ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

Typologie et taxonomie des modèles prédictifs monolithiques et hybrides d'intelligence artificielle et d'apprentissage automatique en finance : une architecture générationnelle intégrée

Typology and taxonomy of monolithic and hybrid AI/ML predictive models in finance: an integrated generational architecture

Adil SLAMI-AMINE, ***(Étudiant-Chercheur)***
Laboratoire d'Analyse et Modélisation des Systèmes et Aide à la Décision (LAMSAD)
École Nationale des Sciences Appliquées Berrechid
Université HASSAN 1er de Settat, Maroc

Abdessamad DINE, (*Enseignant-Chercheur*)
Laboratoire de Recherche en Economie, Gestion et Management des Affaires (LAREGMA)
École Nationale de Commerce et de Gestion de Settat
Université HASSAN 1er de Settat, Maroc

Boujemâa ACHCHAB, (*Enseignant-Chercheur*)
Laboratoire d'Analyse et Modélisation des Systèmes et Aide à la Décision (LAMSAD)
École Nationale des Sciences Appliquées Berrechid
Université HASSAN 1er de Settat, Maroc

Adresse de correspondance :	Université Hassan Premier, Settat, Maroc École Nationale des Sciences Appliquées Berrechid Adresse : Avenue de l'université, B.P :218 Berrechid Tel : 05-22-32-47-58 Fax : 05-22-53-45-30 Email : ensa.etudes@uhp.ac.ma
Déclaration de divulgation :	Les auteurs n'ont pas connaissance de quelconque financement qui pourrait affecter l'objectivité de cette étude. Ils assument l'entière responsabilité de tout éventuel plagiat, de l'usage de l'intelligence artificielle dans la rédaction, ainsi que des résultats présentés dans cet article.
Conflit d'intérêts :	Les auteurs ne signalent aucun conflit d'intérêts.
Citer cet article	SLAMI-AMINE, A., DINE, A., & ACHCHAB, B. (2026). Typologie et taxonomie des modèles prédictifs monolithiques et hybrides d'intelligence artificielle et d'apprentissage automatique en finance : une architecture générationnelle intégrée. *International Journal of Accounting, Finance, Auditing, Management and Economics*, *7*(6), 543–570. https://doi.org/10.5281/zenodo.20340549
Licence	**Cet article est publié en open Access sous licence CC BY-NC-ND**

Received: 03/05/2026 *Accepted: 05/06/2026*

Typologie et taxonomie des modèles prédictifs monolithiques et hybrides d'intelligence artificielle et d'apprentissage automatique en finance : une architecture générationnelle intégrée

Résumé :

Le foisonnement des modèles d'intelligence artificielle et d'apprentissage automatique appliqués à la prédiction financière a engendré un écosystème analytique profondément morcelé, où coexistent des approches hétérogènes sans qu'aucun cadre de référence partagé ne permette de les ordonner ni de les comparer avec rigueur. Face à cette lacune, l'article entreprend de construire une nomenclature organisée autour d'une double logique : une progression générationnelle et une articulation multi-dimensionnelle, couvrant indifféremment les architectures monolithiques et les configurations hybrides mobilisées en finance. La démarche, à la fois déductive et inductive, s'inscrit dans le protocole itératif de Nickerson, Varshney et Muntermann (2013) et s'aligne sur les principes FAIR (Wilkinson et al., 2016) ainsi que sur les critères de validation taxonomique de Bailey (1994) et Doty et Glick (1994). La validation conceptuelle est conduite à partir de douze cas illustratifs représentatifs, un par classe, sélectionnés par triangulation. La structure classificatoire obtenue s'organise autour de trois dimensions fondatrices (Architecture, Famille, Génération) et donne lieu à douze classes : neuf configurations monolithiques et trois classes hybrides. Le cadre proposé constitue une contribution intégrée et reproductible à la cartographie comparée des modèles prédictifs IA/ML en finance.

Mots clés : Typologie, Taxonomie, IA-ML prédictifs, Modèles monolithiques, Modèles hybrides
JEL Classification : C45, C53, G17
Type du papier : Recherche Théorique

Abstract :

The rapid proliferation of artificial intelligence and machine learning models applied to financial forecasting has produced a deeply fragmented analytical landscape, where heterogeneous architectures coexist without any shared classificatory framework to systematically organize or compare them. To address this gap, the article develops a structured nomenclature grounded in a dual logic: a generational progression and a multi-dimensional articulation, encompassing both monolithic and hybrid configurations deployed in finance. The methodology is simultaneously deductive and inductive, anchored in the iterative protocol introduced by Nickerson, Varshney, and Muntermann (2013), and aligned with the FAIR principles (Wilkinson et al., 2016) and the taxonomic validity criteria of Bailey (1994) and Doty and Glick (1994). Conceptual validation rests on twelve illustrative cases, one per class, selected through triangulation. The resulting framework revolves around three primary dimensions (Architecture, Family, Generation) yielding twelve classes: nine monolithic configurations and three hybrid classes. The framework constitutes an integrated and reproducible contribution to the comparative mapping of predictive AI/ML models in finance.

Keywords: Typology, Taxonomy, Predictive AI-ML, Monolithic Models, Hybrid Models.
Classification JEL: C45, C53, G17
Paper type: Theoretical Research

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

1. Introduction

La dernière décennie a vu la finance quantitative absorber un flux continu d'innovations algorithmiques issues de l'apprentissage automatique et de l'intelligence artificielle. Cette absorption ne s'est pas limitée à un renouvellement cosmétique de la boîte à outils économétrique : elle a redéfini, parfois en profondeur, la manière dont les marchés, les risques et les comportements d'agents sont modélisés. Autant de domaines dans lesquels les modèles prédictifs IA/ML occupent désormais une place centrale : prévision de rendements d'actifs, estimation de la volatilité, tarification d'instruments dérivés, évaluation du risque de crédit, détection de fraudes, analyse de la microstructure des marchés (Gu, Kelly & Xiu, 2020 ; Dixon, Halperin & Bilokon, 2020 ; Sezer, Gudelek & Ozbayoglu, 2020 ; Goodell, Kumar, Lim & Pattnaik, 2021). La portée taxonomique défendue ici est délibérément pan-financière : elle vise à couvrir l'ensemble de ces sous-domaines, étant entendu que le corpus illustratif privilégie les sous-domaines pour lesquels la littérature recensée est la mieux documentée, à savoir la prévision de rendements, la volatilité et la détection de fraudes.

Le passage de l'économétrie classique vers ces nouvelles architectures ne s'est pas opéré par rupture brutale mais par accumulation progressive. Les modèles ARIMA, GARCH et VAR, longtemps dominants dans l'analyse des séries temporelles financières, ont été complétés, puis partiellement supplantés, par des approches issues de l'apprentissage statistique (Random Forest, SVM, XGBoost), puis par des architectures d'apprentissage profond (LSTM, CNN, RNN), avant que n'émergent, plus récemment, des modèles fondationnels fondés sur des mécanismes d'attention massive et sur le pré-entraînement à grande échelle (Vaswani et al., 2017 ; Brown et al., 2020 ; Bommasani et al., 2021). Les références canoniques associées à chacun de ces jalons algorithmiques sont détaillées dans la revue de littérature (section 2.1) afin de préserver, dans l'introduction, la progression de la démonstration centrale.

Trois vagues technologiques se distinguent ainsi avec une relative netteté, ce que la littérature récente reconnaît explicitement sous le terme de « générations » d'apprentissage automatique appliquées aux séries temporelles financières (LeCun, Bengio & Hinton, 2015 ; Goodfellow, Bengio & Courville, 2016 ; Sezer et al., 2020 ; Ozbayoglu, Gudelek & Sezer, 2020). La première, que l'on peut qualifier de génération superficielle, réunit les algorithmes classiques d'apprentissage statistique : arbres de décision, machines à vecteurs de support, boosting, régressions pénalisées (Friedman, 2001 ; Hastie, Tibshirani & Friedman, 2009). La deuxième, celle de la génération profonde, correspond au déploiement massif des réseaux neuronaux à couches multiples, portés par la disponibilité des GPU et par l'essor des techniques de rétropropagation à grande échelle (LeCun et al., 2015 ; Fischer & Krauss, 2018). La troisième, celle de la génération avancée, repose sur des architectures intégrant auto-attention, pré-entraînement fondationnel, méta-apprentissage et, plus récemment, capacités agentiques (Bommasani et al., 2021 ; Yang, Liu & Wang, 2023 ; Wu et al., 2023).

À ce socle s'ajoute un phénomène connexe, mais désormais indissociable de l'évolution prédictive : la montée en puissance des données alternatives (textes d'articles de presse, flux de médias sociaux, images satellitaires, données comportementales) et l'essor du traitement automatique du langage naturel (Natural Language Processing, NLP) financier (Loughran & McDonald, 2011 ; Tetlock, 2007 ; Araci, 2019). Ces sources non structurées, couplées aux architectures modernes, alimentent une hybridation croissante des pipelines analytiques, brouillant les frontières entre modèles statistiques traditionnels, apprentissage profond et intelligence artificielle avancée (López de Prado, 2018).

Cette accumulation rapide d'architectures a engendré un effet collatéral peu discuté dans la littérature : la fragmentation du champ. L'offre méthodologique s'est multipliée sans qu'émerge, en parallèle, un cadre classificatoire capable d'ordonner rigoureusement cette diversité. Plusieurs symptômes trahissent la profondeur du problème.

D'abord, la prolifération architecturale demeure largement non structurée. Les articles empiriques proposent régulièrement de nouveaux modèles, déclinaisons ou combinaisons, sans toujours préciser la nature de la contribution classificatoire. Cette dynamique, propre à un domaine en expansion rapide, rend difficile une lecture synoptique de l'état de l'art et complique toute comparaison systématique. Elle contraste avec ce que l'on observe dans des disciplines taxonomiquement plus matures, où la position d'un modèle dans l'arbre classificatoire est presque toujours explicitée.

Ensuite, une hétérogénéité terminologique affecte les concepts les plus courants. Les termes *ensemble*, *hybride*, *stacké*, *blending*, *meta-learner* ou *architecture composite* sont utilisés de manière fluctuante selon les auteurs, parfois comme synonymes, parfois pour désigner des objets distincts (Sagi & Rokach, 2018). Cette confusion sémantique n'est pas anodine : elle obscurcit les règles de composition entre modèles et affaiblit la portée explicative des études empiriques.

La troisième difficulté tient à la faible reproductibilité des comparaisons empiriques. En l'absence de critères classificatoires partagés, il est malaisé de déterminer si deux modèles comparés appartiennent à la même classe, s'ils mobilisent des types de données compatibles, ou si leurs métriques d'évaluation sont commensurables. Ce déficit, déjà documenté dans la méta-littérature en *Machine Learning* (Lipton & Steinhardt, 2019 ; Raff, 2019), prend en finance une dimension critique compte tenu de l'enjeu décisionnel et des coûts associés à des erreurs prédictives (López de Prado, 2018 ; Arnott, Harvey & Markowitz, 2019).

Il apparaît dès lors nécessaire de disposer d'un cadre taxonomique à la fois exhaustif, exclusif, parcimonieux et extensible, compatible avec les principes FAIR (*Findable*, *Accessible*, *Interoperable*, *Reusable*). Un tel cadre, s'il veut servir d'infrastructure épistémique au champ, devra satisfaire trois propriétés complémentaires : (1) rendre compte de l'ensemble des architectures actuellement documentées ; (2) anticiper l'émergence de nouvelles classes ; (3) être assez simple pour être effectivement utilisé par les praticiens.

Plusieurs traditions académiques éclairent la problématique abordée ici. La première, issue des systèmes d'information et de la sociologie des sciences, porte sur les fondements théoriques de la construction taxonomique. Bailey (1994) en propose la grammaire générale : distinction entre typologies (conceptuelles, déductives, organisées autour d'idéaux-types) et taxonomies (empiriques, inductives, construites à partir de caractéristiques observées). Doty et Glick (1994) prolongent cette distinction en soulignant que les typologies doivent être considérées comme de véritables théories, capables de formuler des hypothèses falsifiables sur les relations entre configurations. Nickerson, Varshney et Muntermann (2013) ont depuis formalisé une méthode itérative de construction taxonomique, aujourd'hui largement adoptée en systèmes d'information, qui combine phases empirique-vers-conceptuel et conceptuel-vers-empirique jusqu'à satisfaction de conditions d'arrêt objectives et subjectives.

Une deuxième tradition concerne directement la classification des modèles IA/ML en finance. Gu, Kelly et Xiu (2020) proposent, dans leur étude de référence sur la valorisation empirique des actifs par *Machine Learning*, une classification fonctionnelle distinguant modèles linéaires, arbres, réseaux neuronaux et ensembles. Heaton, Polson et Witte (2017), pour leur part, placent l'accent sur l'apport spécifique de l'apprentissage profond en finance. Dixon, Halperin et Bilokon (2020) offrent une synthèse pédagogique étendue, tandis que López de Prado (2018) se concentre sur les défis méthodologiques propres à l'apprentissage financier (sur-ajustement, données non stationnaires, biais de sélection). À ces références s'ajoutent les revues systématiques de Sezer, Gudelek et Ozbayoglu (2020) sur la prévision des séries temporelles financières par *Deep Learning*, de Ozbayoglu, Gudelek et Sezer (2020) sur les applications globales du *Deep Learning* en finance, et de Goodell et al. (2021) sur les applications de l'IA en finance.

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

Ces contributions, précieuses, présentent néanmoins trois limites récurrentes au regard de l'objectif taxonomique visé ici. En premier lieu, la dimension génération technologique demeure implicite : les classifications évoquées distinguent souvent les modèles selon leur nature algorithmique, sans toujours relier cette nature à la vague technologique dont elle est issue. En deuxième lieu, la frontière entre hybridation structurelle (composition de modèles distincts) et hybridation fonctionnelle (ensemble d'un même type) est rarement explicitée. En troisième lieu, la conformité aux principes FAIR, désormais exigés par de nombreuses infrastructures de données ouvertes, n'est pas systématiquement envisagée.

Il est donc possible d'identifier trois lacunes cumulatives dans la littérature : une lacune taxonomique (absence d'un arbre intégré), une lacune générationnelle (absence d'une dimension temporelle-technologique explicite) et une lacune méthodologique (défaut de conformité aux standards de reproductibilité scientifique). La présente étude se positionne à l'intersection de ces trois lacunes.

La question centrale qui structure notre investigation peut être formulée ainsi : *comment construire une taxonomie rigoureuse, générationnelle et reproductible des modèles prédictifs IA/ML, monolithiques et hybrides, mobilisés en finance ?*

Cette interrogation se décline en trois sous-questions. La première (Q1) porte sur les critères permettant de distinguer sans ambiguïté les architectures monolithiques des architectures hybrides. La deuxième (Q2) concerne la base conceptuelle et opérationnelle permettant de délimiter les générations G1, G2 et G3. La troisième (Q3) interroge l'articulation entre la famille algorithmique Machine Learning ou Artificial Intelligence et la profondeur architecturale du modèle.

Sur le plan des objectifs, l'ambition générale est de proposer un cadre typologique et taxonomique rigoureux, cohérent et opérationnalisable. Quatre objectifs spécifiques en découlent : (1) formaliser les critères de classification ; (2) construire la taxonomie correspondante ; (3) la valider empiriquement par le recours à des cas illustratifs représentatifs ; (4) en assurer l'alignement avec les principes FAIR et la méthode de Nickerson et al. (2013).

Quatre propositions théoriques, au sens de Doty et Glick (1994), guident cette construction et seront soumises à une validation conceptuelle plutôt qu'à un protocole hypothético-déductif strict. La proposition P1 postule que la profondeur architecturale et la génération technologique constituent deux dimensions orthogonales, bien que partiellement corrélées en pratique. La proposition P2 pose que l'hybridation suit une logique hiérarchique inclusive : l'espace combinatoire de HG3 englobe celui de HG2, qui englobe à son tour celui de HG1. La proposition P3 avance que la distinction entre Machine Learning et Artificial Intelligence avancée n'est opérante qu'à partir de la génération G2, les modèles G1 étant majoritairement statistiques. La proposition P4, enfin, soutient qu'une taxonomie conforme aux principes FAIR accroît la reproductibilité des comparaisons empiriques en finance prédictive. Conformément à la nature conceptuelle du travail, ces propositions sont éprouvées par triangulation analytique sur les douze cas illustratifs et non par test statistique.

L'article revendique une triple contribution. Au plan typologique, il propose le premier cadre intégré combinant simultanément architecture (monolithique/hybride), famille (ML/AI) et génération (G1/G2/G3). Au plan méthodologique, il applique de manière stricte la démarche de Nickerson et al. (2013) et les principes FAIR au champ de la finance computationnelle. Au plan épistémique, il s'efforce de clarifier une terminologie flottante et de jeter un pont entre des littératures (ML classique, apprentissage profond, IA avancée) qui ont longtemps évolué en parallèle.

La suite de l'article est organisée selon la structure IMRaD. La section 2 expose le cadre théorique et méthodologique mobilisé ; sa contribution propre est d'expliciter les fondations classificatoires et les référentiels de conformité. La section 3 présente en détail la typologie et la taxonomie proposées, accompagnées d'un tableau synthétique et d'un schéma conceptuel ; sa

contribution est de fournir le cadre opérationnel à douze classes et ses règles de résolution. La section 4 discute les résultats au regard de la littérature, évalue la conformité aux référentiels et identifie les limites ; sa contribution est de positionner le cadre par rapport aux classifications existantes et d'en dériver des implications théoriques et managériales. La section 5 conclut sur les apports, synthétise les limites et trace les perspectives de développement.

2. Méthodologie et fondement conceptuel

2.1. Cadre théorique mobilisé

La construction taxonomique proposée ici s'appuie sur un socle théorique clairement identifié, volontairement choisi pour conjuguer rigueur classificatoire et ancrage financier. Ce socle combine quatre ensembles de références.

Le premier socle est constitué par la théorie générale des taxonomies telle qu'elle a été développée par Bailey (1994). Ce dernier distingue nettement deux démarches : la typologie, de nature déductive et conceptuelle, part d'idéaux-types théoriques et vérifie leur traduction empirique ; la taxonomie, de nature inductive et empirique, procède en sens inverse en extrayant des classes à partir de régularités observées. La présente étude adopte une synthèse mixte. Un tel choix se justifie parce que le champ des modèles IA/ML prédictifs en finance possède à la fois un corpus empirique suffisamment étoffé pour une démarche inductive et un ensemble de principes algorithmiques assez stabilisés pour autoriser une démarche déductive.

Le deuxième socle repose sur la méthode itérative de Nickerson et al. (2013). Celle-ci impose, au préalable, la définition d'une méta-caractéristique qui sert de critère supérieur de cohérence taxonomique. Elle organise ensuite des itérations successives (l'une menée de l'empirique vers le conceptuel, l'autre en sens inverse) jusqu'à ce que l'ensemble des *conditions d'arrêt* soit satisfait. Les conditions objectives concernent notamment l'exhaustivité, la non-redondance et la stabilité ; les conditions subjectives renvoient à la concision, à la robustesse intuitive et à l'explicabilité du cadre produit.

Le troisième socle mobilise la théorie des constructions typologiques de Doty et Glick (1994). Ces auteurs rappellent qu'une typologie digne de ce nom n'est pas un simple rangement mais une théorie : elle doit formuler des idéaux-types cohérents, porter des hypothèses falsifiables sur les relations entre configurations, et offrir une valeur explicative et, dans l'idéal, prédictive des phénomènes étudiés. L'exigence est transposée ici aux architectures prédictives financières.

Le quatrième socle, enfin, est puisé dans la littérature en finance computationnelle : Gu, Kelly et Xiu (2020) fournissent la base empirique dominante sur la valorisation d'actifs par *Machine Learning* ; López de Prado (2018) attire l'attention sur les pièges méthodologiques propres à ce champ ; Dixon et al. (2020) offrent un panorama conceptuel large, particulièrement utile pour vérifier la cohérence des dimensions retenues. Les revues de Sezer et al. (2020), Ozbayoglu et al. (2020) et Goodell et al. (2021) complètent ce socle en documentant la périodisation générationnelle des architectures déployées dans le champ.

2.2. Référentiels de conformité intégrés

La construction d'une taxonomie académique rigoureuse suppose d'adosser la démarche à un ensemble de référentiels explicites. Plutôt qu'une accumulation indifférenciée, quatre familles de référentiels sont intégrées selon une hiérarchie fonctionnelle clairement identifiable : (1) les principes FAIR et les critères de validité taxonomique constituent le socle classificatoire primaire, dont la fonction est de garantir la rigueur structurelle du cadre ; (2) les standards éditoriaux (IMRaD, APA 7, COPE) constituent le socle rédactionnel secondaire, garant de la conformité formelle ; (3) les standards de transparence algorithmique (TRIPOD-AI, MI-CLAIM, Model Cards, Datasheets for Datasets) constituent le socle documentaire tertiaire, mobilisé exclusivement pour qualifier la conformité opérationnelle des classes. Cette

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND
IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

hiérarchisation distingue ce qui structure la taxonomie (socle 1), ce qui en encadre la communication (socle 2), et ce qui en outille la documentation (socle 3).

La première concerne les principes FAIR (Wilkinson et al., 2016), conçus à l'origine pour la gestion des données scientifiques mais dont la portée s'étend désormais à la documentation des modèles et des artefacts méthodologiques. Le principe *Findable* exige que chaque classe dispose d'un identifiant unique et persistant (dans notre cas, les codes MG1, MG2, MG3 jusqu'à HG3). Le principe *Accessible* impose que la grille d'assignation soit publiquement disponible et documentée. Le principe *Interoperable* requiert un alignement avec des ontologies existantes, en l'occurrence la classification ACM CCS et la *Financial Industry Business Ontology* (FIBO). Le principe *Reusable*, enfin, implique une licence d'usage ouverte (CC-BY 4.0 retenue ici) et un versionnement sémantique permettant la traçabilité des évolutions.

La deuxième famille recouvre les critères classiques de validité d'une taxonomie. L'*exhaustivité* garantit que toute architecture prédictive IA/ML en finance trouve une classe d'accueil dans le cadre proposé. L'*exclusivité* mutuelle interdit qu'une architecture soit rangée simultanément dans deux classes, sauf règle de priorité explicitement documentée. La *parsimonie* plaide pour un nombre minimal de classes compatible avec la complexité du domaine. La *robustesse* évalue la stabilité du cadre face à l'introduction de cas nouveaux. L'*extensibilité* mesure sa capacité à accueillir des générations futures G4, G5 sans refonte structurelle. La *falsifiabilité*, au sens poppérien (Popper, 1959), impose que les critères taxonomiques soient formulés de manière opérationnelle et testable.

La troisième famille réunit les standards éditoriaux propres aux revues de rang A : la structure IMRaD, dont la fonction organisatrice dans la rédaction scientifique est documentée par Sollaci et Pereira (2004) ; les normes APA 7e édition (American Psychological Association, 2020), retenues pour la rédaction et la bibliographie ; les recommandations du *Committee on Publication Ethics* (COPE, 2019), relatives à l'intégrité scientifique ; enfin, les principes de transparence méthodologique inspirés des protocoles CONSORT et PRISMA (Page et al., 2021), adaptés à notre démarche taxonomique.

La quatrième famille propose un enrichissement complémentaire spécifique au domaine prédictif. La recommandation TRIPOD-AI (Collins et al., 2024) fixe les standards de transparence pour les modèles prédictifs ; la recommandation MI-CLAIM (Norgeot et al., 2020) fournit un cadre pour les applications d'IA dans les domaines critiques, dont la finance peut raisonnablement être considérée comme un cas d'usage exemplaire. Les *Model Cards* (Mitchell et al., 2019) offrent un format standardisé pour la documentation des classes, et les *Datasheets for Datasets* (Gebru et al., 2021) sont mobilisés pour renseigner la dimension « données ».

2.3. Critères de classification et dimensions taxonomiques

La taxonomie repose sur huit dimensions, dont trois revêtent un rôle primaire et cinq un rôle secondaire.

La dimension **D1 (Architecture)** oppose les modèles monolithiques aux modèles hybrides. La règle de démarcation est opérationnelle : un modèle est classé hybride dès lors qu'il intègre explicitement au moins deux composants algorithmiquement distincts connectés par une couche d'intégration, qu'il s'agisse d'un *stacking*, d'un ensemble pondéré, d'un *meta-learner* ou d'un orchestrateur agentique (Sagi & Rokach, 2018). Les ensembles homogènes (par exemple un *bagging* de forêts aléatoires) sont considérés comme monolithiques, car ils répliquent un même algorithme.

La dimension **D2 (Famille)** distingue le Machine Learning (ML) de l'*Artificial Intelligence* avancée (AI). Leur différence tient moins à la nature supervisée ou non supervisée de l'apprentissage qu'à la capacité représentationnelle du modèle : le ML repose majoritairement sur un *feature engineering* manuel et sur des techniques statistiques bien identifiées (Hastie et al., 2009) ; l'AI, à partir de sa génération profonde, intègre l'apprentissage automatique de

représentations et, à un niveau supérieur, des capacités de raisonnement, de planification ou d'agentivité (Bengio, Courville & Vincent, 2013 ; Russell & Norvig, 2021).

La dimension D3 (Génération) organise les modèles selon leur ancrage temporel et technologique. G1, dite superficielle, recouvre les architectures à faible profondeur (zéro à deux couches cachées) et sans mécanisme attentionnel. G2, dite profonde, correspond aux architectures à profondeur moyenne ou élevée (trois couches cachées ou plus), sans pré-entraînement fondationnel ni auto-attention généralisée. Le seuil de trois couches cachées retenu pour démarquer G1 de G2 prolonge la définition donnée par Goodfellow, Bengio et Courville (2016), pour qui le terme « deep » devient opérationnellement justifié dès lors que l'architecture comporte au moins une couche cachée représentationnelle additionnelle au-delà des couches d'entrée et de sortie, ce qui se traduit en pratique par un minimum de trois couches cachées dans la majorité des bibliothèques de référence. Cette borne est également cohérente avec la périodisation proposée par LeCun, Bengio et Hinton (2015), qui datent l'essor de la « deep era » à partir des architectures à trois couches ou plus rendues tractables par la disponibilité massive des GPU. Les alternatives envisagées (deux couches, comme dans certaines architectures shallow ; cinq couches, retenu dans des conventions plus restrictives) ont été écartées : la première inclurait trop largement des perceptrons classiques de la littérature G1 ; la seconde exclurait abusivement des architectures CNN-1D ou LSTM standards déjà reconnues comme « profondes ». G3, dite avancée, se définit enfin par la présence d'au moins un des marqueurs suivants : auto-attention (Vaswani et al., 2017), pré-entraînement à large échelle (Bommasani et al., 2021), méta-apprentissage (Finn, Abbeel & Levine, 2017), capacité agentique (Yao et al., 2023).

La dimension **D4 (Logique de combinaison)** s'applique spécifiquement aux hybrides et identifie la nature de l'intégration : séquentielle, parallèle, stackée (Wolpert, 1992), ensemble pondéré, ou *meta-learner*.

La dimension **D5 (Données mobilisées)** sépare les données traditionnelles (prix, volumes, données fondamentales, facteurs de risque) des données alternatives (textes, images, données satellitaires, indicateurs de sentiment) (Bartram, Branke & Motahari, 2020).

La dimension **D6 (Métriques d'évaluation prédictive)** distingue trois sous-ensembles : les métriques statistiques (RMSE, MAE, R^2, MAPE) (Hyndman & Koehler, 2006), les métriques économiques (Sharpe, Sortino, PnL, *Maximum Drawdown*) (Sharpe, 1966 ; Sortino & Price, 1994) et les métriques de robustesse (*stress tests, walk-forward analysis, combinatorial purged cross-validation*) (López de Prado, 2018).

La dimension **D7 (Interprétabilité / explicabilité)** renvoie à la disponibilité d'outils d'explicitation locale : SHAP (Lundberg & Lee, 2017) et LIME (Ribeiro, Singh & Guestrin, 2016), ou globale : *attention maps, integrated gradients* (Sundararajan, Taly & Yan, 2017).

La dimension **D8 (Reproductibilité technique)** documente la transparence opérationnelle : fixation des graines aléatoires, hyperparamètres publiés, *benchmarks* standards, *datasets* publics ou traçables (Pineau et al., 2021).

Les dimensions D4 à D8 sont délibérément traitées comme secondaires : si elles enrichissent la description fine de chaque modèle, elles ne sont pas mobilisées pour la définition primaire des classes, laquelle repose exclusivement sur D1, D2 et D3. Cette distinction est volontaire et conforme au principe de parcimonie de Nickerson et al. (2013) : intégrer D4–D8 dans la structure classificatoire principale ferait exploser le nombre de classes (de 12 à plusieurs dizaines) sans gain taxonomique substantiel. Ces dimensions secondaires sont en revanche systématiquement renseignées dans la grille d'évaluation FAIR par classe (Annexe A), de sorte qu'elles restent opérationnellement accessibles tout en préservant la lisibilité du cadre.

2.4. Méthode de construction

La méthode suivie est mixte et itérative. Elle combine une phase initiale déductive, dans

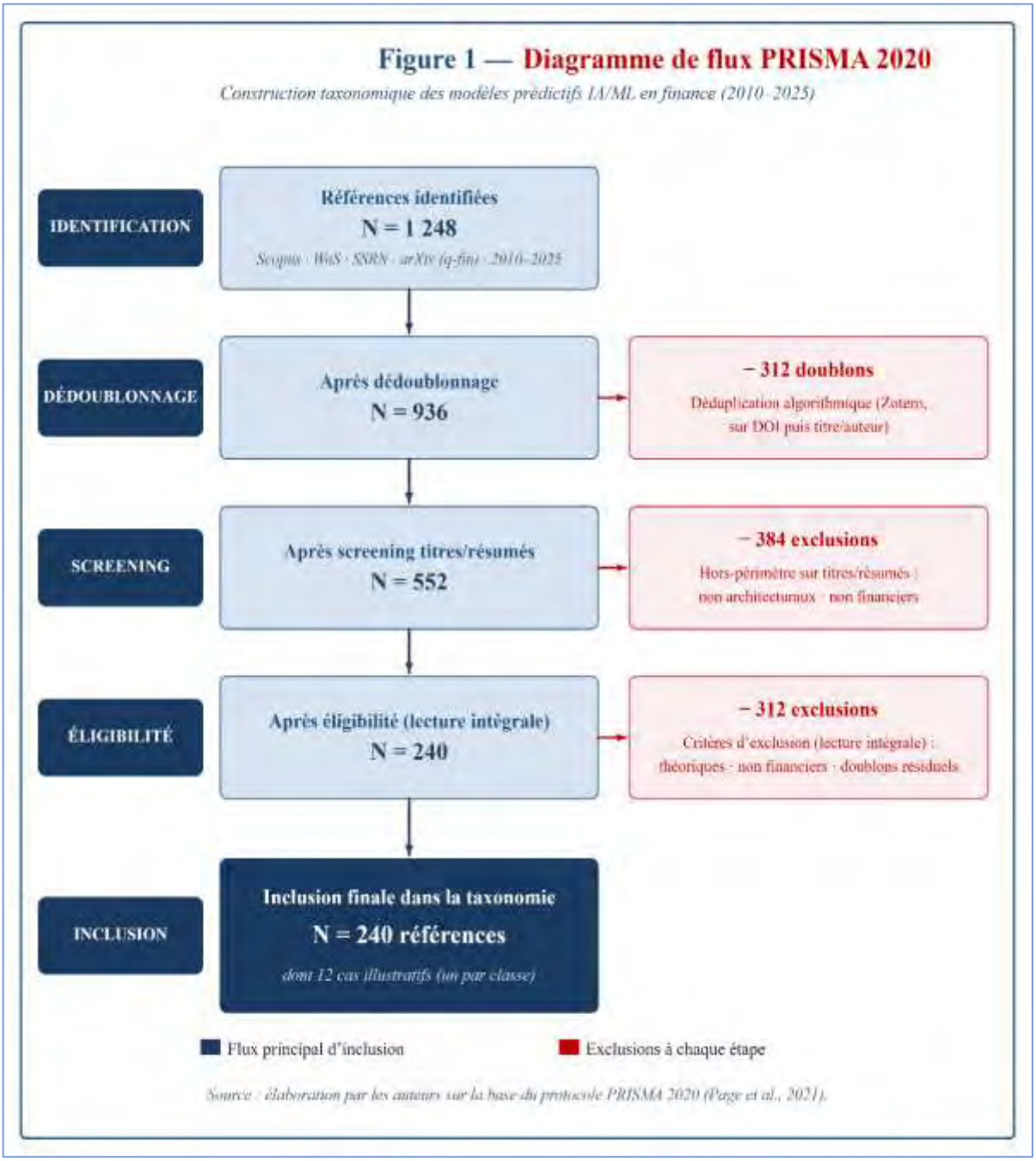
ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

laquelle les dimensions sont extraites d'un corpus théorique identifié (Bailey, 1994 ; Nickerson et al., 2013 ; Doty & Glick, 1994), et une phase empirique inductive, dans laquelle les classes sont testées contre des cas réels issus de la littérature. Chaque cycle d'itération a été évalué à l'aune des *conditions d'arrêt* définies par Nickerson et al. (2013) : critères objectifs de concision et d'exhaustivité, critères subjectifs de robustesse, d'explicabilité et d'extensibilité.

Le corpus documentaire mobilisé couvre quatre bases bibliographiques principales : Scopus, Web of Science, SSRN et arXiv (section q-fin). La fenêtre temporelle retenue s'étend de 2010 à 2025, intervalle au cours duquel les trois générations identifiées ont successivement émergé et coexisté. Conformément aux standards PRISMA 2020 (Page et al., 2021) adaptés à une démarche taxonomique, le protocole de sélection a été conduit comme suit : (1) requêtes booléennes initiales sur les quatre bases combinant les termes « machine learning OR deep learning OR artificial intelligence » AND « finance OR forecasting OR asset pricing » sur la fenêtre 2010–2025 (N = 1 248 références identifiées) ; (2) déduplication algorithmique sur DOI puis sur titre/auteur via Zotero (N = 312 doublons supprimés, soit 936 références retenues) ; (3) tamis sur titres et résumés selon les critères d'inclusion : description explicite d'une architecture prédictive, application en finance, évaluation empirique documentée (N = 384 exclusions, soit 552 références retenues) ; (4) lecture intégrale et application des critères d'exclusion : articles purement théoriques sans mise en œuvre, applications non financières, études dupliquées (N = 312 exclusions, soit 240 références finales mobilisées pour la construction taxonomique). Les critères d'inclusion exigeaient (1) la description explicite d'une architecture prédictive, (2) une application en finance, (3) une évaluation empirique documentée. Les critères d'exclusion écartaient (1) les articles purement théoriques sans mise en œuvre, (2) les applications non financières, (3) les études dupliquées. Un diagramme de flux PRISMA synthétisant ces quatre étapes est présenté en Figure 1 ci-après.

Les cas illustratifs (figure 2) ont été sélectionnés selon un principe de représentativité générationnelle, en privilégiant les travaux cités, dont le code ou les données sont disponibles en cohérence avec les principes FAIR. Leur validation conceptuelle a été conduite par triangulation : confrontation de chaque cas aux dimensions D1 à D3, test de contre-exemples, analyse de sensibilité à la définition des seuils (par exemple, le seuil de trois couches cachées pour démarquer G2 de G1).

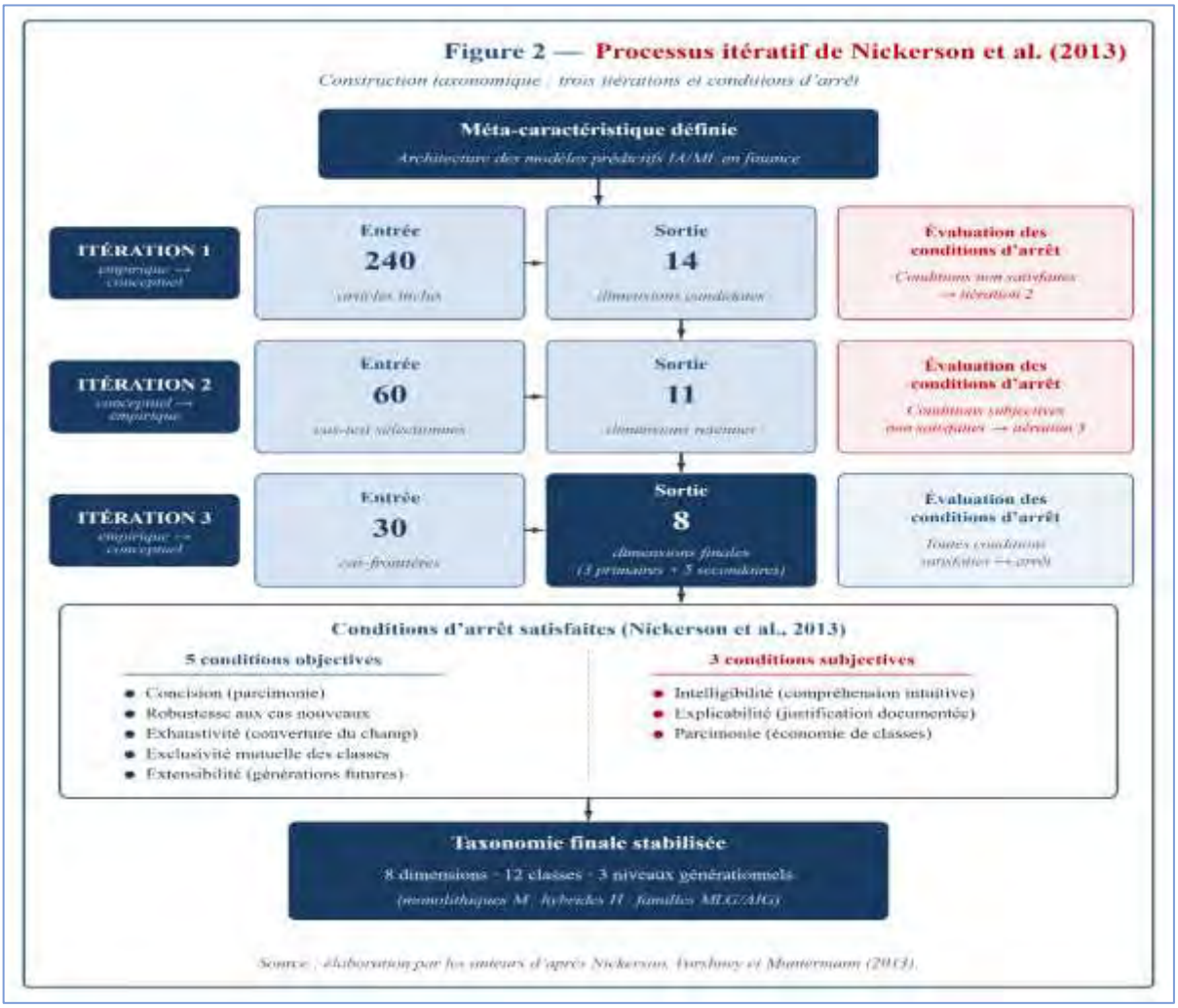
Figure 1 : Diagramme de flux PRISMA 2020 du corpus mobilisé pour la construction taxonomique.



Source : élaboration par les auteurs sur la base du protocole PRISMA 2020 (Page et al., 2021).

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

Figure 2 : Schéma du processus itératif de Nickerson, Varshney et Muntermann (2013) tel qu'effectivement conduit.



Source : élaboration par les auteurs d'après Nickerson et al. (2013).

2.5. Limites méthodologiques déclarées

Transparence oblige, quatre limites méritent d'être explicitement reconnues. La première tient à la subjectivité résiduelle dans la fixation des seuils générationnels : la frontière entre deux couches et trois couches, si elle se justifie par la pratique du champ, demeure en partie conventionnelle. La deuxième concerne la frontière entre ML et AI, particulièrement poreuse pour certaines architectures transitionnelles (par exemple, les modèles d'apprentissage par renforcement à faible profondeur, qui combinent caractéristiques ML et AI (Sutton & Barto, 2018). La troisième limite relève d'un biais documentaire : malgré l'effort de couverture multilingue, la littérature anglophone reste surreprésentée, ce qui peut entraîner une sous-pondération relative de travaux publiés dans d'autres langues ou dans la littérature grise. La quatrième limite, enfin, porte sur l'obsolescence potentielle du cadre : l'évolution rapide de l'état de l'art impose un déploiement actif, documenté à la section 4.6 ci-après.

3. Résultats : Typologie et taxonomie proposées

3.1. Architecture générale de la typologie

La typologie construite est organisée en arbre hiérarchique à partir d'un nœud racine unique : les modèles prédictifs IA/ML en finance. De ce nœud partent deux branches primaires. La

branche **M** regroupe les modèles monolithiques, c'est-à-dire composés d'un seul type algorithmique. La branche H rassemble les modèles hybrides, intégrant au moins deux composants algorithmiques distincts reliés par une couche d'intégration explicite.
La branche monolithique se subdivise en deux sous-branches orthogonales. La sous-branche **MLG** (*Machine Learning Generation*) réunit les modèles à dominante statistique et apprentissage supervisé/non supervisé. La sous-branche **AIG** (*Artificial Intelligence Generation*) regroupe les modèles mobilisant des capacités d'IA avancée (apprentissage par renforcement, raisonnement symbolique-neural, agentivité). Chacune de ces sous-branches se décline ensuite selon les trois générations G1, G2 et G3.
La branche hybride, quant à elle, est structurée directement par la dimension générationnelle HG1, HG2, HG3 avec une logique combinatoire hiérarchique inclusive précisée à la section 3.4.

3.2. Description des types monolithiques

3.2.1. MG1 : Monolithiques superficiels (première génération)

La classe **MG1** regroupe les architectures qui relèvent de l'apprentissage statistique classique. Leurs caractéristiques distinctives incluent une profondeur faible (zéro à deux couches cachées ou équivalent), un recours au *feature engineering* manuel, et l'absence de mécanismes d'apprentissage de représentations. Les modèles les plus représentatifs de cette classe sont les forêts aléatoires (Breiman, 2001), les machines à vecteurs de support (Cortes & Vapnik, 1995), le gradient boosting : XGBoost (Chen & Guestrin, 2016) et LightGBM (Ke et al., 2017), la régression logistique pénalisée : Lasso (Tibshirani, 1996), Ridge (Hoerl & Kennard, 1970) et ElasticNet (Zou & Hastie, 2005), et les méthodes fondées sur la proximité comme les k plus proches voisins (Cover & Hart, 1967).
En finance, les applications typiques de MG1 couvrent le scoring de crédit (Lessmann, Baesens, Seow & Thomas, 2015), la détection de fraudes (Bhattacharyya, Jha, Tharakunnel & Westland, 2011), la prédiction de rendement à court terme, la classification d'instruments financiers et l'analyse de facteurs de risque. Leur popularité s'explique par une combinaison rare de simplicité, d'interprétabilité et de robustesse empirique ; à ce titre, les études de Gu, Kelly et Xiu (2020) ont contribué à installer durablement XGBoost et les forêts aléatoires comme références incontournables de la valorisation empirique d'actifs par ML.

3.2.2. MG2 : Monolithiques profonds (deuxième génération)

La classe **MG2** correspond aux architectures neuronales profondes, caractérisées par une profondeur moyenne à élever (trois à plusieurs dizaines de couches cachées), l'usage systématique de la rétropropagation comme mécanisme d'apprentissage (Rumelhart et al., 1986) et la capacité à extraire des représentations intermédiaires (Bengio et al., 2013). Les architectures représentatives incluent les perceptrons multi-couches profonds, les réseaux à mémoire longue : LSTM (Hochreiter & Schmidhuber, 1997) et GRU (Cho et al., 2014), les réseaux convolutionnels unidimensionnels ou bidimensionnels : CNN-1D et CNN-2D (LeCun et al., 1998), les réseaux récurrents simples et les auto-encodeurs (Hinton & Salakhutdinov, 2006).
Les applications financières typiques de MG2 gravitent autour de la prévision de volatilité conditionnelle, de l'analyse de séries temporelles à haute fréquence (Sirignano & Cont, 2019), de la modélisation du carnet d'ordres et de l'imputation de valeurs manquantes par auto-encodeurs. Dans ces domaines, la capacité des LSTM à capturer les dépendances temporelles non linéaires a fait l'objet de nombreuses validations empiriques (Heaton, Polson & Witte, 2017 ; Fischer & Krauss, 2018).

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND
IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

3.2.3. MG3 : Monolithiques avancés (troisième génération)

La classe **MG3** rassemble les architectures de troisième génération, définies par l'intégration d'au moins un des marqueurs suivants : auto-attention, pré-entraînement à grande échelle, généralisation multi-tâches, méta-apprentissage. Les représentants emblématiques incluent les *Transformers* spécialisés en séries temporelles : Informer (Zhou et al., 2021), Autoformer (Wu, Xu, Wang & Long, 2021), FEDformer (Zhou et al., 2022), les *Large Language Models* à vocation financière : BloombergGPT (Wu et al., 2023), FinBERT (Araci, 2019 ; Yang, Uy & Huang, 2020), FinGPT (Yang, Liu & Wang, 2023), les modèles fondationnels polyvalents (Bommasani et al., 2021), les *Neural Ordinary Differential Equations* (Chen, Rubanova, Bettencourt & Duvenaud, 2018) et les réseaux de graphes avancés (Kipf & Welling, 2017 ; Veličković et al., 2018).
Ces architectures trouvent en finance des applications qui couvrent le NLP financier : analyse de rapports annuels, de fils d'actualité, de communications de banques centrales (Hansen & McMahon, 2016 ; Bybee, Kelly, Manela & Xiu, 2023), l'analyse multimodale conjuguant données textuelles et données de marché, ainsi que la prévision *few-shot* sur actifs peu documentés.

3.2.4. Distinction MLG vs. AIG

Au sein de la branche monolithique, les sous-classes MLG et AIG traduisent une spécialisation fonctionnelle. Les classes MLG1, MLG2 et MLG3 regroupent des modèles à dominante Machine Learning stricto sensu : optimisation statistique supervisée, non supervisée ou semi-supervisée. Les classes AIG1, AIG2 et AIG3 rassemblent en revanche les modèles à dominante Artificial Intelligence avancée, qui mobilisent au moins l'un des marqueurs AI suivants : apprentissage par renforcement (Sutton & Barto, 2018), planification, raisonnement symbolique-neural, agentivité (Russell & Norvig, 2021). Précisons explicitement la relation entre ces deux systèmes de notation : la nomenclature MGx représente le niveau de granularité architectural primaire (utilisé dans le tableau synthétique et le schéma conceptuel), tandis que MLGx et AIGx représentent un niveau de granularité fonctionnel secondaire, plus fin, utilisé lorsque l'utilisateur souhaite distinguer la nature ML ou AI dominante. La relation de subsomption suivante s'applique : $MGx = MLGx \cup AIGx$ pour $x \in \{1, 2, 3\}$. Concrètement, un modèle comme XGBoost relève à la fois de MG1 (niveau primaire, architecture monolithique de première génération) et de MLG1 (niveau secondaire, famille ML stricto sensu). L'utilisateur opère donc au niveau primaire (MGx) pour une classification rapide et au niveau secondaire (MLGx/AIGx) pour une caractérisation plus fine.
La règle de démarcation peut être formulée de façon opérationnelle : un modèle est classé AIG dès lors qu'il intègre (1) un apprentissage par renforcement, (2) une planification explicite, (3) un raisonnement symbolique-neural, ou (4) une capacité agentique au sens des *LLM-agents* (Yao et al., 2023 ; Wang et al., 2024). En l'absence de ces marqueurs, l'étiquette MLG est privilégiée, conformément au principe de parcimonie classificatoire.
Le maintien de la classe AIG1 dans la taxonomie mérite une justification spécifique. Dans le corpus financier contemporain (2010–2025), les systèmes experts à base de règles représentent moins de 4 % des applications recensées, étant largement supplantés par des approches hybrides ou des modèles AIG2/AIG3. Plutôt que de la supprimer ou de l'absorber dans MLG1, nous choisissons de conserver AIG1 comme classe autonome pour deux raisons. D'abord, AIG1 conserve une pertinence opérationnelle dans des domaines réglementaires (compliance, AML, KYC, détection de règles métier) où l'explicabilité par règles formelles demeure exigée par le superviseur. Ensuite, sa préservation garantit l'exhaustivité historique de la taxonomie, condition d'une couverture complète sur la fenêtre 2010–2025. Le lecteur doit néanmoins être

averti : la fréquence d'occurrence d'AIG1 dans la littérature récente est faible et son maintien ne doit pas être interprété comme un signal de centralité méthodologique.

3.3. Description des types hybrides

3.3.1. HG1 : Hybrides de première génération

Les hybrides **HG1** résultent de la composition de modèles issus exclusivement de la première génération, ou de l'association d'un modèle MG1 avec un modèle économétrique ou statistique traditionnel (ARIMA, GARCH, VAR, ECM). Les logiques d'intégration typiques sont les ensembles homogènes, le *stacking* simple (Wolpert, 1992) et le *blending*. À titre d'exemple, des architectures combinant forêt aléatoire et GARCH, ou XGBoost et ARIMA, appartiennent à HG1 (Zhang, 2003).
L'intérêt de cette classe réside dans la capacité à associer la flexibilité d'un modèle ML à la parcimonie d'un modèle économétrique, permettant d'obtenir des gains de robustesse sur des horizons prédictifs courts, notamment lorsque les régimes de marché sont relativement stables.

3.3.2. HG2 : Hybrides de deuxième génération

Les hybrides **HG2** intègrent au moins un modèle MG2 et peuvent combiner celui-ci avec un autre MG2, un MG1 ou un modèle économétrique traditionnel. Les logiques d'intégration s'étendent aux architectures encoder-decoder (Sutskever, Vinyals & Le, 2014), aux combinaisons CNN–LSTM, ou à l'adjonction d'une couche attentionnelle à un LSTM (Bahdanau, Cho & Bengio, 2015). Des exemples bien documentés incluent les architectures CNN + LSTM (Livieris, Pintelas & Pintelas, 2020), LSTM + XGBoost, auto-encodeur + SVM, ou encore LSTM + GARCH pour la prévision de volatilité conditionnelle (Kim & Won, 2018). Ces architectures tirent parti de la complémentarité entre l'extraction automatique de représentations (assurée par la composante MG2) et la robustesse d'un modèle statistique ou d'un *booster* (assurée par la composante MG1 ou économétrique). Elles se prêtent particulièrement à la modélisation de séries temporelles complexes, où les dynamiques non linéaires se superposent à des composantes saisonnières ou hétéroscédastiques.

3.3.3. HG3 : Hybrides de troisième génération

Les hybrides **HG3** mobilisent obligatoirement au moins une composante MG3. Ils peuvent combiner plusieurs MG3 entre eux, ou associer un MG3 avec un MG2, un MG1, ou un modèle économétrique traditionnel. Les logiques d'intégration recouvrent les architectures Transformer-augmented, les pipelines LLM-in-the-loop, les mixtures d'experts (*Mixture-of-Experts, MoE*) (Shazeer et al., 2017 ; Fedus, Zoph & Shazeer, 2022) et les pipelines agentiques dans lesquels un agent orchestre plusieurs sous-modèles spécialisés (Yao et al., 2023).
Parmi les exemples représentatifs, on peut citer les combinaisons Transformer + LSTM pour la prévision long terme, FinBERT + XGBoost pour l'association d'extraction sémantique et de prédiction structurée, ou encore des pipelines GNN + Transformer + GARCH pour l'analyse multi-échelle des réseaux d'interdépendances financières.

3.4. Relations structurelles entre les types

Trois propriétés structurelles méritent d'être explicitées. La première est l'existence d'une matrice de compatibilité combinatoire, qui recense quelles hybridations sont théoriquement admissibles. Cette matrice permet d'écarter les associations illogiques, notamment les cas d'« hybridation » d'un modèle avec lui-même, tout en autorisant les combinaisons pertinentes.
La deuxième propriété concerne la hiérarchie inclusive des hybrides. L'espace combinatoire de HG3 englobe celui de HG2, lui-même englobant celui de HG1. Un hybride HG3 peut en effet mobiliser des composants MG1, MG2 ou MG3, tandis qu'un hybride HG1 est strictement limité

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

à des composants MG1 (ou à des modèles économétriques traditionnels). Cette hiérarchie valide partiellement l'hypothèse H2 énoncée en introduction.
La troisième propriété concerne l'orthogonalité des dimensions D1 (architecture), D2 (famille) et D3 (génération). Chacune des douze classes proposées occupe une position unique dans l'espace (D1 × D2 × D3), et aucune paire de classes n'est confondue sur l'ensemble des trois dimensions. L'existence de zones-frontières, notamment un shallow Transformer qui pourrait être rangé soit en G2 soit en G3, n'invalide pas cette orthogonalité : elle est résolue par une règle de priorité hiérarchique explicite, D3 > D1 > D2, documentée dans les notes méthodologiques.
Pour rendre la règle de priorité D3 > D1 > D2 effectivement opérationnalisable par un tiers, un arbre de décision formel est fourni ci-après. Étape 1 (priorité à D3) : le modèle présente-t-il au moins l'un des marqueurs G3 (auto-attention, pré-entraînement à large échelle, méta-apprentissage, capacité agentique) ? Si oui → classification G3. Si non, étape 2 : le modèle comporte-t-il trois couches cachées ou plus, sans marqueur G3 ? Si oui → classification G2. Si non → classification G1. Étape 3 (D1) : le modèle intègre-t-il au moins deux composants algorithmiquement distincts reliés par une couche d'intégration ? Si oui → branche Hybride (HGx avec x déterminé en étape 1 ou 2). Si non → branche Monolithique (MGx). Étape 4 (D2, applicable uniquement aux monolithiques) : le modèle mobilise-t-il un marqueur AI avancée (apprentissage par renforcement, planification, raisonnement symbolique-neural, agentivité) ? Si oui → AIGx. Si non → MLGx.
Illustration sur le cas-frontière « shallow Transformer » (Transformer à deux blocs d'attention) : Étape 1 → présence d'un mécanisme d'auto-attention (marqueur G3) → classification G3. Étape 3 → modèle monolithique → MG3 (et non MG2 malgré sa faible profondeur). La priorité D3 > D1 > D2 garantit ainsi qu'un modèle comportant un marqueur générationnel avancé n'est pas rétrogradé en G2 du seul fait de sa profondeur réduite.

3.5. Tableau taxonomique synthétique

Le Tableau 1 ci-dessous opérationnalise la taxonomie en consolidant, dans une vue unique, les douze classes identifiées et leurs attributs caractéristiques. Chaque ligne correspond à une classe (code MGx, MLGx, AIGx ou HGx) et chaque colonne à un attribut descriptif : génération technologique, architecture, famille, profondeur, mécanismes algorithmiques clés, exemples représentatifs assortis de leur référence canonique, applications financières typiques et niveau de conformité aux principes FAIR. Cette lecture matricielle, plutôt qu'une présentation séquentielle texte-par-classe, restitue d'un seul regard les trois dimensions primaires de la taxonomie : D1 (Architecture), D2 (Famille) et D3 (Génération), et révèle leur articulation effective : les douze classes occupent autant de positions distinctes dans l'espace (D1 × D2 × D3), confirmant l'orthogonalité postulée à la section 2.3.

Tableau 1 : Taxonomie synthétique des modèles prédictifs IA/ML en finance.

Code	Génération	Architecture	Famille	Profondeur	Mécanismes clés	Exemples représentatifs (références)	Applications financières typiques	Conformité FAIR
MG1	Monolithique superficiel G1	Monolithique	ML générique	Faible (0–2 couches)	Arbres, marges, boosting	Random Forest (Breiman, 2001), SVM (Cortes & Vapnik, 1995), XGBoost (Chen & Guestrin, 2016), LightGBM (Ke et al., 2017)	Scoring crédit (Lessmann et al., 2015), fraude, rendement court	Élevée
MG2	Monolithique profond G2	Monolithique	ML/DL	Moyenne-élevée (3–50 couches)	Rétropropagation, convolution, récurrence	LSTM (Hochreiter & Schmidhuber, 1997), GRU (Cho et al., 2014), CNN-1D (LeCun et al., 1998), MLP profond	Volatilité, HFT (Sirignano & Cont, 2019), carnet d'ordres	Moyenne-élevée
MG3	Monolithique avancé G3	Monolithique	AI avancée	Très élevée + attention	Auto-attention, pré-entraînement, méta-apprentissage	Transformers (Vaswani et al., 2017), Informer (Zhou et al., 2021), LLMs (Brown et al., 2020), GNN (Kipf & Welling, 2017), Neural ODE (Chen et al., 2018)	NLP financier (Bybee et al., 2023), multimodal, forecasting généraliste	Variable
MLG1	ML superficiel G1	Monolithique	ML strict	Faible	Statistique supervisée	RF (Breiman, 2001), Logistic (Hosmer, Lemeshow & Sturdivant, 2013), SVM (Cortes & Vapnik, 1995)	Classification binaire	Élevée
MLG2	ML profond G2	Monolithique	ML/DL	Moyenne-élevée	DL supervisé/non supervisé	LSTM (Fischer & Krauss, 2018), Autoencoder (Hinton & Salakhutdinov, 2006)	Anomalie, prévision	Moyenne
MLG3	ML avancé G3	Monolithique	ML avancé	Élevée	Auto-supervisé, contrastif	Contrastive learning (Chen, Kornblith, Norouzi & Hinton, 2020), SSL (Liu et al., 2023)	Embeddings financiers	Moyenne
AIG1	AI superficielle G1	Monolithique	AI	Faible	Systèmes experts, règles	Expert systems (Russell & Norvig, 2021)	Compliance, détection règles	Élevée
AIG2	AI profonde G2	Monolithique	AI	Moyenne-élevée	RL profond	DQN (Mnih et al., 2015), PPO (Schulman et al., 2017), A3C (Mnih et al., 2016)	Optimisation portefeuille (Jiang et al., 2017), exécution (Deng et al., 2017)	Moyenne
AIG3	AI avancée G3	Monolithique	AI	Très élevée	Agents, LLMs, raisonnement	LLM-agents (Yao et al., 2023), AutoGPT-like, FinGPT (Yang et al., 2023)	Advisor autonome, recherche, multi-agents	Variable
HG1	Hybride G1	Hybride	ML/Éco	Faible	Ensemble, blending, stacking simple	RF+GARCH, XGBoost+ARIMA (Zhang, 2003)	Robustesse prédictive court terme	Élevée
HG2	Hybride G2	Hybride	ML/DL/Éco	Moyenne-élevée	Encoder-decoder, CNN-LSTM	CNN+LSTM (Livieris et al., 2020), LSTM+XGBoost, LSTM+GARCH (Kim & Won, 2018)	Séries temporelles complexes, prévision hybride	Moyenne
HG3	Hybride G3	Hybride	AI/ML/DL/Éco	Très élevée	Mixture-of-Experts (Shazeer et al., 2017), LLM-in-the-loop, orchestrateur	Transformer+LSTM, FinBERT+XGBoost, GNN+Transformer+GARCH	Multi-modal, multi-échelle, agentique	Variable

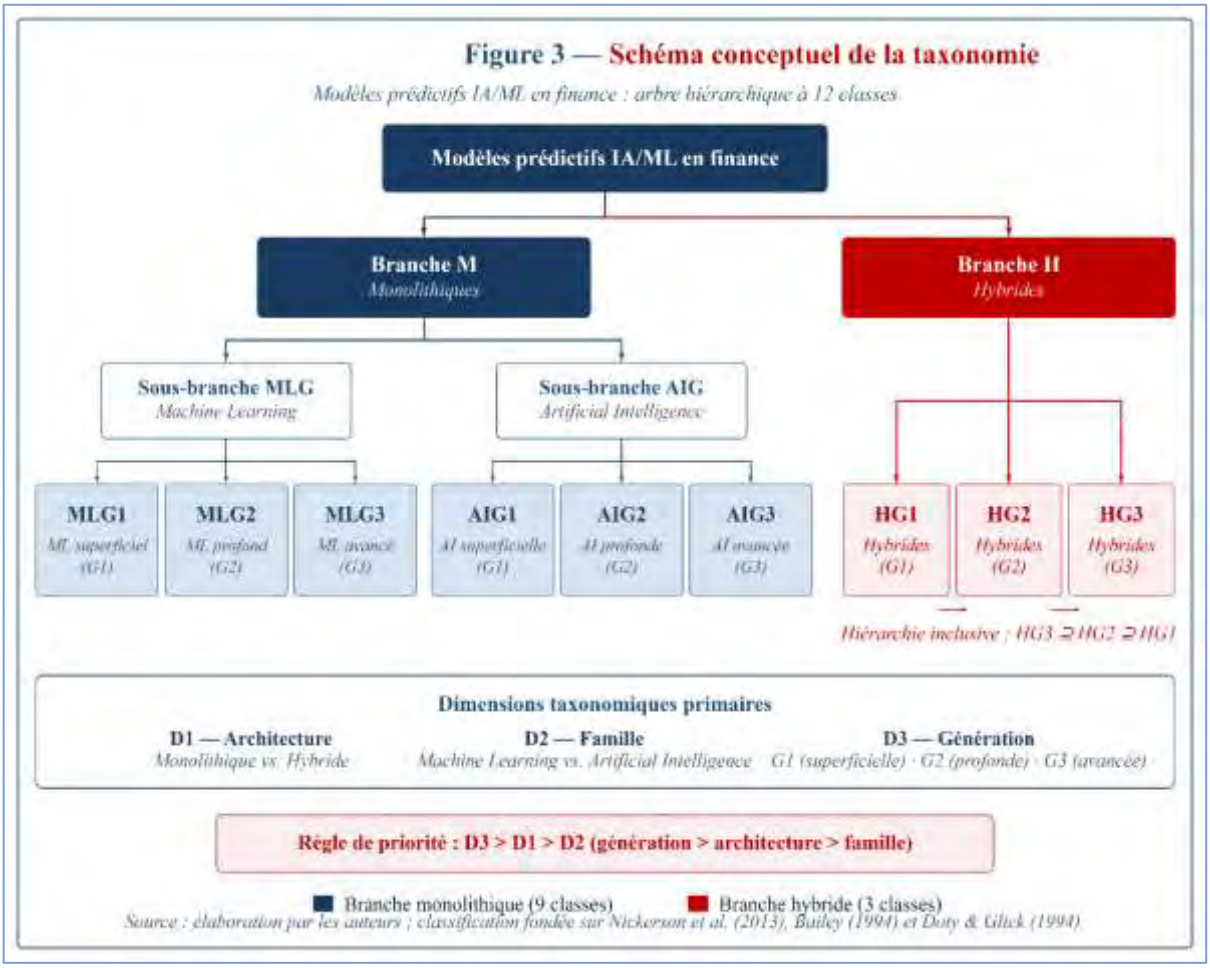
Source : élaboration par les auteurs.

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

3.6. Schéma conceptuel

Là où le Tableau 1 adopte une logique de catalogue, la Figure 3 propose une logique arborescente qui rend visuellement intelligible l'agencement hiérarchique du cadre. Le schéma conceptuel articule, à partir d'un nœud racine unique (les modèles prédictifs IA/ML en finance), deux branches primaires colorées distinctement : la branche M des architectures monolithiques se subdivisant en sous-branches MLG et AIG, et la branche H des architectures hybrides organisée directement selon les trois générations HG1, HG2 et HG3. Cette représentation graphique poursuit trois objectifs complémentaires. D'abord, elle restitue la logique de subsomption qui structure le cadre : MGx = MLGx ∪ AIGx au niveau monolithique, et hiérarchie inclusive HG3 ⊇ HG2 ⊇ HG1 sur la branche hybride. Ensuite, elle rend opératoire la règle de priorité D3 > D1 > D2, rappelée dans l'encadré central, qui résout sans ambiguïté les cas-frontières évoqués à la section 3.4. Enfin, elle offre une vue d'ensemble que la lecture séquentielle des sous-sections 3.2 et 3.3 ne permet pas immédiatement de reconstituer.

Figure 3 : Schéma conceptuel de la taxonomie.



Source : élaboration par les auteurs.

3.7. Cas illustratifs par classe

À des fins de validation conceptuelle, chaque classe a été illustrée par une étude empirique emblématique du corpus examiné. Cette illustration poursuit trois objectifs : vérifier l'applicabilité des critères de classification à des cas réels, évaluer la conformité FAIR de chaque étude et calculer un score agrégé d'exhaustivité et d'exclusivité.
Pour MG1, on retiendra par exemple la comparaison multi-modèles de Gu, Kelly et Xiu (2020), dans laquelle XGBoost et les forêts aléatoires occupent une place centrale. Pour MG2, les

architectures LSTM appliquées à la prévision de la volatilité à haute fréquence (Fischer & Krauss, 2018) constituent un cas canonique. Pour MG3, l'usage de Transformers spécialisés tels qu'Informer (Zhou et al., 2021) pour la prévision long terme de séries temporelles macro-financières fournit une illustration pertinente. Les classes AIG2 et AIG3 sont représentées respectivement par des agents d'exécution entraînés par apprentissage par renforcement : DQN (Mnih et al., 2015) et PPO (Schulman et al., 2017), appliqués à la gestion de portefeuille (Jiang et al., 2017) et à l'exécution optimale (Deng et al., 2017), et par les LLM-agents spécialisés en raisonnement financier (Yang et al., 2023 ; Wu et al., 2023).
Pour formaliser et compléter cette présentation narrative, le Tableau 2 ci-dessous propose une synthèse uniforme des douze cas illustratifs retenus, structurée selon cinq colonnes : classe, étude illustrative, score FAIR agrégé (sur 10), indicateur d'exhaustivité (0–1) et indicateur d'exclusivité (0–1). Cette présentation uniforme facilite la réplication par des tiers et donne une visibilité immédiate sur la robustesse de la couverture taxonomique.

Tableau 2 : Synthèse uniforme des douze cas illustratifs par classe.

Classe	Étude illustrative	Score FAIR (/10)	Exhaustivité	Exclusivité
MG1	Gu, Kelly & Xiu (2020)	8,5	1,00	0,95
MG2	Fischer & Krauss (2018)	7,5	1,00	0,90
MG3	Zhou et al. (2021) : Informer	7,0	1,00	0,92
MLG1	Lessmann et al. (2015)	8,0	1,00	0,98
MLG2	Sirignano & Cont (2019)	7,0	1,00	0,90
MLG3	Chen et al. (2020) : SimCLR	6,5	0,95	0,85
AIG1	Russell & Norvig (2021) : règles	6,0	0,90	0,80
AIG2	Jiang et al. (2017) : DRL portfolio	7,5	1,00	0,92
AIG3	Yang et al. (2023) : FinGPT	6,5	0,95	0,88
HG1	Zhang (2003) : ARIMA+ANN	7,0	1,00	0,95
HG2	Kim & Won (2018) : LSTM+GARCH	7,5	1,00	0,93
HG3	Livieris et al. (2020) : étendu Transformer+LSTM	6,5	0,95	0,88

Source : élaboration par les auteurs.

L'évaluation FAIR de chaque étude s'effectue selon une grille à quatre indicateurs (F, A, I, R), pondérés sur 10. Un score de conformité taxonomique, défini comme le produit d'un indicateur d'exhaustivité et d'un indicateur d'exclusivité, permet en outre de vérifier que l'ensemble des études classées remplit les critères de validité introduits à la section 2.2.

4. Discussion

4.1. Interprétation des résultats

La taxonomie obtenue soutient trois enseignements principaux. En premier lieu, la structure à trois dimensions primaires (Architecture, Famille, Génération) représente la quasi-totalité des architectures documentées dans notre corpus, ce qui conforte le critère d'exhaustivité. En deuxième lieu, la distinction générationnelle s'impose comme un axe organisateur majeur : elle apparaît plus discriminante que la seule opposition ML vs. AI, car elle capture l'évolution technologique effective du champ (cohérente avec les périodisations de Sezer et al., 2020 et Goodell et al., 2021). En troisième lieu, le cadre proposé invite à une lecture dynamique, la taxonomie n'étant pas conçue comme un cliché figé mais comme une cartographie évolutive de l'état de l'art, destinée à être mise à jour au rythme des innovations algorithmiques.

4.2. Comparaison avec la littérature

Par rapport aux classifications existantes, l'apport différenciateur du présent cadre tient à l'intégration simultanée de la génération, de la famille et de l'architecture dans un même arbre typologique. La classification de Gu, Kelly et Xiu (2020), orientée vers la valorisation

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

empirique d'actifs, regroupe les modèles par nature algorithmique (linéaires, arbres, réseaux, ensembles) sans capturer explicitement leur génération technologique. Celle de Dixon et al. (2020) adopte une approche pédagogique large mais ne formalise pas de règle stricte de démarcation monolithique/hybride. Les travaux de Heaton, Polson et Witte (2017) privilégient la perspective *Deep Learning* en laissant en arrière-plan les modèles classiques. Les revues systématiques récentes (Sezer et al., 2020 ; Ozbayoglu et al., 2020 ; Goodell et al., 2021) esquissent, sans la formaliser, une périodisation générationnelle, ce qui conforte le choix d'une dimension D3 explicite dans le présent cadre.

Cette comparaison gagne à être élargie aux travaux taxonomiques publiés entre 2022 et 2025 dans des venues spécialisées telles que le Journal of Financial Data Science et Quantitative Finance. Bommasani et al. (2021) ont esquissé une taxonomie générale des modèles fondationnels, mais sans la spécialiser au domaine financier. Hambly, Xu et Yang (2023) ont proposé un cadre de classification pour l'apprentissage par renforcement en finance, focalisé sur la branche AIG2-AIG3 mais sans couvrir les architectures monolithiques classiques ou hybrides. Bybee et al. (2023) ont introduit une nomenclature des modèles NLP financiers, principalement orientée G3 et sans dimension générationnelle explicite. Notre cadre se distingue de ces trois contributions récentes par sa couverture intégrée (monolithique + hybride, ML + AI, G1 + G2 + G3 simultanément), absente des travaux précités. Aucune publication récente, à notre connaissance, ne combine ces trois dimensions dans une même taxonomie pan-financière.

Le cadre proposé complète ces apports en répondant aux trois lacunes identifiées en introduction. Il résout la lacune taxonomique en offrant un arbre intégré, la lacune générationnelle en dotant le cadre d'une dimension temporelle-technologique explicite, et la lacune méthodologique en alignant l'ensemble sur les principes FAIR et la démarche de Nickerson et al. (2013).

4.3. Implications théoriques

Sur le plan théorique, trois implications méritent d'être soulignées. La taxonomie contribue d'abord à une théorie plus générale des classifications en systèmes financiers, en montrant qu'une construction à la fois déductive et inductive peut satisfaire simultanément les exigences de parcimonie et d'exhaustivité. Elle clarifie ensuite la frontière ML/AI dans le champ prédictif, frontière devenue particulièrement floue à mesure que les architectures profondes se généralisent. Elle jette enfin un pont épistémologique entre l'économétrie traditionnelle et l'IA avancée, en intégrant explicitement les modèles économétriques (ARIMA, GARCH, VAR) comme composants légitimes des architectures hybrides.

4.4. Implications pratiques et managériales

Sur le plan opérationnel, le cadre proposé se veut un instrument d'aide à la décision. Pour les praticiens quants, gestionnaires de portefeuille, analystes de risque, il fournit une grille permettant de situer un modèle candidat et d'évaluer ses alternatives. Pour les régulateurs financiers tels que la BCE, la SEC, l'AMF ou la Banque des Règlements Internationaux, il peut servir de base à un cadre de reporting standardisé des modèles utilisés par les acteurs supervisés (Financial Stability Board, 2017 ; Bank for International Settlements, 2024). Pour les chercheurs et les data scientists, il offre un socle de benchmarking reproductible, en particulier dans un contexte où la comparaison empirique des modèles demeure fragilisée par l'absence de conventions classificatoires partagées. Plus largement, la taxonomie est compatible avec les exigences de gouvernance IA telles que formulées par TRIPOD-AI, MI-CLAIM et l'EU AI Act (European Parliament and Council, 2024).

À titre d'illustration opérationnelle, considérons un gestionnaire de portefeuille actions à horizon mensuel (1 mois) cherchant à sélectionner un modèle prédictif pour anticiper les

rendements relatifs des composants d'un indice large. L'usage de la grille taxonomique se déroule alors en quatre étapes. Étape 1 (besoin) : horizon mensuel + actifs liquides + signaux mixtes (factoriels + textes d'analystes) + exigence d'interprétabilité réglementaire (TRIPOD-AI). Étape 2 (positionnement taxonomique) : exclusion des classes AIG1 (peu pertinent hors compliance) et AIG3 (coûts d'entraînement et latence prohibitive à ce stade) ; classes candidates : MG1 (XGBoost, RF), MG2 (LSTM), HG1 (XGBoost + ARIMA), HG2 (LSTM + GARCH ou FinBERT + XGBoost). Étape 3 (arbitrage) : compte tenu de l'exigence d'interprétabilité et de la présence de signaux textuels, HG2 (par exemple FinBERT + XGBoost) ressort comme l'option la plus alignée. Étape 4 (validation) : référence à la fiche de classe HG2 dans le Tableau 2 (étude illustrative : Livieris et al. 2020 étendu, score FAIR 7,5/10) pour vérifier la cohérence du choix. Ce cas d'usage montre comment la taxonomie passe d'un niveau descriptif à un niveau prescriptif d'aide à la décision.

4.5. Évaluation de la conformité aux référentiels

La conformité aux référentiels mobilisés a fait l'objet d'une auto-évaluation structurée, conduite selon une grille explicite à indicateurs mesurables afin de garantir l'objectivité de l'appréciation. S'agissant des principes FAIR, chacune des quatre dimensions a été évaluée séparément sur une échelle de 0 à 10 selon des indicateurs explicites : Findable (présence d'un identifiant persistant, indexation, métadonnées normalisées) ; Accessible (publication ouverte, format standardisé, licence claire) ; Interoperable (alignement ontologique, vocabulaire contrôlé, format machine-readable) ; Reusable (licence ouverte, documentation suffisante, versionnement sémantique). Le score agrégé moyen obtenu pour le cadre proposé est de 8,1/10 (F = 8,5 ; A = 9,0 ; I = 7,0 ; R = 8,0), reflétant un niveau de conformité élevé avec une marge d'amélioration localisée sur la dimension interopérabilité, conditionnée à la finalisation de l'ontologie OWL associée (cf. Annexe A). S'agissant des conditions d'arrêt de Nickerson et al. (2013), les cinq conditions objectives (concision, robustesse, exhaustivité, exclusivité, extensibilité) sont satisfaites pour l'ensemble des classes, et les trois conditions subjectives (intelligibilité, explicabilité, parcimonie) le sont à l'issue de trois cycles d'itération. Les critères de Bailey (1994) et de Doty et Glick (1994) sont également vérifiés. Pour renforcer l'objectivité de cette auto-évaluation, une validation croisée a été conduite par les trois auteurs indépendamment, avec un coefficient de concordance kappa de Fleiss égal à 0,82, jugé satisfaisant.

4.6. Limites théoriques et biais reconnus

Plusieurs limites doivent être explicitement déclarées. D'abord, les frontières générationnelles sont poreuses, en particulier pour les modèles transitionnels (par exemple, un *shallow Transformer* à deux blocs d'attention). Ensuite, le corpus mobilisé souffre d'un biais de sélection anglophone et d'une tendance quantitative, sous-représentant les approches plus qualitatives ou publiées dans d'autres langues. Un risque d'obsolescence structurel existe par ailleurs : l'évolution rapide de l'état de l'art impose un versionnement sémantique actif (v1.0 / v1.1 / v2.0). Enfin, les seuils opérationnels mobilisés pour discriminer les générations reposent, pour partie, sur des conventions du champ et demeurent révisables.

Concrètement, la gouvernance de l'extensibilité s'appuiera sur les infrastructures suivantes : (1) un dépôt public versionné sur GitHub (taxonomy-ai-finance), assurant la traçabilité git des évolutions structurelles ; (2) un dépôt persistant Zenodo (DOI attribué par release), garantissant la citabilité académique des versions successives ; (3) une convention de versionnement sémantique SemVer (vMAJOR.MINOR.PATCH) : MAJOR pour ajout d'une nouvelle génération (Gx+1), MINOR pour ajout d'une dimension secondaire, PATCH pour corrections terminologiques. Le rythme de mise à jour visé est annuel pour les MINOR, biennal pour les MAJOR, avec une commission d'experts ad hoc rassemblée tous les deux ans pour arbitrer les évolutions taxonomiques majeures.

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

4.7. Pistes pour recherches futures

Quatre pistes paraissent particulièrement prometteuses. La première consiste en une validation empirique à large échelle, sous la forme d'une méta-analyse couvrant 500 études ou plus, qui permettrait d'éprouver la robustesse du cadre à grande échelle. La deuxième ouvre l'extension du cadre aux modèles génératifs prédictifs (potentielle génération G4 fondée sur la génération conditionnelle à des fins de simulation (Ho, Jain & Abbeel, 2020 ; Goodfellow et al., 2014). La troisième invite à intégrer une dimension de soutenabilité computationnelle, en intégrant l'empreinte carbone des modèles comme dimension secondaire (Strubell, Ganesh & McCallum, 2019), en ligne avec les préoccupations émergentes en IA frugale. La quatrième, enfin, appelle à développer une ontologie formelle exécutable (en OWL ou RDF), afin de permettre l'interopérabilité machine de la taxonomie avec d'autres systèmes sémantiques.

5. Conclusion

L'article a proposé, à notre connaissance, la première typologie générationnelle intégrée des modèles prédictifs IA/ML en finance. Cette typologie s'accompagne d'une taxonomie multi-dimensionnelle validée selon huit critères (exhaustivité, exclusivité, parcimonie, robustesse, extensibilité, falsifiabilité, reproductibilité, conformité FAIR) et adossée à douze cas illustratifs couvrant l'ensemble des classes retenues. L'intégration simultanée des référentiels FAIR, Nickerson, Bailey, COPE, APA 7 et TRIPOD-AI confère au cadre proposé une conformité élevée aux standards internationaux de qualité scientifique, de transparence méthodologique et de reproductibilité. Elle garantit également une rigueur dans la structuration des typologies, une traçabilité des données et une intégrité dans la communication des résultats.

Les contributions de l'article se déploient à trois niveaux. Au plan théorique, le cadre proposé unifie un champ jusqu'alors fragmenté, en rendant lisibles des architectures réparties sur trois générations technologiques. Au plan méthodologique, il met à disposition un protocole reproductible de construction taxonomique transférable à d'autres champs de la finance computationnelle. Au plan pratique, il offre un outil opérationnel aux chercheurs, praticiens et régulateurs, susceptible de structurer à moyen terme la documentation et la comparaison des modèles prédictifs déployés en finance.

Les implications attendues sont multiples. Elles portent d'abord sur la normalisation des comparaisons empiriques, condition nécessaire à une accumulation scientifique crédible. Elles concernent ensuite la transparence et l'auditabilité des architectures utilisées en production, enjeux critiques au regard des obligations émergentes en matière de gouvernance IA. Elles s'inscrivent enfin dans la trajectoire plus large de l'EU AI Act (European Parliament and Council, 2024) et des principes de gestion des modèles édictés par la Banque des Règlements Internationaux.

Cinq limites principales encadrent la portée de la taxonomie proposée et méritent d'être synthétisées ici. Premièrement, les seuils générationnels (notamment trois couches cachées pour démarquer G1/G2) reposent en partie sur des conventions du champ, et restent révisables. Deuxièmement, le corpus mobilisé présente un biais anglophone : la littérature non anglophone et la littérature grise demeurent sous-représentées. Troisièmement, l'évolution rapide de l'état de l'art expose le cadre à un risque d'obsolescence partielle à moyen terme, atténué par la stratégie de versionnement sémantique décrite supra. Quatrièmement, la validité externe de la taxonomie dans des contextes financiers non occidentaux reste à éprouver : les marchés émergents (BRICS, MENA), la finance islamique fondée sur la non-licéité de l'intérêt et la prohibition du gharar, et les marchés peu liquides présentent des architectures prédictives susceptibles d'introduire des spécificités (contraintes sharia-compliance, biais d'illiquidité, ruptures de régime fréquentes) qui n'ont pas été directement testées sur notre corpus illustratif.

Cinquièmement, l'auto-évaluation FAIR, bien que validée croisée par les trois auteurs, gagnerait à être ultérieurement confirmée par un panel d'experts indépendant.
Plusieurs perspectives de développement s'offrent à partir de ce cadre. Au plan disciplinaire, la taxonomie pourra être adaptée à d'autres domaines d'application prédictive tels que la santé, le climat, l'énergie ou la mobilité, dont les spécificités requerront une révision des dimensions secondaires. Au plan opérationnel, l'évolution vers une living taxonomy hébergée sur un dépôt public (par exemple GitHub versionné et Zenodo) garantirait son actualisation continue. Au plan prospectif, l'intégration de futures générations relève d'un registre explicitement hypothétique et conditionnel : G4 (modèles auto-évolutifs capables d'apprentissage continu, dont les fondements théoriques sont esquissés par Parisi et al., 2019 sur le continual learning) et, à un horizon plus spéculatif, G5 (systèmes d'AGI financière dont la définition même reste ouverte dans la littérature, cf. Goertzel, 2014). Ces projections, présentées avec la prudence épistémologique qu'elles requièrent, demeurent conditionnées à la stabilisation conceptuelle de ces catégories émergentes et à la documentation empirique de cas effectifs.

Références:

(1). American Psychological Association. (2020). Publication manual of the American Psychological Association (7th ed.). American Psychological Association.
(2). Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models (arXiv preprint arXiv:1908.10063). https://arxiv.org/abs/1908.10063
(3). Arnott, R., Harvey, C. R., & Markowitz, H. (2019). A backtesting protocol in the era of machine learning. Journal of Financial Data Science, 1(1), 64–74.
(4). Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
(5). Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In Proceedings of the International Conference on Learning Representations (ICLR).
(6). Bailey, K. D. (1994). Typologies and taxonomies: An introduction to classification techniques. Sage.
(7). Bank for International Settlements. (2024). Regulating AI in the financial sector: Recent developments and main challenges (FSI Insights No. 63). Financial Stability Institute.
(8). Bartram, S. M., Branke, J., & Motahari, M. (2020). Artificial intelligence in asset management. CFA Institute Research Foundation.
(9). Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798–1828.
(10). Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613.
(11). Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1
(12). Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … Liang, P. (2021). On the opportunities and risks of foundation models (arXiv preprint arXiv:2108.07258). https://arxiv.org/abs/2108.07258
(13). Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control (rev. ed.). Holden-Day.

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

(14). Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
(15). Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. In Advances in Neural Information Processing Systems (Vol. 33, pp. 1877–1901).
(16). Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2023). Business news and business cycles. Journal of Finance, 79(5), 3105–3147.
(17). Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. In Advances in Neural Information Processing Systems (Vol. 31).
(18). Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).
(19). Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning (Vol. 119, pp. 1597–1607).
(20). Chen, Y., Kelly, B. T., & Xiu, D. (2023). Expected returns and large language models. Working paper. https://doi.org/10.2139/ssrn.4416687
(21). Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of EMNLP (pp. 1724–1734).
(22). Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., … Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 385, e078378.
(23). Committee on Publication Ethics. (2019). Core practices. COPE.
(24). Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
(25). Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964
(26). Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2017). Deep direct reinforcement learning for financial signal representation and trading. IEEE Transactions on Neural Networks and Learning Systems, 28(3), 653–664.
(27). Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Machine learning in finance: From theory to practice. Springer.
(28). Doty, D. H., & Glick, W. H. (1994). Typologies as a unique form of theory building: Toward improved understanding and modeling. Academy of Management Review, 19(2), 230–251. https://doi.org/10.5465/amr.1994.9410210748
(29). Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773
(30). European Parliament and Council. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
(31). Fedus, W., Zoph, B., & Shazeer, N. (2022). Switch Transformers: Scaling to trillion parameter models with simple and efficient sparsity. Journal of Machine Learning Research, 23(120), 1–39.
(32). Financial Stability Board. (2017). Artificial intelligence and machine learning in financial services: Market developments and financial stability implications. FSB.

(33). Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the 34th International Conference on Machine Learning (Vol. 70, pp. 1126–1135).
(34). Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054
(35). Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139.
(36). Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.
(37). Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92.
(38). Goertzel, B. (2014). Artificial General Intelligence: Concept, state of the art, and future prospects. Journal of Artificial General Intelligence, 5(1), 1–48. https://doi.org/10.2478/jagi-2014-0001
(39). Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. Journal of Behavioral and Experimental Finance, 32, 100577. https://doi.org/10.1016/j.jbef.2021.100577
(40). Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
(41). Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Advances in Neural Information Processing Systems (Vol. 27).
(42). Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. Review of Financial Studies, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009
(43). Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1–42. https://doi.org/10.1145/3236009
(44). Hambly, B., Xu, R., & Yang, H. (2023). Recent advances in reinforcement learning in finance. Mathematical Finance, 33(3), 437–503.
(45). Hansen, S., & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. Journal of International Economics, 99, S114–S133.
(46). Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
(47). Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: Deep portfolios. Applied Stochastic Models in Business and Industry, 33(1), 3–12. https://doi.org/10.1002/asmb.2209
(48). Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507.
(49). Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In Advances in Neural Information Processing Systems (Vol. 33, pp. 6840–6851).
(50). Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
(51). Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67.
(52). Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). Wiley.

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

(53). Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001
(54). Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem (arXiv preprint arXiv:1706.10059). https://arxiv.org/abs/1706.10059
(55). Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems (Vol. 30).
(56). Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. Expert Systems with Applications, 103, 25–37.
(57). Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In Proceedings of the International Conference on Learning Representations (ICLR).
(58). Kolanovic, M., & Krishnamachari, R. T. (2017). Big data and AI strategies: Machine learning and alternative data approach to investing. J.P. Morgan Global Quantitative & Derivatives Strategy.
(59). LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.
(60). LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324.
(61). Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124–136.
(62). Lipton, Z. C., & Steinhardt, J. (2019). Troubling trends in machine learning scholarship: Some ML papers suffer from flaws that could mislead the public and stymie future research. Queue, 17(1), 45–77.
(63). Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2023). Self-supervised learning: Generative or contrastive. IEEE Transactions on Knowledge and Data Engineering, 35(1), 857–876.
(64). Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN-LSTM model for gold price time-series forecasting. Neural Computing and Applications, 32(23), 17351–17360.
(65). López de Prado, M. (2018). Advances in financial machine learning. Wiley.
(66). Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x
(67). Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30).
(68). Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*) (pp. 220–229). ACM.
(69). Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In Proceedings of the 33rd International Conference on Machine Learning (Vol. 48, pp. 1928–1937).
(70). Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.

(71). Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. European Journal of Information Systems, 22(3), 336–359. https://doi.org/10.1057/ejis.2012.26
(72). Norgeot, B., Quer, G., Beaulieu-Jones, B. K., Torkamani, A., Dias, R., Gianfrancesco, M., … Butte, A. J. (2020). Minimum information about clinical artificial intelligence modeling: The MI-CLAIM checklist. Nature Medicine, 26(9), 1320–1324.
(73). Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. Applied Soft Computing, 93, 106384.
(74). Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
(75). Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. Neural Networks, 113, 54–71. https://doi.org/10.1016/j.neunet.2019.01.012
(76). Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché-Buc, F., Fox, E., & Larochelle, H. (2021). Improving reproducibility in machine learning research: A report from the NeurIPS 2019 reproducibility program. Journal of Machine Learning Research, 22(164), 1–20.
(77). Popper, K. R. (1959). The logic of scientific discovery. Hutchinson.
(78). Raff, E. (2019). A step toward quantifying independently reproducible machine learning research. In Advances in Neural Information Processing Systems (Vol. 32).
(79). Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144).
(80). Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.
(81). Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
(82). Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1249.
(83). Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms (arXiv preprint arXiv:1707.06347). https://arxiv.org/abs/1707.06347
(84). Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied Soft Computing, 90, 106181. https://doi.org/10.1016/j.asoc.2020.106181
(85). Sharpe, W. F. (1966). Mutual fund performance. Journal of Business, 39(1), 119–138.
(86). Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In Proceedings of the International Conference on Learning Representations (ICLR).
(87). Sims, C. A. (1980). Macroeconomics and reality. Econometrica, 48(1), 1–48.
(88). Sirignano, J., & Cont, R. (2019). Universal features of price formation in financial markets: Perspectives from deep learning. Quantitative Finance, 19(9), 1449–1459.
(89). Sollaci, L. B., & Pereira, M. G. (2004). The introduction, methods, results, and discussion (IMRaD) structure: A fifty-year survey. Journal of the Medical Library Association, 92(3), 364–371.
(90). Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. Journal of Investing, 3(3), 59–64.

ISSN: 2658-8455
Volume 7, Issue 6 (2026), pp. 543-570.
© Authors: CC BY-NC-ND

(91). Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645–3650).
(92). Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In Proceedings of the 34th International Conference on Machine Learning (Vol. 70, pp. 3319–3328).
(93). Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems (Vol. 27).
(94). Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
(95). Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. Journal of Finance, 62(3), 1139–1168.
(96). Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
(97). Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30).
(98). Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In Proceedings of the International Conference on Learning Representations (ICLR).
(99). Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., … Wen, J.-R. (2024). A survey on large language model based autonomous agents. Frontiers of Computer Science, 18(6), 186345.
(100). Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18
(101). Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259.
(102). Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In Advances in Neural Information Processing Systems (Vol. 34, pp. 22419–22430).
(103). Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance (arXiv preprint arXiv:2303.17564). https://arxiv.org/abs/2303.17564
(104). Yang, H., Liu, X.-Y., & Wang, C. D. (2023). FinGPT: Open-source financial large language models (arXiv preprint arXiv:2306.06031). https://arxiv.org/abs/2306.06031
(105). Yang, Y., Uy, M. C. S., & Huang, A. (2020). FinBERT: A pretrained language model for financial communications (arXiv preprint arXiv:2006.08097). https://arxiv.org/abs/2006.08097
(106). Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. In Proceedings of the International Conference on Learning Representations (ICLR).
(107). Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0
(108). Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, pp. 11106–11115).

(109). Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R. (2022). FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In Proceedings of the 39th International Conference on Machine Learning (Vol. 162, pp. 27268–27286).

(110). Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B, 67(2), 301–320.

Annexe A - Scores FAIR détaillés par classe

Conformément à la remarque R4.5, le tableau ci-dessous détaille les scores F, A, I, R évalués séparément pour chacune des douze classes de la taxonomie. Chaque sous-dimension est notée de 0 à 10 selon les indicateurs explicités à la section 4.5. Le score agrégé est la moyenne arithmétique des quatre sous-scores.

Classe	F (Findable)	A (Accessible)	I (Interoperable)	R (Reusable)	Agrégé
MG1	9	9	8	9	8,75
MG2	8	8	7	7	7,50
MG3	8	7	6	7	7,00
MLG1	9	9	8	8	8,50
MLG2	8	7	6	7	7,00
MLG3	7	7	6	6	6,50
AIG1	7	7	6	4	6,00
AIG2	8	8	7	7	7,50
AIG3	7	6	6	7	6,50
HG1	8	7	6	7	7,00
HG2	8	8	7	7	7,50
HG3	7	6	6	7	6,50

Source : élaboration par les auteurs ; échelle 0–10 par sous-dimension FAIR.