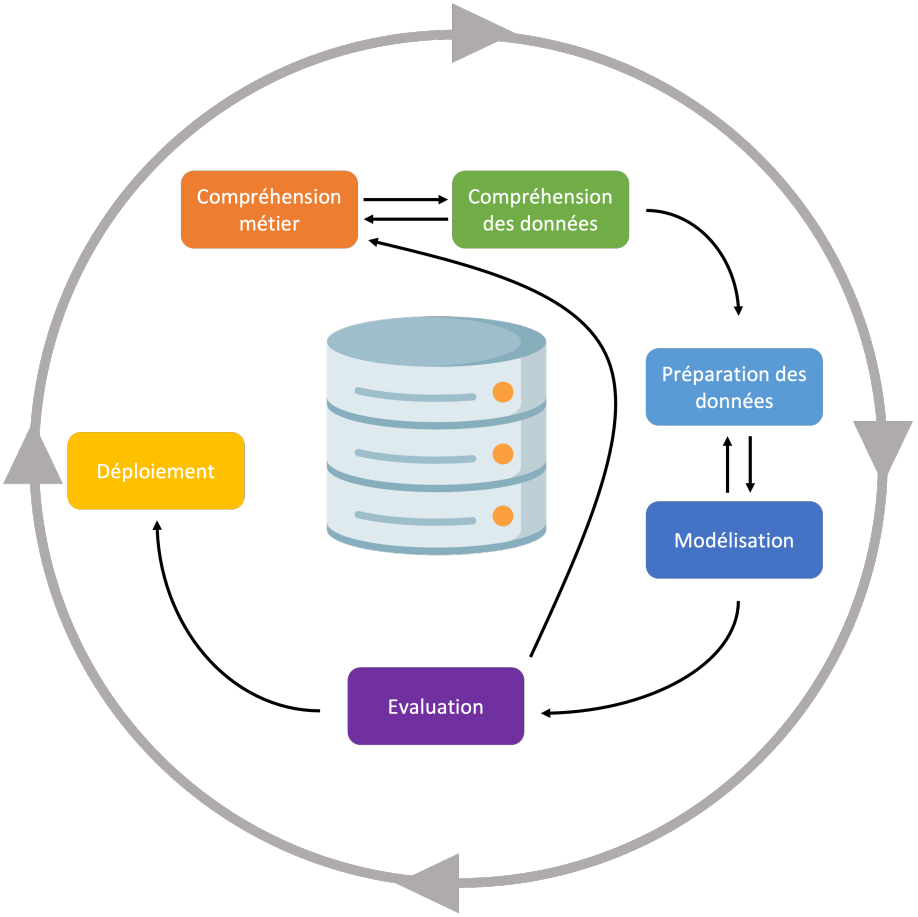

FIGURE 4.3 – Processus normalisé interprofessionnel pour l'exploration de données (CRISP-DM) en apprentissage automatique d'après [243]

Tout comme le principe du *test-driven development* (TDD) [11] est fondamental pour s'assurer de la qualité d'un logiciel et limiter les erreurs [246], l'évaluation des modèles doit également faire l'objet d'une réflexion approfondie en amont et s'adapter au contexte de chaque cas d'application.

4.3.2.2/ LES ÉTAPES D'ÉVALUATION D'UN MODÈLE

Le processus d'évaluation d'un modèle lorsqu'il est mis en place dans son intégralité se décompose en cinq étapes distinctes [247], présentées sur la Figure 4.4.

La première étape est de déterminer ce qui doit être évalué pour considérer que le modèle est qualitatif et qu'il répondra bien aux besoins initiaux. Il peut s'agit de vérifier l'exactitude, comme c'est souvent le cas en priorité, par le biais de plusieurs métriques décrites dans la section suivante. Mais également la robustesse, qui permet entre autre de vérifier que le modèle n'est pas trop sensible aux données bruitées ou aberrantes. Ainsi que

11. Processus de développement d'un logiciel piloté par les tests pour s'assurer que ce dernier respecte l'ensemble des spécifications fournies par le cahier des charges.

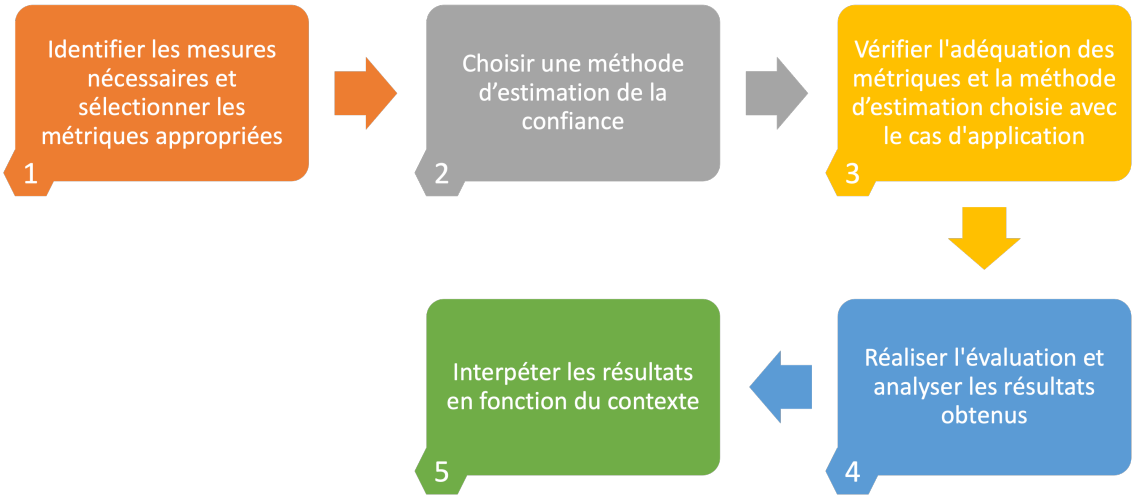

FIGURE 4.4 – Procédure d'évaluation d'un algorithme d'apprentissage automatique

toutes les autres exigences fonctionnelles et non fonctionnelles citées précédemment. Cette étape est cruciale, elle conditionne la conception du modèle en lui-même puisque c'est au regard des résultats obtenus via les métriques appliquées que le modèle est validé ou ajusté. Le choix d'un modèle est gouverné par cette évaluation. En effet, si le concepteur choisi d'ajouter une métrique qui mesure l'explicabilité du modèle, il fera en sorte que l'algorithme d'apprentissage automatique créé soit bien explicable. Et si ce n'est pas le cas au départ, alors il modifiera son système jusqu'à ce qu'il soit de plus en plus explicable. Finalement, le choix des critères d'évaluation va guider la conception du modèle, il est par conséquent indispensable de les définir avec soin.

La deuxième étape consiste à sélectionner une méthode pour estimer la confiance à accorder aux résultats des métriques déterminées à l'étape précédente. Ce type de méthode, qualifié d'estimation de l'incertitude, permet de mesurer le degré de confiance associé aux prédictions émises par le modèle. C'est une mesure de plus pour évaluer la fiabilité du système d'apprentissage. Cette estimation peut se faire grâce à un estimateur ponctuel, comme l'erreur quadratique moyenne (MSE) qui est largement répandu en analyse des données [12], ou bien grâce au calcul d'un intervalle de confiance. L'intervalle de confiance donne une estimation de la plage de valeurs (entre minimum et maximum) qui peuvent être prises par les résultats du modèle pour un niveau de confiance donné. L'estimation de la confiance permet de relativiser les résultats obtenus par les différentes

12. cette métrique est d'ailleurs plus amplement détaillée dans la section suivante

métriques d'évaluation.

La troisième étape permet de s'assurer que la méthode d'évaluation créée au cours des deux étapes précédentes est bien adaptée au contexte d'application. Cela permet de se rapprocher de la compréhension métier du cas étudié et de vérifier que l'évaluation faite corresponds bien à la problématique initiale. C'est une question importante car en fonction de la nature du modèle (classifieur ou régresseur par exemple), mais aussi des données d'entrées ou des besoins des clients finaux. En effet, l'équité et l'interprétabilité sont de besoins fondamentaux pour une application de santé, ils ne doivent pas être négligés. En Smart City, tout dépend du cas d'usage précis mais dès lors que les données d'entrée concernent de près les citoyens, il faut veiller à ce que ces deux exigences soit également évaluées.

La quatrième étape est simplement la mise en oeuvre effective de la méthode d'évaluation définie par le concepteur avec l'analyse des résultats obtenus.

La cinquième et dernière étape ressemble beaucoup à la précédente si ce n'est que l'analyse des résultats se fera au regard du contexte qui concerne le cas d'application. Tout comme l'étape numéro trois, elle a pour but de contextualiser l'évaluation par rapport au problème que doit résoudre le modèle. En effet, l'interprétation des résultats peut-être tout à fait différente pour une application de gestion des déchets d'une ville ou pour la conduite d'un véhicule autonome.

Ces deux dernières étapes vont définir si le modèle réalisé est conforme à ce qui est attendu ou non. Dans le cas où les résultats ne sont pas satisfaisant, le modèle pourra ensuite être corrigé pour obtenir de meilleurs performances. L'interprétation des résultats conditionne la réussite d'un projet d'apprentissage automatique. Une mauvaise interprétation peut conduire au déploiement d'un modèle qui ne satisfait pas les besoins des utilisateurs finaux ou pire, devient une source de biais dommageable pour une partie de la population. Quant aux trois premières étapes, elles ont tout intérêt à être longuement réfléchies dès le début de la conception du modèle pour assurer un meilleur processus de conception de celui-ci.

4.3.3/ Limites dans l'évaluation des modèles sans cohérence

Les cinq étapes d'évaluation des modèles d'apprentissage est un canevas peu respecté, la première étape étant souvent faite à la hâte, tandis que les deuxième, troisième et cinquième étapes sont tout simplement absentes [247]. En définitive, seul la mise en oeuvre effective de l'évaluation subsiste, et elle consiste dans la plupart des cas à vérifier la précision du modèle en calculant la différence entre les données connues et les résultats produits, ce qui est réducteur.

Les évaluations proposées sont spécifiques à la tâche effectuée par le modèle [13] et se décomposent en trois types : l'évaluation par discrimination humaine, l'évaluation grâce à des benchmarks (ou métriques types) et l'évaluation par confrontation des pairs [238]. Cette dernière est plus répandue dans l'évaluation de modèles d'apprentissage par renforcement, en particulier pour la mise au point d'IA dédiée aux jeux vidéos par exemple, c'est pourquoi nous ne développerons pas cet aspect dans les lignes suivantes. La discrimination humaine permet de vérifier que le modèle prédit des résultats proches voir équivalent à ceux d'un humain. En pratique ce type d'évaluation n'est pas des plus courant, bien que le test de Turing (ou assimilé) soit encore parfois utilisé notamment pour les chatbot [14]. L'autre principe d'évaluation est d'utiliser un benchmark ou tout simplement une collection de métriques standards. Le principe du benchmark est de proposer un test d'évaluation standard pour un problème donné (et donc une tâche bien spécifique). C'est ainsi que les grands modèles de langage on leur propre benchmark comme MMLU [248] ou ARC [249], et qu'il existe des benchmark spécifiques pour l'évaluation des véhicules autonomes comme CAVBench [250]. Les compétitions organisées par différents organismes comme Kaggle [15] ou la DARPA [16] font également office de benchmark puisque les critères d'évaluation des modèles sont définis en amont par les organisateurs de la compétition. Bien que l'évaluation grâce aux benchmark soit de plus en plus répandue, ils ne sont toutefois pas exempts d'inconvénients.

L'un d'entre eux rejoint le principal reproche qui peut être fait aux méthodes d'évaluation actuelles, qui se focalisent souvent sur un ensemble limité de comportements prédictifs [17], sans tenir compte du contexte global [251]. Or, il est crucial de prendre en considération le contexte lorsque les modèles sont utilisés dans des applications concrètes du monde réel. Cette différence de point de vue est illustrée par deux paradigmes d'évaluation décrits par [251] : l'évaluation *application-centric* et l'évaluation *learner-centric*. Le principe d'une évaluation *application-centric* est de prendre en compte le contexte de l'application pour s'assurer que le modèle y sera performant et que c'est bien le meilleur qui puisse être utilisé dans ce cas. C'est également une manière plus solide d'examiner la fiabilité des modèles d'apprentissage automatique. L'évaluation *learner-centric*, a contrario, est réalisée uniquement sur les performances de l'algorithme d'apprentissage en lui-même, elle est donc indépendante du contexte. Bien qu'elle soit limitée, c'est le type d'évaluation le plus communément utilisée à l'heure actuelle dans un système d'apprentissage. Cette indifférence d'évaluation par rapport au contexte néglige plusieurs aspects comme les aspects sociaux ou éthiques.

13. en anglais le terme utilisé est *task-oriented*
14. à noter que le mécanisme du CAPTCHA se sert de ce principe pour vérifier que l'opérateur n'est pas une IA justement
15. https://www.kaggle.com/competitions
16. https://web.archive.org/web/20120803132113/http://archive.darpa.mil/grandchallenge/index.asp
17. la plupart du temps lié à une vérification de la précision ou de la robustesse du modèle

De plus, le choix des métriques utilisées, et ce même dans le cas des benchmarks ou des compétitions, est souvent limité à moins de cinq ce qui est peu. Il faut se rappeler la loi de Goodhart : “When a measure becomes a target, it ceases to be a good measure” [252] pour comprendre qu'un trop petit nombre de métriques d'évaluation du modèle n'est pas quelque chose de souhaitable. En effet, les modèles sont entraînés pour performer au mieux sur cet ensemble limité de mesures, il y a donc un grand risque que ces dernières deviennent des objectifs à atteindre pour le modèle.

L'ensemble de ces limites est illustré par le fait qu'actuellement, la majorité des travaux menés en apprentissage automatique évalue la cohérence entre les résultats du modèle et les données d'entrée, sans prendre en compte la cohérence avec le contexte de l'application qui peut être formalisé sous forme de connaissance.

4.3.4/ Évaluation des différents types de cohérence

Les protocoles d'évaluation actuels se portent en général sur la performance de prédiction d'un modèle par rapport aux données d'entraînement et sa capacité à généraliser, i.e. à produire des résultats corrects sur des données jusque là inconnues.

Toutefois, obtenir un modèle fiable et efficace ne peut se résumer à la recherche d'une simple cohérence des résultats avec les données d'entraînement. Il est important d'élargir les protocoles d'évaluation pour s'assurer que le modèle produit soit également cohérent avec des connaissances spécifiques à son domaine d'application.

4.3.4.1/ La cohérence avec les données d'entraînement

La définition de la cohérence donnée par Mitchell [33] étant à ce jour encore prépondérante dans les travaux d'apprentissage automatique, il est normal que la cohérence avec les données d'entraînement soit le type d'évaluation le plus répandu et bien souvent le seul utilisé. Toutefois, il existe de nombreuses métriques capables d'évaluer ce type de cohérence sous des différents angles en fonction de chaque cas d'application.

Le concepteur utilise une métrique, voir bien souvent une combinaison de métriques, dans le but d'améliorer les modèles qu'il produit [18]. Lorsqu'il s'agit d'apprentissage supervisé, les métriques indiquent en majorité le taux d'erreurs du modèle par rapport au données d'entraînement ainsi que sa capacité à généraliser. Les métriques indiquant la capacité de généralisation d'un modèle s'obtiennent en dissimulant une petite partie des données d'entraînement (appelé jeu de données test ou jeu de données d'exclusion) lors de l'apprentissage de l'algorithme. Ensuite, il convient de vérifier si les métriques

18. L'évaluation permettant d'orienter les choix d'architectures utilisées et d'affiner les hyperparamètres des algorithmes

obtenues sur les résultats des données d'entraînement ne diffèrent pas trop des données d'exclusion. Lorsqu'il s'agit d'apprentissage non supervisé, les métriques évaluent l'aptitude du modèle à séparer les données en clusters ayant une forte variance inter-classes [19] et une faible variance intra-classes [20]. L'apprentissage non supervisé n'étant pas l'objet principal de cette thèse, ces métriques ne seront pas développées dans les paragraphes suivants.

L'efficacité d'un classifieur, i.e. un modèle dont la fonction est d'attribuer des étiquettes ou des catégories à des données d'entrée, peut être mesurée par différentes métriques. Bien souvent, l'évaluateur commence par calculer la matrice de confusion (voir tableau 4.1) qui permet de détailler les performances de chaque classe, révélant les vrais positifs, les faux positifs, les vrais négatifs et les faux négatifs. Cette matrice peut être adaptée dans le cas de l'évaluation d'un classifieur multiclasses [21].

	Classe réelle vraie	Classe réelle fausse
Classe prédite vraie	Vrai positif (VP)	Faux positif (FP)
Classe prédite fausse	Faux Négatif (FN)	Vrai négatif (VN)

TABLE 4.1 – Matrice de confusion

L'exactitude [22], en tant que métrique fondamentale, mesure la proportion d'échantillons correctement prédits (vrais positifs et vrais négatifs) parmi l'ensemble total des prédictions. Un important taux de faux positifs et de faux négatifs aura pour effet de faire baisser cette métrique. La précision quant à elle, permet d'évaluer le taux d'échantillons positifs réels correctement identifiés (vrais positifs) parmi l'ensemble des échantillons positifs prédit par le modèle (vrais positifs et faux positif). Le rappel [23] mesure la proportion d'échantillons positifs réels correctement identifiés (vrais positifs) parmi tous les échantillons réellement positifs prédit par le modèle (vrais positifs et faux négatifs). Le F1-Score combine ces deux mesures en un score unique qui équilibre la précision et le rappel comme le montre le tableau 4.2.

La spécificité est une mesure intéressante dans le cas de problèmes déséquilibrés [24] puisqu'elle évalue la capacité du modèle à identifier correctement les échantillons négatifs parmi toutes les prédictions. La courbe ROC (*Receiver Operating Characteristic*) et l'aire sous la courbe ROC (AUC-ROC) offrent une vue complète de la capacité discriminatoire du modèle, en traçant le taux de vrais positifs par rapport au taux de faux positifs pour différents seuils de décision. La précision équilibrée (*Balanced Accuracy*) calcule la moyenne des taux de rappel pour chaque classe, en prenant en compte les déséqui-

19. différence entre classes aussi appelée variance expliquée
20. variabilité à l'intérieur des classes aussi appelée variance résiduelle
21. ayant plus de deux classes en sortie
22. en anglais *accuracy*
23. qui est aussi appelé sensibilité, en particulier dans le domaine médical
24. i.e. avec une sur-représentation de certaines classes vis-à-vis des autres

libres de classe, ce qui en fait une métrique idéale pour les problèmes multiclasses. Ces différentes métriques sont également exposées plus en détail dans le tableau 4.2.

Métrique	Formule	Description	Utilisation
Exactitude (*Accuracy*)	$\frac{VP+VN}{VP+FP+VN+FN}$	L'*accuracy* mesure le rapport entre les prédictions correctes et le nombre total d'instances évaluées.	Classification binaire et multi-classe (si adaptation de la formule)
Taux d'erreur (1 - *Accuracy*)	$\frac{VP+VN}{VP+FP+VN+FN}$	Le taux d'erreur évalue la proportion d'échantillons incorrectement classés par un modèle de classification.	Classification binaire et multi-classe (si adaptation de la formule)
Précision (p)	$\frac{VP}{VP+FP}$	La précision évalue la proportion de prédictions positives correctes parmi toutes les prédictions positives faites par un modèle.	Classification binaire et multi-classe (si adaptation de la formule)
Rappel (r)	$\frac{VP}{VP+FN}$	Le rappel évalue la proportion de prédictions positives correctes parmi tous les échantillons réellement positifs dans un ensemble de donnée.	Classification binaire et multi-classe (si adaptation de la formule)
F1-score (*F-Measure*)	$\frac{2\times p\times r}{p+r}$	La F1-Score calcule la moyenne harmonique de la précision et du rappel.	Classification binaire et multi-classe (si adaptation de la formule) Utile si classes déséquilibrées
Sensibilité (S_e)	$\frac{VP}{VP+FN}$	La sensibilité, également appelée taux de vrais positifs (TPR) ou taux de détection, mesure la capacité du modèle à détecter correctement les échantillons positifs réels.	Classification binaire et multi-classe (si adaptation de la formule)
Spécificité (S_p)	$\frac{VN}{VN+FP}$	La spécificité évalue la capacité d'un modèle de classification à identifier correctement les échantillons négatifs réels. Souvent associée à la sensibilité.	Classification binaire et multi-classe (si adaptation de la formule)
Balanced Accuracy (BA)	$\frac{1}{n}\sum_{i=1}^{n} r_i$ n = nombre de classes	La BA est utilisée pour évaluer la capacité d'un modèle de classification à bien performer sur des ensembles de données déséquilibrés, où le nombre d' échantillons dans chaque classe diffère considérablement.	Classification binaire et multi-classe (si adaptation de la formule) Utile si classes déséquilibrées
Aire sous la courbe ROC ($AUC-ROC$)	$\int_0^1 h(t_1)\,dt_1$ $t_1 : 1 - S_p$ $t_2 : S_e$ $h : t_1 \rightarrow t_2$	L'AUC-ROC, ou aire sous la courbe ROC (Receiver Operating Characteristic), quantifie la capacité d'un modèle à discriminer entre les classes positives et négatives en variant le seuil de décision.	Classification binaire et multi-classe (si adaptation de la formule)

TABLE 4.2 – Principales métriques d'évaluation des classifieurs

Les performances d'un régresseur, i.e. un modèle capable de prédire des valeurs numériques continues en fonction de données d'entrée, sont estimées à partir d'autres métriques. Cependant, l'objectif reste similaire : connaître l'aptitude de ces modèles à capturer les relations entre les variables et à généraliser sur de nouvelles données.

Le coefficient de détermination (R^2), largement répandu, évalue la proportion de la variance totale des données expliquée par le modèle. Ce qui revient à mesurer à quel point les valeurs prédites par le modèle s'approchent des valeurs réelles. Un R^2 élevé indique que le modèle explique parfaitement la variance des données, donc toutes les prédictions correspondent exactement aux valeurs réelles. Tandis qu'un R^2 proche de 0 indique une incapacité du modèle à expliquer la variance des données, les résultats produits sont par conséquent aléatoires.

L'erreur quadratique moyenne (MSE [25]), quantifie la moyenne des carrés des différences entre les valeurs prédites et les valeurs réelles. Cette mesure reflète la distance moyenne entre les prédictions du modèle et les données observées, fournissant une indication précise de l'ampleur des erreurs de prédiction. L'erreur quadratique moyenne racine (RMSE [26]) est ensuite obtenue en prenant la racine carrée du MSE. Le RMSE est généralement préféré au MSE pour l'évaluation des modèles de régression, car il a l'avantage d'être exprimé dans la même unité que la variable cible ce qui facilite son interprétation.

Une autre métrique importante est l'erreur absolue moyenne (MAE [27]), qui mesure la moyenne des valeurs absolues des différences entre les valeurs prédites et les valeurs réelles. Contrairement au RMSE, le MAE accorde un poids uniforme à toutes les erreurs, il est par conséquent moins sensible aux valeurs aberrantes. Là encore, une variation permet d'obtenir une autre métrique : l'erreur absolue moyenne en pourcentage (MAPE [28]) qui donne une indication de la précision du modèle en termes de pourcentage d'erreur par rapport aux valeurs réelles. Facilement interprétable - plus le MAPE est bas, meilleure est la performance du modèle - cette métrique permet de mieux comprendre à quel point les prédictions du modèle s'écartent en moyenne des valeurs réelles. Néanmoins, le calcul du MAPE peut être difficile lorsque les valeurs réelles sont négatives ou proches de zéro car cela peut occasionner une division par zéro ou une déformation des pourcentages d'erreur.

L'erreur moyenne logarithmique (MSLE [29]) calcule la moyenne des carrés des écarts entre les logarithmes des valeurs prédites et les logarithmes des valeurs réelles. Son utilisation est particulièrement bénéfique dans le cas où les valeurs de la variable cible ont une grande amplitude et peuvent varier sur plusieurs ordres de grandeur. Similaire au MAPE, cette métrique est plus difficile à interpréter. En effet, une diminution du MSLE ne signifie pas nécessairement que le modèle prédit mieux les valeurs réelles en termes d'erreurs brutes, mais plutôt en termes d'erreurs relatives sur l'échelle logarithmique. De plus, de part l'utilisation du logarithme dans sa formule, elle ne peut être appliquée qu'à des valeurs strictement positives.

Le tableau 4.3 présente plus en détails quelques métriques utilisées dans l'évaluation habituelle des régresseurs.

Ces métriques sont souvent adaptées à un type d'apprentissage en particulier - les classifieurs et les régresseurs étant évalués différemment- mais elles ne sont pas spécifiques au type de données d'entrée et encore moins au domaine d'application. Par exemple, une matrice de confusion (voir tableau 4.1) associée à des mesures de rappel, de précision

25. *Mean Square Error*
26. *Root Mean Square Error*
27. *Mean Absolute Error*
28. *Mean Absolute Percentage Error*
29. *Mean Squared Logarithmic Error*

Métrique	Formule	Description
Coefficient de détermination (R^2)	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$ y valeur connue de y $\hat{y}$ valeur prédite de y $\bar{y}$ valeur moyenne de y	Le coefficient de détermination est une métrique utilisée pour évaluer la qualité de l'ajustement d'un modèle de régression aux données observées.
Erreur quadratique moyenne (MSE)	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Le MSE mesure la moyenne des carrés des différences entre les valeurs prédites et les valeurs réelles, ce qui permet de quantifier l'ampleur des erreurs de prédiction.
Erreur quadratique moyenne des logarithmes ($MSLE$)	$\frac{1}{n}\sum_{i=1}^{n}(\log(1 + y_i) - \log(1 + \hat{y}_i))^2$	Le MSLE mesure la moyenne des carrés des différences entre les logarithmes des valeurs prédites et des valeurs réelles.
Erreur absolue moyenne (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	Le MAE mesure la moyenne des valeurs absolues des différences entre les valeurs prédites par le modèle et les valeurs réelles observées.
Erreur absolue moyenne en pourcentage ($MAPE$)	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i - \hat{y}_i}{y_i}\right\rvert$	Le MAPE mesure la moyenne des pourcentages absolus des erreurs de prédiction par rapport aux valeurs réelles observées dans un ensemble de données.
Racine Carrée de l'Erreur Quadratique Moyenne ($RMSE$)	$\sqrt{\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{n}}$	Le RMSE est une version modifiée du MSE, où les erreurs sont élevées au carré, puis la racine carrée est prise pour ramener la métrique à la même unité que la variable cible.

TABLE 4.3 – Principales métriques d'évaluation des modèles de régression

et un F1-score sont mis en place à la fois pour évaluer un modèle de prédiction de faux billets (dont les données en entrée sont des valeurs numériques) mais également pour évaluer un modèle capable de classer des images de chiens et de chats. Dans le cas d'un problème de régression comme la prédiction de consommation d'énergie de bâtiments ou bien d'îlots de chaleur [30] le RMSE est en capacité d'évaluer les performances de chaque modèle alors que leur finalité diffère.

L'évaluation des modèles requiert une approche holistique, ainsi il est nécessaire de combiner plusieurs métriques d'évaluation pour s'assurer d'avoir un modèle performant et d'éviter le piège illustré par la loi de Goodhart (cf. Section 4.3). Si les concepteurs ne se fient qu'au coefficient de détermination ou à la précision (accuracy) cela peut conduire à des modèles défectueux, tels qu'illustré par le Quartet d'Anscombe [253], incapables de généraliser sur de nouvelles données.

Toutefois, même en utilisant un ensemble varié des métriques énoncées ci-dessus, il n'est pas certain que le modèle réalisé soit cohérent avec son contexte d'utilisation. En effet, ces métriques permettent seulement de mesurer la cohérence des résultats avec les données d'entraînement mais aucune d'entre elles ne vérifie la cohérence entre les résultats et des connaissances préalables (cf. définition de Yu et al. [34]) capables de prendre en compte l'environnement de l'application.

30. prédiction des différences de température sur tout le territoire d'une ville

4.3.4.2/ La cohérence avec les connaissances

Dans l'état actuel de nos connaissances, il n'existe pas à ce jour, de métrique "universelle" permettant d'évaluer la cohérence entre les résultats d'un modèle et des connaissances préalables. Cette métrique n'existera d'ailleurs probablement pas un jour puisque l'évaluation de cette cohérence dépend très fortement du type de problème, des données d'entrée et même du type de cohérence que l'on veut évaluer. Il existe de nombreux types de cohérence qu'il est possible d'évaluer : la cohérence logique, la cohérence sémantique, la cohérence factuelle, la cohérence physique, la cohérence temporelle, la cohérence spatiale, la cohérence hiérarchique ou encore la cohérence entre pairs. Pour chaque nouvelle application il est important de déterminer au départ la et les types de cohérence qu'il convient de respecter. En effet, d'après la définition de la cohérence faite par [34], les systèmes d'intelligence artificielle doivent être capables de produire des résultats cohérents avec les connaissances initiales. La nature des connaissances diffère d'un problème à l'autre, par conséquent les règles [31] qui en découlent ne sont pas toutes du même acabit.

L'évaluation de la cohérence linguistique est un sujet fondamental dans l'évaluation des grands modèles linguistiques (LLM [32]) puisque cela permet de vérifier la capacité du modèle à générer du texte qui ne se contredit pas [41]. De même, vérifier la cohérence linguistique d'un classifieur de textes permet entre autres de ne pas avoir des résultats complètement différents pour deux textes avec un sens très similaire. Ce type de cohérence en regroupe en réalité six autres : la cohérence négative, la cohérence symétrique, la cohérence transitive, la cohérence additive, la cohérence sémantique et la cohérence factuelle. Le but est de garder une certaine homogénéité dans le texte pour favoriser sa vraisemblance (voir sa véracité dans certains cas).

En logique, la cohérence négative correspond à la règle si p est vrai $\Leftrightarrow \neg p$ est faux [254]. Plus précisément, la cohérence négative signifie que si une proposition p est vraie, alors sa négation $\neg p$ doit être fausse, et inversement, si une proposition p est fausse, alors sa négation $\neg p$ doit être vraie. Il ne peut y avoir de situation où une proposition et sa négation soient toutes les deux vraies ou fausses simultanément. En cohérence linguistique, elle vérifie que le texte n'affirme pas tout et son contraire ; ainsi si l'on écrit "tous les oiseaux peuvent voler" on ne peut pas écrire que "les hirondelles ne savent pas voler" plus tard dans le texte [255].

La cohérence symétrique se réfère à un concept de logique et de relation entre les éléments, dans lequel une fonction ou une relation entre deux variables reste la même lorsque les positions de ces variables sont échangées. Elle correspond à la règle $f(x, y) = f(y, x)$ [41], cela signifie que les résultats ou les prédictions d'une fonction

31. qu'on peut également appeler contraintes
32. pour *Large Language Model*

ou d'un modèle doivent être les mêmes, peu importe l'ordre dans lequel les variables sont présentées. En linguistique, elle permet de reconnaître des propos identiques peut importe l'ordre d'apparition des termes. Ainsi, "par un groupe d'insurgés dans les Philippines" et "dans les Philippines par un groupe d'insurgés" sont vues comme des propositions équivalentes en termes de signification [256]. En somme, la cohérence symétrique assure que les relations ou les prédictions restent constantes et ne changent pas en fonction de l'ordre des éléments impliqués.

La cohérence transitive est un concept qui se réfère à la relation logique entre trois éléments ou propositions. Si le premier élément est en relation avec le deuxième, et le deuxième élément est en relation avec le troisième, alors il en découle que le premier élément est également en relation avec le troisième élément. En d'autres termes, en considérant les trois déclarations X, Y et Z, lorsqu'on sait que X entraîne Y et que Y entraîne Z, il faut en déduire que X entraîne Z. Cette proposition est illustrée par la formule $(X \rightarrow Y) \wedge (Y \rightarrow Z)$ alors $X \rightarrow Z$ [41]. Si un LLM prédit qu'un dauphin est un cétacé et qu'un cétacé est un mammifère, alors il doit aussi prédire qu'un dauphin est un mammifère [257].

La cohérence additive, telle que décrite par [41], stipule que si une fonction attribue la même valeur à deux variables distinctes, alors elle attribuera également la même valeur à la somme de ces variables. Cela signifie que si une fonction attribue la même valeur c à deux variables indépendantes x et y (c'est-à-dire $f(x) = f(y) = c$), alors cette fonction attribuera également la même valeur c à la somme des variables ($f(x + y) = c$), lorsque c est une étiquette prédite. La cohérence additive garantit que la prédiction d'une fonction reste stable lorsque des valeurs indépendantes sont combinées, pour maintenir la cohérence logique des prédictions.

Les cohérences négatives, symétriques, transitives et additives sont en réalité des cohérences relatives au domaine de la logique. Elles impliquent l'utilisation et la manipulation d'opérateurs logiques pour établir des relations logiques correctes entre différentes propositions, déclarations ou prémisses. Les opérateurs logiques tels que "et" (conjonction), "ou" (disjonction), "non" (négation), "si...alors" (implication), etc., sont utilisés pour exprimer et analyser les relations logiques entre les éléments d'un raisonnement. En utilisant ces opérateurs logiques, il est possible de construire des expressions logiques qui représentent différentes connaissances souvent identifiées comme étant des règles métier. La cohérence logique est atteinte lorsque les résultats d'un modèle ne sont pas en contradiction avec l'ensemble de ces règles définies au préalable. Elles ne s'appliquent donc pas uniquement au traitement du langage mais sont mobilisables dans de nombreux autres domaines.

La cohérence logique en robotique fait référence à la capacité d'un robot ou d'un système robotique à maintenir des relations logiques cohérentes entre les informations perçues

de son environnement et les croyances qu'il forme à partir de ces informations. Si le robot perçoit des informations qui semblent contradictoires ou en désaccord avec son modèle interne du monde, il devrait être capable de détecter ces incohérences et de réviser ses croyances ou sa perception pour parvenir à une compréhension cohérente et logique de son environnement [258]. Dans le domaine ferroviaire, de nombreuses normes doivent être respectées lors de la sélection d'un itinéraire pour garantir la sécurité des êtres humains, du matériel ainsi que de l'environnement. La cohérence logique peut-être mobilisée sous forme de règles métier appliquées au cours d'un processus de vérification de différents scénarios comme réalisé par [259].

Un modèle est sémantiquement cohérent s'il est capable de comprendre et d'analyser le sens global d'une phrase ou d'un paragraphe, indépendamment des variations dans la manière dont ce sens est exprimé en utilisant différents mots ou structures linguistiques [41]. La cohérence sémantique est cruciale pour les tâches de traitement du langage naturel, car elle garantit que le modèle peut généraliser son apprentissage au-delà des exemples spécifiques qu'il a rencontrés lors de son entraînement. Cela permet au modèle de fournir des réponses cohérentes et appropriées même lorsque les formulations des requêtes ou des énoncés sont modifiées tout en conservant le même sens. Pour résumer, la cohérence sémantique reflète la capacité d'un système de traitement du langage naturel à capturer le sens profond des textes et à produire des résultats cohérents pour des contenus équivalents, même si les expressions linguistiques diffèrent.

La cohérence sémantique dans le domaine de la vision par ordinateur à une signification quelque peu différente de celle donnée précédemment. En effet, la cohérence sémantique des images fait référence à la propriété selon laquelle les éléments visuels, tels que les segments ou parties spécifiques d'une image, ont une signification et une apparence similaires à travers différentes instances d'objets, même lorsque ces objets présentent des variations d'apparence et de posture [260]. Considérant différentes images d'oiseaux, la cohérence sémantique garantie que les ailes, le bec ou toute autre partie auront une apparence similaire et conserveront leur rôle d'une image à l'autre, même si les oiseaux sont dans des poses différentes ou que leur apparence varie [261]. La vérification de cette propriété facilite et améliore les travaux de reconstitution d'image en 3D comme présenté par [261, 262].

Un modèle est considéré comme factuellement cohérent s'il est capable de produire des informations exactes et précises, en évitant toute contradiction avec les connaissances généralement acceptées et le contexte donné [41, 263]. La cohérence factuelle est particulièrement pertinente pour les tâches de génération de langage, tels que le résumé de texte, les réponses aux questions et la génération de dialogues, où les modèles doivent produire des contenus qui reflètent correctement la réalité et ne contredisent pas les informations bien établies. La cohérence factuelle est cruciale pour éviter la propagation

d'informations incorrectes ou trompeuses et pour assurer que les utilisateurs reçoivent des réponses précises et pertinentes de la part des systèmes de traitement du langage naturel.

La cohérence de mouvement [33] fait référence à la propriété d'une séquence d'images à présenter des mouvements ou des déplacements d'objets cohérents et fluides au cours du temps c'est-à-dire à la cohérence spatiotemporelle. Elle implique que les mouvements observés dans une séquence d'images suivent des modèles réguliers et prévisibles, sans brusques changements ou sauts dans les mouvements des objets [264, 265]. La cohérence spatiotemporelle garantie que les lois du temps et de l'espace sont bien respectées par le modèle et de se fait elle permet d'obtenir des modèles capables de mieux généraliser. La cohérence temporelle n'est pas cantonnée aux problèmes liés aux séquences vidéos mais à l'ensemble des modèles de séries temporelles. Dans le domaine des réseaux sociaux ou des données financières, elle peut se référer aux tendances ou aux comportements qui évoluent de manière cohérente au fil du temps. La cohérence spatiale fait-elle souvent référence à des dynamiques existantes entre objets dans un ou plusieurs plans de l'espace. Un exemple illustrant la cohérence spatiale consiste à établir que si deux individus pénètrent dans une salle qui n'a qu'une seule porte d'accès, ils doivent nécessairement sortir par cette même porte. Si ce n'est pas le cas, le bon sens en déduit qu'ils n'ont pas quitté la pièce et ce même si une seule personne est visible à la caméra [34] [266]. L'analyse de réseaux dynamiques doit également prendre en compte la cohérence spatiotemporelle comme décrit par [267] afin d'améliorer la prédiction de l'évolution des liens d'un réseau au cours du temps.

La cohérence spatiotemporelle peut-être vue comme une sous-catégorie de la cohérence physique qui implique que les événements, les interactions ou les comportements qui se produisent dans un contexte physique sont conformes aux règles et aux principes fondamentaux qui gouvernent le monde physique [36]. Bien sûr, cette forme de cohérence ne se limite pas aux considérations sur l'espace-temps mais bien sûr une multitude de lois physiques connues des scientifiques et parfois révisées au fur et à mesure des dernières découvertes. En hydrologie, la cohérence physique fait référence à la propriété selon laquelle les modèles, les données et les analyses utilisés pour étudier les phénomènes hydrologiques respectent les principes physiques et les lois qui gouvernent le cycle de l'eau et les processus hydrologiques [268]. Les résultats, les prédictions et les interprétations obtenus à partir des modèles et analyses doivent en accord avec notre compréhension des phénomènes physiques réels qui se produisent dans les systèmes hydrologiques.

En science des matériaux, la cohérence physique se réfère à la concordance et à l'harmonie entre les propriétés physiques et les comportements observés d'un matériau avec

33. *motion consistency* en anglais
34. la seconde étant probablement masquée par un élément du mobilier

les principes et les lois physiques qui en régissent la structure [37]. L'analyse de la cohérence physique est fondamentale en science des matériaux car elle permet de comprendre en profondeur les matériaux, d'élaborer de nouvelles formulations, structures et applications basées sur des principes solides de physique. Elle joue un rôle essentiel dans le développement de matériaux innovant avec des propriétés spécifiques pour répondre aux besoins technologiques et industriels.

Dans le domaine de la mécanique, la cohérence physique signifie que les mouvements d'un objet obéissent aux lois du mouvement de Newton. Dans le domaine de la thermodynamique, cela implique que les échanges d'énergie dans un système respectent les principes de conservation de l'énergie [269]. En astronomie, la cohérence physique garantie que les mouvements des planètes et des étoiles sont conformes aux lois de la gravité. Et cette liste pourrait encore s'étendre considérablement s'il fallait énumérer de manière exhaustive les nombreux domaines de la physique où il est crucial d'assurer la cohérence.

La cohérence hiérarchique implique que les systèmes d'apprentissage sont conçus de manière à refléter et à exploiter les relations hiérarchiques présentes dans les données ou dans les tâches à résoudre. Dans le domaine biologique, la cohérence hiérarchique peut être observée dans la manière dont les organismes sont classifiés en taxonomies, en respectant les relations d'ascendance et de parenté comme c'est le cas pour les fonctions des protéines. La prédiction des fonctions associées aux protéines fait l'objet de nombreuses études impliquant l'usage de modèles prédictifs. Or, les prédictions pour être cohérentes doivent respecter la taxonomie représentant les différentes fonctions sans quoi les résultats ne seront pas le reflet de la réalité et ne donneront que des informations partielles [223, 270].

La dernier type de cohérence présenté ici est la cohérence entre pairs qui fait référence à un système où les différents éléments ou parties sont en accord les uns avec les autres, et où les informations partagées ou échangées entre les éléments sont cohérentes et non contradictoires [35]. En apprentissage automatique, la cohérence entre pairs implique que plusieurs modèles, même s'ils peuvent avoir des architectures, des paramètres ou des méthodes d'apprentissage différents, parviennent à des conclusions similaires ou cohérentes pour un ensemble donné de données ou de tâches [271] [36]. Dans certaines situations spécifiques, il est possible de confronter les résultats obtenus avec des prédictions générées par des individus humains, dans le but d'évaluer si les prédictions du

35. Cette définition fait penser à celle du consensus

36. La cohérence entre pairs est souvent recherchée dans des systèmes distribués, des réseaux informatiques, des bases de données partagées et d'autres environnements où plusieurs entités interagissent et partagent des données. Dans ces contextes, elle garantit que chaque élément dispose d'une vue cohérente des informations partagées, ce qui permet d'éviter les erreurs, les incohérences et les conflits. Par exemple, dans un système de base de données distribuée, la cohérence entre pairs signifie que chaque nœud du réseau possède une copie à jour et cohérente des données partagées, de sorte que les opérations de lecture et d'écriture se déroulent sans confusion ni contradictions.

modèle correspondent à ce que des humains auraient produit [272]. Cette forme particulière de comparaison est désignée par le terme "cohérence humaine" (ou *humain consistency* en anglais).

4.4/ Méthodologie d'évaluation de la cohérence

Avant de mettre au point une méthode d'évaluation de la cohérence, nous avons effectué une analyse de la littérature en sélectionnant des travaux de recherche en intelligence artificielle hybride pour déterminer si leurs auteurs utilisaient des bases de connaissances dans leurs évaluations.

4.4.1/ Analyse de l'existant

Une analyse des articles de la catégorie *Informed Machine Learning* provenant de la SLR permet d'illustrer davantage les propos tenus précédemment. Au total, quarante-six études ont été examinées, quarante-quatre utilisaient des classifieurs et deux des régresseurs. La figure 4.5 présentent les différentes métriques d'évaluation utilisées en fonction du type de modèle d'apprentissage.

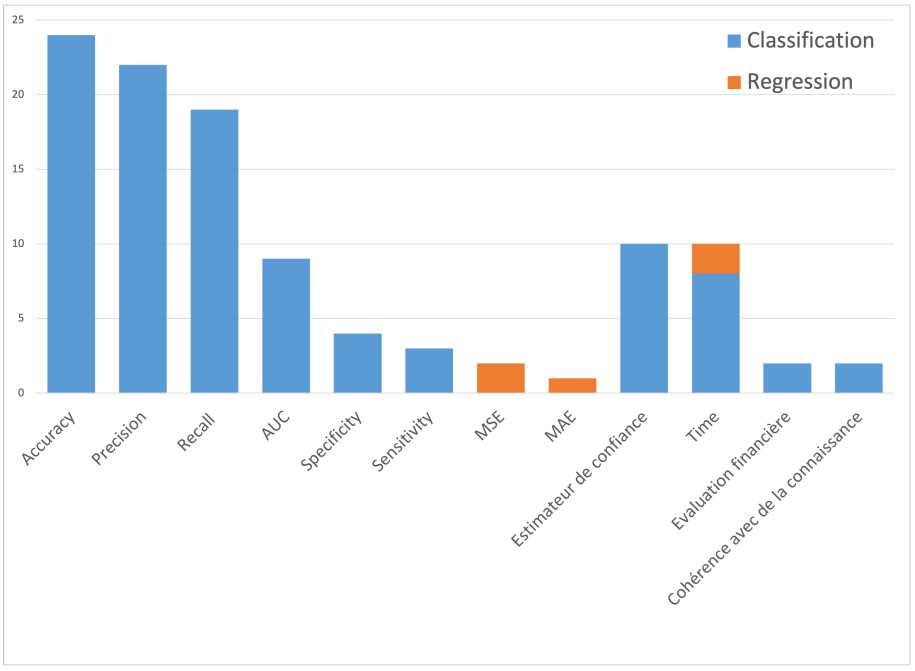

Figure 4.5 – Fréquence d'usage des métriques d'évaluation en fonction du type de modèle d'apprentissage automatique

Nous pouvons remarquer qu'une grande majorité d'études basent leurs évaluations sur l'exactitude des résultats vis-à-vis des données avec lesquelles le modèle a pu être en-

traîné. Les métriques sont sélectionnées en fonction de la nature de la tâche réalisée : classification ou régression. La plupart des articles font mention de plusieurs métriques pour réaliser leur test de validation, ce qui est une bonne chose puisque cela permet d'évaluer un panel de caractéristiques. Comme le rappel [247] dans un contexte de classification supervisée, une seule mesure comme l'exactitude [37] ne permet pas de rendre compte de l'ensemble des défauts d'un modèle. Il faut impérativement utiliser plusieurs métriques pour diminuer les faiblesses de chacunes.

En règle générale, les études suivent un principe quasiment fondamental en science des données : l'évaluation doit se faire sur un jeu de test inconnu du modèle. En effet, afin d'améliorer la généralisation de ce dernier, il est primordial de ne pas l'évaluer sur des données avec lesquelles il a été entraîné. Les concepteurs séparent en général leur ensemble de données d'origine en deux : les données d'entraînement (souvent 70-80% de l'ensemble) et les données de test (souvent 20-30% de l'ensemble). Les données d'entraînement comme leur nom l'indique servent à entraîner le modèle. Les données de test servent uniquement à l'évaluer afin de ne pas biaiser l'examen du modèle. Une autre technique peut s'ajouter pour permettre une généralisation encore meilleure : la cross-validation [273]. Plutôt que de diviser l'ensemble seulement en deux, la validation croisée consiste à diviser les données en plusieurs morceaux de données d'entraînement, appelés "*fold*" [38]. Ensuite, le modèle vas être entraîné et testé à plusieurs reprises. À chaque itération, un *fold* différent sert d'ensemble de test, tandis que les autres *fold* servent d'ensemble d'entraînement. La performance est mesurée à chaque itération. Enfin, pour obtenir une évaluation globale de la performance, il faut faire la moyenne des mesures de performance obtenues lors de ces itérations. Cette technique garantie une évaluation plus fiable de la performance du modèle, car il est testé sur plusieurs sous-ensembles de données lors de son entraînement, en théorie cela permet d'estimer plus précisément sa capacité à généraliser avec de nouvelles données. De plus, une dernière évaluation est souvent réalisée avec le jeu de données de test qui est lui complètement inconnu du modèle. Toutefois, cela ne garantie pas toujours la capacité du modèle à généraliser dans un cas d'utilisation réel. Cela dépend en grande partie des données avec lesquelles il a été entraîné et évalué : elles doivent être très similaires à celles qu'il consommera en réalité.

Cette analyse a également permis de mettre en avant le fait que peu d'études utilisaient des estimateurs de confiance lors des évaluations. Dans le cas d'une régression, l'intervalle de confiance peut éventuellement être calculé à partir du RMSE (e.g. avec ± 2 * RMSE). En revanche, dans le cas d'une classification il est préférable d'étudier l'écart type de certaines métriques comme le F1-score [97]. La confiance d'un modèle peut aussi être estimée vis-à-vis des résultats obtenus par d'autres modèles pour vérifier s'ils

37. *accuracy*
38. en français on peut traduire cela par le mot “plis”

sont bien indépendants les uns des autres. Un test d'indépendance comme l'*ANalysis Of VAriance* (ANOVA) [39] ou McNemar (ou t-test, etc.) permettent d'évaluer si les différents modèles ont bien des résultats significativement différents les uns des autres.

Quelques études s'intéressent également à des mesures plus variées comme le temps d'entraînement d'un modèle [112] ou sa capacité à générer un profit économique [110] ou à limiter le coût de stockage des données [53]. Un seul article présente une étude de la robustesse de son modèle où la variation de caractéristiques est fréquente (les données d'entrée sont des images), dans le cadre de l'évaluation d'un classifieur capable de reconnaître différents types de connecteurs [94]. Deux études qui mesurent la cohérence de leur modèle avec des connaissances (issues d'une ontologie ou d'un *knowledge graph*) sont également présentes [108, 151]. En proportion elles représentent moins de 5% des articles analysés qui mobilisent pourtant tous des connaissances dans la création de leur modèle d'apprentissage automatique. L'évaluation systématique de la cohérence entre les modèles et la *prior knowledge* n'est donc pas encore un standard même dans le domaine des intelligences artificielles hybrides.

Le bilan de cette analyse met en évidence le fait que les mesures d'évaluation, même pour des travaux d'*Informed Machine Learning* ne sont pas très variés. Les métriques utilisées sont bien souvent celles qui vérifient la cohérence entre les résultats et les données d'entraînement. De plus, cette analyse permet de faire des conclusions similaires à celles observées par [247], à savoir que le choix d'un estimateur de confiance est souvent absent et que le contexte est lui aussi souvent ignoré. Il n'est donc pas possible de relativiser correctement les résultats obtenus, encore moins d'assurer que les modèles mis au point sont capables de bien performer sur des données futures. La mise en place d'évaluation contextuelle à pour objectif de ce rapprocher de la réalité et de prendre en considérations différents facteurs d'environnements capables d'influencer le modèle (notamment des contraintes physiques connues).

4.4.2/ Approche d'évaluation des systèmes d'apprentissage automatique informés

L'évaluation d'un système d'apprentissage n'est pas toujours aisée, et s'il existe des cadres méthodologiques relatifs à la construction d'un modèle dans sa globalité, il n'y a aujourd'hui pas de méthode qui soit fréquemment employée pour l'étape d'évaluation. De plus le terme d'évaluation de la cohérence est couramment utilisé en référence à la cohérence entre pairs (une forme de consensus entre les différents modèles) et non pas la cohérence avec des connaissances préalables comme décrit par [36]. La cohérence définit par l'équation 2.1, peut-être vue comme une forme de congruence entre les résultats d'un modèle et les connaissances et/ou lois qui doivent être respectées.

39. en français analyse de la variance

Toutefois, la cohérence n'est pas la seule caractéristique à prendre en compte dans l'évaluation d'un modèle. Karpatne et al. [36] affirment que la performance résulte de la combinaison de l'exactitude, de la simplicité du modèle [40] et de la cohérence. Pour nous, cette définition est réductrice car en fonction du contexte, il peut exister d'autres dimensions à prendre en compte comme la robustesse (résistance face aux données bruitées) ou le respect de la vie privée par exemple.

L'objectif principal de l'évaluation est de garantir la qualité du modèle final, c'est-à-dire un modèle qui soit à la fois performant tout en étant en mesure de combler les attentes de l'utilisateur final. C'est pourquoi le protocole d'évaluation doit entre autres permettre d'évaluer le modèle en fonction du contexte. Pour ce faire nous proposons une méthode d'évaluation en sept étapes présentées par la figure 4.6.

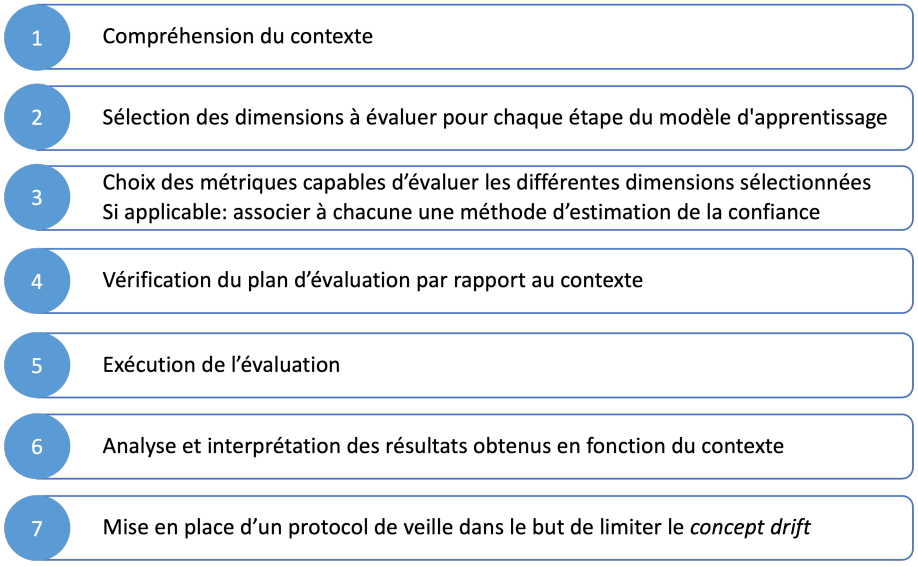

FIGURE 4.6 – Le nouveau protocole d'évaluation mis au point

En premier lieu, une compréhension approfondie du contexte d'application est nécessaire pour appréhender l'environnement étudié, i.e les circonstances, les antécédents et les facteurs pertinents qui influencent une situation donnée. Cette étape permet également de prendre connaissance des différentes contraintes qui s'appliquent au problème, leur étude permettra ensuite de mettre en place un protocole d'évaluation adéquat. Dès lors, la connaissance métier à un rôle à jouer dans la mise en place de l'examen du modèle d'apprentissage automatique.

Par la suite, il faut déterminer quelles parties du processus d'apprentissage doivent être évaluées : les données d'entraînement et/ou l'architecture du modèle et/ou le mécanisme d'apprentissage et/ou les résultats finaux. Dans l'examen des données, il est essentiel de vérifier la qualité des données en contrôlant leur respect des règles métier, mais il est

40. par simplicité les auteurs entendent modèle le moins complexe possible

également possible de s'intéresser à leur anonymisation, à l'équilibre des classes[41] ou la pertinence des caractéristiques. L'évaluation de l'architecture du modèle implique de considérer entre autres sa complexité, l'application de techniques de régularisation et le choix d'une structure adaptée à la tâche. Lors de l'étape d'apprentissage, il est important de sélectionner judicieusement la méthode d'optimisation et dans certains cas de choisir une fonction de perte capable d'assimiler les contraintes que le modèle doit respecter. Enfin, la phase de résultats nécessite l'évaluation de la performance du modèle, sa robustesse face à des données inconnues, la cohérence de ses prédictions dans différentes situations ainsi que toutes les autres dimensions devant être évaluées en fonction du contexte. Cette approche globale de l'évaluation vise à contribuer à la création de modèles d'apprentissage automatique plus fiables et plus efficaces en favorisant l'examen multidimensionnel systématique des modèles d'apprentissage.

Évaluer ces différentes dimensions aux endroits clés nécessite de choisir des métriques adaptées à chacun de ces aspects. Pour évaluer la performance d'un classifieur, des métriques telles que la précision, le rappel, le F1-score et l'AUC-ROC sont couramment utilisées (cf. Section 4.3.4). Toutefois, ces métriques ne sont pas les seules, pour évaluer la cohérence avec les connaissances, un score de cohérence peut également être mis en place par exemple. D'autres mesures peuvent évaluer la robustesse, le niveau d'interprétabilité ou de complexité du modèle, etc. Il est dommage de se limiter uniquement à l'examen de la performance d'un modèle à l'aide de métriques génériques. En outre, il est essentiel d'associer des méthodes d'estimation de la confiance dès lors que cela est possible. L'intervalle de confiance, pour un niveau de confiance spécifique tel que 95% ou 99%, permet de quantifier la précision de l'estimation en indiquant une fourchette de valeurs plausibles autour du résultat du modèle.

La vérification du plan d'évaluation par rapport au contexte consiste à prendre du recul par rapport aux objectifs et à l'environnement du projet pour s'assurer que le plan d'évaluation est aligné avec les attentes et les besoins réels. Cela implique de s'interroger sur les dimensions, les métriques et autres critères de succès choisis, en les évaluant à la lumière des objectifs du projet et des exigences du domaine. Cette démarche permet de garantir que les résultats de l'évaluation seront pertinents et informatifs pour les parties prenantes, tout en évitant des évaluations inutiles ou trop limitées.

La mise en œuvre du plan préalablement défini représente l'étape 5 de méthodologie. C'est une phase bien sûr nécessaire, c'est sur elle que va reposer le reste de l'examen du modèle. Toutefois l'étape suivante est celle qui a le plus d'impact final puisqu'il s'agit d'analyser l'ensemble des résultats obtenus tout en les interprétant en fonction du contexte. Les résultats doivent être interprétés à la lumière des spécificités du domaine, des contraintes opérationnelles ainsi que des préoccupations des parties prenantes.

41. il faut veiller à ce que les données ne soient pas biaisées en particulier lorsqu'elles doivent représenter une population d'être vivants

Cette démarche permet de déterminer si les résultats sont significatifs, si des ajustements sont nécessaires, et si les conclusions peuvent être généralisées au contexte plus large. L'interprétation des résultats permet de mesurer les performances des modèles, de comparer différents modèles sur la base de critères précédemment établis, parfois d'optimiser les hyperparamètres des modèles pour les améliorer, de détecter un manque de généralisation ou de robustesse, de vérifier que les modèles soient cohérents avec des connaissances préalables, etc. En résumé, cette phase permet de garantir la qualité d'un modèle avant son déploiement.

A tout cela s'ajoute une septième et dernière étape, dans le cas où le modèle est déployé en production : la mise en place d'un protocole de veille capable d'alerter si une dérive de concept [42] a lieu. La dérive de concept est une situation où la relation sous-jacente entre les facteurs et la cible des prédictions du modèle évolue au fil du temps, c'est-à-dire qu'au bout d'un certain temps, le modèle n'est plus en capacité de fournir des prédictions pertinentes sur les données qui lui sont fournies [274]. Cet incident est souvent en lien avec la dérive des données [43] phénomène où la distribution des données utilisées pour l'entraînement d'un modèle évolue avec le temps [274]. Lorsque des anomalies sont repérées, une mise à jour du modèle est nécessaire, entrant de facto un nouveau protocole d'évaluation s'adaptant au nouveau contexte.

L'évaluation est un moment important dans la conception d'un modèle d'apprentissage, elle conditionne notamment son déploiement et sa capacité à répondre aux utilisateurs. Pourtant, la question de l'évaluation est souvent reléguée à une analyse presque mécanique de certaines métriques largement utilisées en fonction du type d'algorithme. Certaines d'entre elles sont parfois mobilisées dans la phase d'apprentissage par nécessité (pour les algorithmes requérant une fonction de perte ou une fonction de coût par exemple) mais elles sont rarement exploitées dans les autres phases comme la préparation des données ou le choix de l'architecture. En plus d'adapter l'évaluation au contexte, il est important de s'assurer que l'évaluation réalisée correspond à chaque étape du processus d'apprentissage.

Ainsi, la cohérence avec les connaissances peut être évaluée sur les résultats mais également sur les données d'entrée. Cette première évaluation, avant même d'opérer l'étape d'apprentissage, peut s'avérer fort utile pour qualifier des données issues de capteurs par exemple. Ce type de méthode s'apparente à de l'ingénierie des fonctionnalités bien sûr, mais aborder cette procédure sous l'angle d'une évaluation permet de la systématiser dans le processus d'élaboration des systèmes d'apprentissage automatique. L'évaluation de l'architecture de l'algorithme peut simplement se limiter à vérifier qu'il est bien conçu pour répondre au problème initial (utiliser un algorithme de classification pour réaliser une régression est inapproprié). Cependant, pour des problèmes plus complexes, tels que

42. *concept drift* en anglais
43. *data drift* en anglais

ceux qui exigent le respect de la cohérence hiérarchique, il est tout à fait approprié d'évaluer l'architecture du modèle pour vérifier qu'elle respecte bien cette contrainte [223]. L'évaluation doit servir la construction du modèle d'apprentissage à chaque étape pour s'assurer que les résultats répondront bien aux exigences pour lesquelles il a été conçu. L'évaluation des résultats doit, quant à elle, être systématiquement multidimensionnelle et tenir compte efficacement du contexte. En effet, certaines dimensions telles que le coût d'utilisation du modèle, le temps d'exécution, la capacité à être interprété ou l'équité sont parfois des conditions nécessaires pour le déploiement réussi d'un modèle en production. Il convient donc d'établir un protocole d'évaluation qui soit en mesure de prendre en compte l'intégralité de ces aspects.

Enfin, il est indispensable d'évaluer les modèles en utilisant à la fois des métriques quantitatives et qualitatives. Les métriques quantitatives bien connues fournissent une évaluation objective de la performance, mais il ne faut pour autant pas négliger les métriques qualitatives seules garantes de la satisfaction des utilisateurs. L'équilibre entre ces deux types de métriques garantit la qualité des modèles qui doivent être à la fois efficaces du point de vue des résultats et adaptés aux besoins réels des utilisateurs.

4.5/ Conclusion

Actuellement le protocole d'évaluation des systèmes d'apprentissage automatique n'est pas toujours conçu et exécuté avec la rigueur requise. En plus des mesures de performance habituelles qui visent à vérifier l'adéquation entre les résultats et les données d'entrées, une évaluation de la cohérence entre les résultats et les connaissances préalables est fortement recommandée. D'autres dimensions sont également à prendre en compte en fonction du contexte de l'application, certains domaines étant contraints par des obligations légales ou bien dictées par les lois de la physique. Ainsi, accorder une attention plus rigoureuse à l'étape d'évaluation et de test des modèles, à l'instar de la philosophie employée dans le génie logiciel, va devenir de plus en plus incontournable à mesure que les systèmes d'apprentissage sont déployés en production. L'impact majeur que ces systèmes ont sur les utilisateurs finaux oblige les concepteurs à vérifier que les systèmes mis au point ne comportent pas des biais qui seraient contraires à l'éthique ou qui ne garantirait pas la vie privée [44].

La diversification dans l'évaluation des dimensions est également un moyen de se prémunir des problèmes induits par un panel de métriques trop limité, comme l'illustre bien le Quartet d'Anscombe [253] [45]. C'est également un argument démontré par [247] pour

44. en Europe c'est une obligation légale depuis la mise en place du RGPD
45. Ce célèbre quartet montre qu'une métrique comme le coefficient de détermination ne peut à lui seul vérifier la performance d'un modèle

différentes métriques utilisées pour évaluer les modèles de classification. La création de plateforme de tests ou de benchmarks plus complets peut aider à standardiser les évaluations tout en ayant une vision plus complète des performances des modèles conçus. Il faut néanmoins faire attention à ce que le benchmark ne soit pas trop générique, il doit être cohérent avec le cas d'utilisation ou être complété par d'autres métriques appropriées au contexte. Une évaluation multidimensionnelle permet en outre de ne pas tomber dans le piège mis en lumière par la loi de Goodhart : si les concepteurs utilisent une seule métrique (ou un nombre très limité), elle deviendrait dès lors un objectif à atteindre pour le modèle ce qui n'est une bonne chose.

À ce jour, la communauté des modèles d'apprentissage automatique est encore confrontée à d'importants défis. Tout d'abord, le manque d'outils dédiés à l'élaboration et la mise en œuvre d'évaluations constituent un questionnement majeur. Pourquoi existe-t-il tant d'outils dans le génie logiciel tandis que l'évaluation des algorithmes d'apprentissage se résume encore souvent à l'usage d'un nombre très limité de métriques évaluant leur performance par rapport aux données avec lesquelles ils ont été entraînés ? De plus, la gestion des phénomènes de *data drift* et *concept drift* oblige les concepteurs à mettre en place une surveillance de leurs modèles en production pour prévenir l'obsolescence du modèle initialement élaboré. Les données tout comme la base de connaissance peuvent évoluer, il faut pouvoir garantir la mise à jour des systèmes d'apprentissage automatique.

C'est pourquoi, une mesure de la cohérence entre les résultats prédits et les connaissances préalables à tout à fait sa place dans l'évaluation des systèmes d'apprentissage automatique. Toutefois, nous avons pu voir dans ce chapitre qu'il existait différentes formes de cohérence. La métrique ou la méthodologie d'évaluation est très dépendante de la forme de cohérence étudiée. Il n'est donc pas possible d'utiliser une métrique universelle de la cohérence, cette mesure doit résulter d'une démarche spécifique à chaque cas d'application. La prise en compte du protocole d'évaluation et la mise en place d'une métrique capable de rendre compte de la cohérence d'un algorithme d'apprentissage sont l'objet du prochain chapitre.

5

Transformation des connaissances en contraintes d'apprentissage

5.1 Introduction . . . 119

5.2 Exemple basé sur des facteurs prévisibles . . . 120

5.2.1 L'oscillateur harmonique amorti . . . 120

5.2.2 Prédire sans connaissance . . . 122

5.2.3 Prédire avec des connaissances . . . 124

5.3 Scénario Réel : Complexité et Chaos . . . 125

5.3.1 Modélisation de la durée de vie des matériaux . . . 126

5.3.2 Ajout de connaissances dans un réseau de neurones . . . 129

5.3.3 Bilan et limites . . . 131

5.4 L'apprentissage automatique informé par une ontologie . . . 133

5.4.1 Ontology-based Physics-Informed Machine Learning . . . 133

5.4.1.1 Détermination de règles pour un domaine spécifique . . . 134

5.4.1.2 Association de règles à une loi physique . . . 134

5.4.1.3 Formalisation des lois physiques en équations mathématiques . . . 136

5.4.2 Mise en oeuvre de l'expérimentation . . . 138

5.4.2.1 Description des données . . . 139

5.4.2.2 Règles spécifiques au contexte . . . 139

5.4.3 Évaluation . . . 141

5.4.3.1 Protocole d'évaluation . . . 141

5.4.3.2 Résultats 143
5.5 Conclusion 147
5.5.1 Challenges 147

Ce chapitre explique comment ajouter des contraintes dans un algorithme d'apprentissage automatique tout en vérifiant que ses prédictions sont bien cohérentes avec les connaissances initiales. La démarche est progressive, d'abord sur un exemple d'oscillateur harmonique puis sur un exemple lié au problème de prescription du point de rupture dans le domaine de la fatigue des matériaux. Enfin, ce chapitre présente une formalisation des connaissances dans deux ontologies. La première a pour vocation de déterminer les règles que l'application doit suivre dans un contexte particulier, et la seconde d'associer ces règles à la loi physique générique appropriée traduite ensuite en équation mathématique. Cette équation est alors transformée pour compléter la fonction de perte d'un réseau de neurones. Cet ensemble forme un framework de *Physics-Informed Machine Learning* appelé OIML. L'évaluation des résultats est faite par un score d'incohérence, qui a démontré que les prédictions respectaient bien les réalités de terrain. Le chapitre se conclut par une analyse critique du travail présenté.

5.1/ Introduction

Il existe de nombreux modèles d'apprentissage automatique pouvant être informés par la *prior knowledge* et ce de bien diverses façons comme expliqué précédemment. Notre objectif dans le présent chapitre est en premier lieu d'étudier la transformation des connaissances en contrainte, en particulier durant le processus d'apprentissage d'un modèle.

Les expérimentations réalisées dans ce chapitre s'intéressent en particulier aux réseaux de neurones puisque ce type d'algorithmes est de plus en plus utilisé et qu'ils se prêtent bien à l'ajout de contraintes dans leur structure. Ils peuvent être informés avec divers types de connaissances, toutefois ce chapitre se concentre essentiellement sur la vérification de la cohérence d'un modèle avec des lois physiques sous forme d'équation différentielles partielles.

Ce chapitre est également l'occasion de mettre en place le protocole d'évaluation présenté au chapitre précédent. En incluant bien évidemment une mesure de la cohérence entre les résultats et les connaissances physiques relatives au cas d'application.

En guise d'introduction dans la première section, nous présentons la prédiction des mouvements d'un oscillateur harmonique pour expliquer le principe d'ajout de contraintes physiques dans un réseau de neurones. Dans la deuxième section, nous décrivons un cas plus sophistiqué, celui de l'estimation de la durée de vie d'un matériau. Cette application permet d'illustrer l'apport de contraintes dans un cas réel tout en montrant les limites de conception actuelles, notamment en ce qui concerne l'évaluation du modèle. Dans la troisième section, nous proposons une nouvelle méthode pour formaliser l'ensemble des connaissances dans une ontologie afin de les exploiter plus facilement dans un système

d'apprentissage automatique [1]. Là encore, une évaluation de cette nouvelle approche est réalisée.

5.2/ Exemple basé sur des facteurs prévisibles

L'exemple qui permet d'illustrer le mécanisme derrière une grande partie des réseaux de neurones informés est celui de l'oscillateur harmonique. Proposé par Ben Mosley [2], le cas de l'oscillateur harmonique permet de mettre en avant les bénéfices que peut avoir l'ajout d'une équation différentielle partielle (PDE) dans la fonction perte (*loss*) d'un réseau de neurones. Bien sûr, cet exemple reste éloigné des cas réels qui seront présentés dans les sections suivantes. L'usage d'un réseau de neurones pour produire une solution à une équation différentielle partielle déjà connue n'est bien sûr optimal. Toutefois, cet exemple est utile pour faire comprendre étape par étape comment l'ajout de connaissances améliorent l'entraînement du modèle. Le choix d'une problématique basique et très déterministe comme celle de l'oscillateur harmonique est donc réalisé à des fins pédagogiques.

5.2.1/ L'oscillateur harmonique amorti

L'oscillateur est un système qui, à l'instar d'un ressort, produit des oscillations qui peuvent être d'origine mécanique, acoustique ou bien électrique. En mécanique, l'oscillateur harmonique (voir Figure 5.1) est un système qui lorsqu'il a subit une perturbation par rapport à sa position d'équilibre, subit une force de rappel F proportionnelle à son déplacement x tel que

$$\overrightarrow{F} = -kx$$

où k est une constante positive.

Il peut être vu comme une version idéale d'un oscillateur où les oscillations suivent une fonction sinusoïdale au cours du temps, voir Figure 5.2. Le coefficient de friction F_f qui s'exerce sur l'oscillateur harmonique s'écrit de la manière suivante [3] :

$$F_f = -\mu \frac{dx}{dt}$$

1. Cette dernière partie est une contribution ayant fait l'objet d'une publication dans la conférence SITIS 2023 : S. Ghidalia, O. Labbani Narsis, A. Bertaux, and C. Nicolle, 'Automating Physical Knowledge Integration in Machine Learning', in 2023 17th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, Nov. 2023.
2. Le dépôt Github se trouve ici : https://github.com/benmoseley/harmonic-oscillator-pinn
3. Tous les détails de cette démonstration sont disponibles à l'adresse suivante : https://beltoforion.de/en/harmonic_oscillator/

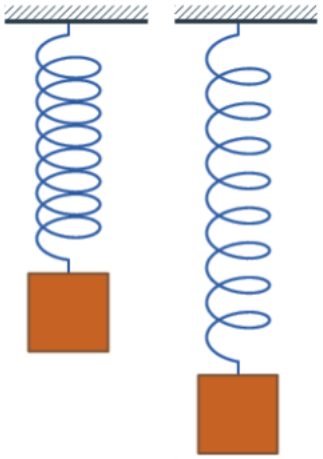

FIGURE 5.1 – Ressort qui symbolise le mouvement d'oscillation d'un oscillateur harmonique amorti

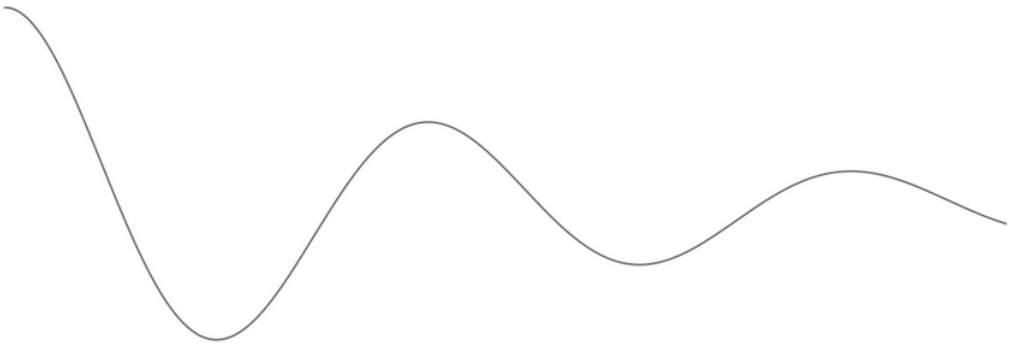

FIGURE 5.2 – Fonction sinusïdale qui décrit le mouvement d'un oscillateur harmonique amorti

Or, la seconde loi de Newton, l'inertie F_i, i.e. la position à laquelle le ressort aspire à revenir, est définie comme suit :

$$F_i = m\frac{d^2x}{dt^2}$$

Dans le cas de l'oscillateur harmonique, la force d'inertie est opposée aux deux forces que sont le rappel et le coefficient de friction. Par conséquent, l'équation simplifiée qui représente ce système peut s'écrire de la manière suivante :

$$m\frac{d^2x}{dt^2} + \mu\frac{dx}{dt} + kx = 0 \tag{5.1}$$

Pouvant être ré-arrangée comme ceci :

$$\frac{d^2x}{dt^2} + \delta\frac{dx}{dt} + w_0x = 0 \tag{5.2}$$

avec $\delta = \frac{\mu}{2m}$ et $w_0 = \sqrt{\frac{k}{m}}$. La solution de cette équation peut dès lors être trouvée grâce à l'ansatz exponentiel qui trouve un mouvement d'oscillation [4] décrit par la fonction :

$$s = -\delta \pm \sqrt{\delta^2 - {w_0}^2} \tag{5.3}$$

Il existe donc trois cas différents pour l'oscillateur harmonique amorti :

— le sous-amortissement qui a lieu lorsque $\delta < w0$
— l'amortissement critique qui a lieu lorsque $\delta = w0$
— le sur-amortissement qui a lieu lorsque $\delta > w0$

5.2.2/ Prédire sans connaissance

Dans le cas étudié ici, il s'agit d'un oscillateur harmonique amorti ayant un δ égal à 2 et un w_0 égal à 15 dont la solution est représentée par la sinusoïde bleue de la Figure 5.3. Il s'agit donc d'un oscillateur harmonique sous-amorti. La sinusoïde qui le représente servira de "vérité de terrain" (*ground truth*) permettant d'évaluer la capacité d'un réseau de neurones à prévoir sa position future. Dix points ont été choisis aléatoirement au début de cette courbe pour servir de données d'entraînement au modèle. L'objectif du modèle est d'être capable de prédire la position prise par le ressort dans le futur on s'appuyant uniquement sur les données d'entraînement peu nombreuses en orange sur la Figure 5.3.

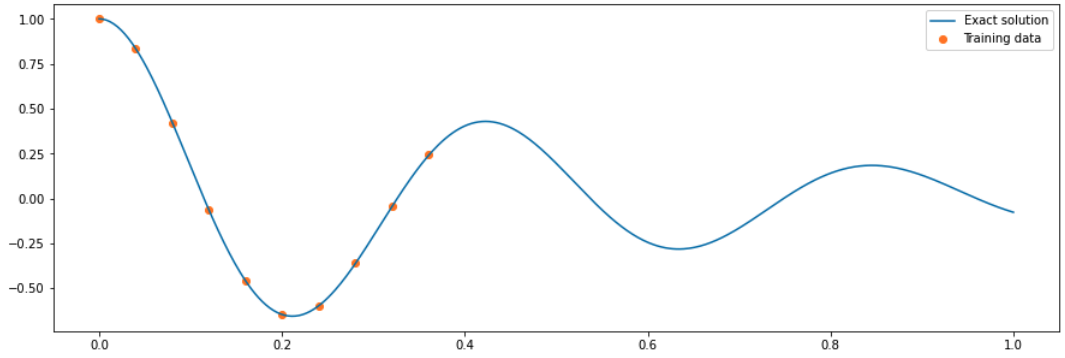

Figure 5.3 – *Ground truth* de l'oscillateur harmonique amorti et données d'entraînement

Le réseau de neurones *fully-connected* utilisé pour créer le modèle comporte une couche d'entrée à une variable, trois couches cachées de 32 neurones également activées par la fonction *tanh* ainsi qu'une couche de sortie linéaire comme montré sur la Figure 5.4. Entraîné sur 6000 epochs à l'aide de l'optimiseur Adam, un taux d'apprentissage de 0.001 et une fonction de perte basée sur l'erreur quadratique moyenne [5] (score MSE vu dans le paragraphe 4.3.4) comme le montre l'équation 5.4.

4. comme celui effectué par un ressort
5. aussi appelée L2 loss

$$L_{data} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_i - Y_i)^2 \tag{5.4}$$

$$Loss = L_{data}$$

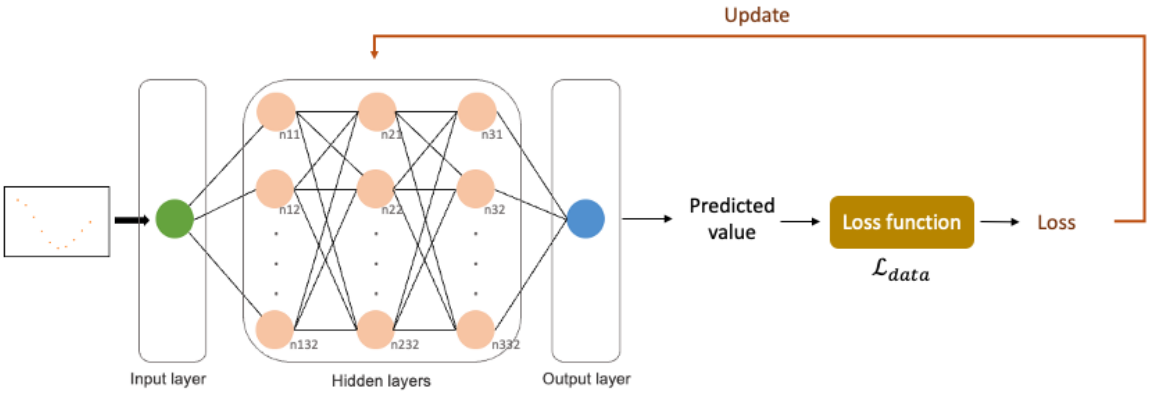

FIGURE 5.4 – Oscillateur harmonique : architecture du réseau de neurones sans apport de connaissance

Le modèle créé peine à prédire les prochaines positions prises par le ressort comme cela est illustré par la Figure 5.5 qui montre les résultats obtenus. Ce résultat est loin d'être surprenant au regard du peu de données utilisées et du type de réseau de neurones choisi qui n'est pas adapté aux séries temporelles [6].

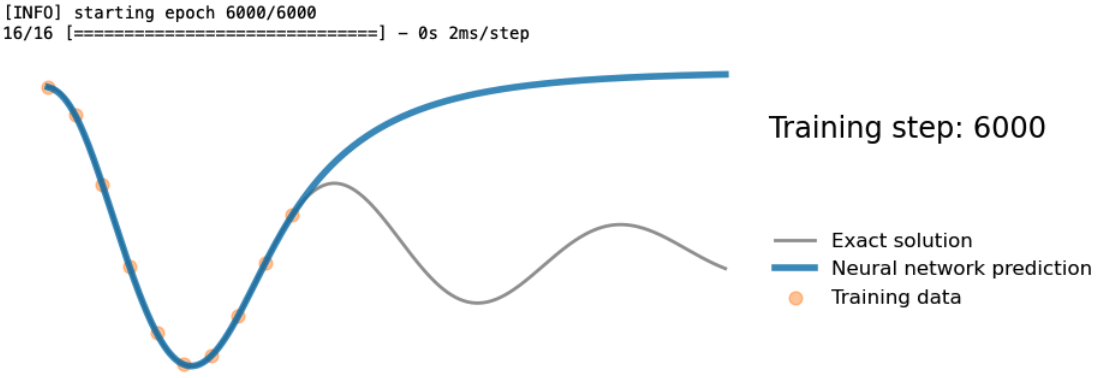

FIGURE 5.5 – Oscillateur harmonique : prédiction du réseau de neurones sans apport de connaissance

6. Contrairement aux réseaux de neurones récurrents comme LSTM (Long-Short Term Memory) ou les GRU (Gated Recurrent Units)

5.2.3/ Prédire avec des connaissances

Cependant, en conservant exactement les mêmes paramètres et la même architecture et en ajoutant l'équation 5.1 de la PDE au niveau de la fonction de perte, des résultats bien différents sont obtenus (cf Figure 5.6). Cette fois-ci, le réseau de neurones est guidé vers une solution de l'équation différentielle partielle et ce malgré le faible nombre de données d'entraînement. Dans le cas d'un réseau de neurones informé, sa fonction de perte totale est souvent formulée comme une combinaison pondérée de deux termes : un terme qui corresponds à la cohérence du modèle avec les données et un terme qui corresponds à la cohérence du modèle avec les connaissances préalables. La formule générale de la fonction de perte peut dès lors être exprimée comme suit [37, 40, 275] :

$$\mathcal{L} = \mathcal{L}_{data} + \lambda \mathcal{L}_{knowledge} \tag{5.5}$$

où $\mathcal{L}_{data}$ est la partie qui mesure l'erreur entre les prédictions du modèle et les données d'apprentissage, $\mathcal{L}_{knowledge}$ est la partie qui mesure le non respect des contraintes par le modèle, et λ est un coefficient de pondération qui contrôle l'importance relative de chaque contrainte.

La fonction de perte de l'oscillateur harmonique est une combinaison entre $L_{knowledge}$ (cf. équation 5.6), L_{data} (cf. équation 5.4) et un coefficient de pondération λ égal à 1. Le réseau de neurones ne connaît pas la solution exacte attendue [7] mais il est informé de l'équation différentielle partielle à résoudre dans sa fonction de perte. L'ajout de connaissance dans le processus d'entraînement du modèle améliore considérablement ses résultats (cf Figure 5.7) vis-à-vis du *ground truth*.

$$L_{knowledge} = \frac{1}{m} \sum_{j=1}^{m} \left(\left[m \frac{d^2}{dt^2} + \mu \frac{d}{dt} + k \right] \widehat{Y}_j \right)^2 \tag{5.6}$$

$$Loss = L_{data} + L_{knowledge} \tag{5.7}$$

Finalement, avec un simple réseau de neurones de type perceptron multicouche (MLP) et dix données d'entraînement situées uniquement au début de la sinusoïde à prédire, l'ajout d'une d'équation différentielle partielle dans la fonction de perte a permis d'obtenir des résultats très proches de ce qui était attendu. Au regard de cette expérience, il est permis de penser que l'ajout de connaissance dans la fonction de perte à un intérêt en cas de données d'apprentissage limitées (low data). De même, il semblerait que ce soit une méthode intéressante pour faire respecter certaines règles inhérentes au système au réseau de neurones et donc par conséquent de s'assurer qu'un modèle lors de son

7. i.e. la sinusoïde présentée par l'équation 5.3

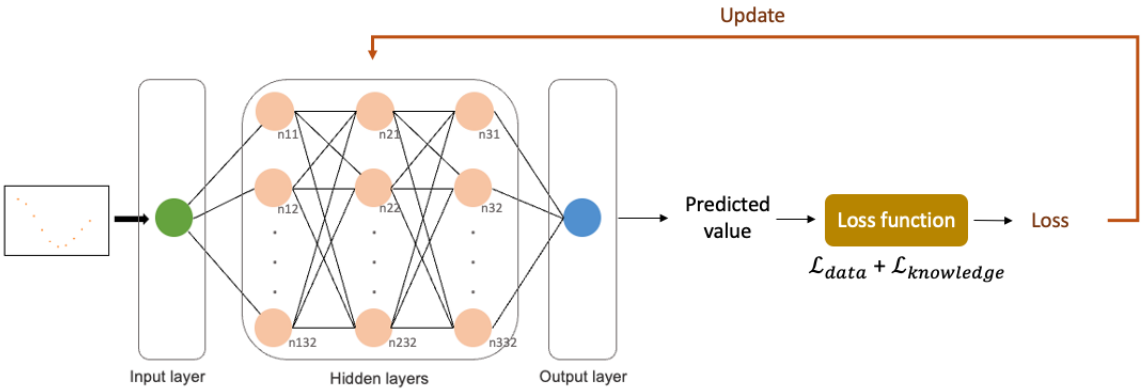

FIGURE 5.6 – Oscillateur harmonique : architecture du réseau de neurones avec apport de connaissance

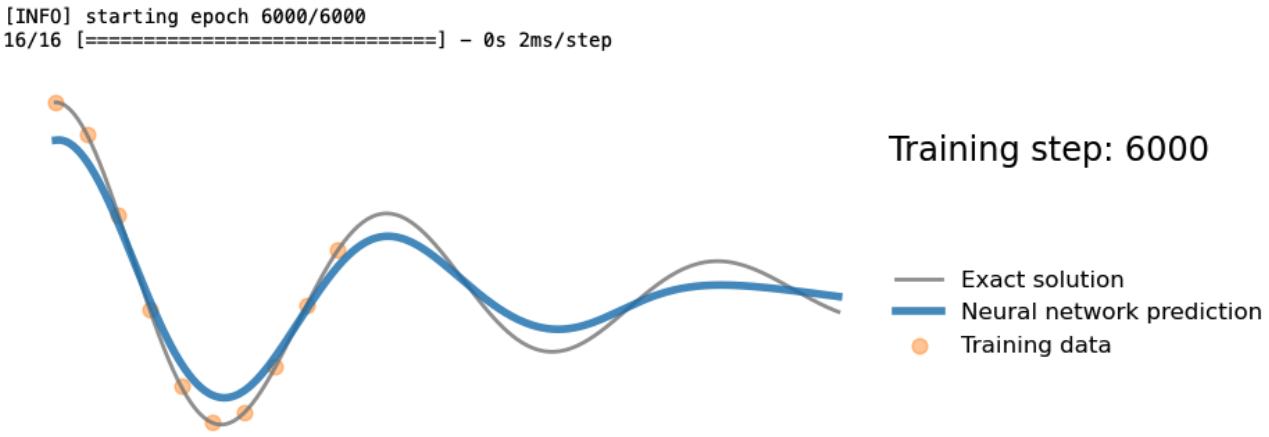

FIGURE 5.7 – Oscillateur harmonique : prédiction du réseau de neurones avec apport de connaissance

entraînement est poussé à rester cohérent avec celles-ci.

5.3/ SCÉNARIO RÉEL : COMPLEXITÉ ET CHAOS

Dans le domaine de l'étude des matériaux, il existe des problèmes plus complexes que celui de l'oscillateur harmonique. Notamment l'estimation de la durée de résistance d'un matériau mis sous contrainte avant de se rompre complètement[8]. C'est n'est pas une question triviale puisque différents fils de métal, confectionnés par la même entreprise, issus du même processus de fabrication et soumis à une même pression (amplitude de stress) ne casseront pas exactement au même moment. Le décalage pourra être plus au

8. aussi appelé durée de vie

moins long. Pourquoi ? Parce que ces différents fils sont des individus non identiques : ils n'ont pas toujours le même vécu ou le même processus d'élaboration [9] et ce malgré toutes les précautions prises lors de leurs conception. C'est pour cela que les fabricants mentionnent des caractéristiques générales (valeurs moyennes) associées à des écart-types : pour rendre compte de la part d'aléatoire spécifique à chaque individu.

5.3.1/ Modélisation de la durée de vie des matériaux

La durée de vie des matériaux est un domaine étudié depuis longtemps, des connaissances ont donc été acquises par les experts du domaine. En particulier pour les métaux les plus utilisés comme l'acier ou l'aluminium. Ainsi, la courbe de Wöhler représentée sur la Figure 5.8, aussi appelée courbe S-N [10], est utilisée pour représenter les résultats d'essais de fatigue [11]. Cette courbe théorique ne représente pas chacun des matériaux spécifiquement, et encore moins les individus, mais elle permet de connaître les caractéristiques communes aux différents matériaux lorsqu'il s'agit de leur durée de vie. Elle illustre le fait que lorsque l'amplitude de stress diminue la durée de vie du matériau en moyenne augmente et la courbure de la courbe de Wöhler diminue. Les experts du domaine constatent également que l'application de faibles amplitudes de stress entraînent une augmentation de l'écart-type de la durée de vie, ce qui rend les prédictions plus complexes pour de telles situations.

Déterminer la courbe particulière pour un matériau nécessite des expérimentations à différentes amplitude de stress qui peuvent être coûteuses puisqu'il faut sacrifier de nombreux matériaux et mobiliser du personnel durant toute la durée de l'étude. En particulier lorsqu'il s'agit de nouveaux alliages difficiles à mettre au point et à produire. Toutefois, à l'aide d'éléments mathématiques il est possible de pouvoir réduire le nombre d'expérimentations tout en ayant une courbe qui représente bien les spécificités du matériau étudié. C'est là que l'apprentissage automatique peut être utilisé pour retrouver la courbe de Wöhler à partir de quelques exemples d'apprentissage. Cette expérience, à l'instar de celle menée par [37] à partir de l'échantillon de données du fil d'acier [12] [276], a été reproduite ici.

Afin de pouvoir les utiliser dans un algorithme d'apprentissage automatique, les données de [276] ont été séparées en données d'entraînement et en données d'évaluation du modèle comme on peut le voir sur la Figure 5.9. Les données d'entraînement sont transmises à un réseau de neurones *fully-connected* comportant une couche d'entrée, une couche cachée de 16 neurones activée par la fonction *tanh* et une couche de sortie de

9. parfois ils ont même des défauts de conception, ce qui arrive par exemple quand de l'air est ajoutée fortuitement au matériau lors de sa création
10. pour *Stress* vs *Number of cycles*
11. Le mot fatigue est ici utilisé pour dire "durée de vie"
12. *Steel Wire dataset*

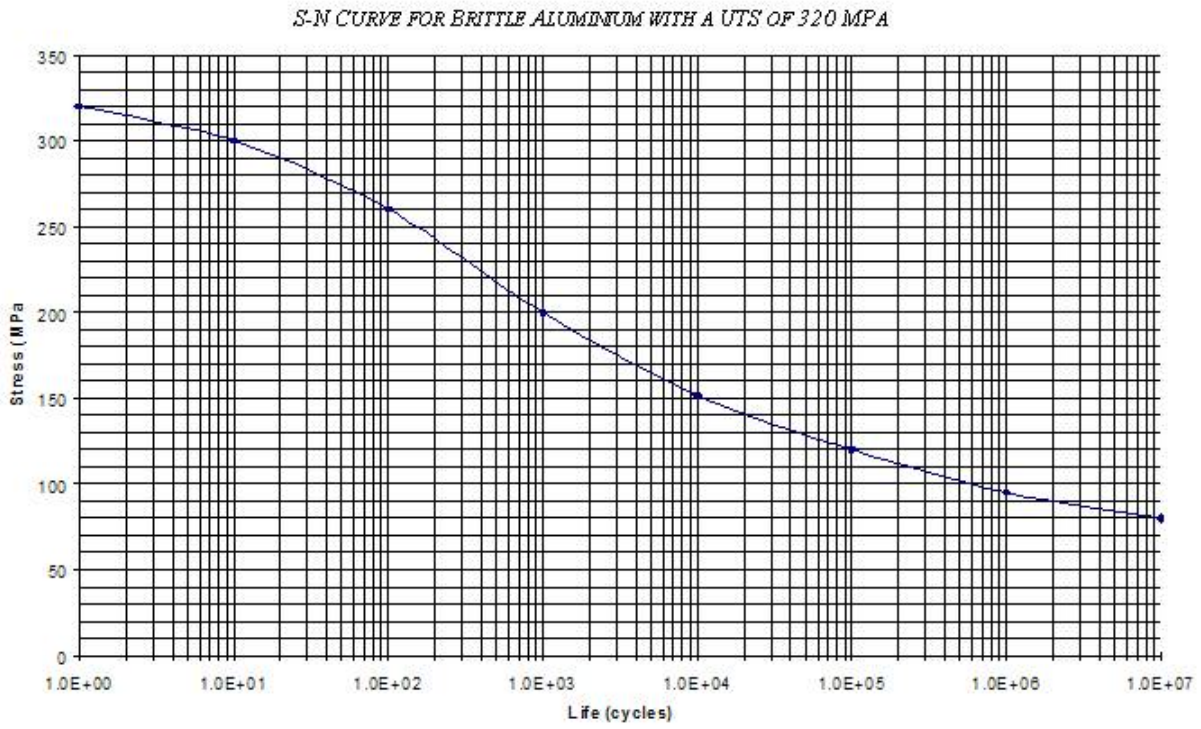

FIGURE 5.8 – Courbe de Wöhler ou courbe S-N

deux neurones (pour obtenir à la fois la moyenne et l'écart type de chaque point de la courbe) activée par la fonction *softplus*. L'architecture globale de ce réseau de neurones est représentée sur la Figure 5.10. L'optimiseur utilisé pour l'entraînement est Adam et la fonction de perte est calculée à partir de la fonction de vraisemblance négative [13] présentée par l'équation 5.8.

$$\mathcal{L}_{data} = \mathcal{L}_{negloglike} = -\sum_{i=1}^{n} \left(Y_i \log\left(\widehat{Y_i}\right) + (1 - Y_i) \log\left(1 - \widehat{Y_i}\right) \right) \tag{5.8}$$

Le réseaux de neurones créé permet d'obtenir la courbe de Wöhler pour le fil d'acier. Cette courbe représente l'individu moyen et son écart-type montre l'intervalle de confiance de 95% comme présenté en Figure 5.11. Malheureusement, cette courbe ainsi modélisée ne respecte pas toujours les règles reconnues des experts en conception des matériaux. En effet, l'écart-type au niveau de la plus grande amplitude de stress appliquée est plus grand que l'écart-type pour une amplitude de stress inférieure (cf. Figure 5.11). De même la courbure de la courbe ne semble pas toujours diminuer alors que le stress est de plus en plus faible ce qui est également montré par la Figure 5.11. Ces quelques détails confirment que le modèle élaboré ne représente pas forcément bien la

13. En anglais cette fonction est souvent appelée *negloglike* pour *negative log likelihood*, elle est d'ailleurs disponible directement dans le framework TensorFlow

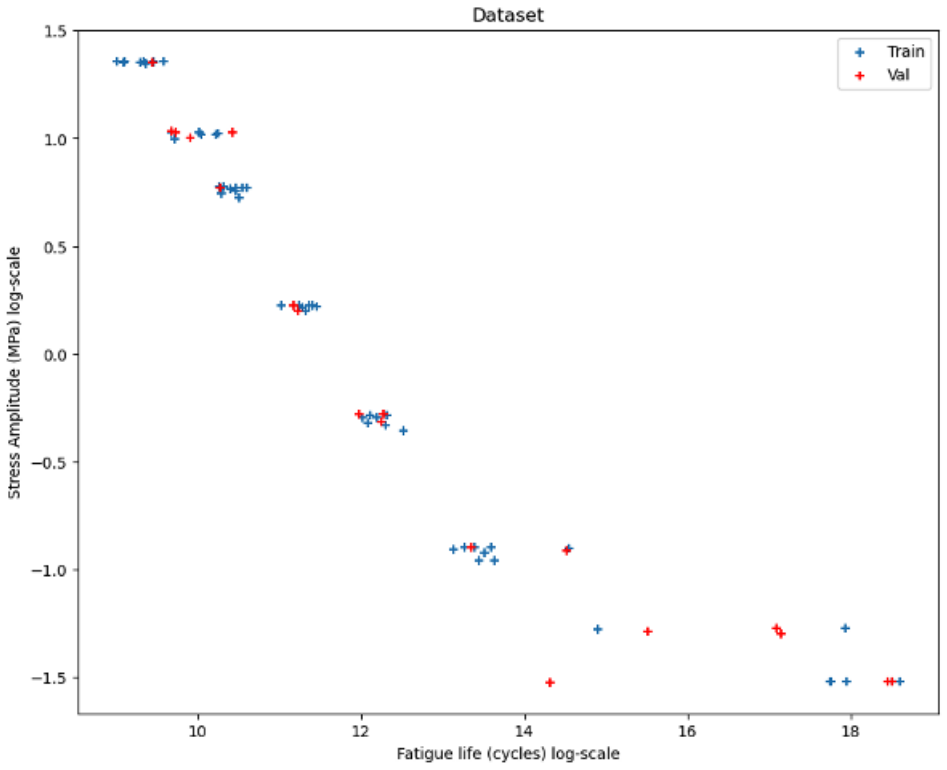

FIGURE 5.9 – Durée de vie des matériaux : données d'entraînement et d'évaluation

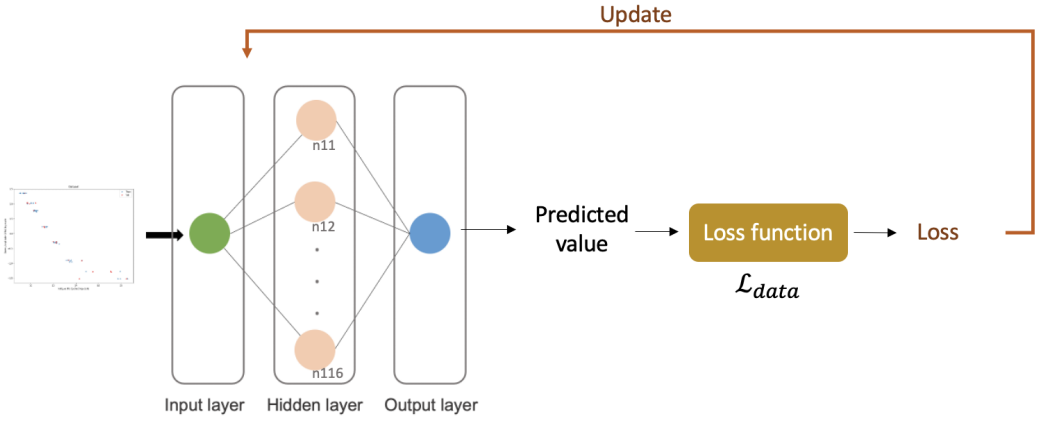

FIGURE 5.10 – Durée de vie des matériaux : architecture du réseau de neurones sans apport de connaissance

réalité en dépit de ses scores de performance raisonnables [14].

L'ajout de connaissances relatives à la durée de vie des matériaux, notamment les trois règles vu précédemment, pourrait-elles le rendre plus performant ?

14. Le score *Root Mean Squared Error* étant de 1.3719

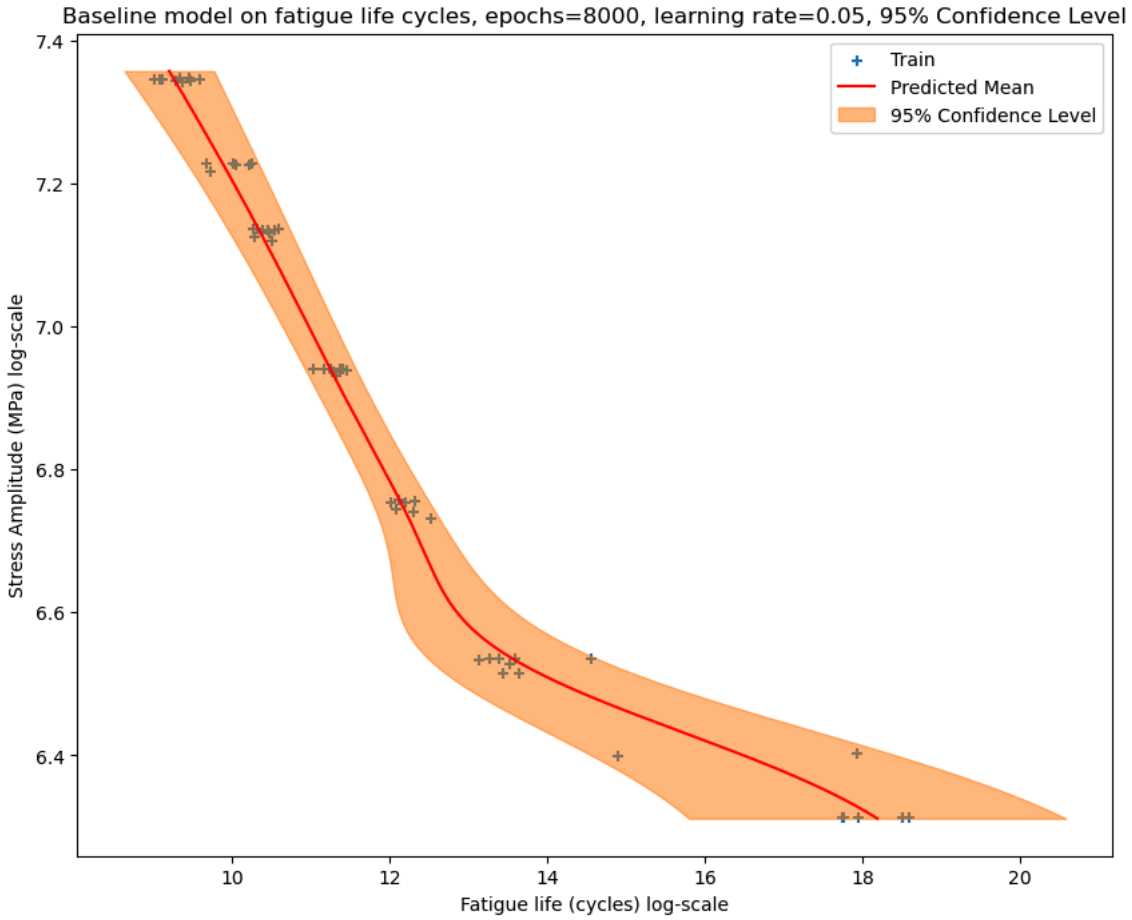

FIGURE 5.11 – Durée de vie des matériaux : prédiction du réseau de neurones sans apport de connaissance

5.3.2/ AJOUT DE CONNAISSANCES DANS UN RÉSEAU DE NEURONES

Les trois connaissances données par les experts en conception des matériaux peuvent être transformées en contraintes puis intégrées dans le mécanisme d'apprentissage du réseau de neurones par le biais de la fonction de perte. En effet, chacune des ces règles peut-être représentée par une équation en utilisant les fonctions dérivées comme détaillé dans le tableau 5.1. Ainsi, une connaissance exprimée en langage naturel est transformée en contrainte formalisée à l'aide d'une équation mathématique.

Ces équations aux dérivées partielles vont être vérifiées lors de l'application de la fonction de perte dans le réseaux de neurones. À chaque entraînement, le respect des contraintes est vérifié. Comme l'objectif est de minimiser la fonction de perte, le score de celle-ci augmentera en fonction du nombre de fois où les règles n'ont pas été satisfaites. Ce mécanisme à pour but de guider le réseau de neurones vers le respect de l'ensemble des contraintes ajoutées. Cette fois encore, la fonction de perte du réseau de neurones prendra la forme de l'équation 5.5 avec un $\mathcal{L}_{data}$ correspondant à l'équation 5.8 et un

Connaissance	Traduction mathématique	Contrainte (PDE)
Lorsque l'amplitude de stress diminue, la durée de vie du matériau en moyenne augmente	La dérivée première de la prédiction de la durée de vie moyenne Y_μ est négative	$\frac{dY_\mu}{dx} < 0$
Lorsque l'amplitude de stress diminue, l'écart-type de la durée de vie du matériau augmente	La dérivée première de l'écart-type Y_σ est négative	$\frac{dY_\sigma}{dx} < 0$
Lorsque l'amplitude de stress diminue, la courbure de la courbe de Wöhler diminue	La dérivée seconde de la prédiction de vie moyenne Y_μ est positive	$\frac{d^2Y_\mu}{dx^2} > 0$

TABLE 5.1 – Tableau des connaissances et contraintes associées d'après [37]

$\mathcal{L}_{knowledge}$ égale à :

$$\mathcal{L}_{knowledge} = \mathcal{L}_k^1 + \mathcal{L}_k^2 + \mathcal{L}_k^3 \tag{5.9}$$

où $\mathcal{L}_k^1$, $\mathcal{L}_k^2$ et $\mathcal{L}_k^3$ représente les trois contraintes associés aux trois différentes règles que doit respecter le réseau de neurones. Ici, le coefficient de pondération λ est égal à 1000 et il est valable pour l'ensemble de la partie relative à la vérification du respect des contraintes puisque les trois règles sont aussi importantes les unes que les autres.

Les termes $\mathcal{L}_k^1$, $\mathcal{L}_k^2$ et $\mathcal{L}_k^3$ sont définis de telle façon que le réseau de neurones est uniquement pénalisé lorsqu'une contrainte n'a pas été bien respectée. Pour s'assurer de cela, une fonction d'état $I(A > B)$ est mise à 0 si la condition $A > B$ est respectée et à 1 sinon, comme montré sur l'équation 5.10. Dans le cas étudié, il n'existe pas de *ground truth* contrairement à l'oscillateur harmonique, il faut déterminer des points de collocations X_r^c, où $c = 1, 2, ..., N_r$, auxquels les contraintes doivent être satisfaites. Les points de collocations correspondent dans cet exemple à 1000 points générés espacés linéairement dans la plage d'amplitude de stress étudiée. Ainsi, les trois termes de régularisation de $\mathcal{L}_{knowledge}$ s'écrivent [37] :

$$\begin{cases} \mathcal{L}_k^1 = \frac{1}{N_r} \sum_{c=1}^{N_r} I\left(\frac{\widehat{dY_\mu^{r,c}}}{dX_r^c} > 0\right)\left(\frac{\widehat{dY_\mu^{r,c}}}{dX_r^c}\right)^2 \\ \mathcal{L}_k^2 = \frac{1}{N_r} \sum_{c=1}^{N_r} I\left(\frac{\widehat{dY_\sigma^{r,c}}}{dX_r^c} > 0\right)\left(\frac{\widehat{dY_\sigma^{r,c}}}{dX_r^c}\right)^2 \\ \mathcal{L}_k^2 = \frac{1}{N_r} \sum_{c=1}^{N_r} I\left(0 > \frac{\widehat{d^2Y_\sigma^{r,c}}}{dX_r^{c2}}\right)\left(\frac{\widehat{d^2Y_\sigma^{r,c}}}{dX_r^{c2}}\right)^2 \end{cases} \tag{5.10}$$

Le réseau de neurones s'entraîne en tentant de minimiser les termes de régularisation $\mathcal{L}_k^1$, $\mathcal{L}_k^2$ et $\mathcal{L}_k^3$ ce qui le force à trouver une solution qui respecte le plus possible les contraintes mais sans garantie qu'elles soit toujours respectées à chacun des points de

collocations. Mise à part la fonction de perte, l'architecture du réseau de neurones ne subit pas de changement majeur comme on peut le voir sur la Figure 5.12. L'ajout de contraintes dans l'entraînement du réseau vas permettre d'obtenir une meilleure généralisation du modèle comme on peut le voir sur la Figure 5.13.

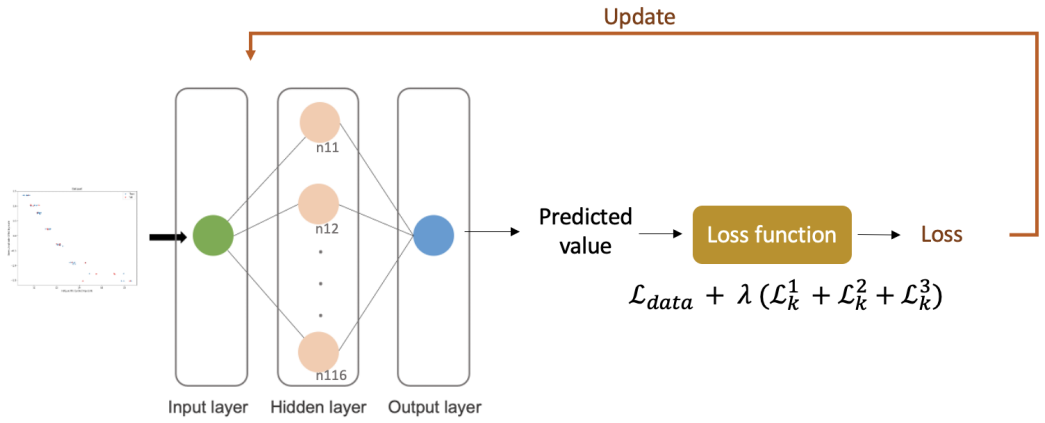

FIGURE 5.12 – Durée de vie des matériaux : architecture du réseau de neurones avec apport de connaissance

5.3.3/ BILAN ET LIMITES

Le principal atout de l'apport de connaissance est une meilleure généralisation du modèle créé. La généralisation fait référence à la capacité du modèle à bien performer sur des données qu'il n'a pas rencontrées auparavant. En d'autres termes, un modèle bien généralisé est capable de fournir des prédictions précises et fiables même pour des situations inédites. L'apport de connaissance permet de contre-balancer le phénomène d'overfitting souvent présent dans les réseaux de neurones, d'autant plus lorsque les données d'entrée sont limitées comme c'est le cas ici. En effet, si le modèle n'a pas été exposé à suffisamment de scénarios différents lors de l'entraînement, il peut ne pas être en mesure de généraliser correctement. Une stratégie efficace consiste à s'assurer que l'ensemble d'entraînement couvre une large gamme de paramètres pertinents et de cas d'utilisation. Ce qui ne peut être le cas ici, puisque l'objectif est justement de limiter le nombre d'expériences réelles faites sur les matériaux. Un modèle capable d'être juste et suffisamment précis sans avoir accès à de nombreuses données d'entraînement est donc nécessaire.

Toutefois, la conception de ce type de modèle n'est pas aisé. Pour être intégrées dans la fonction de perte, les lois physiques doivent souvent être présentées sous la forme d'équations différentielles partielles, qui doivent ensuite être traduites en code exécutable dans le cadre de la programmation d'un réseau de neurones. Pour accomplir cela, il est essentiel de faire appel à des experts du domaine qui possèdent une solide connais-

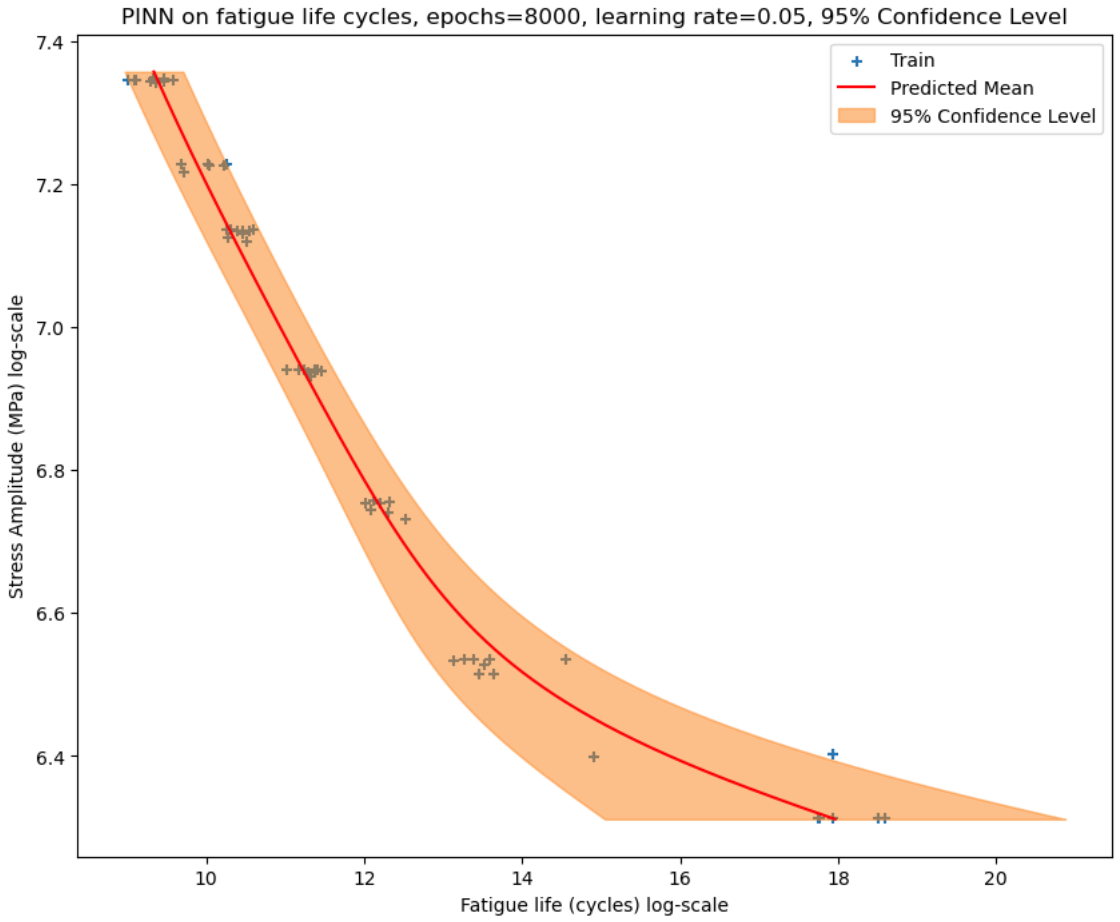

FIGURE 5.13 – Durée de vie des matériaux : prédiction du réseau de neurones avec apport de connaissance

sance des lois physiques applicables à une situation donnée et qui sont en mesure de développer le code nécessaire pour la fonction de perte. Or, la création d'une fonction de perte appropriée, dans un framework particulier (Tensorflow ou PyTorch par exemple) est souvent dévolue à un ingénieur en apprentissage automatique (ou *Data Scientist*) spécialisé en programmation peu familier avec les lois de physiques. La complexité de cette tâche nous a fait réfléchir à une solution capable de faciliter le travail aux experts du domaine grâce à une meilleure formalisation de leurs connaissances.

De plus, la qualité des connaissances préalables peut varier considérablement en fonction de leur source, de leur pertinence et de leur fiabilité. Si la connaissance est incorrecte ou inappropriée, elle peut conduire à des erreurs dans les prédictions du modèle. Une meilleure formalisation des connaissances est nécessaire pour permettre la réutilisation des connaissances dans diverses applications et faciliter la conception de toutes sortes d'algorithmes d'apprentissage automatique guidés par les connaissances [35, 36]. Quoi de mieux qu'une ontologie pour formaliser la connaissance et la mobiliser de manière plus automatisée dans les algorithmes d'apprentissage automatique ?

5.4/ L'apprentissage automatique informé par une ontologie

Les contraintes de conception des algorithmes d'apprentissage automatique informés actuels, combinées au manque de réutilisation des connaissances, nous ont incités à concevoir un système capable de résoudre ces problèmes. En s'inspirant des travaux analysés au chapitre 3, un nouveau framework de conception d'*Informed Machine Learning* est proposé.

5.4.1/ Ontology-based Physics-Informed Machine Learning

L'objectif principal est de formaliser les connaissances physiques pour les intégrer plus facilement dans la fonction de perte d'un réseau de neurones. Pour ce faire, il est nécessaire de (1) déterminer les règles que l'application doit suivre dans un domaine particulier, (2) associer ces règles à la loi physique appropriée et (3) les formaliser dans une équation mathématique pour déterminer le terme $\mathcal{L}_{knowledge}$. Pour atteindre cet objectif, nous avons conçu un framework d'apprentissage automatique informé par une ontologie modélisant des lois physiques appelé *Ontology-based Physics-Informed Machine Learning* (OPIML), décrit dans la Figure 5.14. Ce framework incorpore deux ontologies spécialisées réalisant les étapes (1) et (2) ainsi qu'un code Python dédié à l'étape (3).

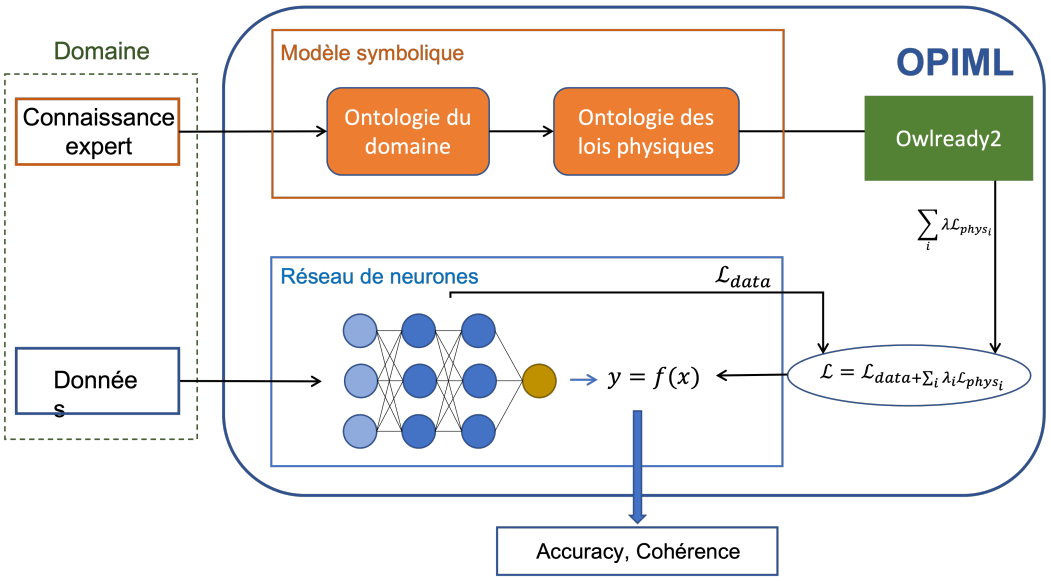

Figure 5.14 – Architecture et fonctionnement du framework *Ontology-based Physics-Informed Machine Learning* (OPIML)

L'exemple d'application est une nouvelle fois réalisé sur l'estimation de la durée de vie des matériaux, cependant comme expliqué dans la prochaine section, les lois de physique formalisées dans l'ontologie sont génériques et peuvent donc s'appliquer à différents domaines sans être modifiées.

5.4.1.1/ Détermination de règles pour un domaine spécifique

La première ontologie (“ontologie du domaine” dans la Figure 5.14) concerne les connaissances liées à l’application étudiée, ainsi que les règles qui lui sont associées associées. Comme décrit dans la section 5.3, dans l’étude de la fatigue des matériaux, on observe une relation où l’augmentation de l’amplitude de stress appliquée au matériau (S) entraîne une diminution de sa durée de vie (L). Cette relation se matérialise dans le calcul de la dérivée partielle $\frac{\partial L}{\partial S} < 0$ qui formalise mathématique la relation monotone négative entre la contrainte de stress appliquée (S) et la durée de vie du matériau (L). Les deux autres règles présentes dans le Tableau 5.1 peuvent également être représentées d’une manière analogue.

Dans un tout autre domaine d’étude, celui de la thermodynamique des bâtiments, on peut retrouver une même relation monotone négative cette fois entre la climatisation (C) et la température d’une pièce (T) : plus la climatisation augmente, plus la température de la pièce diminue [269]. Là encore, le calcul de la dérivée partielle $\frac{\partial T}{\partial C} < 0$ permet de formaliser mathématiquement la relation monotone négative entre la climatisation (C) et la température d’une pièce (T).

Dès lors, il est intéressant de noter que, dans ces deux domaines distincts que sont la modélisation des informations du bâtiment (BIM) et la science des matériaux, le même type de relation monotone négative est appliqué. Cette observation nous a conduit à créer une seconde ontologie plus abstraite, capable de s’appliquer dans des situations très différentes comme peuvent l’être les deux exemples ci-dessus.

5.4.1.2/ Association de règles à une loi physique

En reprenant les exemples donnés précédemment, on observe que dans les deux cas la règle appliquée est "si A augmente, alors B diminue", appelée règle de relation monotone négative, A et B étant des objets différents selon le domaine d’application. Cette règle abstraite est formalisée dans la seconde ontologie, nommée "ontologie des lois physiques" dans la Figure 5.14. Elle est également associée à la dérivée partielle abstraite $\frac{\partial B}{\partial A} < 0$. En partant de ce constat, nous savons que lorsque nous devons appliquer une règle inverse à deux objets, A et B, il est nécessaire d’incorporer cette équation aux dérivées partielles pendant la phase d’apprentissage de l’algorithme.

Ainsi, chaque règle spécifique présente dans l’"ontologie de domaine" est associée à au moins une loi physique abstraite représentée dans l’"ontologie des lois physiques". L’association se fait par un lien de subsomption, c’est-à-dire que la règle spécifique "lorsque l’air conditionné augmente, la température de la pièce diminue" est une instance de la classe "si A augmente, alors B diminue". Finalement, chaque règle physique, spécifique

à un contexte particulier et associée à des variables particulières, est en fait une instanciation d'une loi physique plus générique.

Notre ontologie de la physique (présentée dans la Figure 5.15) est capable de représenter une situation avec plusieurs règles qui peuvent être appliquées dans le contexte de cette situation. Ces règles sont des règles abstraites qui représentent des connaissances génériques telles que la loi de relation monotone positive ("si A augmente, alors B augmente") ou la loi de relation monotone négative ("si A augmente, alors B diminue") comme le montre la Figure 5.16.

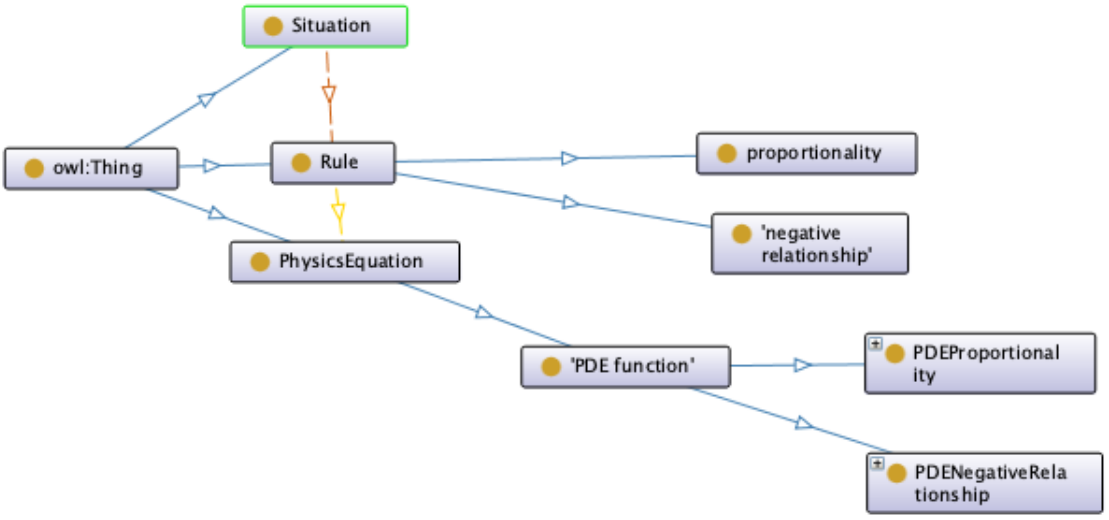

FIGURE 5.15 – Ontologie des lois physiques

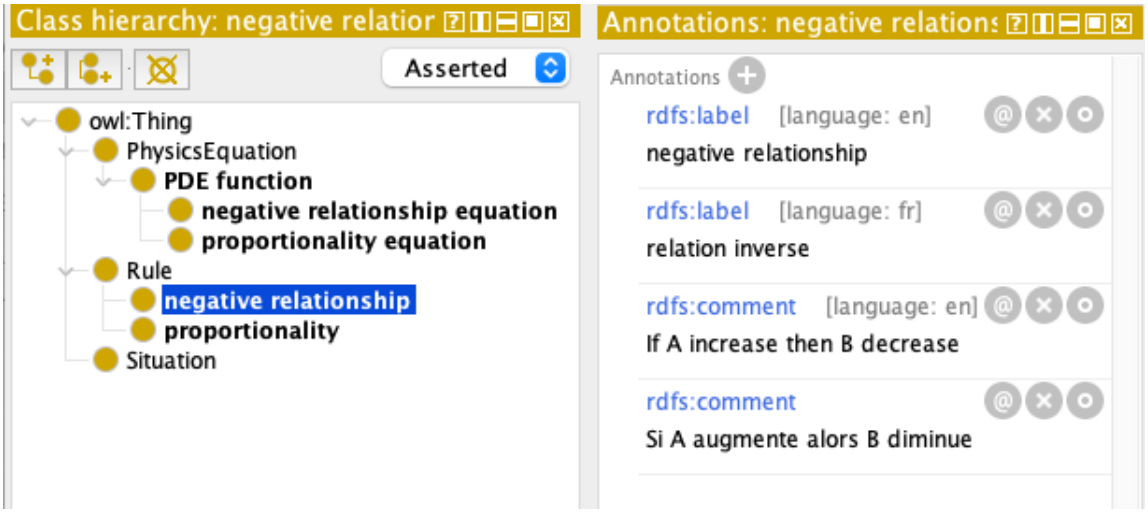

FIGURE 5.16 – Règle de relation monotone négative

Ces règles sont associées à des fonctions spécifiques qui correspondent à leur transformation en vue de leur utilisation dans une fonction de perte. Ainsi, la loi de relation monotone positive devient "la dérivée de B par rapport à A est positive" et la loi de relation monotone négative devient "la dérivée de B par rapport à A est négative" comme

le montre la Figure 5.17. Le lien entre les règles et leurs fonctions correspondantes est réalisé grâce aux règles SWRL, de sorte qu'il peut être déduit par un moteur d'inférence comme HermiT [190] ou Pellet [189]. L'ontologie est basée sur le langage OWL-Lite, et correspond donc à la logique de description $\mathcal{SHIF}$.

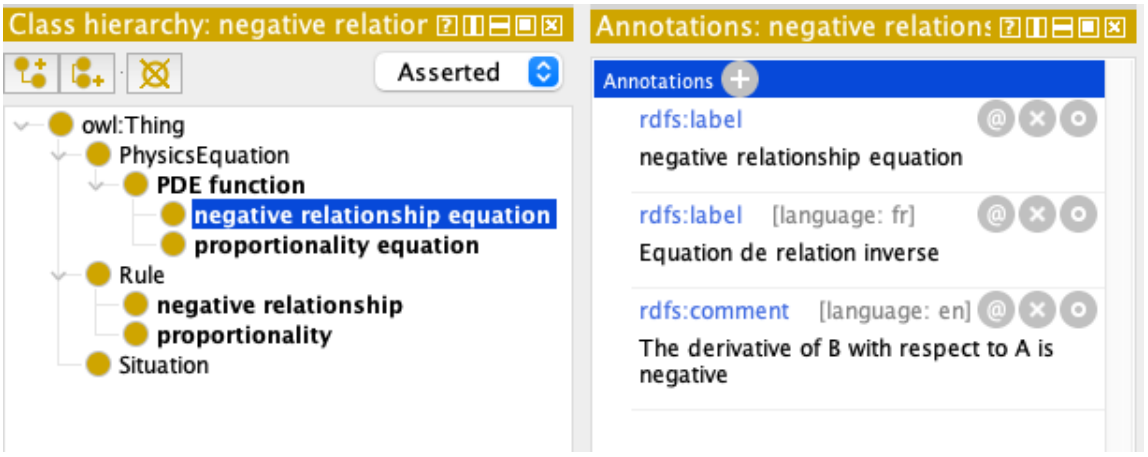

FIGURE 5.17 – Équation de la relation monotone négative

L'ontologie du domaine utilise cette seconde ontologie des lois physiques pour spécifier la loi générale qui contraindra le réseau de neurone par le biais de la fonction de perte.

5.4.1.3/ FORMALISATION DES LOIS PHYSIQUES EN ÉQUATIONS MATHÉMATIQUES

Les règles associées à chaque application sont ensuite récupérées à l'aide de la bibliothèque OwlReady2 [277]. Le moteur d'inférence Pellet, accessible via Owlready2, transmet au code Python l'ensemble des règles associées à une application, en précisant pour chaque règle la fonction physique abstraite qui lui est attribuée. Les paramètres de contexte étant formalisés dans l'ontologie, ils peuvent être facilement transmis à la fonction Python. Ce processus garantit que toutes les connaissances extraites de l'ontologie sont effectivement converties pour être utilisées dans le code de programmation essentiel à la mise en œuvre de l'algorithme d'apprentissage automatique.

Pour intégrer les règles physiques dans la fonction de perte par le biais du terme $\mathcal{L}_{knowledge}$ (c.f Équation 5.9 dans la section précédente), il est impératif d'exprimer ces règles sous forme d'équations mathématiques, qui prennent souvent la forme d'équations aux dérivées partielles. Pour chaque loi physique abstraite formalisée dans l'ontologie de la physique, une fonction Python correspondante a été développée pour la représenter. Cette fonction accepte des variables contextuelles de l'ontologie du domaine comme paramètres d'entrée pour s'adapter à chaque règle spécifique. Par transitivité, cette fonction peut calculer le terme $\mathcal{L}_k^i$ correspondant à une règle i en fonction des paramètres donnés par l'ontologie du domaine (c.f Équation 5.10). La fonction de perte finale peut donc cou-

vrir plusieurs règles différentes, chacune d'entre elles étant ajoutée à la perte physique.

Par exemple, une fonction Python peut calculer l'équation $\frac{\partial B}{\partial A} < 0$ correspondant à la règle de la relation monotone négative ("si A augmente, alors B diminue") lorsque les paramètres A et B sont connus. L'algorithme 1 est le pseudo-code de cette fonction, c'est-à-dire le terme de perte calculé pour chaque règle de relation monotone négative.

Algorithm 1 Loi sur les relations négatives

Require: Tenseurs x et y
Ensure: $\mathcal{L}_{NegRel}$
procedure PENALTY(x, y) ▷ La pénalité associée à x
  Initialize *tape* to record operations
  Record operations on x
  $y \leftarrow$ ModelOutput(x)
  $dx \leftarrow$ ComputeGradient(y, x) *//Première dérivée de y par rapport à x*
  $n \leftarrow$ SizeOfFirstAxis(x)
  state_indicator $\leftarrow$ IndicatorTensors($dx > 0$)
  loss $\leftarrow \frac{\text{state\_indicator} \times dx^2}{n}$
  $\mathcal{L}_{NegRel} \leftarrow$ ReduceToSumOfTensors($loss$)
  return $\mathcal{L}_{NegRel}$ ▷ La pénalité associée à la relation monotone négative
end procedure

Pour expliquer le pseudo-code, considérons ce qui suit :

1. **Initialize** *tape* **to record operations** : Cette ligne permet de garder une trace des opérations effectuées sur les tenseurs (via la variable *tape*). Ceci est essentiel pour la différenciation automatique utilisée plus tard pour calculer les gradients.
2. **Record operations on** x : Cette ligne spécifie que les opérations effectuées sur le tenseur x doivent être enregistrées via la variable *tape*. Ceci est nécessaire pour calculer les gradients de x.
3. $y \leftarrow$ **ModelOutput**(x) : Cette ligne calcule la sortie y d'un modèle d'apprentissage automatique en fonction de l'entrée x.
4. $dx \leftarrow$ **ComputeGradient**(y, x) : Cette ligne calcule la dérivée première dx de la sortie y par rapport à x, en utilisant les opérations enregistrées sur la variable *tape*.
5. $n \leftarrow$ **SizeOfFirstAxis**(x) : Cette ligne calcule n, la taille du tenseur x. Elle est utilisée ultérieurement pour normaliser la perte.
6. **state indicator** $\leftarrow$ **IndicatorTensor**($dx > 0$) : Cette ligne crée un tenseur qui a la même forme que dx appelé indicateur d'état. L'élément correspondant à celui de dx est défini à 1 s'il est supérieur à 0, sinon il est défini à 0. L'objectif de cet indicateur d'état est de refléter si la règle a été correctement respectée ou non pour chaque valeur de x. En effet, lorsque la règle "si A augmente, B diminue" est observée, alors dx est inférieur à 0 (i.e. $\frac{\partial B}{\partial A} < 0$). Or, il est logique de ne pas ajouter de pénalité lorsque cette contrainte est bien respectée d'où la mise de l'élément à 0.

7. **loss** $\leftarrow \frac{\textbf{state indicator} \times dx^2}{n}$: Cette ligne calcule la perte en élevant au carré dx, en le multipliant par l'indicateur d'état (soit 0, soit 1), puis en le divisant par n. Si l'indicateur d'état est fixé à 1, cela aura pour effet d'ajouter une pénalité, égale au carré de la dérivée, à la fonction de perte, puisque la règle n'est pas respectée. Si l'indicateur d'état est fixé à 0, aucune pénalité ne sera ajoutée à la fonction de perte car on ne veut pas pénaliser un modèle qui respecte bien les règles.
8. $\mathcal{L}_{NegRel} \leftarrow$ **SumOfTensor**(**loss**) : Cette ligne additionne tous les éléments du tenseur $loss$ pour obtenir une valeur scalaire unique, qui représente la perte totale pour cette règle (i.e. $\mathcal{L}_k^i$).
9. **return** $\mathcal{L}_{NegRel}$: Cette ligne renvoie la perte totale calculée d'une instance de la loi de relation monotone négative en tant que résultat de l'algorithme.

Grâce à cette méthodologie, les fonctions de perte physique ne sont écrites qu'une seule fois en Python et peuvent être réutilisées dans différents projets. L'objectif final est de disposer d'une ontologie de la physique suffisamment complète pour que les modifications ne soient nécessaires que dans le cadre de l'ontologie du domaine, qui dépend de chaque cas d'application.

5.4.2/ Mise en oeuvre de l'expérimentation

Le framework OPIML proposé est développé en utilisant Python v3.9, Keras v2.10, Protégé v5.5, et OwlReady2 v0.4 sur un Mac mini avec un processeur Apple M1 et 8 Go de RAM.

Pour appliquer OPIML sur les données de durée de vie des matériaux, nous avons tout d'abord dû identifier les fonctions abstraites associées au sujet (la règle de relation monotone positive et sa relation inverse) et les formaliser dans l'ontologie des lois physiques. Ces deux lois sont particulièrement intéressantes puisqu'elles peuvent s'appliquer à d'autres domaines comme le BIM [269].

Ces fonctions abstraites peuvent ensuite être appliquées à des règles spécifiques au contexte grâce à l'usage d'une ontologie de domaine. En effet, c'est dans cette dernière que sont représentées toutes les règles associées à chaque cas d'usage.

Les fonctions équations correspondantes sont établies à partir de l'ensemble de ces connaissances formalisées récupérées en Python grâce à la bibliothèque Owlready2 pour chaque cas d'application spécifique. Elles peuvent dès lors être mobilisées dans le code Python qui représente la fonction de perte d'un réseau de neurones.

Grâce à cette approche, les connaissances formellement documentées dans l'ontologie sont automatiquement traduites en Python, ce qui permet de les intégrer facilement dans la phase d'apprentissage d'un algorithme de machine learning (par le biais de la fonction

de perte).

5.4.2.1/ Description des données

Contrairement à la première démonstration présentée dans la section 5.3, l'expérience menée ici n'est pas restreinte à l'utilisation d'une base de données sur le fil d'acier (Steel Wire) mais étudie également l'alliage d'aluminium 2024-T4 (2024-T4) et le fil d'aluminium recuit (AAW) présenté par [37]. Le nombre de données disponibles pour chaque ensemble est très limité. Il existe 75 individus dans le jeu du fil d'acier, 252 individus pour l'alliage d'aluminium 2024-T4 et 200 individus pour le fil d'aluminium recuit. Ces trois jeux de données vont nous permettre de créer un modèle capable d'estimer la durée de vie moyenne d'un matériau (courbe de Wöhler) tout en rendant compte des diverses spécificités liées à chaque individu.

L'objectif du modèle est de construire une fonction capable de prédire le moment où le matériau va se rompre en fonction de chaque individu tout en respectant les trois lois physiques exprimées dans le Tableau 5.1. À noter que l'ensemble des données sont transformées en échelle logarithmique et normalisées selon les recommandations de [37].

5.4.2.2/ Règles spécifiques au contexte

L'ontologie du domaine permet d'instancier une nouvelle situation (par exemple, les contraintes liées à la prédiction de la durée de vie en fatigue de l'acier) avec les règles associées, comme l'illustre la figure 5.18.

Plusieurs règles peuvent être associées à chaque situation. Nous utilisons ici les règles présentées dans le Tableau 5.1. La Figure 5.19 illustre comment est représentée dans l'ontologie la première règle que le modèle doit vérifier (i.e. à mesure que l'amplitude de stress augmente, la durée de vie moyenne d'un matériau diminue). Cette règle est une instance de la classe "negative relationship" [15] dont le facteur A (ici "contrainte (Mpa) log") et la valeur cible B (ici "durée de vie log") sont données. Cette règle s'applique à la valeur moyenne de B ce qui est matérialisé par l'*ObjectProperty* "statisticalFunctionToApply" [16]. Les variables A et B correspondent aux noms des colonnes de l'ensemble de données. Enfin, comme cette première règle est identifiée comme appartenant à la classe "negative relationship", le moteur d'inférence en déduit qu'il doit utiliser la fonction "negative relationship equation" [17] pour appliquer cette contrainte dans la fonction de perte.

L'ontologie de domaine, qui renferme les situations et leurs règles correspondantes, est

15. en français "relation monotone négative"
16. Cet *ObjectProperty* est un paramètre optionnel, dans le cas où il n'est pas renseigné, la valeur cible n'est pas altéré par ce paramètre
17. en français "équation de la relation monotone négative"

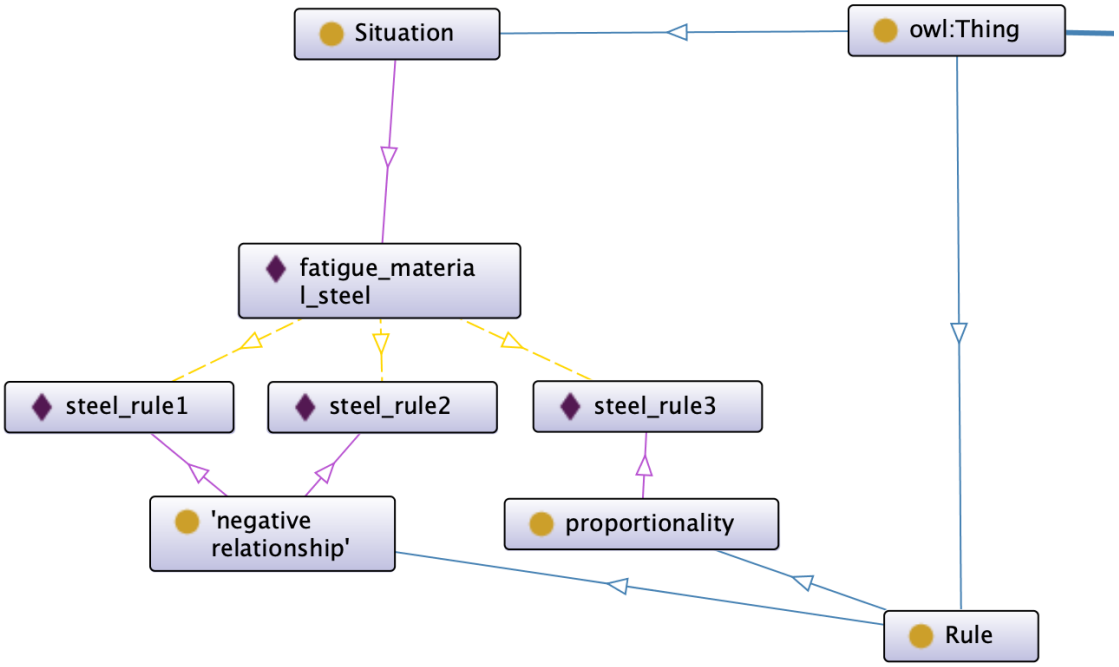

FIGURE 5.18 – Ensemble des règles associées au contexte de l'estimation de la durée de vie d'un matériau

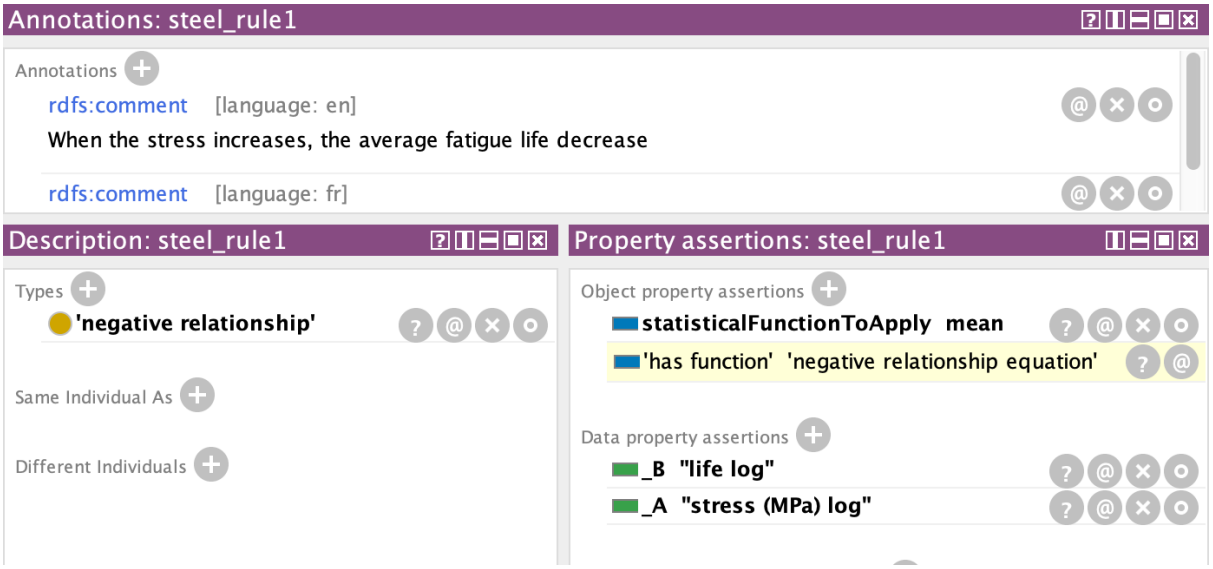

FIGURE 5.19 – Première règle pour estimer la durée de vie de l'acier en fonction d'un stress appliqué

par la suite utilisée dans le code de programmation d'un réseau de neurones à l'aide de la bibliothèque OwlReady2. Cette bibliothèque gère les inférences en utilisant le moteur d'inférence Pellet pour les relier aux fonctions physiques abstraites, puis, par transitivité, aux fonctions Python employées pour calculer le terme $\mathcal{L}_k^i$ de chacune (cf. section 5.4.1.3).

Ce processus permet à l'expert de décrire ses règles dans l'ontologie du domaine sans avoir à se préoccuper de la programmation de chacune des fonctions Python associées.

5.4.3/ Évaluation

5.4.3.1/ Protocole d'évaluation

Un protocole d'évaluation du modèle est mis en place en suivant les différentes étapes recommandées dans la Section 4.4. À l'exception de la septième étape puisque notre exemple n'a pas pour objectif d'être déployé en production et la nature peu altérable des données dans le temps.

L'évaluation de ce modèle de prédiction du moment de rupture d'un matériau en fonction du stress, nécessite l'examen de plusieurs dimensions comme l'exactitude, la pertinence du modèle, la cohérence, la robustesse, l'efficacité, la réutilisabilité. D'autres dimensions, dans ce contexte, ne sont en revanche pas nécessaires à l'évaluation du modèle. L'équité n'est pas une propriété qui a besoin d'être vérifiée, les données n'étant pas relatives aux êtres humains. Pour ces mêmes raisons, la confidentialité des données et leur sécurité ne seront pas évaluées. La facilité d'utilisation n'est pas non plus évaluée car notre expérimentation ne comporte pas d'interface homme-machine. Quant à l'interprétabilité, elle est ici très limitée puisqu'il n'y a qu'un seul facteur dans nos jeux de données qui puisse expliquer la valeur cible.

Exactitude Le modèle d'apprentissage est une régression supervisée. Les métriques permettant d'évaluer l'exactitude de ce type de problématique sont bien connues et déjà mentionnées dans le Chapitre 4. Pour notre contexte spécifique, nous en avons sélectionné trois afin d'avoir un panel d'indicateurs différents : le coefficient de détermination (R^2), l'erreur moyenne absolue en pourcentage ($MAPE$) et l'erreur quadratique moyenne ($RMSE$). Les prédictions obtenues, c'est-à-dire la valeur moyenne du moment de rupture du matériau en fonction du stress subit doivent également montrer l'intervalle de confiance associé au seuil de 95%.

Pertinence du modèle La pertinence du modèle s'évalue bien souvent à l'aide d'un jeu de donnée dédié à la validation du modèle. Ces données ne sont pas utilisées lors de l'entraînement du modèle, elles sont gardées inconnues pour pouvoir servir lors de l'évaluation. L'écart entre les mesures d'exactitude obtenues sur les données d'entraînement et les données de validation est un indicateur clé de la performance du modèle. Si le modèle s'entraîne parfaitement sur les données d'entraînement mais obtient des résultats nettement inférieurs sur les données de validation, cela peut indiquer un surapprentis-

sage, c'est-à-dire que le modèle est trop spécifique aux données d'entraînement et ne généralise pas correctement. L'utilisation de données de validation permet donc d'évaluer la capacité d'un modèle à généraliser sur des exemples qu'il n'a pas encore vus [18].

Dans le cas présent, le nombre de données dans chaque ensemble étant limité nous avons opté pour la mise en place d'une validation manuelle pendant l'entraînement du modèle via Keras. Le jeu de validation est composé en prenant environ 20% des données de chaque ensemble de manière aléatoire.

Cohérence L'évaluation de la cohérence entre les résultats et les connaissances préalables a été permise par la création d'un score d'incohérence que le modèle s'efforce de minimiser lors de son entraînement. En effet, pour ce type de cohérence, c'est-à-dire celle la cohérence physique, le plus simple est de créer un score capable de mesurer le nombre de fois où une règle physique n'a pas été respectée et lorsque c'est le cas d'ajouter une pénalité au score reflétant cette erreur. On peut même pondérer chaque règle ou bien l'ensemble des règles pour refléter leur importance dans un système.

En ce qui concerne cette expérimentation, le score d'incohérence mis au point est celui présenté l'équation 5.9. Il s'agit du terme global $L_{knowledge}$ détaillée plus haut dans le paragraphe 5.3. Plus ce score d'incohérence est proche de zéro, plus le modèle respecte les lois physiques.

Robustesse Le critère de robustesse fait référence à la capacité d'un modèle à maintenir des performances stables, cohérentes et fiables face à des variations, des perturbations ou des incertitudes dans les données ou l'environnement. Afin de vérifier ce critère de robustesse, nous avons introduit un léger bruit dans les données servant à l'entraînement du modèle, puis nous avons observé la modification que cela engendre sur le score $RMSE$ de chaque modèle.

Efficacité L'efficacité de chaque modèle peut se mesurer grâce au temps d'entraînement de chacun. Le nombre de données d'entraînement étant très limité, nous n'avons pas rencontré de problématique particulière concernant un temps d'entraînement beaucoup trop long. En effet, la formation de chaque modèle s'est achevée en moins de 2 minutes, comme en témoigne la section suivante qui expose les résultats. Lorsque l'on parle de modèles d'apprentissage, il est courant de constater que le temps d'exécution est négligeable pour chacun d'entre eux et ce malgré un temps d'entraînement parfois

18. Dans certains cas, le jeu de validation est utilisé dans le cadre de la validation croisée pour obtenir une estimation plus robuste de la performance d'un modèle. La validation croisée divise les données en plusieurs ensembles de formation et de validation, permettant ainsi de mieux évaluer la capacité de généralisation du modèle.

long. C'est pour cette raison que nous n'avons pas jugé pertinent de nous intéresser au temps d'exécution dans notre cas.

Réutilisabilité Le critère de réutilisabilité est plus complexe à évaluer car il est difficile de mettre en place une métrique capable de donner une mesure objective de ce critère. Il faut donc tout d'abord déterminer ce qu'on entend par réutilisabilité du modèle dans ce contexte. La plupart des éléments ont été réutilisés pour les trois jeux de données sur la fatigue des matériaux puisque ces derniers se ressemblent très fortement. Ce qui nous intéresse ici est de savoir dans quelle mesure certains éléments peuvent également être utilisés dans un contexte totalement différent, comme la prédiction de la température dans un bâtiment par exemple.

5.4.3.2/ Résultats

Afin de pouvoir comparer les résultats obtenus, trois modèles utilisant la même architecture de réseau de neurones, à savoir un réseau entièrement connecté contenant deux couches cachées avec 16 neurones pour chacune, ont été mis au point.

Le premier est un modèle de base, appelé "baseline", dont la fonction de perte ne comprend pas de terme $\mathcal{L}_{knowledge}$, il n'est donc pas informé par de la connaissance. L'objectif est de vérifier si les autres modèles grâce à l'apport de connaissance performent mieux (sur différentes métriques) ou non et dans quelle proportion. Le deuxième est un modèle informé par la connaissance, nommé PIML, avec une fonction de perte définie par l'équation 5.9 dont la fonction de perte a été entièrement codée en Python. Enfin, le troisième est un modèle OPIML identique au précédent, mais pour lequel la connaissance physique est entièrement fournie par une ontologie comme décrit dans les paragraphes précédents. En théorie, le fait que les connaissances soient issues d'une ontologie, et non purement codées en Python, ne devrait pas impacter les performances obtenues par rapport au deuxième modèle. La comparaison entre ces deux modèles informés permet de vérifier cette hypothèse.

Exactitude Dans le cas des trois premières métriques (R^2, $RMSE$ et $MAPE$), les deux modèles informés par la connaissance (modèle PIML et modèle OPIML) pour les ensembles de données Wire Steel et 2024-T4 sont légèrement plus performants que le modèle de base (cf. Tableaux 5.2, 5.3). En revanche, ce n'est pas le cas pour l'ensemble de données AAW, pour lequel le modèle de base est cette fois-ci un peu plus performant (cf. Tableau 5.4). Les écarts de performances ne sont pas très élevés et il reste difficile dans ces conditions d'affirmer qu'un des deux paradigmes de programmation (modèle informé ou non) est meilleur que l'autre. Le score d'incohérence est sûrement plus utile

pour les départager.

TABLE 5.2 – Mesures d'évaluation pour le jeu de données sur le fil d'acier

Steel Wire	**Baseline**	**PIML** ($\lambda = 770$)	**OPIML** ($\lambda = 770$)
R^2	0.76	**0.77**	**0.77**
RMSE	1.428	**1.415**	**1.415**
MAPE	**0.058**	0.059	0.059
Pertinence	0.7844	**0.7267**	**0.7267**
$L_{knowledge}$	2.25	**0**	**0**
Robustesse	0.0013	**-0.0061**	**-0.0061**
time (en secondes)	**20**	40	37

TABLE 5.3 – Mesures d'évaluation pour le jeu de données sur 2024-T4

2024-T4	**Baseline**	**PIML** ($\lambda = 500$)	**OPIML** ($\lambda = 500$)
R^2	0.92	**0.93**	**0.93**
RMSE	0.512	**0.459**	**0.459**
MAPE	0.02	**0.02**	**0.02**
Pertinence	0.0371	**0.0218**	**0.0218**
$L_{knowledge}$	3.9	**0**	**0**
Robustesse	0.022	**0.0115**	**0.0115**
time (en secondes)	**24**	71	68

TABLE 5.4 – Mesures d'évaluation pour le jeu de données sur AAW

AAW	**Baseline**	**PIML** ($\lambda = 1000$)	**OPIML** ($\lambda = 1000$)
R^2	**0.94**	0.93	0.93
RMSE	**0.374**	0.386	0.386
MAPE	**0.027**	0.028	0.028
Pertinence	**-0.0314**	-0.0166	-0.0166
$L_{knowledge}$	0.51	**0**	**0**
Robustesse	**-0.0056**	0.0077	0.0077
time (en secondes)	**54**	103	104

Pertinence du modèle La pertinence du modèle a été mesurée par l'écart du score *RMSE* entre les jeux de validation et les jeux d'entraînement. Pour calculer cette pertinence, nous avons établi la métrique suivante :

$$Pertinence = \frac{1}{n}\sum_{i=0}^{n} RMSE_{validation} - \frac{1}{n}\sum_{i=0}^{n} RMSE_{training} \tag{5.11}$$

Où n est le nombre d'epoch d'entraînement du modèle, $RMSE_{validation}$ est le score $RMSE$ pour chaque epoch sur le jeu de donnée de validation et $RMSE_{training}$ est le score $RMSE$ pour chaque epoch sur le jeu de donnée d'entraînement.

En règle générale, le score $RMSE$ est plus élevé sur le jeu de donnée de validation que sur le jeu de donnée d'entraînement puisque le modèle ne connaît pas les données du jeu de validation. Plus le score de $Pertinence$ est proche de 0, moins l'écart entre $RMSE_{validation}$ et $RMSE_{training}$ est grand. Cela signifie aussi qu'il y a moins de surapprentissage.

Le score de $Pertinence$ de chaque modèle est présenté dans chacun des trois Tableaux 5.2, 5.3 et 5.4. Pour "Steel Wire" et "2024-T4", ce sont les modèles informés qui affichent le moins de surapprentissage. Néanmoins, les modèles basés sur l'ensemble de données "Steel Wire" semblent davantage enclins au surapprentissage que les autres, ce phénomène pouvant être attribué au nombre limité d'individus présents (75 au total) dans ce jeu-là.

Cohérence Dans le cas des deux modèles informés, le terme $L_{knowledge}$ est toujours égal à 0, ce qui signifie que ces modèles sont compatibles avec les règles dont la fonction de perte est informée. En revanche, ce même score est toujours supérieur à 0 dans le cas du modèle de base, ce qui signifie qu'il n'est pas totalement compatible avec les lois physiques relatives au contexte. Ce score, permet de conclure que les deux modèles informés sont susceptibles de mieux se généraliser sur de nouvelles données, résultat par ailleurs démontré en détail par [37].

Robustesse La robustesse d'un modèle se manifeste par sa capacité à ne pas être significativement affecté par l'ajout de bruit aux données. Pour effectuer cette évaluation, nous avons ré-entraîné chaque modèle en utilisant des données d'entraînement contenant du bruit, généré à l'aide de la fonction numpy.random.normal(loc=0, scale=0.05, size=n), où n représente la taille de l'ensemble de données d'entraînement.

Ensuite, nous avons calculé le score $RMSE$ sur les données de test non bruitées, que nous désignons ici par $RMSE_{noise}$. Pour créer un score de robustesse, nous avons calculé la différence entre le score $RMSE$ initial et le score $RMSE_{noise}$ via l'équation suivante :

$$Robustesse = RMSE_{noise} - RMSE \tag{5.12}$$

Ainsi, plus le score de $Robustesse$ est minime, moins un modèle est perturbé par l'ajout

de bruit dans les données. Si le score est inférieur à zéro, cela signifie que le modèle est même meilleur sur le jeu de test avec un entraînement réalisé sur des données bruitées.

Ce score de *Robustesse* est présenté dans les Tableaux 5.2, 5.3 et 5.4. Dans l'ensemble les modèles ne sont pas très perturbés par l'ajout de bruit dans les données d'entraînement. Une fois encore, ce score est légèrement meilleur pour le modèle "Baseline" du jeu de données "AAW" ce qui n'est pas le cas pour les deux autres jeux de données.

Efficacité En ce qui concerne les temps d'entraînement, tous relativement court puisqu'inférieurs à deux minutes, c'est toujours les modèles dits "baseline" qui l'emportent. Effectivement, comme indiqué dans les tableaux synthétisant les évaluations effectuées, le temps d'apprentissage est réduit de moitié pour les modèles utilisant des connaissances par rapport à ceux qui n'en utilisent pas. L'ajout de contrainte dans les modèles augmente le temps d'apprentissage de ceux-ci. En revanche, il n'y a pas une différence significative entre les réseaux de neurones informés via du code Python et ceux informés par l'ontologie.

Réutilisabilité La réutilisabilité d'un modèle, comme discuté précédemment, est plus difficile à réaliser de manière objective. Toutefois, si on se base sur la réutilisation d'éléments du modèle dans un contexte différent de celui de l'étude de la fatigue des matériaux, seuls les modèles OPIML ont un élément réutilisable : les connaissances physiques.

En effet, le framework OPIML a été créé dans le but de permettre une réutilisation des connaissances issues de la physique. Par conséquent les équations différentielles partielles peuvent s'adapter sur d'autres problématiques (comme celle de la gestion de température dans les bâtiments). Concernant les autres modèles, la réutilisabilité est beaucoup plus limitée, le code Python nécessitant beaucoup de modifications pour s'adapter à un domaine différent de celui étudié ici.

L'objectif principal de cette expérimentation est de proposer une méthode générique de formalisation des connaissances. De ce point de vue, nous sommes en mesure de présenter des résultats identiques pour les modèles PIML et OPIML, comme le montrent les tableaux 5.2, 5.3 et 5.4. Cela confirme notre hypothèse : l'utilisation d'ontologies comme source de connaissances n'influe pas sur les résultats obtenus. Le framework OPIML peut dès lors être utilisé pour ajouter plus simplement des contraintes dans un algorithme d'apprentissage. La formalisation des connaissances facilite leur réutilisation dans différents contextes, automatise l'ajout de contraintes physiques dans les réseaux de neurones pour les non-spécialistes de l'apprentissage automatique et limites ainsi les erreurs de programmation.

5.5/ Conclusion

Ce chapitre présente OPIML, un framework de *Physics-Informed Machine Learning* dans lequel la connaissance est formalisée sous la forme d'ontologies. Une approche générique modélisant l'ensemble des règles physiques requises dans deux ontologies, en utilisant l'estimation de durée de vie des matériaux comme domaine d'application pour notre expérimentation [37] a été présentée.

Ces deux ontologies nous permettent (1) de déterminer les règles que l'application doit suivre dans un contexte particulier, et (2) d'associer ces règles à la loi physique générique appropriée traduite ensuite en équation mathématique. Cette équation peut alors être automatiquement transformée en code de programmation Python pour être ajoutée à la fonction de perte d'un réseau de neurones via le terme $\mathcal{L}_{knowledge}$.

Il est important de noter que ces règles abstraites peuvent être utilisées dans d'autres domaines, comme la prévision de la température dans un bâtiment [269].

L'objectif principal de ce travail de formalisation est de faciliter la conception de modèles de réseaux de neurones informés en faisant gagner du temps aux concepteurs, pour qui la transformation des règles métier en contraintes peut parfois s'avérer compliquée.

Cette expérimentation a également été l'objet d'une évaluation des résultats. En particulier d'une évaluation de la cohérence, par le biais d'un score d'incohérence, qui a démontré que les réseaux de neurones informés par les connaissances, contrairement aux autres, respectaient bien les lois physiques. Cette évaluation a également permis de mettre en avant l'intérêt d'utiliser des ontologies pour formaliser la connaissance afin d'accroître la réutilisabilité de certains éléments du système d'apprentissage automatique.

5.5.1/ Challenges

Le travail présenté ici est néanmoins amené à être étendu ultérieurement car il présente plusieurs limites. Tout d'abord, nous avons utilisé une architecture simple de réseau neuronal entièrement connecté qui contient deux couches cachées (avec 16 neurones pour chacune des couches cachées) comme décrit par [37]. Il appartient souvent au scientifique des données de choisir cette architecture, qui peut changer en fonction du contexte. Une évolution possible consisterait à décrire formellement cette architecture dans l'ontologie du domaine, afin qu'elle puisse être modifiée plus facilement. Il s'agit d'une technique qui pourrait être développée à l'avenir. De plus, il serait dommage de se limiter uniquement aux réseaux de neurones dans ce framework, d'autres algorithmes d'apprentissage automatique nécessitant une fonction de perte comme SVM pourrait également être utilisée.

Jusqu'à présent, un poids global a été attribué à l'ensemble des règles relatives aux contraintes physiques. Il est nécessaire de modifier le code que nous avons développé pour pouvoir ajuster individuellement le poids associé à chaque règle dans la fonction de perte, d'ailleurs ce poids devrait être défini directement dans l'ontologie du domaine pour laisser aux experts métiers le soin de le déterminer.

De plus, dans l'état actuel des choses le score de chaque contrainte ajoutée à la fonction de perte lorsqu'elle est violée est arbitrairement égal au carré de sa dérivée. À l'avenir, il devrait être possible de définir le mode de calcul de ce score dans l'ontologie, encore une fois pour pouvoir l'adapter en fonction de chaque contexte particulier.

L'ontologie physique existante doit également être complétée par des règles physiques supplémentaires afin d'améliorer son applicabilité dans divers projets. Ce processus implique l'identification et la compréhension de nouvelles règles physiques, qui servent de base au développement de règles abstraites. Par exemple en ajoutant les lois relatives à la cinétique chimique qui ont des applications dans de nombreux domaines comme la qualité de l'eau ou de l'air ainsi que la prévision du trafic routier [278]. Ce type de lois sont particulièrement utiles lorsqu'on étudie l'évolution dynamique de fluides au cours du temps.

Enfin, nous souhaitons étendre notre champ d'application au-delà de l'incorporation de règles physiques ; nous visons à incorporer divers autres types de règles, y compris des règles juridiques, éthiques ou relatives à certains aspects métier. Pour ce faire, nous devons explorer en profondeur toutes les transformations possibles [19] pour obtenir des contraintes qui peuvent être incorporées dans un système d'apprentissage automatique, et réaliser ces transformations par le biais d'ontologies.

19. telle que la t-norm utilisée en logique floue

III

Conclusion

6

Conclusion générale

6.1 Travaux effectués 153
6.2 Perspectives 154
6.2.1 Vers une IA de confiance 155
6.2.1.1 L'explicabilité des modèles d'IA 155
6.2.1.2 Les autres paramètres de la confiance 155
6.2.2 Simulations de phénomènes physiques complexes 156
6.2.2.1 L'IA pour la Smart City et l'industrie 156
6.2.2.2 Les modèles 3D 157
6.2.3 Modularisation des systèmes d'IA 157

Aujourd'hui, l'intelligence artificielle impact le quotidien de tous les citoyens, et pas uniquement en leur permettant d'ajouter des filtres vidéos sur leurs réseaux sociaux préférés. Depuis les cinq dernières années l'intérêt pour l'IA hybride, des systèmes combinant de l'apprentissage automatique avec du symbolique, ne cesse de croître. En particulier car l'intelligence artificielle ne se cantonne plus à réaliser des prédictions sur des séries temporelles uni-factorielles mais bien à prendre des décisions dans des environnements beaucoup plus complexes avec des enjeux importants comme la santé, l'industrie ou la Smart City.

Déployer des modèles d'intelligence artificielle dans le monde réel implique que ces derniers soient en capacité de respecter les différentes contraintes inhérentes à la réalité physique. Les concepteurs doivent également veiller à ce que leurs systèmes respectent bien les lois juridiques en vigueurs pour des questions de légalité. L'intelligence artificielle a désormais une responsabilité sociale qui lui demande d'avoir une certaine éthique et de garantir entre autres l'équité entre les personnes, le respect de la vie privée [1] et une certaine transparence.

Le raisonnement inductif, représenté par l'apprentissage automatique en intelligence artificielle, ne peut à lui seul, en l'état actuel des choses, résoudre l'ensemble de ces problématiques. L'apprentissage automatique est très dépendant de la qualité des données avec lesquelles l'algorithme est entraîné. Il est rare d'avoir des données adaptées en quantité suffisante pour représenter tous les cas possibles dans le monde réel, tout en étant parfaitement correctes, récentes et cohérentes entre elles.

L'usage d'un raisonnement déductif, comme ceux utilisés en logique grâce aux ontologies, peut apporter la rigueur et la connaissance dont les systèmes d'apprentissage automatique ont besoin. L'ajout de connaissances dans les raisonnements inductifs leur permet dès lors de respecter certaines contraintes reflétant notre univers matériel. L'intention d'améliorer l'apprentissage en incorporant des connaissances relève d'un domaine spécifique appelé "Apprentissage Automatique Informé" (*Informed Machine Learning*).

L'avènement de l'intelligence artificielle hybride mêlant raisonnement inductif et déductif va très certainement permettre d'améliorer les systèmes d'apprentissage automatique actuels. La question de leur évaluation est centrale, car pour être plus largement utilisée dans des applications complexes du monde réel, l'intelligence artificielle doit être en capacité de prouver sa qualité. Au-delà de la performance par rapport aux données, une étude de la cohérence par rapport aux connaissances préalables est nécessaire. En effet, l'usage de connaissance est souvent lié à l'ajout de contrainte dans les modèles, or il est important de s'assurer qu'elles ont bien été respectées pour garantir une plus grande fiabilité de ce dernier.

L'évaluation de la cohérence occupe une place centrale au sein de cette thèse. Par le

1. qui est aussi une obligation en Europe via le RGPD

biais de cette évaluation, nous avons également mis en lumière la question de l'évaluation des systèmes d'apprentissage automatique de manière plus générale, souvent axée sur la performance sans accorder suffisamment d'attention au respect des contraintes du monde réel.

6.1/ Travaux effectués

Le Chapitre 2 explique le contexte des travaux de recherches qui ont été menés. Il met l'accent sur les besoins de la Smart City en termes de système d'information en précisant pourquoi l'usage d'intelligence artificielle hybride devient nécessaire dans cet environnement hétérogène fortement contraint. Cette nécessité découle de la diversité des données, exigeant des méthodes d'alignement structurel, schématique et sémantique pour formaliser la compréhension du contexte de production des données, essentiel pour lever toute ambiguïté dans l'exploitation de ces données dans des contextes politiques ou sociaux. De plus, l'évaluation de l'apprentissage automatique implique l'intégration de connaissances préalables du contexte en vue d'améliorer la cohérence des prédictions, notamment pour la vérification de la conformité aux contraintes.

Le Chapitre 3, présente une revue de la littérature systématique des travaux de recherches utilisant des techniques d'apprentissage automatique combinées avec des ontologies. Ce travail a permis de mettre en évidence trois catégories principales d'hybridation des raisonnements inductifs et déductifs ainsi que les sous-catégories qui leur sont associées. Une étude analytique des différents types d'algorithmes d'apprentissage automatique utilisés pour créer ces systèmes a été également présentée. Enfin, ce classement a été mis en perspective par rapport aux modèles de conception présentés par van Bekkum et al. [195] et aux six types de systèmes neuro-symboliques présentés par Kautz [196].

Le Chapitre 4, s'appuyant sur les résultats de la SLR, met en évidence les différents endroits où la connaissance peut s'intégrer aux systèmes d'apprentissage automatique. À partir de cette information, il devient plus facile de comprendre où et comment évaluer la cohérence entre les modèles d'apprentissage et les cohérences préalables. Cette question épineuse de l'évaluation des modèles d'intelligence artificielle hybrides a mis en avant les manquements actuels en termes d'appréciation de la performance globale d'un modèle. Sur la base d'une analyse de plus de 45 articles de la catégorie *Informed Machine Learning*, nous avons mis en évidence que l'examen de la cohérence entre les résultats du modèle final et les connaissances n'est que trop peu réalisé. La plupart des travaux se contentent d'évaluer leurs modèles uniquement à l'aune de la cohérence entre les résultats et les données, sans même s'adapter au contexte de leur cas d'application. De plus, l'évaluation concerne principalement les sorties du modèle final, l'évaluation

des autres étapes de l'apprentissage n'est pas mise en avant. Or, une évaluation systématisée des différentes étapes de conception d'un modèle pourrait permettre de mieux qualifier la donnée, d'utiliser une structure d'algorithme approprié au contexte et de mieux maîtriser l'entraînement. Pour ce faire, nous avons présenté une nouvelle méthodologie d'évaluation d'un modèle d'apprentissage automatique comportant sept étapes. L'objectif est de concevoir des évaluations qui considèrent le contexte du cas d'usage final pour maximiser la qualité du modèle déployé.

Enfin, le Chapitre 5 est une mise en œuvre d'une transformation de connaissance en contrainte dans un modèle d'apprentissage automatique par le biais d'une ontologie. L'intelligence artificielle hybride n'est pas encore largement répandue, elle manque encore d'outils et de méthode pour formaliser les connaissances afin de les mobiliser de manière rapide, efficace et systématique dans les algorithmes d'apprentissage. Plutôt que de formaliser les connaissances dans une base de connaissance, elles sont plus souvent ajoutées directement dans la phase de préparation des données ou directement dans l'algorithme d'apprentissage grâce à du code informatique. Cette manière de faire présuppose que le concepteur de l'algorithme soit capable de comprendre parfaitement l'expertise métier et de la traduire sous forme de code informatique. Ce processus doit être répété pour chaque application et à chaque fois que des connaissances doivent être ajoutées ou mises à jour, ce qui n'est pas optimal en termes de maintenance. C'est pourquoi nous avons mis au point un cadre de conception qui permet aux experts d'exprimer leurs connaissances dans une ontologie, transformée ensuite en équation mathématique intégrée en Python dans la fonction de perte d'un réseau de neurones. Cette expérience a été menée en utilisant deux lois physiques que sont la loi de proportionnalité et la loi de relation inverse avec pour domaine d'application la prédiction de la fatigue d'un matériau. Cette première implémentation pourra par la suite être étendue à d'autres lois physiques, voire à des règles métier exprimée en logique de premier ordre.

6.2/ Perspectives

Les domaines de l'intelligence artificielle hybride et de l'évaluation de ces nouveaux systèmes sont relativement récents, offrant ainsi de nombreuses perspectives d'évolution future. Dans cette section, nous souhaitons en présenter trois qui nous semblent particulièrement intéressantes. La première concerne les avancées dans le domaine de l'intelligence artificielle de confiance (ou *Trustworthy AI* en anglais), où la contribution des connaissances peut jouer un rôle significatif. La seconde aborde les systèmes physiques plus complexes que ceux traités dans cette thèse. Enfin, la dernière soulève la question de la modularisation dans les composants d'un système d'intelligence artificielle.

6.2.1/ Vers une IA de confiance

L'évaluation du respect de ces contraintes, par le biais de l'examen de la cohérence avec des connaissances préalables est un premier pas vers une meilleure assurance de la qualité des modèles. Cependant, la cohérence n'est pas la seule dimension qui mérite d'être mieux étudiée lors de l'évaluation.

L'IA de confiance est une intelligence artificielle que les êtres humains considèrent comme fiable. Pour ce faire, l'IA doit s'assurer qu'un ensemble de dimension comme la robustesse, la généralisation, l'explicabilité, la reproductibilité, l'équité, la préservation de la vie privée et la responsabilité sont bien respectées [279]. Cela rejoint fortement les propos sur l'évaluation tenus dans le Chapitre 4 de cette thèse.

6.2.1.1/ L'explicabilité des modèles d'IA

La question de l'explicabilité des systèmes n'a pas été traitée dans cette thèse, toutefois ce sujet a également un rôle à jouer dans l'adoption de l'intelligence artificielle dans des environnements complexes du monde réel. La branche de l'intelligence artificielle explicable (XAI) aborde conjointement les aspects d'interprétabilité et de transparence des modèles. Cette discipline a gagné en importance ces dernières années, en grande partie en réponse aux pressions légales de plus en plus strictes imposées aux systèmes informatiques en général.

Les problèmes de "boîte noire" (*black-box* en anglais), des modèles peu interprétables souvent générés par l'apprentissage profond, sont fréquemment au centre des préoccupations liées au manque de confiance en l'IA [280]. L'incompréhension vis-à-vis des résultats et le manque de transparence des modèles conduisent les utilisateurs à se méfier des prédictions réalisées.

En incorporant des connaissances préexistantes à travers l'utilisation d'ontologies, un cadre logique et cohérent est mis en place pour expliquer les modèles, assurant ainsi leur conformité avec les connaissances existantes dans le domaine. Cette intégration non seulement renforce la crédibilité des explications, mais elle facilite également une compréhension plus approfondie du processus de raisonnement du modèle [72, 77].

6.2.1.2/ Les autres paramètres de la confiance

L'explicabilité des modèles, bien que très importante, n'est pas la seule dimension qui mérite d'être étudiée si l'on souhaite avoir un système d'IA digne de confiance.

Une évaluation plus approfondie de la robustesse, la généralisation, la reproductibilité, l'équité, la préservation de la vie privée et la responsabilité ne pourra qu'être bénéfique

et faire avancer la recherche vers des IA plus fiables. Chacune de ces dimensions mérite une attention particulière, notamment en encourageant le développement de techniques d'évaluation partagées au sein de la communauté de l'apprentissage automatique.

6.2.2/ Simulations de phénomènes physiques complexes

L'étude et la prise de décision sur des cas réels sont souvent confrontées à la gestion de phénomènes physiques complexes. L'ajout de connaissances relatives aux lois de la thermodynamique ou à la mécanique des fluides peut-être essentielle à la création de simulations fiables.

6.2.2.1/ L'IA pour la Smart City et l'industrie

L'étude de la mécanique des fluides numériques est au cœur de nombreux travaux en *Physics-Informed Machine Learning* [40]. En s'appuyant sur les équations d'Euler ou les équations de Navier-Stokes il est possible de réaliser des simulations dans de nombreux domaines comme l'aéronautique, l'hydrogéologie ou les réseaux électriques intelligents (*Smart Grids* en anglais).

Dans le domaine de l'aéronautique, la modélisation des écoulements d'air est essentielle pour la conception de structures aérodynamiques efficaces. Un enjeu important lorsqu'on connaît l'impact environnemental du transport aérien. Réduire la consommation de carburant des avions, les émissions de CO2 et les nuisances sonores devient un véritable défi pour les années à venir.

La simulation de flux de fluides souterrains peut bénéficier à l'étude en hydrogéologie pour la gestion durable des ressources en eau. Cartographier et comprendre les mouvements de l'eau souterraine, les transferts de contaminants, et les interactions complexes entre les aquifères permet aux hydrogéologues d'assurer une préservation des nappes phréatiques.

L'amélioration des Smart Grids permettra d'optimiser la distribution d'énergie dans les réseaux électriques. La complexité de cette distribution est accentuée par l'intégration croissante de sources d'énergie renouvelable, telles que l'énergie éolienne et solaire, dont la production est intermittente par nature. La gestion efficace de ces sources d'énergie nécessite une compréhension précise de la manière dont l'électricité se propage dans le réseau électrique, tout comme les fluides circulent dans un système de tuyauterie.

L'étude des simulations de ces phénomènes physiques complexes amène à améliorer l'évaluation des modèles d'apprentissage réalisé. Par exemple, le benchmark LIPS évalue quatre aspects du modèle de simulation que sont les performances liées à l'apprentissage, la facilité d'utilisation dans l'industrie, la bonne généralisation et la conformité

aux lois physiques [281].

C'est également une occasion pour améliorer la formalisation des connaissances en mécaniques des fluides étant donné leur pertinence dans divers secteurs industriels.

6.2.2.2/ Les modèles 3D

L'exploration des modèles en 3D constitue une opportunité pour analyser des éléments visuels, images ou vidéos, au sein desquels les lois physiques sont omniprésentes. Cette approche trouve des applications dans divers domaines, notamment en biomécanique, où l'analyse des mouvements, des structures anatomiques, et des processus physiologiques peut bénéficier de la modélisation 3D pour obtenir une compréhension plus approfondie.

Là encore, une amélioration de la formalisation des connaissances et une évaluation de la cohérence des modèles développés peuvent apporter des avantages substantiels au domaine de la santé.

6.2.3/ Modularisation des systèmes d'IA

La modularisation est une décomposition d'un système complexe en modules autonomes interconnectés, une stratégie qui simplifie le développement, la maintenance et favorise la réutilisation de code. Elle offre également une gestion plus fluide des équipes de développement et une meilleure adaptation aux besoins de mise à jour réguliers. Cette méthodologie est un outil fondamental dans la conception de systèmes informatiques plus agiles et adaptables.

Les applications d'intelligence artificielle sont encore bien souvent très monolithiques, leurs différents composants n'étant pas réutilisables dans d'autres cas d'application pourtant similaires. Notre expérimentation faite au Chapitre 5 vise à favoriser la réutilisation de connaissance formelle dans diverses applications, c'est un premier pas vers une meilleure modularisation des systèmes d'intelligence artificielle.

Toutefois, il reste encore beaucoup de travail à réaliser pour les rendre plus combinables entre elles et plus réutilisables. Pour y parvenir, il pourrait être judicieux de s'inspirer de ce qu'a déjà été mis en place dans le domaine du génie logiciel.

Bibliographie

[1] N. J. Nilsson, *The Quest for Artificial Intelligence*. Cambridge University Press, 1 ed., 2009.

[2] P. Wang, "On Defining Artificial Intelligence," *Journal of Artificial General Intelligence*, vol. 10, pp. 1–37, Jan. 2019.

[3] M. Minsky, *The Society of Mind*. New York etc. : Simon and Schuster, Jan. 1986.

[4] D. Dobrev, "A definition of artificial intelligence," 2012.

[5] J. McCarthy, M. L. Minsky, N. Rochester, I. B. M. Corporation, and C. E. Shannon, "A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE," *AI Magazine*, 1955.

[6] A. Kaplan and M. Haenlein, "Siri, Siri, in my hand : Who's the fairest in the land ? On the interpretations, illustrations, and implications of artificial intelligence," *Business Horizons*, vol. 62, pp. 15–25, Feb. 2019.

[7] P. Wang, "Three fundamental misconceptions of Artificial Intelligence," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 19, pp. 249–268, Sept. 2007. Publisher : Taylor & Francis _eprint : https ://doi.org/10.1080/09528130601143109.

[8] S. Russell and P. Norvig, *Artificial Intelligence : A Modern Approach*. Pearson, 3rd edition ed., 2009.

[9] Plato, *The Republic of Plato*. Oxford University Press, 1888.

[10] R. Andrea, C. Stefania, A. Omicini, *et al.*, "Position paper : On the role of abductive reasoning in semantic image segmentation," in *CEUR WORKSHOP PROCEEDINGS*, vol. 3419, pp. 75–84, Sun SITE Central Europe, RWTH Aachen University, 2023.

[11] F. Roudaut, "Comment on invente les hypothèses : Peirce et la théorie de l'abduction," *Cahiers philosophiques*, vol. 150, pp. 45–65, Dec. 2017. 00002 Bibliographie_available : 0 Cairndomain : www.cairn.info Cite Par_available : 0 Publisher : Vrin.

[12] R. Smith, "Aristotle's Logic," in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Metaphysics Research Lab, Stanford University, fall 2020 ed., 2020.

[13] J. Haugeland, *Artificial Intelligence : The Very Idea*. MIT press, 1989.

[14] N. Guarino, D. Oberle, and S. Staab, "What Is an Ontology ?," in *Handbook on Ontologies* (S. Staab and R. Studer, eds.), International Handbooks on Information Systems, pp. 1–17, Berlin, Heidelberg : Springer, 2009.

[15] R. P. Dameri, “Searching for smart city definition : a comprehensive proposal,” *International Journal of computers & technology*, vol. 11, no. 5, pp. 2544–2551, 2013.

[16] A. Camero and E. Alba, “Smart city and information technology : A review,” *cities*, vol. 93, pp. 84–94, 2019.

[17] L. Anthopoulos, “Defining smart city architecture for sustainability,” in *Proceedings of 14th electronic government and 7th electronic participation conference (IFIP2015)*, pp. 140–147, 2015.

[18] C. Knudsen, E. Moreno, B. Arimah, R. Otieno, and O. Ogunsanya, “World Cities Report 2020, The Value of Sustainable Urbanization, Key Findings and Messages,” tech. rep., United Nations Human Settlements Programme, 2020.

[19] Publications Office of the European Union, “Urban europe : Statistics on cities, towns and suburbs : 2016 edition.,” tech. rep., European Union, sep 2016.

[20] A. Cocchia, “Smart and digital city : A systematic literature review,” *Smart city : How to create public and economic value with high technology in urban space*, pp. 13–43, 2014.

[21] E. S. Vergini and P. P. Groumpos, “A review on Zero Energy Buildings and intelligent systems,” in *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–6, July 2015.

[22] A. Kylili and P. A. Fokaides, “European smart cities : The role of zero energy buildings,” *Sustainable Cities and Society*, vol. 15, pp. 86–95, July 2015.

[23] O. Lindholm, H. u. Rehman, and F. Reda, “Positioning Positive Energy Districts in European Cities,” *Buildings*, vol. 11, p. 19, Jan. 2021.

[24] “SET-Plan ACTION n°3.2 Implementation Plan,” June 2018.

[25] M. Pal, A. A. Alyafi, S. Ploix, P. Reignier, and S. Bandyopadhyay, “Unmasking the causal relationships latent in the interplay between occupant’s actions and indoor ambience : A building energy management outlook,” *Applied Energy*, vol. 238, pp. 1452–1470, Mar. 2019.

[26] D. Kolokotsa, D. Rovas, E. Kosmatopoulos, and K. Kalaitzakis, “A roadmap towards intelligent net zero- and positive-energy buildings,” *Solar Energy*, vol. 85, pp. 3067–3084, Dec. 2011.

[27] A. A. Alyafi, J. Guillbaud, P. Reignier, and S. Ploix, “Explanations engine for energy management systems in buildings,” in *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems : Technology and Applications (IDAACS)*, vol. 2, pp. 722–729, Sept. 2017.

[28] P. Cardullo and R. Kitchin, “Being a ‘citizen’ in the smart city : up and down the scaffold of smart citizen participation in Dublin, Ireland,” *GeoJournal*, vol. 84, pp. 1–13, Feb. 2019.

[29] R. Kitchin, P. Cardullo, and F. C. Di, "Citizenship, Justice, and the Right to the Smart City," in *The Right to the Smart City* (P. Cardullo, C. Di Feliciantonio, and R. Kitchin, eds.), pp. 1–24, Emerald Publishing Limited, Jan. 2019.

[30] M. d. Waal and M. Dignum, "The citizen in the smart city. How the smart city could transform citizenship," *it - Information Technology*, vol. 59, pp. 263–273, Dec. 2017.

[31] S. Joss, M. Cook, and Y. Dayot, "Smart Cities : Towards a New Citizenship Regime ? A Discourse Analysis of the British Smart City Standard," *Journal of Urban Technology*, vol. 24, pp. 29–49, Oct. 2017.

[32] J. K. Day and D. E. Gunderson, "Understanding high performance buildings : The link between occupant knowledge of passive design systems, corresponding behaviors, occupant comfort and environmental satisfaction," *Building and Environment*, vol. 84, pp. 114–124, Jan. 2015.

[33] T. M. Mitchell, *Machine Learning*. USA : McGraw-Hill, Inc., 1 ed., 1997.

[34] T. Yu, S. Simoff, and T. Jan, "VQSVM : A case study for incorporating prior domain knowledge into inductive machine learning," *Neurocomputing*, vol. 73, pp. 2614–2623, Aug. 2010.

[35] L. von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy, M. Walczak, J. Garcke, C. Bauckhage, and J. Schuecker, "Informed Machine Learning – A Taxonomy and Survey of Integrating Knowledge into Learning Systems," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2021.

[36] A. Karpatne, G. Atluri, J. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, "Theory-guided data science : A new paradigm for scientific discovery from data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, pp. 2318–2331, Oct. 2017.

[37] T. Zhou, S. Jiang, T. Han, S.-P. Zhu, and Y. Cai, "A physically consistent framework for fatigue life prediction using probabilistic physics-informed neural network," *International Journal of Fatigue*, vol. 166, p. 107234, Jan. 2023.

[38] Z. Cui, T. Gao, K. Talamadupula, and Q. Ji, "Knowledge-augmented deep learning and its applications : A survey." http://arxiv.org/abs/2212.00017, Nov. 2022.

[39] the Business Group Rule, "Manifeste pour les règles métiers - Principes de l'indépendance des règles," 2003.

[40] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, pp. 422–440, May 2021.

[41] M. Jang, D. S. Kwon, and T. Lukasiewicz, "BECEL : Benchmark for Consistency Evaluation of Language Models," *Proceedings of the 29th ...*, 2022. Publisher : aclanthology.org.

[42] Y. Wang, “Concept Algebra : A Denotational Mathematics for Formal Knowledge Representation and Cognitive Robot Learning,” *Journal of Advanced Mathematics and Applications*, vol. 4, pp. 62–87, June 2015.

[43] B. Choi and H. Lee, “An empirical investigation of KM styles and their effect on corporate performance,” *Information & Management*, vol. 40, pp. 403–417, May 2003.

[44] R. Fikes and T. Kehler, “The role of frame-based representation in reasoning,” *Communications of the ACM*, vol. 28, pp. 904–920, Sept. 1985.

[45] P. Hudak, J. Hughes, S. Peyton Jones, and P. Wadler, “A history of Haskell : being lazy with class,” in *Proceedings of the third ACM SIGPLAN conference on History of programming languages*, (San Diego California), ACM, June 2007.

[46] S. Heymans, L. Ma, D. Anicic, Z. Ma, N. Steinmetz, Y. Pan, J. Mei, A. Fokoue, A. Kalyanpur, A. Kershenbaum, E. Schonberg, K. Srinivas, C. Feier, G. Hench, B. Wetzstein, and U. Keller, “Ontology Reasoning with Large Data Repositories,” in *Ontology Management : Semantic Web, Semantic Web Services, and Business Applications* (M. Hepp, P. De Leenheer, A. De Moor, and Y. Sure, eds.), Computing for Human Experience, pp. 89–128, Boston, MA : Springer US, 2008.

[47] J. Mylopoulos, “An overview of Knowledge Representation,” *ACM SIGART Bulletin*, pp. 5–12, Jan. 1981.

[48] J. McCarthy, “Recursive functions of symbolic expressions and their computation by machine, Part I,” *Communications of the ACM*, vol. 3, pp. 184–195, Apr. 1960.

[49] B. Kitchenham, R. Pretorius, D. Budgen, O. Pearl Brereton, M. Turner, M. Niazi, and S. Linkman, “Systematic literature reviews in software engineering – A tertiary study,” *Information and Software Technology*, vol. 52, pp. 792–805, Aug. 2010.

[50] M. Gusenbauer and N. R. Haddaway, “Which academic search systems are suitable for systematic reviews or meta-analyses ? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources,” *Research Synthesis Methods*, vol. 11, no. 2, pp. 181–217, 2020. 00102 _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/jrsm.1378.

[51] M. Usman, N. b. Ali, and C. Wohlin, “A Quality Assessment Instrument for Systematic Literature Reviews in Software Engineering,” *arXiv :2109.10134 [cs]*, Sept. 2021.

[52] S. Palazzo, F. Murabito, C. Pino, F. Rundo, D. Giordano, M. Shah, and C. Spampinato, “Exploiting structured high-level knowledge for domain-specific visual classification,” *PATTERN RECOGNITION*, vol. 112, Apr. 2021. 00001 Place : THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, OXON, ENGLAND Publisher : ELSEVIER SCI LTD Type : Article.

[53] Z. Kuang, X. Zhang, J. Yu, Z. Li, and J. Fan, “Deep embedding of concept ontology for hierarchical fashion recognition,” *NEUROCOMPUTING*, vol. 425, pp. 191–206,

Feb. 2021. 00001 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[54] X. Wang, Y. Mao, X. Wu, Q. Xu, W. Jiang, and S. Yin, "An ATC instruction processing-based trajectory prediction algorithm designing," *NEURAL COMPUTING & APPLICATIONS*, Jan. 2021. 00000 Place : 236 GRAYS INN RD, 6TH FLOOR, LONDON WC1X 8HL, ENGLAND Publisher : SPRINGER LONDON LTD Type : Article ; Early Access.

[55] L. Hong, H. Xu, and X. Shi, "Constructing Ontology of Brain Areas and Autism to Support Domain Knowledge Exploration and Discovery," *INTERNATIONAL JOURNAL OF COMPUTATIONAL INTELLIGENCE SYSTEMS*, vol. 14, no. 1, pp. 834–846, 2021. 00000 Place : 29 AVENUE LAUMIERE, PARIS, 75019, FRANCE Publisher : ATLANTIS PRESS Type : Article.

[56] A. S. Patel, G. Merlino, D. Bruneo, A. Puliafito, O. P. Vyas, and M. Ojha, "Video representation and suspicious event detection using semantic technologies," *SEMANTIC WEB*, vol. 12, no. 3, pp. 467–491, 2021. 00000 Place : NIEUWE HEMWEG 6B, 1013 BG AMSTERDAM, NETHERLANDS Publisher : IOS PRESS Type : Article.

[57] F. Ali, S. El-Sappagh, S. M. R. Islam, D. Kwak, A. Ali, M. Imran, and K.-S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *INFORMATION FUSION*, vol. 63, pp. 208–222, Nov. 2020. 00000 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[58] N. Gomathi and N. P. Karlekar, "Ontology and Hybrid Optimization Based SVNN for Privacy Preserved Medical Data Classification in Cloud," *INTERNATIONAL JOURNAL ON ARTIFICIAL INTELLIGENCE TOOLS*, vol. 28, May 2019. 00000 Place : 5 TOH TUCK LINK, SINGAPORE 596224, SINGAPORE Publisher : WORLD SCIENTIFIC PUBL CO PTE LTD Type : Article.

[59] J.-R. Ruiz-Sarmiento, C. Galindo, J. Monroy, F.-A. Moreno, and J. Gonzalez-Jimenez, "Ontology-based conditional random fields for object recognition," *KNOWLEDGE-BASED SYSTEMS*, vol. 168, pp. 100–108, Mar. 2019. 00000 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[60] W. Zhang, M. Wang, Y. Zhu, J. Wang, and N. Ghei, "A hybrid neural network approach for fine-grained emotion classification and computing," *JOURNAL OF INTELLIGENT & FUZZY SYSTEMS*, vol. 37, no. 3, pp. 3081–3091, 2019. 00000 Place : NIEUWE HEMWEG 6B, 1013 BG AMSTERDAM, NETHERLANDS Publisher : IOS PRESS Type : Article.

[61] B. Makni and J. Hendler, "Deep learning for noise-tolerant RDFS reasoning," *SEMANTIC WEB*, vol. 10, no. 5, pp. 823–862, 2019. 00000 Place : NIEUWE HEM-

WEG 6B, 1013 BG AMSTERDAM, NETHERLANDS Publisher : IOS PRESS Type : Article.

[62] G. Petrucci, M. Rospocher, and C. Ghidini, "Expressive ontology learning as neural machine translation," *JOURNAL OF WEB SEMANTICS*, vol. 52-53, pp. 66–82, Oct. 2018. 00000 Place : PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS Publisher : ELSEVIER SCIENCE BV Type : Article.

[63] Z. Kuang, J. Yu, Z. Li, B. Zhang, and J. Fan, "Integrating multi-level deep learning and concept ontology for large-scale visual recognition," *PATTERN RECOGNITION*, vol. 78, pp. 198–214, June 2018. 00000 Place : THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, OXON, ENGLAND Publisher : ELSEVIER SCI LTD Type : Article.

[64] W. Gao, Y. Chen, A. Q. Baig, and Y. Zhang, "Ontology geometry distance computation using deep learning technology," *JOURNAL OF INTELLIGENT & FUZZY SYSTEMS*, vol. 35, no. 4, pp. 4517–4524, 2018. 00000 Place : NIEUWE HEMWEG 6B, 1013 BG AMSTERDAM, NETHERLANDS Publisher : IOS PRESS Type : Article.

[65] O. Yilmaz, "Matching points of interest with user context : an ANN approach," *TURKISH JOURNAL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCES*, vol. 25, no. 4, pp. 2784–2795, 2017. 00000 Place : ATATURK BULVARI NO 221, KAVAKLIDERE, ANKARA, 00000, TURKEY Publisher : TUBITAK SCIENTIFIC & TECHNICAL RESEARCH COUNCIL TURKEY Type : Article.

[66] I. Donadello and L. Serafini, "Integration of numeric and symbolic information for semantic image interpretation," *INTELLIGENZA ARTIFICIALE*, vol. 10, no. 1, pp. 33–47, 2016. 00000 Place : NIEUWE HEMWEG 6B, 1013 BG AMSTERDAM, NETHERLANDS Publisher : IOS PRESS Type : Article.

[67] K. Pancerz and A. Lewicki, "Encoding symbolic features in simple decision systems over ontological graphs for PSO and neural network based classifiers," *NEUROCOMPUTING*, vol. 144, pp. 338–345, Nov. 2014. 00000 Place : PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS Publisher : ELSEVIER SCIENCE BV Type : Article.

[68] I. Gabriel, V. Negru, and D. Zaharie, "Neuroevolution Based Multi-Agent System with Ontology Based Template Creation for Micromanagement in Real-Time Strategy Games," *INFORMATION TECHNOLOGY AND CONTROL*, vol. 43, no. 1, pp. 98–109, 2014. 00000 Place : KAUNAS UNIV TECHNOL, DEPT ELECTRONICS ENGINEERING, STUDENTU STR 50, KAUNAS, LT-51368, LITHUANIA Publisher : KAUNAS UNIV TECHNOLOGY Type : Article.

[69] Y. Jang, J. Ham, B.-J. Lee, and K.-E. Kim, "Cross-Language Neural Dialog State Tracker for Large Ontologies Using Hierarchical Attention," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 26, pp. 2072–2082, Nov. 2018.

[70] M. Rubiolo, M. L. Caliusco, G. Stegmayer, M. Coronel, and M. G. Fabrizi, "Knowledge discovery through ontology matching : An approach based on an Artificial Neural Network model," *Information Sciences*, vol. 194, pp. 107–119, 2012.

[71] D. Rosaci, "CILIOS : Connectionist inductive learning and inter-ontology similarities for recommending information agents," *Information Systems*, vol. 32, no. 6, pp. 793–825, 2007.

[72] R. Confalonieri, T. Weyde, T. R. Besold, and F. M. d. P. Martín, "Using ontologies to enhance human understandability of global post-hoc explanations of black-box models," *Artificial Intelligence*, vol. 296, p. 103471, 2021.

[73] G. Foo, S. Kara, and M. Pagnucco, "Screw detection for disassembly of electronic waste using reasoning and re-training of a deep learning model," *Procedia CIRP*, vol. 98, pp. 666–671, 2021.

[74] A. Ayadi, A. Samet, F. d. B. d. Beuvron, and C. Zanni-Merk, "Ontology population with deep learning-based NLP : a case study on the Biomolecular Network Ontology," *Procedia Computer Science*, vol. 159, pp. 572–581, 2019.

[75] G. J. Shannon, N. Rayapati, S. M. Corns, and D. C. Wunsch, "Comparative study using inverse ontology cogency and alternatives for concept recognition in the annotated National Library of Medicine database," *Neural Networks*, vol. 139, pp. 86–104, 2021.

[76] H. Hassanzadeh, S. Karimi, and A. Nguyen, "Matching patients to clinical trials using semantically enriched document representation," *Journal of Biomedical Informatics*, vol. 105, p. 103406, 2020.

[77] C. Panigutti, A. Perotti, and D. Pedreschi, "Doctor XAI : An Ontology-Based Approach to Black-Box Sequential Data Classification Explanations," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, (New York, NY, USA), pp. 629–639, Association for Computing Machinery, 2020.

[78] X. Huang, C. Zanni-Merk, and B. Crémilleux, "Enhancing Deep Learning with Semantics : an application to manufacturing time series analysis," *Procedia Computer Science*, vol. 159, pp. 437–446, 2019.

[79] J. Wang, Y. Wu, X. Liu, and X. Gao, "Knowledge acquisition method from domain text based on theme logic model and artificial neural network," *Expert Systems with Applications*, vol. 37, no. 1, pp. 267–275, 2010.

[80] M. S. Zarchi, A. Monadjemi, and K. Jamshidi, "A semantic model for general purpose content-based image retrieval systems," *Computers & Electrical Engineering*, vol. 40, no. 7, pp. 2062–2071, 2014.

[81] F. Ali, S. El-Sappagh, S. M. R. Islam, A. Ali, M. Attique, M. Imran, and K.-S. Kwak, "An intelligent healthcare monitoring framework using wearable sensors and social networking data," *Future Generation Computer Systems*, vol. 114, pp. 23–43, 2021.

[82] Z. Wang, L. Li, M. Song, J. Yan, J. Shi, and Y. Yao, “Evaluating the Traditional Chinese Medicine (TCM) Officially Recommended in China for COVID-19 Using Ontology-Based Side-Effect Prediction Framework (OSPF) and Deep Learning,” *Journal of Ethnopharmacology*, vol. 272, p. 113957, 2021.

[83] A. M. Rinaldi, C. Russo, and C. Tommasino, “A semantic approach for document classification using deep neural networks and multimedia knowledge graph,” *Expert Systems with Applications*, vol. 169, p. 114320, 2021.

[84] E. Amador-Domínguez, E. Serrano, D. Manrique, P. Hohenecker, and T. Lukasiewicz, “An ontology-based deep learning approach for triple classification with out-of-knowledge-base entities,” *Information Sciences*, vol. 564, pp. 85–102, 2021.

[85] H. Liu, Q. Gao, and P. Ma, “Photovoltaic generation power prediction research based on high quality context ontology and gated recurrent neural network,” *Sustainable Energy Technologies and Assessments*, vol. 45, p. 101191, 2021.

[86] M. Keyarsalan and G. A. Montazer, “Designing an intelligent ontological system for traffic light control in isolated intersections,” *Engineering Applications of Artificial Intelligence*, vol. 24, no. 8, pp. 1328–1339, 2011.

[87] M. Mao, Y. Peng, and M. Spring, “An adaptive ontology mapping approach with neural network based constraint satisfaction,” *Journal of Web Semantics*, vol. 8, no. 1, pp. 14–25, 2010.

[88] S. Nayak, A. Zaveri, P. H. Serrano, and M. Dumontier, “Experience : Automated Prediction of Experimental Metadata from Scientific Publications,” *J. Data and Information Quality*, vol. 13, Aug. 2021. Place : New York, NY, USA Publisher : Association for Computing Machinery.

[89] G. Alexandridis, J. Aliprantis, K. Michalakis, K. Korovesis, P. Tsantilas, and G. Caridakis, “A Knowledge-Based Deep Learning Architecture for Aspect-Based Sentiment Analysis,” *INTERNATIONAL JOURNAL OF NEURAL SYSTEMS*, vol. 31, Oct. 2021.

[90] K. Akila, S. I. Priyadharshini, P. Ulaganathan, P. Prempriya, B. Yuvasri, T. S. Praba, and Veeramuthuvenkatesh, “Ontology based multiobject segmentation and classification in sports videos,” *JOURNAL OF INTELLIGENT & FUZZY SYSTEMS*, vol. 41, no. 5, pp. 5399–5409, 2021.

[91] R. Messaoudi, F. Jaziri, A. Mtibaa, F. Gargouri, and A. Vacavant, “Ontology-Driven Approach for Liver MRI Classification and HCC Detection,” *INTERNATIONAL JOURNAL OF PATTERN RECOGNITION AND ARTIFICIAL INTELLIGENCE*, vol. 35, Sept. 2021.

[92] R. Cheng and J. Chen, “A location conversion method for roads through deep learning-based semantic matching and simplified qualitative direction knowledge representation,” *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104400, 2021.

[93] K. Niu, Y. Lu, X. Peng, and J. Zeng, "Fusion of sequential visits and medical ontology for mortality prediction," *Journal of Biomedical Informatics*, vol. 127, p. 104012, 2022.

[94] D. Zhao, D. Xue, X. Wang, and F. Du, "Adaptive vision inspection for multi-type electronic products based on prior knowledge," *Journal of Industrial Information Integration*, p. 100283, 2021.

[95] A. Ahani, M. Nilashi, W. A. Zogaan, S. Samad, N. O. Aljehane, A. Alhargan, S. Mohd, H. Ahmadi, and L. Sanzogni, "Evaluating medical travelers' satisfaction through online review analysis," *Journal of Hospitality and Tourism Management*, vol. 48, pp. 519–537, 2021.

[96] G. Deepak, D. Surya, I. Trivedi, A. Kumar, A. Lingampalli, and S. vijayan, "An artificially intelligent approach for automatic speech processing based on triune ontology and adaptive tribonacci deep neural networks," *Computers & Electrical Engineering*, vol. 98, p. 107736, 2022.

[97] M. Abdollahi, X. Gao, Y. Mei, S. Ghosh, J. Li, and M. Narag, "Substituting clinical features using synthetic medical phrases : Medical text data augmentation techniques," *Artificial Intelligence in Medicine*, vol. 120, p. 102167, 2021.

[98] G. Ye, Y. Li, H. Xu, D. Liu, and S.-F. Chang, "EventNet : A Large Scale Structured Concept Library for Complex Event Detection in Video," in *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, (New York, NY, USA), pp. 471–480, Association for Computing Machinery, 2015.

[99] S. Albukhitan, T. Helmy, and A. Alnazer, "Arabic Ontology Learning Using Deep Learning," in *Proceedings of the International Conference on Web Intelligence*, WI '17, (New York, NY, USA), pp. 1138–1142, Association for Computing Machinery, 2017.

[100] L. Serafini, I. Donadello, and A. d. Garcez, "Learning and Reasoning in Logic Tensor Networks : Theory and Application to Semantic Image Interpretation," in *Proceedings of the Symposium on Applied Computing*, SAC '17, (New York, NY, USA), pp. 125–130, Association for Computing Machinery, 2017.

[101] D. Moussallem, A.-C. Ngonga Ngomo, P. Buitelaar, and M. Arcan, "Utilizing Knowledge Graphs for Neural Machine Translation Augmentation," in *Proceedings of the 10th International Conference on Knowledge Capture*, K-CAP '19, (New York, NY, USA), pp. 139–146, Association for Computing Machinery, 2019.

[102] M. Gaur, A. Alambo, J. P. Sain, U. Kursuncu, K. Thirunarayan, R. Kavuluru, A. Sheth, R. Welton, and J. Pathak, "Knowledge-Aware Assessment of Severity of Suicide Risk for Early Intervention," in *The World Wide Web Conference*, WWW '19, (New York, NY, USA), pp. 514–525, Association for Computing Machinery, 2019.

[103] J. Chakraborty, S. K. Bansal, L. Virgili, K. Konar, and B. Yaman, "OntoConnect : Unsupervised Ontology Alignment with Recursive Neural Network," in *Proceedings*

of the 36th Annual ACM Symposium on Applied Computing, pp. 1874–1882, New York, NY, USA : Association for Computing Machinery, 2021.

[104] D. Jurkevičius and O. Vasilecas, "Ontology Creation Using Formal Concepts Approach," in *Proceedings of the 11th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies*, CompSysTech '10, (New York, NY, USA), pp. 64–70, Association for Computing Machinery, 2010.

[105] B. Zhou, Y. Svetashova, A. Gusmao, A. Soylu, G. Cheng, R. Mikut, A. Waaler, and E. Kharlamov, "SemML : Facilitating development of ML models for condition monitoring with semantics," *Journal of Web Semantics*, vol. 71, p. 100664, 2021.

[106] K. Chung, H. Yoo, and D.-E. Choe, "Ambient context-based modeling for health risk assessment using deep neural network," *JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING*, vol. 11, pp. 1387–1395, Apr. 2020. 00000 Place : TIERGARTENSTRASSE 17, D-69121 HEIDELBERG, GERMANY Publisher : SPRINGER HEIDELBERG Type : Article.

[107] P. Hohenecker and T. Lukasiewicz, "Ontology Reasoning with Deep Neural Networks," *JOURNAL OF ARTIFICIAL INTELLIGENCE RESEARCH*, vol. 68, pp. 503–540, 2020. 00000 Place : USC INFORMATION SCIENCES INST, 4676 ADMIRALITY WAY, MARINA DEL REY, CA 90292-6695 USA Publisher : AI ACCESS FOUNDATION Type : Article.

[108] J. Ren, H. Wang, and T. Liu, "Information Retrieval Based on Knowledge-Enhanced Word Embedding Through Dialog : A Case Study," *INTERNATIONAL JOURNAL OF COMPUTATIONAL INTELLIGENCE SYSTEMS*, vol. 13, no. 1, pp. 275–290, 2020. 00000 Place : 29 AVENUE LAUMIERE, PARIS, 75019, FRANCE Publisher : ATLANTIS PRESS Type : Article.

[109] R. Kumar, H. S. Pannu, and A. K. Malhi, "Aspect-based sentiment analysis using deep networks and stochastic optimization," *NEURAL COMPUTING & APPLICATIONS*, vol. 32, pp. 3221–3235, Apr. 2020. 00000 Place : 236 GRAYS INN RD, 6TH FLOOR, LONDON WC1X 8HL, ENGLAND Publisher : SPRINGER LONDON LTD Type : Article.

[110] S. Evert, P. Heinrich, K. Henselmann, U. Rabenstein, E. Scherr, M. Schmitt, and L. Schroeder, "Combining Machine Learning and Semantic Features in the Classification of Corporate Disclosures," *JOURNAL OF LOGIC LANGUAGE AND INFORMATION*, vol. 28, pp. 309–330, June 2019. 00000 Place : VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS Publisher : SPRINGER Type : Article.

[111] Q. Zhou, P. Yan, H. Liu, and Y. Xin, "A hybrid fault diagnosis method for mechanical components based on ontology and signal analysis," *JOURNAL OF INTELLIGENT MANUFACTURING*, vol. 30, pp. 1693–1715, Apr. 2019. 00000 Place : VAN GODE-

WIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS Publisher : SPRINGER Type : Article.

[112] Z. H. Pozveh, A. Monadjemi, and A. Ahmadi, “FNLP-ONT : A feasible ontology for improving NLP tasks in Persian,” *EXPERT SYSTEMS*, vol. 35, Aug. 2018. 00000 Place : 111 RIVER ST, HOBOKEN 07030-5774, NJ USA Publisher : WILEY Type : Article.

[113] B. Song, Z. Jiang, and L. Liu, “Automated experiential engineering knowledge acquisition through Q&A contextualization and transformation,” *ADVANCED ENGINEERING INFORMATICS*, vol. 30, pp. 467–480, Aug. 2016. 00000 Place : THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, OXON, ENGLAND Publisher : ELSEVIER SCI LTD Type : Article.

[114] S. Rubrichi, S. Quaglini, A. Spengler, P. Russo, and P. Gallinari, “A system for the extraction and representation of summary of product characteristics content,” *ARTIFICIAL INTELLIGENCE IN MEDICINE*, vol. 57, pp. 145–154, Feb. 2013. 00000 Place : PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS Publisher : ELSEVIER SCIENCE BV Type : Article.

[115] C. D. Emele, T. J. Norman, M. Sensoy, and S. Parsons, “Learning strategies for task delegation in norm-governed environments,” *AUTONOMOUS AGENTS AND MULTI-AGENT SYSTEMS*, vol. 25, pp. 499–525, Nov. 2012. 00000 Place : VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS Publisher : SPRINGER Type : Article.

[116] I. Santos, C. Laorden, B. Sanz, and P. G. Bringas, “Enhanced Topic-based Vector Space Model for semantics-aware spam filtering,” *EXPERT SYSTEMS WITH APPLICATIONS*, vol. 39, pp. 437–444, Jan. 2012. 00000 Place : THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, ENGLAND Publisher : PERGAMON-ELSEVIER SCIENCE LTD Type : Article ; Proceedings Paper.

[117] W. V. Woensel, P. C. Roy, S. S. R. Abidi, and S. R. Abidi, “Indoor location identification of patients for directing virtual care : An AI approach using machine learning and knowledge-based methods,” *Artificial Intelligence in Medicine*, vol. 108, p. 101931, 2020.

[118] H. Ko, P. Witherell, Y. Lu, S. Kim, and D. W. Rosen, “Machine learning and knowledge graph based design rule construction for additive manufacturing,” *Additive Manufacturing*, vol. 37, p. 101620, 2021.

[119] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery, “Learning to construct knowledge bases from the World Wide Web,” *Artificial Intelligence*, vol. 118, no. 1, pp. 69–113, 2000.

[120] L. Castillo, E. Armengol, E. Onaindía, L. Sebastiá, J. González-Boticario, A. Rodríguez, S. Fernández, J. D. Arias, and D. Borrajo, “samap : An user-oriented adaptive

system for planning tourist visits," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1318–1332, 2008.

[121] Y. Jia, Y. Qi, H. Shang, R. Jiang, and A. Li, "A Practical Approach to Constructing a Knowledge Graph for Cybersecurity," *Engineering*, vol. 4, no. 1, pp. 53–60, 2018.

[122] M. J. d. Silva, C. E. Pereira, and M. Götz, "Context-Aware Recommendation for Industrial Alarm System," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 229–234, 2018.

[123] R. Thomopoulos, S. Destercke, B. Charnomordic, I. Johnson, and J. Abécassis, "An iterative approach to build relevant ontology-aware data-driven models," *Information Sciences*, vol. 221, pp. 452–472, 2013.

[124] M. B. Messaoud, P. Leray, and N. B. Amor, "SemCaDo : A serendipitous strategy for causal discovery and ontology evolution," *Knowledge-Based Systems*, vol. 76, pp. 79–95, 2015.

[125] Q. Rajput and S. Haider, "BNOSA : A Bayesian network and ontology based semantic annotation framework," *Journal of Web Semantics*, vol. 9, no. 2, pp. 99–112, 2011.

[126] A. G. Valarakos, V. Karkaletsis, D. Alexopoulou, E. Papadimitriou, C. D. Spyropoulos, and G. Vouros, "Building an allergens ontology and maintaining it using machine learning techniques," *Computers in Biology and Medicine*, vol. 36, no. 10, pp. 1155–1184, 2006.

[127] J. Bock, U. Lösch, and H. Wang, "Automatic Reasoner Selection Using Machine Learning," in *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, WIMS '12, (New York, NY, USA), Association for Computing Machinery, 2012.

[128] F. Wu and D. S. Weld, "Automatically Refining the Wikipedia Infobox Ontology," in *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, (New York, NY, USA), pp. 635–644, Association for Computing Machinery, 2008.

[129] A. S. Rath, D. Devaurs, and S. N. Lindstaedt, "UICO : An Ontology-Based User Interaction Context Model for Automatic Task Detection on the Computer Desktop," in *Proceedings of the 1st Workshop on Context, Information and Ontologies*, CIAO '09, (New York, NY, USA), Association for Computing Machinery, 2009.

[130] O. Mabrouk, L. Hlaoua, and M. N. Omri, "Exploiting ontology information in fuzzy SVM social media profile classification," *APPLIED INTELLIGENCE*, Nov. 2020. 00000 Place : VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS Publisher : SPRINGER Type : Article ; Early Access.

[131] R. Mehri, V. Haarslev, and H. Chinaei, "A machine learning approach for optimizing heuristic decision-making in Web Ontology Language reasoners," *COMPUTATIONAL INTELLIGENCE*, vol. 37, pp. 273–314, Feb. 2021. 00000 Place : 111 RIVER ST, HOBOKEN 07030-5774, NJ USA Publisher : WILEY Type : Article.

[132] R. M. Ghoniem, N. Alhelwa, and K. Shaalan, “A Novel Hybrid Genetic-Whale Optimization Model for Ontology Learning from Arabic Text,” *ALGORITHMS*, vol. 12, Sept. 2019. 00000 Place : ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND Publisher : MDPI Type : Article.

[133] A. G. Salguero, J. Medina, P. Delatorre, and M. Espinilla, “Methodology for improving classification accuracy using ontologies : application in the recognition of activities of daily living,” *JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING*, vol. 10, pp. 2125–2142, June 2019. 00000 Place : TIERGARTENSTRASSE 17, D-69121 HEIDELBERG, GERMANY Publisher : SPRINGER HEIDELBERG Type : Article.

[134] S. Wan and M.-W. Mak, “Predicting subcellular localization of multi-location proteins by improving support vector machines with an adaptive-decision scheme,” *INTERNATIONAL JOURNAL OF MACHINE LEARNING AND CYBERNETICS*, vol. 9, pp. 399–411, Mar. 2018. 00000 Place : TIERGARTENSTRASSE 17, D-69121 HEIDELBERG, GERMANY Publisher : SPRINGER HEIDELBERG Type : Article.

[135] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh, and A. Hussain, “Concept-Level Sentiment Analysis with Dependency-Based Semantic Parsing : A Novel Approach,” *COGNITIVE COMPUTATION*, vol. 7, pp. 487–499, Aug. 2015. 00000 Place : ONE NEW YORK PLAZA, SUITE 4600, NEW YORK, NY, UNITED STATES Publisher : SPRINGER Type : Article.

[136] M. Manuja and D. Garg, “Intelligent text classification system based on self-administered ontology,” *TURKISH JOURNAL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCES*, vol. 23, no. 5, pp. 1393–1404, 2015. 00000 Place : ATATURK BULVARI NO 221, KAVAKLIDERE, ANKARA, 00000, TURKEY Publisher : TUBITAK SCIENTIFIC & TECHNICAL RESEARCH COUNCIL TURKEY Type : Article.

[137] I. Markievicz, J. Kapociute-Dzikiene, M. Tamosiunaite, and D. Vitkute-Adzgauskiene, “Action Classification in Action Ontology Building Using Robot-Specific Texts,” *INFORMATION TECHNOLOGY AND CONTROL*, vol. 44, no. 2, pp. 155–164, 2015. 00000 Place : KAUNAS UNIV TECHNOL, DEPT ELECTRONICS ENGINEERING, STUDENTU STR 50, KAUNAS, LT-51368, LITHUANIA Publisher : KAUNAS UNIV TECHNOLOGY Type : Article.

[138] P. Kordjamshidi and M.-F. Moens, “Global machine learning for spatial ontology population,” *JOURNAL OF WEB SEMANTICS*, vol. 30, pp. 3–21, Jan. 2015. 00000 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[139] J. M. del Rincon, M. J. Santofimia, and J.-C. Nebel, “Common-sense reasoning for human action recognition,” *PATTERN RECOGNITION LETTERS*, vol. 34, pp. 1849–

1860, Nov. 2013. 00000 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[140] N. R. Greenbaum, Y. Jernite, Y. Halpern, S. Calder, L. A. Nathanson, D. A. Sontag, and S. Horng, "Improving documentation of presenting problems in the emergency department using a domain-specific ontology and machine learning-driven user interfaces," *International Journal of Medical Informatics*, vol. 132, p. 103981, 2019.

[141] A. Annane, Z. Bellahsene, F. Azouaou, and C. Jonquet, "Building an effective and efficient background knowledge resource to enhance ontology matching," *JOURNAL OF WEB SEMANTICS*, vol. 51, pp. 51–68, Aug. 2018. 00000 Place : PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS Publisher : ELSEVIER SCIENCE BV Type : Article.

[142] J. Z. Pan, C. Bobed, I. Guclu, F. Bobillo, M. J. Kollingbaum, E. Mena, and Y.-F. Li, "Predicting Reasoner Performance on ABox Intensive OWL 2 EL Ontologies," *INTERNATIONAL JOURNAL ON SEMANTIC WEB AND INFORMATION SYSTEMS*, vol. 14, pp. 1–30, Mar. 2018. 00000 Place : 701 E CHOCOLATE AVE, STE 200, HERSHEY, PA 17033-1240 USA Publisher : IGI GLOBAL Type : Article ; Proceedings Paper.

[143] G. Rizzo, C. d'Amato, N. Fanizzi, and F. Esposito, "Tree-based models for inductive classification on the Web Of Data," *JOURNAL OF WEB SEMANTICS*, vol. 45, pp. 1–22, Aug. 2017. 00000 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[144] A. Lawrynowicz and J. Potoniec, "Pattern Based Feature Construction in Semantic Data Mining," *INTERNATIONAL JOURNAL ON SEMANTIC WEB AND INFORMATION SYSTEMS*, vol. 10, no. 1, pp. 27–65, 2014. 00000 Place : 701 E CHOCOLATE AVE, STE 200, HERSHEY, PA 17033-1240 USA Publisher : IGI GLOBAL Type : Article.

[145] A. Khan, J. A. Doucette, and R. Cohen, "Validation of an Ontological Medical Decision Support System for Patient Treatment Using a Repository of Patient Data : Insights into the Value of Machine Learning," *ACM TRANSACTIONS ON INTELLIGENT SYSTEMS AND TECHNOLOGY*, vol. 4, Sept. 2013. 00000 Place : 2 PENN PLAZA, STE 701, NEW YORK, NY 10121-0701 USA Publisher : ASSOC COMPUTING MACHINERY Type : Article.

[146] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling, "Never-Ending Learning," *Commun. ACM*, vol. 61, pp. 103–115, Apr. 2018. Place : New York, NY, USA Publisher : Association for Computing Machinery.

[147] E. Akand, M. Bain, and M. Temple, “Learning from Ontological Annotation : An Application of Formal Concept Analysis to Feature Construction in the Gene Ontology,” in *Proceedings of the Third Australasian Workshop on Advances in Ontologies - Volume 85*, AOW ’07, (AUS), pp. 15–23, Australian Computer Society, Inc., 2007.

[148] N. Mihindukulasooriya, M. R. A. Rashid, G. Rizzo, R. García-Castro, O. Corcho, and M. Torchiano, “RDF Shape Induction Using Knowledge Base Profiling,” in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC ’18, (New York, NY, USA), pp. 1952–1959, Association for Computing Machinery, 2018.

[149] M. Rico, N. Mihindukulasooriya, D. Kontokostas, H. Paulheim, S. Hellmann, and A. Gómez-Pérez, “Predicting Incorrect Mappings : A Data-Driven Approach Applied to DBpedia,” in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC ’18, (New York, NY, USA), pp. 323–330, Association for Computing Machinery, 2018.

[150] L. Zhang, T. Wu, X. Chen, B. Lu, C. Na, and G. Qi, “Auto Insurance Knowledge Graph Construction and Its Application to Fraud Detection,” in *The 10th International Joint Conference on Knowledge Graphs*, IJCKG’21, (New York, NY, USA), pp. 64–70, Association for Computing Machinery, 2021.

[151] J. Chen, F. Lecue, J. Z. Pan, S. Deng, and H. Chen, “Knowledge graph embeddings for dealing with concept drift in machine learning,” *JOURNAL OF WEB SEMANTICS*, vol. 67, Feb. 2021. 00000 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[152] S. Sabra, K. M. Malik, M. Afzal, V. Sabeeh, and A. C. Eddine, “A hybrid knowledge and ensemble classification approach for prediction of venous thromboembolism,” *EXPERT SYSTEMS*, vol. 37, Feb. 2020. 00000 Place : 111 RIVER ST, HOBOKEN 07030-5774, NJ USA Publisher : WILEY Type : Article.

[153] F. Ali, D. Kwak, P. Khan, S. El-Sappagh, A. Ali, S. Ullah, K. H. Kim, and K.-S. Kwak, “Transportation sentiment analysis using word embedding and ontology-based topic modeling,” *KNOWLEDGE-BASED SYSTEMS*, vol. 174, pp. 27–42, June 2019. 00000 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[154] N.-C. Hsieh, K.-C. Lee, and W. Chen, “The transformation of surgery patient care with a clinical research information system,” *EXPERT SYSTEMS WITH APPLICATIONS*, vol. 40, pp. 211–221, Jan. 2013. 00000 Place : THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, ENGLAND Publisher : PERGAMON-ELSEVIER SCIENCE LTD Type : Article.

[155] N. Fanizzi, C. d’Amato, and F. Esposito, “Composite Ontology Matching with Uncertain Mappings Recovery,” *SIGAPP Appl. Comput. Rev.*, vol. 11, pp. 17–29, Mar. 2011. Place : New York, NY, USA Publisher : Association for Computing Machinery.

[156] K. McGlinn, B. Yuce, H. Wicaksono, S. Howell, and Y. Rezgui, "Usability evaluation of a web-based tool for supporting holistic building energy management," *Automation in Construction*, vol. 84, pp. 154–165, 2017.

[157] G. Zhao and X. Zhang, "Domain-Specific Ontology Concept Extraction and Hierarchy Extension," in *Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval*, NLPIR 2018, (New York, NY, USA), pp. 60–64, Association for Computing Machinery, 2018.

[158] A. Meroño-Peñuela, R. Pernisch, C. Guéret, and S. Schlobach, "Multi-Domain and Explainable Prediction of Changes in Web Vocabularies," in *Proceedings of the 11th on Knowledge Capture Conference*, K-CAP '21, (New York, NY, USA), pp. 193–200, Association for Computing Machinery, 2021. 00000 event-place : Virtual Event, USA.

[159] S. Bischof, A. Harth, B. Kaempgen, A. Polleres, and P. Schneider, "Enriching integrated statistical open city data by combining equational knowledge and missing value imputation," *JOURNAL OF WEB SEMANTICS*, vol. 48, pp. 22–47, Jan. 2018. 00000 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[160] S. Radovanovic, B. Delibasic, M. Jovanovic, M. Vukicevic, and M. Suknovic, "A Framework for Integrating Domain Knowledge in Logistic Regression with Application to Hospital Readmission Prediction," *INTERNATIONAL JOURNAL ON ARTIFICIAL INTELLIGENCE TOOLS*, vol. 28, Sept. 2019. 00000 Place : 5 TOH TUCK LINK, SINGAPORE 596224, SINGAPORE Publisher : WORLD SCIENTIFIC PUBL CO PTE LTD Type : Article.

[161] J. Shi, W. Ji, Z. Gao, Y. Gao, Y. Wang, X. Liao, and F. Shi, "Ontology-based code snippets management in a cloud environment," *JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING*, vol. 10, pp. 2971–2985, Aug. 2019. 00000 Place : TIERGARTENSTRASSE 17, D-69121 HEIDELBERG, GERMANY Publisher : SPRINGER HEIDELBERG Type : Article.

[162] O. P. Patri, A. V. Panangadan, V. S. Sorathia, and V. K. Prasanna, "Sensors to Events : Semantic Modeling and Recognition of Events from Data Streams," *INTERNATIONAL JOURNAL OF SEMANTIC COMPUTING*, vol. 10, pp. 461–501, Dec. 2016. 00000 Place : 5 TOH TUCK LINK, SINGAPORE 596224, SINGAPORE Publisher : WORLD SCIENTIFIC PUBL CO PTE LTD Type : Article.

[163] J. Ye, G. Stevenson, and S. Dobson, "USMART : An Unsupervised Semantic Mining Activity Recognition Technique," *ACM TRANSACTIONS ON INTERACTIVE INTELLIGENT SYSTEMS*, vol. 4, Jan. 2015. 00000 Place : 2 PENN PLAZA, STE 701, NEW YORK, NY 10121-0701 USA Publisher : ASSOC COMPUTING MACHINERY Type : Article.

[164] K. Radinsky, S. Davidovich, and S. Markovitch, "Learning to Predict from Textual Data," *JOURNAL OF ARTIFICIAL INTELLIGENCE RESEARCH*, vol. 45, pp. 641–684, 2012. 00000 Place : USC INFORMATION SCIENCES INST, 4676 ADMIRALITY WAY, MARINA DEL REY, CA 90292-6695 USA Publisher : AI ACCESS FOUNDATION Type : Article.

[165] C. Djellali, "Using Hamming Similarity to Map Ontology Learning : A New Data Mining System," in *Proceedings of the 2013 Research in Adaptive and Convergent Systems*, RACS '13, (New York, NY, USA), pp. 82–87, Association for Computing Machinery, 2013.

[166] F. Getahun and K. Woldemariyam, "Integrated Ontology Learner : Towards Generic Semantic Annotation Framework," in *Proceedings of the 9th International Conference on Management of Digital EcoSystems*, MEDES '17, (New York, NY, USA), pp. 142–149, Association for Computing Machinery, 2017.

[167] L. Oliveira, R. Rocha Silva, and J. Bernardino, "Wine Ontology Influence in a Recommendation System," *BIG DATA AND COGNITIVE COMPUTING*, vol. 5, June 2021.

[168] X. Xue, H. Wang, and W. Liu, "Matching sensor ontologies with unsupervised neural network with competitive learning," *PEERJ COMPUTER SCIENCE*, vol. 7, Nov. 2021.

[169] M. Pérez-Pérez, G. Igrejas, F. Fdez-Riverola, and A. Lourenço, "A framework to extract biomedical knowledge from gluten-related tweets : The case of dietary concerns in digital era," *Artificial Intelligence in Medicine*, vol. 118, p. 102131, 2021.

[170] J. R. Mendez, T. R. Cotos-Yanez, and D. Ruano-Ordas, "A new semantic-based feature selection method for spam filtering," *APPLIED SOFT COMPUTING*, vol. 76, pp. 89–104, Mar. 2019. 00000 Place : RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS Publisher : ELSEVIER Type : Article.

[171] M. Rani, A. K. Dhar, and O. P. Vyas, "Semi-automatic terminology ontology learning based on topic modeling," *ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE*, vol. 63, pp. 108–125, Aug. 2017. 00000 Place : THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, ENGLAND Publisher : PERGAMON-ELSEVIER SCIENCE LTD Type : Article.

[172] B. Yang, "Construction of logistics financial security risk ontology model based on risk association and machine learning," *Safety Science*, vol. 123, p. 104437, 2020.

[173] C. Russo, K. Madani, and A. M. Rinaldi, "An Unsupervised Approach for Knowledge Construction Applied to Personal Robots," *IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS*, vol. 13, pp. 6–15, Mar. 2021. 00000 Place : 445 HOES LANE, PISCATAWAY, NJ 08855-4141 USA Publisher : IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC Type : Article.

[174] T. L. Packer and D. W. Embley, “Cost-Effective Information Extraction from Lists in OCRed Historical Documents,” in *Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing*, HIP ’15, (New York, NY, USA), pp. 23–30, Association for Computing Machinery, 2015.

[175] Q. Qiu, Z. Xie, L. Wu, and W. Li, “Geoscience keyphrase extraction algorithm using enhanced word embedding,” *EXPERT SYSTEMS WITH APPLICATIONS*, vol. 125, pp. 157–169, July 2019. 00000 Place : THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, ENGLAND Publisher : PERGAMON-ELSEVIER SCIENCE LTD Type : Article.

[176] P. Zhou and N. El-Gohary, “Semantic information alignment of BIMs to computer-interpretable regulations using ontologies and deep learning,” *Advanced Engineering Informatics*, vol. 48, p. 101239, 2021.

[177] A. Benarab, F. Rafique, and J. Sun, “An Ontology Embedding Approach Based on Multiple Neural Networks,” in *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, ICMLC ’19, (New York, NY, USA), pp. 186–190, Association for Computing Machinery, 2019.

[178] J. Fu, T. Mei, K. Yang, H. Lu, and Y. Rui, “Tagging Personal Photos with Transfer Deep Learning,” in *Proceedings of the 24th International Conference on World Wide Web*, WWW ’15, (Republic and Canton of Geneva, CHE), pp. 344–354, International World Wide Web Conferences Steering Committee, 2015.

[179] S. Mohan, R. Angell, N. Monath, and A. McCallum, “Low Resource Recognition and Linking of Biomedical Concepts from a Large Ontology,” in *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, BCB ’21, (New York, NY, USA), Association for Computing Machinery, 2021. 00000 event-place : Gainesville, Florida.

[180] D. L. Rubin, N. H. Shah, and N. F. Noy, “Biomedical ontologies : a functional perspective,” *Briefings in Bioinformatics*, vol. 9, pp. 75–90, Jan. 2008.

[181] W. Wong, W. Liu, and M. Bennamoun, “Ontology learning from text : A look back and into the future,” *ACM Computing Surveys*, vol. 44, pp. 20 :1–20 :36, Sept. 2012.

[182] F. N. Al-Aswadi, H. Y. Chan, and K. H. Gan, “Automatic ontology construction from text : a review from shallow to deep learning trend,” *Artificial Intelligence Review*, vol. 53, pp. 3901–3928, Aug. 2020.

[183] A. C. Khadir, H. Aliane, and A. Guessoum, “Ontology learning : Grand tour and challenges,” *Computer Science Review*, vol. 39, p. 100339, 2021.

[184] M. N. Asim, M. Wasim, M. U. G. Khan, W. Mahmood, and H. M. Abbasi, “A survey of ontology learning techniques and applications,” *Database*, vol. 2018, Jan. 2018.

[185] A. Maedche and S. Staab, “Ontology learning for the Semantic Web,” *IEEE Intelligent Systems*, vol. 16, pp. 72–79, Mar. 2001. 03204 Conference Name : IEEE Intelligent Systems.

[186] X. Liu, Y. Song, S. Liu, and H. Wang, "Automatic taxonomy construction from keywords," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, (New York, NY, USA), pp. 1433–1441, Association for Computing Machinery, 2012.

[187] G. Petasis, V. Karkaletsis, G. Paliouras, A. Krithara, and E. Zavitsanos, "Ontology Population and Enrichment : State of the Art," in *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution : Bridging the Semantic Gap* (G. Paliouras, C. D. Spyropoulos, and G. Tsatsaronis, eds.), Lecture Notes in Computer Science, pp. 134–166, Berlin, Heidelberg : Springer, 2011.

[188] Y. KALFOGLOU and M. SCHORLEMMER, "Ontology mapping : the state of the art," *The Knowledge Engineering Review*, vol. 18, no. 1, p. 1–31, 2003.

[189] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, "Pellet : A practical owl-dl reasoner," *Journal of Web Semantics*, vol. 5, no. 2, pp. 51–53, 2007.

[190] R. D. Shearer, B. Motik, and I. Horrocks, "Hermit : A highly-efficient owl reasoner.," in *Owled*, vol. 432, p. 91, 2008.

[191] D. Dou, H. Wang, and H. Liu, "Semantic data mining : A survey of ontology-based approaches," in *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*, pp. 244–251, Feb. 2015.

[192] P. Duboue, *The Art of Feature Engineering : Essentials for Machine Learning*. Cambridge University Press, June 2020. 00035 Google-Books-ID : _BzhDwAAQBAJ.

[193] P. Donís Ebri, "Combining ontologies and Machine Learning for Explainable Artificial Intelligence," Master's thesis, E.T.S. de Ingenieros Informáticos (UPM), 2021.

[194] J. Liebowitz, *The Handbook of Applied Expert Systems*. CRC Press, 1997.

[195] M. van Bekkum, M. de Boer, F. van Harmelen, A. Meyer-Vitali, and A. t. Teije, "Modular Design Patterns for Hybrid Learning and Reasoning Systems : a taxonomy, patterns and use cases," *arXiv :2102.11965 [cs]*, Mar. 2021.

[196] Henry Kautz, "The Third AI Summer, Henry Kautz, AAAI 2020 Robert S. Engelmore Memorial Award Lecture," Feb. 2020.

[197] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You ?" : Explaining the Predictions of Any Classifier," *arXiv :1602.04938 [cs, stat]*, Aug. 2016. 08107 arXiv : 1602.04938.

[198] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.

[199] A. Bennetot, J.-L. Laurent, R. Chatila, and N. Díaz-Rodríguez, "Towards Explainable Neural-Symbolic Visual Reasoning," *arXiv :1909.09065 [cs]*, Oct. 2019.

[200] A. Campagner and F. Cabitza, "Back to the Feature : A Neural-Symbolic Perspective on Explainable AI," in *Machine Learning and Knowledge Extraction* (A. Holzinger,

P. Kieseberg, A. M. Tjoa, and E. Weippl, eds.), Lecture Notes in Computer Science, (Cham), pp. 39–55, Springer International Publishing, 2020.

[201] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, *Learning from data : a short course*. S.l. : AMLbook.com, 2012.

[202] P. C. Soto, N. Ramzy, F. Ocker, and B. Vogel-Heuser, "An ontology-based approach for preprocessing in machine learning," in *2021 IEEE 25th International Conference on Intelligent Engineering Systems (INES)*, pp. 000133–000138, July 2021.

[203] K. Pearson, "VII. Note on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, Dec. 1895.

[204] C. Spearman, "The Abilities of Man their Nature and Measurement," *Nature*, vol. 120, pp. 181–183, Aug. 1927.

[205] M. G. Kendall, "A New Measure of Rank Correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.

[206] R. A. Fisher, "XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance.," *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, vol. 52, pp. 399–433, Jan. 1919.

[207] S. M. Stigler, "Karl Pearson's Theoretical Errors and the Advances They Inspired," *Statistical Science*, vol. 23, no. 2, pp. 261–271, 2008.

[208] S. Kullback, *Information Theory and Statistics*. Courier Corporation, July 1997.

[209] K. Pearson, "LIII. *On lines and planes of closest fit to systems of points in space*," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, pp. 559–572, Nov. 1901.

[210] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9, pp. 2812–2831, 2014. Publisher : Royal Society of Chemistry.

[211] R. Shen, S. Bubeck, and S. Gunasekar, "Data Augmentation as Feature Manipulation," in *Proceedings of the 39th International Conference on Machine Learning*, pp. 19773–19808, PMLR, June 2022.

[212] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular Value Decomposition and Principal Component Analysis," Mar. 2003.

[213] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Sept. 2013.

[214] M. Xu, "Understanding Graph Embedding Methods and Their Applications," *SIAM Review*, vol. 63, pp. 825–853, Jan. 2021.

[215] H. Ren, R. Stewart, J. Song, V. Kuleshov, and S. Ermon, "Adversarial Constraint Learning for Structured Prediction," May 2018.

[216] J. Pfrommer, C. Zimmerling, J. Liu, L. Kärger, F. Henning, and J. Beyerer, "Optimisation of manufacturing process parameters using deep neural networks as surrogate models," *Procedia CIRP*, vol. 72, pp. 426–431, Jan. 2018.

[217] A. Daw, A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided Neural Networks (PGNN) : An Application in Lake Temperature Modeling," Sept. 2021.

[218] M. Hilario, A. Kalousis, P. Nguyen, and A. Woznica, "A data mining ontology for algorithm selection and meta-mining," in *The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, (Bled, Slovenia), pp. 76–87, Jan. 2009.

[219] G. I. Diaz, A. Fokoue-Nkoutche, G. Nannicini, and H. Samulowitz, "An effective algorithm for hyperparameter optimization of neural networks," *IBM Journal of Research and Development*, vol. 61, pp. 9 :1–9 :11, July 2017.

[220] B. G. Humm and A. Zender, "An Ontology-Based Concept for Meta AutoML," in *Artificial Intelligence Applications and Innovations* (I. Maglogiannis, J. Macintyre, and L. Iliadis, eds.), IFIP Advances in Information and Communication Technology, (Cham), pp. 117–128, Springer International Publishing, 2021.

[221] A. M. Kalet, J. N. Doctor, J. H. Gennari, and M. H. Phillips, "Developing Bayesian networks from a dependency-layered ontology : A proof-of-concept in radiation oncology," *Medical Physics*, vol. 44, no. 8, pp. 4350–4359, 2017. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/mp.12340.

[222] C. Vens, J. Struyf, L. Schietgat, S. Džeroski, and H. Blockeel, "Decision trees for hierarchical multi-label classification," *Machine Learning*, vol. 73, pp. 185–214, Nov. 2008.

[223] S. Feng, P. Fu, and W. Zheng, "A hierarchical multi-label classification method based on neural networks for gene function prediction," *Biotechnology & Biotechnological Equipment*, vol. 32, pp. 1613–1621, Nov. 2018.

[224] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The Graph Neural Network Model," *IEEE Transactions on Neural Networks*, vol. 20, pp. 61–80, Jan. 2009.

[225] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A Comprehensive Survey on Graph Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, pp. 4–24, Jan. 2021.

[226] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A Comprehensive Survey of Loss Functions in Machine Learning," *Annals of Data Science*, vol. 9, pp. 187–212, Apr. 2022.

[227] M. Medina Grespan, A. Gupta, and V. Srikumar, "Evaluating Relaxations of Logic for Neural Networks : A Comprehensive Study," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, (Montreal, Canada), pp. 2812–2818, International Joint Conferences on Artificial Intelligence Organization, Aug. 2021.

[228] X. Hu and K. Liu, "Structural Deterioration Knowledge Ontology towards Physics-Informed Machine Learning for Enhanced Bridge Deterioration Prediction," *Journal of Computing in Civil Engineering*, vol. 37, p. 04022051, Jan. 2023.

[229] C.-H. Chou, H. Wu, J.-L. Kang, D. S.-H. Wong, Y. Yao, Y.-C. Chuang, S.-S. Jang, and J. D.-Y. Ou, "Physically Consistent Soft-Sensor Development Using Sequence-to-Sequence Neural Networks," *IEEE Transactions on Industrial Informatics*, vol. 16, pp. 2829–2838, Apr. 2020.

[230] B. Vinayagasu and S. Srivatsa, "Software Quality in Artificial Intelligence System," *Information Technology Journal*, vol. 6, pp. 835–842, Aug. 2007.

[231] J. M. Zhang, M. Harman, L. Ma, and Y. Liu, "Machine Learning Testing : Survey, Landscapes and Horizons," *IEEE Transactions on Software Engineering*, vol. 48, pp. 1–36, Jan. 2022. Conference Name : IEEE Transactions on Software Engineering.

[232] I. Standards, "Ieee standard glossary of software engineering terminology," *IEEE Std 610.12-1990*, pp. 1–84, 1990.

[233] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," in *International Conference on Learning Representations*, 2015.

[234] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Adversarial Attacks on Deep Neural Networks for Time Series Classification," in *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, July 2019. arXiv :1903.07054 [cs, stat].

[235] T. Gu, B. Dolan-Gavitt, and S. Garg, "BadNets : Identifying Vulnerabilities in the Machine Learning Model Supply Chain," Mar. 2019.

[236] C. O'Neil, *Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy*. USA : Crown Publishing Group, 2016.

[237] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining Explanations : An Overview of Interpretability of Machine Learning," Feb. 2019. arXiv :1806.00069 [cs, stat].

[238] J. Hernández-Orallo, "Evaluation in artificial intelligence : from task-oriented to ability-oriented measurement," *Artificial Intelligence Review*, vol. 48, pp. 397–447, Oct. 2017.

[239] S. Amershi, A. Begel, C. Bird, R. DeLine, H. Gall, E. Kamar, N. Nagappan, B. Nushi, and T. Zimmermann, "Software Engineering for Machine Learning : A Case Study," in *2019 IEEE/ACM 41st International Conference on Software Engineering : Software Engineering in Practice (ICSE-SEIP)*, (Montreal, QC, Canada), pp. 291–300, IEEE, May 2019.

[240] W. J. Frawley, G. Piatetsky-Shapiro, and C. J. Matheus, "Knowledge Discovery in Databases : An Overview," *AI Magazine*, vol. 13, pp. 57–57, Sept. 1992.

[241] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Magazine*, vol. 17, pp. 37–37, Mar. 1996.

[242] D. W. W. Rovce, “MANAGING THE DEVELOPMENT OF LARGE SOFTWARE SYSTEMS,” in *Technical Papers of Western Electronic Show and Convention (WesCon)*, (Los Angeles), 1970.

[243] R. Wirth and J. Hipp, “Crisp-dm : towards a standard process modell for data mining,” in *4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 2000.

[244] marktab, “Qu’est-ce que le processus Team Data Science Process ? - Azure Architecture Center.”

[245] J. Saltz, “CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects,” Nov. 2020.

[246] N. Nagappan, E. M. Maximilien, T. Bhat, and L. Williams, “Realizing quality improvement through test driven development : results and experiences of four industrial teams,” *Empirical Software Engineering*, vol. 13, no. 3, pp. 289–302, 2008.

[247] N. Japkowicz, “Why Question Machine Learning Evaluation Methods ? An Illustrative Review of the Shortcomings of Current Methods,” in *AAAI workshop on evaluation methods for machine learning*, (Boston), AAAI, 2006.

[248] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt, “Measuring Massive Multitask Language Understanding,” in *International Conference on Learning Representations*, Oct. 2020.

[249] F. Chollet, “On the Measure of Intelligence,” Nov. 2019.

[250] Y. Wang, S. Liu, X. Wu, and W. Shi, “CAVBench : A Benchmark Suite for Connected and Autonomous Vehicles,” in *2018 IEEE/ACM Symposium on Edge Computing (SEC)*, pp. 30–42, Oct. 2018.

[251] B. Hutchinson, N. Rostamzadeh, C. Greer, K. Heller, and V. Prabhakaran, “Evaluation Gaps in Machine Learning Practice,” in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’22, (New York, NY, USA), pp. 1859–1876, Association for Computing Machinery, 2022.

[252] M. Strathern, “‘Improving ratings’ : audit in the British University system,” *European Review*, vol. 5, pp. 305–321, July 1997. Publisher : Cambridge University Press.

[253] F. J. Anscombe, “Graphs in Statistical Analysis,” *The American Statistician*, vol. 27, pp. 17–21, Feb. 1973. Publisher : Taylor & Francis _eprint : https ://www.tandfonline.com/doi/pdf/10.1080/00031305.1973.10478966.

[254] L. Aina, R. Bernardi, and R. Fernández, “A distributional study of negated adjectives and antonyms,” in *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it 2018* (E. Cabrio, A. Mazzei, and F. Tamburini, eds.), pp. 7–13, Accademia University Press, 2018.

[255] N. Kassner and H. Schütze, “Negated and Misprimed Probes for Pretrained Language Models : Birds Can Talk, But Cannot Fly,” in *Proceedings of the 58th Annual*

Meeting of the Association for Computational Linguistics, (Online), pp. 7811–7818, Association for Computational Linguistics, 2020.

[256] A. Kumar and A. Joshi, "Striking a Balance : Alleviating Inconsistency in Pre-trained Models for Symmetric Classification Tasks," in *Findings of the Association for Computational Linguistics : ACL 2022*, (Dublin, Ireland), pp. 1887–1895, Association for Computational Linguistics, 2022.

[257] R. Lin and H. T. Ng, "Does BERT know that the IS-a relation is transitive ?," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2 : Short Papers)*, (Dublin, Ireland), pp. 94–99, Association for Computational Linguistics, May 2022.

[258] H. Qu and S. M. Veres, "Verification of logical consistency in robotic reasoning," *Robotics and Autonomous Systems*, vol. 83, pp. 44–56, Sept. 2016.

[259] K.-M. Wang, X. Wang, Z. Wang, G.-F. Wu, and Y. Xu, "Logical consistency verification of state sensing in safety-critical decision : A case study of train routing selection," *IET Intelligent Transport Systems*, vol. 16, no. 8, pp. 1042–1057, 2022.

[260] W.-C. Hung, V. Jampani, S. Liu, P. Molchanov, M.-H. Yang, and J. Kautz, "SCOPS : Self-Supervised Co-Part Segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Long Beach, CA, USA), pp. 869–878, IEEE, June 2019.

[261] X. Li, S. Liu, K. Kim, S. De Mello, V. Jampani, M.-H. Yang, and J. Kautz, "Self-supervised Single-View 3D Reconstruction via Semantic Consistency," in *Computer Vision – ECCV 2020* (A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds.), vol. 12359, (Cham), pp. 677–693, Springer International Publishing, 2020. Series Title : Lecture Notes in Computer Science.

[262] C.-H. Yao, W.-C. Hung, Y. Li, M. Rubinstein, M.-H. Yang, and V. Jampani, "LASSIE : Learning Articulated Shapes from Sparse Image Ensemble via 3D Part Discovery," July 2022.

[263] O. Honovich, R. Aharoni, J. Herzig, H. Taitelbaum, and ..., "TRUE : Re-evaluating Factual Consistency Evaluation," *arXiv preprint arXiv . . .*, 2022. Publisher : arxiv.org.

[264] T. Xu, Z.-H. Feng, X.-J. Wu, and J. Kittler, "Learning Adaptive Discriminative Correlation Filters via Temporal Consistency Preserving Spatial Feature Selection for Robust Visual Object Tracking," *IEEE Transactions on Image Processing*, vol. 28, pp. 5596–5609, Nov. 2019. Conference Name : IEEE Transactions on Image Processing.

[265] O. C. Mutlu, M. Honarmand, S. Surabhi, and D. P. Wall, "TempT : Temporal consistency for Test-time adaptation," Apr. 2023.

[266] R. Marroquin, J. Dubois, and C. Nicolle, "Ontology for a Panoptes building : Exploiting contextual information and a smart camera network," *Semantic Web*, vol. 9, Aug. 2018.

[267] W. Yu, W. Cheng, C. C. Aggarwal, H. Chen, and W. Wang, "Link prediction with spatial and temporal consistency in dynamic networks," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 3343–3349, Aug. 2017.

[268] K. Li, G. Huang, S. Wang, and S. Razavi, "Development of a physics-informed data-driven model for gaining insights into hydrological processes in irrigated watersheds," *Journal of Hydrology*, vol. 613, p. 128323, Oct. 2022.

[269] L. Di Natale, B. Svetozarevic, P. Heer, and C. N. Jones, "Physically Consistent Neural Networks for building thermal modeling : Theory and analysis," *Applied Energy*, vol. 325, p. 119806, Nov. 2022.

[270] G. Valentini, "Hierarchical Ensemble Methods for Protein Function Prediction," *ISRN Bioinformatics*, vol. 2014, p. 901419, May 2014.

[271] Q. Wang, Z. Wang, L. Zhang, P. Liu, and ..., "A novel consistency evaluation method for series-connected battery systems based on real-world operation data," *IEEE Transactions on . . .*, 2020.

[272] S. Kannappan, Y. Liu, and B. Tiddeman, "Human consistency evaluation of static video summaries," *Multimedia Tools and Applications*, 2019.

[273] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation," in *Encyclopedia of Database Systems* (L. LIU and M. T. ÖZSU, eds.), pp. 532–538, Boston, MA : Springer US, 2009.

[274] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under Concept Drift : A Review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, pp. 2346–2363, Dec. 2019. Conference Name : IEEE Transactions on Knowledge and Data Engineering.

[275] Y. Chen, Q. Yang, Z. Chen, C. Yan, S. Zeng, and M. Dai, "Physics-informed neural networks for building thermal modeling and demand response control," *Building and Environment*, vol. 234, p. 110149, Apr. 2023.

[276] C.-l. Shen, *The statistical analysis of fatigue data*. PhD thesis, The University of Arizona, 1994.

[277] J.-B. Lamy, "Owlready : Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies," *Artificial Intelligence in Medicine*, vol. 80, pp. 11–28, July 2017.

[278] Y. Liang, K. Ouyang, Y. Wang, Z. Pan, Y. Yin, H. Chen, J. Zhang, Y. Zheng, D. S. Rosenblum, and R. Zimmermann, "Mixed-Order Relation-Aware Recurrent Neural Networks for Spatio-Temporal Forecasting," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–15, 2022. Conference Name : IEEE Transactions on Knowledge and Data Engineering.

[279] B. Li, P. Qi, B. Liu, S. Di, J. Liu, J. Pei, J. Yi, and B. Zhou, “Trustworthy AI : From Principles to Practices,” *ACM Computing Surveys*, vol. 55, pp. 177 :1–177 :46, Jan. 2023.

[280] W. J. von Eschenbach, “Transparency and the Black Box Problem : Why We Do Not Trust AI,” *Philosophy & Technology*, vol. 34, pp. 1607–1622, Dec. 2021.

[281] M. Leyli Abadi, A. Marot, J. Picault, D. Danan, M. Yagoubi, B. Donnot, S. Attoui, P. Dimitrov, A. Farjallah, and C. Etienam, “LIPS - Learning Industrial Physical Simulation benchmark suite,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 28095–28109, Dec. 2022.

Table des figures

2.1 Les six domaines de la Smart City [16] 12

2.2 Comment est utilisée la donnée dans une Smart City ? 12

2.3 Briques d'évaluation de la cohérence 19

2.4 Les quatre types de gestion des connaissances [43] 27

3.1 Méthodologie de la SLR 34

3.2 Selection des articles 40

3.3 Catégories de combinaisons d'ontologie et d'apprentissage automatique présentées dans cette SLR 41

3.4 Types d'algorithmes d'apprentissage automatique présents 43

3.5 Proportion de raisonnement déductif dans les études par catégorie 44

3.6 Thème de l'intelligence artificiel par catégorie 45

3.7 Domaines d'application présents dans cette SLR 47

3.8 Évolution du nombre de publications sur la combinaison des ontologies et de l'apprentissage automatique 49

3.9 Pays où est basé le premier auteur 49

3.10 Continent où est basé le premier auteur 50

3.11 Mécanisme de l'*Ontology learning* 51

3.12 Mécanisme de l'*Ontology mapping* 55

3.13 Mécanisme du raisonnement déductif basé sur l'apprentissage 57

3.14 Mécanisme de l'*Informed machine learning* 59

3.15 Mécanisme de l'*Ontology explain black-box* 63

3.16 Mécanisme d'un système expert intégrant de l'apprentissage automatique . 65

3.17 Mécanisme de l'application hybride 67

3.18 Différences entre les trois paradigmes d'étude de l'IA Hybride 72

4.1 Organisation d'un modèle d'apprentissage automatique d'après [201] . . . 80

4.2 Les trois techniques utilisées en ingénierie des caractéristiques : la sélection, l'extraction et l'augmentation 84

4.3 Processus normalisé interprofessionnel pour l'exploration de données (CRISP-DM) en apprentissage automatique d'après [243] 95

4.4 Procédure d'évaluation d'un algorithme d'apprentissage automatique . . . 96

4.5 Fréquence d'usage des métriques d'évaluation en fonction du type de modèle d'apprentissage automatique 109

4.6 Le nouveau protocole d'évaluation mis au point 112

5.1 Ressort qui symbolise le mouvement d'oscillation d'un oscillateur harmonique amorti 121

5.2 Fonction sinusïdale qui décrit le mouvement d'un oscillateur harmonique amorti 121

5.3 *Ground truth* de l'oscillateur harmonique amorti et données d'entraînement 122

5.4 Oscillateur harmonique : architecture du réseau de neurones sans apport de connaissance 123

5.5 Oscillateur harmonique : prédiction du réseau de neurones sans apport de connaissance 123

5.6 Oscillateur harmonique : architecture du réseau de neurones avec apport de connaissance 125

5.7 Oscillateur harmonique : prédiction du réseau de neurones avec apport de connaissance 125

5.8 Courbe de Wöhler ou courbe S-N 127

5.9 Durée de vie des matériaux : données d'entraînement et d'évaluation . . . 128

5.10 Durée de vie des matériaux : architecture du réseau de neurones sans apport de connaissance 128

5.11 Durée de vie des matériaux : prédiction du réseau de neurones sans apport de connaissance 129

5.12 Durée de vie des matériaux : architecture du réseau de neurones avec apport de connaissance 131

5.13 Durée de vie des matériaux : prédiction du réseau de neurones avec apport de connaissance 132

5.14 Architecture et fonctionnement du framework *Ontology-based Physics-Informed Machine Learning* (OPIML) 133

5.15 Ontologie des lois physiques 135

5.16 Règle de relation monotone négative 135

5.17 Équation de la relation monotone négative 136

5.18 Ensemble des règles associées au contexte de l'estimation de la durée de vie d'un matériau 140

5.19 Première règle pour estimer la durée de vie de l'acier en fonction d'un stress appliqué 140

Liste des tables

3.1 Requête finale utilisée pour chaque moteur de recherche d'articles scientifiques . . . 39

3.2 Données extraites des études sélectionnées . . . 40

3.3 Utilisation des algorithmes d'apprentissage automatique . . . 43

3.4 Liste des articles proposant un raisonnement déductif non limité aux liens de subsomption par grande catégorie . . . 45

3.5 Liste des articles associés à leur thématique d'IA . . . 46

3.6 Liste des articles associés à leur domaine d'application . . . 47

3.7 Algorithmes d'apprentissage automatique utilisés pour l'*ontology learning* . 53

3.8 Algorithmes d'apprentissage automatique utilisés pour l'alignement d'ontologies . . . 56

3.9 Algorithmes d'apprentissage automatique utilisés pour le raisonnement déductif . . . 58

3.10 Algorithmes d'apprentissage automatique utilisés pour l'*informed machine learning* . . . 61

3.11 Algorithmes d'apprentissage automatique utilisés pour l'*ontology explain black-box* . . . 64

3.12 Algorithmes d'apprentissage automatique utilisés pour les systèmes experts intégrant l'apprentissage . . . 66

3.13 Algorithmes d'apprentissage automatique utilisés pour les applications hybrides . . . 67

3.14 Alignement de nos catégories d'hybridation avec les modèles de conception de Van Bekkum et al. [195] et la taxonomie de Kautz [196] . . . 68

4.1 Matrice de confusion . . . 100

4.2 Principales métriques d'évaluation des classifieurs . . . 101

4.3 Principales métriques d'évaluation des modèles de régression . . . 103

5.1 Tableau des connaissances et contraintes associées d'après [37] 130
5.2 Mesures d'évaluation pour le jeu de données sur le fil d'acier 144
5.3 Mesures d'évaluation pour le jeu de données sur 2024-T4 144
5.4 Mesures d'évaluation pour le jeu de données sur AAW 144

Liste des publications

Conférence internationale

S. Ghidalia, O. Labbani Narsis, A. Bertaux, and C. Nicolle, 'Automating Physical Knowledge Integration in Machine Learning', in 2023 17th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, Nov. 2023.

Chapitre de livre

S. Ghidalia, O. Labbani Narsis, A. Bertaux, and C. Nicolle, 'Ville intelligente : valeur et vérité des données numériques', in Smart City et prise de décision, Mare & Martin, 2023, pp. 47–62.

Article soumis pour publication

Première relecture réalisée, en attente de l'éditeur depuis juin 2023 :

S. Ghidalia, O. Labbani Narsis, A. Bertaux, and C. Nicolle, 'Combining Machine Learning and Ontology : A Systematic Literature Review', Artificial Intelligence Review, Jul. 2022, doi : 10.21203/rs.3.rs-1881512/v1.

Document réalisé avec LaTeX et :
le style LaTeX pour Thèse de Doctorat créé par S. Galland — https://github.com/gallandarakhneorg/tex-templates
la collection de paquets `tex-upmethodology` — http://www.arakhne.org/tex-upmethodology/

école doctorale **sciences pour l'ingénieur et microtechniques**

Titre : Étude sur les mesures d'évaluation de la cohérence entre connaissance et compréhension dans le domaine de l'intelligence artificielle

Mots-clés : intelligence artificielle, apprentissage automatique, ontologie, évaluation, cohérence, intelligence artificielle hybride

Résumé :

Cette thèse traite de la notion de cohérence au sein des systèmes intelligents. Son objectif principal est d'analyser comment la cohérence, en tant que concept, peut être comprise et évaluée dans le domaine de l'intelligence artificielle, en mettant particulièrement l'accent sur les connaissances préalables intégrées dans ces systèmes. Ce travail, financé dans le cadre du projet européen H2020 RESPONSE, repose sur le context applicatif de la Smart City où l'évaluation de la cohérence entre prédictions artificielles et réalités de terrain reste la condition préalable à toute initiative politique. Un examen minutieux de la cohérence en relation avec l'intelligence artificielle, ainsi qu'une exploration approfondie des connaissances préalables fait l'objet de ce travail. Pour cela une revue systématique de la littérature est réalisée pour cartographier le paysage actuel, mettant en lumière l'intersection et l'interaction entre l'apprentissage automatique et les ontologies, avec un focus particulier sur les techniques algorithmiques en usage. Notre analyse comparative positionne également notre recherche par rapport à des œuvres significatives dans le domaine. Une étude approfondie sur les différentes méthodes d'intégration des connaissances analyse comment la cohérence peut être évaluée en fonction des techniques d'apprentissage utilisées. La qualité globale des systèmes d'intelligence artificielle, avec un focus particulier sur l'évaluation de la cohérence, est également examinée. L'ensemble de cette étude est ensuite appliqué sur l'évaluation de la cohérence d'un modèle par rapport aux lois physiques représentées au sein d'ontologies. Deux études de cas, l'une sur la prédiction des mouvements d'un oscillateur harmonique et l'autre sur l'estimation de la durée de vie d'un matériau, sont présentées pour souligner l'importance des contraintes physiques dans l'évaluation de la cohérence. De plus, nous proposons une nouvelle méthode pour formaliser les connaissances dans une ontologie, en évaluant son efficacité. L'objectif de ce travail est d'apporter un nouvel éclairage sur l'évaluation des algorithmes d'apprentissage automatique, en proposant une méthode d'évaluation de la cohérence. Cette thèse aspire à être une contribution significative au domaine de l'intelligence artificielle, en mettant en lumière l'importance de la cohérence dans la construction de systèmes intelligents fiables et pertinents.

Title: Coherence distance between knowledge and understanding in artificial intelligence

Keywords: artificial intelligence, machine learning, ontology, evaluation, consistency, hybrid artificial intelligence

Abstract:

This thesis investigates the concept of coherence within intelligent systems, aiming to assess how coherence can be understood and measured in artificial intelligence, with a particular focus on pre-existing knowledge embedded in these systems. This research is funded as part of the European H2020 RESPONSE project and is set in the context of smart cities, where assessing the coherence between AI predictions and real-world data is a fundamental prerequisite for policy initiatives. The main objective of this work is to examine coherence in the field of artificial intelligence meticulously and to conduct a thorough exploration of prior knowledge. To this end, we conduct a systematic literature review to map the current landscape, focusing on the convergence and interaction between machine learning and ontologies, and highlighting, in particular, the algorithmic techniques employed. In addition, our comparative analysis positions our research in the broader context of important work in the field.

An in-depth study of different knowledge integration methods is undertaken to analyze how coherence can be assessed based on the learning techniques employed. The overall quality of artificial intelligence systems, with particular emphasis on coherence assessment, is also examined. The whole study is then applied to the coherence evaluation of models concerning the representation of physical laws in ontologies. We present two case studies, one on predicting the motion of a harmonic oscillator and the other on estimating the lifetime of materials, to highlight the importance of respecting physical constraints in coherence assessment. In addition, we propose a new method for formalizing knowledge within an ontology and evaluate its effectiveness. This research aims to provide new perspectives in the evaluation of machine learning algorithms by introducing a coherence evaluation method. This thesis aspires to make a substantial contribution to the field of artificial intelligence by highlighting the critical role of coherence in the development of reliable and relevant intelligent systems.

Université Bourgogne-Franche-Comté
32, avenue de l'Observatoire
25000 Besançon, France