

Aspen University

AI-Driven Fair Creditworthiness and Credit Scoring :
A Two-Decade Analysis in the Tri-State Area
(2004–2024)

NAPO Tchín

Janvier 2026

Chapitre 5

Discussion

Introduction

Ce chapitre propose une interprétation des résultats empiriques présentés au Chapitre 4 et les replace dans les débats académiques et institutionnels relatifs à la transformation de l’octroi du crédit hypothécaire sous l’effet de l’intelligence artificielle (IA). Nous articulons les résultats autour des quatre questions de recherche (RQ1–RQ4), en mobilisant la littérature économique, juridique, sociologique et algorithmique. Une interprétation régionale est proposée pour le Tri-State Area (New York, New Jersey, Connecticut), suivie d’une discussion théorique, méthodologique et politique. Le chapitre se conclut sur les contributions, les limites et les perspectives futures.

La transition empirique documentée au Chapitre 4 — d’un régime pré-algorithmique (2004–2017) à un régime structuré par l’usage des modèles IA (2018–2023) — constitue une rupture majeure dans la production, la classification et l’utilisation des informa-

tions sur les emprunteurs. Cette rupture s’inscrit dans un débat académique beaucoup plus large sur l’automatisation du jugement (Kleinberg, Ludwig, Mullainathan, & Sunstein, 2018), les risques de « proxy discrimination » (Barocas & Selbst, 2016 ; Kearns & Roth, 2020), le rôle des données incomplètes (Fuster, Goldsmith-Pinkham, Ramcharan, & Vickery, 2019) et les transformations de la gouvernance financière par les algorithmes (Cowgill & Tucker, 2020).

Les résultats démontrent que :

1. les gradients de revenu se renforcent significativement sous l’ère IA ;
2. la prévisibilité des décisions de crédit s’accroît de manière substantielle ;
3. certains déterminants structurels (type de prêt, motif du prêt) deviennent plus stables ;
4. la fairness algorithmique s’améliore pour certains groupes (notamment Black), mais se détériore pour les profils ayant des informations manquantes.

Ces dynamiques traduisent non seulement une évolution technique, mais également une transformation institutionnelle profonde, inscrite dans une longue histoire de stratification résidentielle et financière aux États-Unis (Squires, 2017 ; Rothstein, 2018). Les sections suivantes développent ces observations en suivant la structure des questions de recherche.

5.1 Synthèse intégrée des résultats du Chapitre 4

Les résultats du Chapitre 4 révèlent quatre lignes de force essentielles pour comprendre l’impact de l’IA sur l’octroi du crédit dans le Tri-State Area.

5.1.1 Renforcement des gradients de revenu

Les analyses logistiques de la période IA (2018–2023) montrent des coefficients négatifs significatifs pour les groupes *Low income* et *Middle income*, indiquant une diminution substantielle des odds d’approbation par rapport aux zones à haut revenu. Ces résultats s’inscrivent dans la continuité des travaux de Fuster et al. (2019) et de Bhutta, Blair, et Dettling (2021), qui documentent l’importance des caractéristiques géographiques comme variables proxy des risques financiers.

L’amplification des gradients constatée sous IA suggère que les algorithmes apprennent — voire accentuent — les régularités socio-économiques situées dans le tissu urbain et résidentiel. La capacité des modèles à systématiser et amplifier ces structures, déjà signalée par Mullainathan et Spiess (2017), confirme que les modèles IA fonctionnent souvent comme des amplificateurs statistiques des réalités préexistantes. À ce titre, les résultats empiriques illustrent de manière concrète le déplacement du pouvoir de qualification du risque, des agents de crédit humains vers des infrastructures computationnelles qui agrègent des signaux issus des marchés du travail, du logement et de la richesse patrimoniale.

5.1.2 Accroissement de la performance prédictive

Le passage de pseudo- R^2 quasi nuls (pré-IA) à $\approx 0,38$ dans l'ère IA traduit une transformation majeure. Les modèles entraînés en période IA capturent plus efficacement les relations structurelles entre caractéristiques du prêt, contexte socio-économique et probabilité d'approbation. Les courbes ROC confirmées par la Figure 4.2 illustrent cette amélioration continue, qui s'inscrit dans les travaux de Kleinberg et al. (2018) sur la réduction du *noise* décisionnel dans les processus humains.

Dans une perspective plus large d'IA et de prise de décision, ces résultats dialoguent avec les contributions issues de la littérature en intelligence artificielle et en apprentissage statistique (par exemple Russell & Norvig), qui défendent l'idée que des systèmes d'aide à la décision bien calibrés peuvent améliorer la cohérence interne d'un processus tout en rendant plus visibles les arbitrages implicites. Ici, l'augmentation de la performance prédictive n'est pas neutre : elle s'accompagne d'une structuration plus nette des perdants et des gagnants du régime de crédit.

5.1.3 Persistance des déterminants structurels

Les variables *loan type* et *loan purpose*, historiquement centrales dans la littérature en finance immobilière (Di Maggio & Kermani, 2022), montrent des effets significatifs, cohérents et plus stables sous IA. L'hétérogénéité inter-institutionnelle — autrefois importante — se réduit substantiellement dans les données 2018–2023. Cela suggère que les modèles IA agissent comme un mécanisme de standardisation des pratiques de crédit, en alignant les barèmes implicites des différents prêteurs sur un socle commun de règles de décision.

Dans cette perspective, l’IA n’introduit pas seulement de nouveaux déterminants : elle hiérarchise et rigidifie les déterminants existants. Les configurations de risque liées au type de prêt (par exemple, prêts FHA vs prêts conventionnels) et au motif du prêt (achat, refinancement, consolidation) sont davantage internalisées dans les modèles, ce qui réduit la marge de manœuvre discrétionnaire au niveau local, mais peut aussi limiter la capacité d’adaptation aux contextes particuliers.

5.1.4 Reconfiguration des patterns de fairness

Les analyses de fairness du Tableau 6.4 révèlent deux dynamiques contrastées :

- convergence presque parfaite entre White et Black sur TPR, FNR et PPV, en contraste avec les résultats plus inégalitaires mis en évidence par Bartlett, Morse, Stanton, et Wallace (2022) dans d’autres contextes de crédit ;
- pénalisation extrême des dossiers « Race Not Available », caractérisés par un DI très faible, une faible sensibilité et une calibration dégradée.

Ce résultat est cohérent avec la littérature sur l’impact des données incomplètes (Buolamwini & Gebru, 2018 ; Barocas & Selbst, 2016). L’IA amplifie les lacunes statistiques et pénalise fortement les profils mal renseignés. Il confirme également l’intuition centrale des travaux de Selbst et collaborateurs sur les systèmes sociotechniques : les propriétés distributives d’un modèle ne peuvent être évaluées sans une analyse fine de la production des données d’entrée. Les inégalités ne résident pas uniquement dans les coefficients ou les prédictions, mais dans la trajectoire complète qui conduit certains profils à être mal ou partiellement observés.

5.2 Discussion par question de recherche

5.2.1 RQ1 : Effets du revenu sous l'ère IA

Les résultats montrent un gradient de revenu très marqué : les zones à faible revenu voient leurs odds d'approbation diminuer d'environ 22 %, même après contrôle des caractéristiques du prêt. L'IA amplifie manifestement des structures socio-économiques historiquement enracinées dans la géographie urbaine du Tri-State, conformément aux mécanismes mis en évidence par Hardy, Logan, et Pustejovsky (2023) dans d'autres marchés de crédit.

Ce renforcement des gradients doit être interprété à la lumière des débats sur la « discrimination par proxy » (Barocas & Selbst, 2016). Les variables de revenu de quartier, de pauvreté ou de chômage, combinées à des informations sur les prix de l'immobilier et la structure du marché local, jouent le rôle de proxies pour des dimensions normativement sensibles (classe sociale, race, stabilité résidentielle). Dans la mesure où les modèles IA sont explicitement optimisés pour la performance prédictive, ils ont tendance à exploiter pleinement ces corrélations, quitte à reproduire des dynamiques de stratification antérieures.

5.2.2 RQ2 : Capacité prédictive et structure des modèles

L'amélioration de l'AUC dans l'ère IA ($\approx 0,61$ à $\approx 0,91$ pour RF/XGBoost) traduit une consolidation structurelle du régime d'octroi. Ce renforcement de la prévisibilité renvoie directement aux travaux en économétrie algorithmique de Mullainathan et Spiess (2017) et aux approches statistiques présentées par (Friedman, Hastie, &

Tibshirani, 2001).

Cette montée en puissance de la prédictivité doit cependant être replacée dans le cadre plus large des débats sur l'IA « fiable » et la gouvernance des systèmes intelligents. Les travaux de Kearns et Roth sur « l'algorithme éthique » insistent sur le fait qu'une amélioration de la précision n'est acceptable que si elle s'accompagne de garanties sur la robustesse, la transparence et la non-discrimination des modèles. De même, la littérature en droit et technologies (par exemple, Citron) rappelle que les systèmes de notation automatisés peuvent renforcer des asymétries de pouvoir si leurs mécanismes internes sont opaques pour les emprunteurs et les régulateurs.

5.2.3 RQ3 : Qualité du screening (approbation vs défaut proxy)

Le screening IA est légèrement plus cohérent : les dossiers approuvés sont 3,6 points moins associés à des signaux de risque. Cependant, l'effet direct de l'indicateur *era_IA* reste modeste ($p \approx 0,059$), indiquant que l'IA améliore la cohérence globale du screening, mais sans révolution complète du processus, en ligne avec les résultats prudents rapportés par Agarwal, Ben-David, et Seru (2015) sur les transformations graduelles des comportements de risque des prêteurs.

D'un point de vue conceptuel, ces résultats suggèrent que l'IA joue un rôle de « raffinement » plutôt que de rupture : la frontière entre bons et mauvais risques devient plus nette, mais l'architecture générale du régime de crédit demeure intacte. La valeur ajoutée se situe dans la réduction des « erreurs grossières » (acceptation de dossiers manifestement risqués, rejet de dossiers très solides), plutôt que dans un changement radical des populations servies. Cette observation est cohérente avec une vision « Ruscellienne » de l'IA comme système d'optimisation de fonctions objectifs définies en

amont : tant que la fonction à optimiser reste centrée sur le risque de défaut à court terme, la structure sociale sous-jacente n'est pas transformée.

5.2.4 RQ4 : Fairness algorithmique

Les résultats indiquent :

- quasi-égalité des performances conditionnelles entre White et Black ;
- pénalisation extrême des dossiers « Race Not Available ».

Ces patterns sont à la fois proches et partiellement distincts des résultats de Bartlett et al. (2022), qui documentent une discrimination persistante dans de nombreux segments du marché du crédit. Ils confirment la centralité des données complètes dans tout système algorithmique, rejoignant les conclusions de Buolamwini et Gebru (2018).

Du point de vue des métriques de fairness, le cas du Tri-State Area illustre également les tensions mises en évidence par la littérature (Kleinberg, Selbst, Kearns) entre différentes définitions de l'équité : égalité des taux d'erreur, calibration inter-groupe, impact disparate, etc. La relative convergence des performances conditionnelles entre White et Black coexiste avec des gradients de revenu renforcés et une pénalisation des données manquantes. Autrement dit, certains critères de fairness s'améliorent, tandis que d'autres dimensions de justice distributive restent problématiques ou se dégradent, ce qui renforce la nécessité d'une approche multi-critères.

5.3 Interprétation régionale : New York, New Jersey, Connecticut

5.3.1 Pourquoi une lecture régionale est essentielle

Le Tri-State Area est un espace socio-économique profondément hétérogène. L’usage des modèles IA — qui reposent largement sur des signaux géographiques — interagit avec cette structure spatiale.

Comme le soulignent Gyourko et Tracy (2014), les marchés immobiliers régionaux ne constituent pas de simples fonds décoratifs : ils structurent profondément la distribution du risque, la composition des emprunteurs et la complétude des données disponibles. En conséquence, lorsque les modèles IA exploitent les caractéristiques des Census tracts, ils reproduisent et systématisent ces réalités territoriales.

L’importance de cette lecture régionale tient également au fait que l’IA opère comme un « amplificateur structurel » (Kearns & Roth, 2020) : elle renforce les régularités statistiques, qu’elles soient équitables ou non. Dans les régions où les écarts socio-économiques et raciaux sont déjà prononcés, les effets redistributifs de l’IA peuvent être démultipliés. À l’inverse, dans des environnements plus homogènes, les mêmes modèles peuvent produire des résultats relativement plus neutres en termes d’inégalités.

5.3.2 New York : un laboratoire extrême des dynamiques IA

New York est l’un des environnements urbains les plus contrastés du pays. Les écarts de revenu entre Manhattan et le Bronx, ou entre Williamsburg et East New York,

dépassent souvent 400 %. Ces disparités se doublent d’une ségrégation résidentielle persistante, documentée de longue date par la littérature sur la ségrégation urbaine (Massey & Denton, 1993).

Dans ce contexte, les modèles IA apprennent des patterns socio-spatiaux extrêmement polarisés. Les variables ACS — revenu médian, taux de pauvreté, composition raciale — deviennent des signaux statistiquement très discriminants. L’analyse du Chapitre 4 montre que les gradients de revenu sont plus marqués en période IA : dans un État où les extrêmes socio-économiques sont si nets, cette amplification est très probable.

Un second élément structurel caractéristique de New York réside dans l’incomplétude des données démographiques. Les dossiers HMDA de certains quartiers urbains denses affichent des taux élevés de « Race Not Available ». Or, comme l’indiquent nos résultats et la littérature sur la qualité des données (Buolamwini & Gebru, 2018), l’incomplétude des données entraîne des pénalités majeures dans les métriques de fairness.

Enfin, les modèles IA interagissent fortement avec les structures spatiales du marché immobilier new-yorkais : les algorithmes reproduisent des réalités micro-locales (îlots de richesse et zones de pauvreté accumulée). L’IA n’y supprime pas les clivages historiques : elle les formalise statistiquement, dans l’esprit des mécanismes décrits par Rothstein (2018). Cette formalisation algorithmique confère toutefois un vernis de neutralité technique à des disparités qui demeurent profondément politiques, ce qui rend d’autant plus cruciale l’intervention des régulateurs locaux.

5.3.3 New Jersey : homogénéité suburbaine et IA plus stable

Le New Jersey se distingue par une homogénéité socio-économique beaucoup plus grande que New York ou Connecticut. L'État est dominé par de vastes zones suburbaines caractérisées par :

- des revenus médians plus homogènes ;
- une plus faible segmentation raciale ;
- un parc immobilier relativement stable et prévisible.

Cette homogénéité réduit l'amplitude des gradients observables dans les données.

Les modèles IA opèrent dans un environnement où :

- la variance intra-tract est plus faible ;
- les données HMDA sont plus complètes ;
- les signaux socio-économiques sont plus continus.

Dans ces conditions, les algorithmes produisent généralement des décisions plus calibrées et moins sujettes aux extrêmes. Les risques de proxy discrimination y sont significativement atténués, car les régularités socio-économiques apprises par les modèles sont plus lisses et moins liées à des phénomènes de ségrégation profonde. New Jersey apparaît donc comme l'environnement où l'IA fonctionne le plus « harmonieusement », confirmant la logique d'interaction entre structure régionale et comportement algorithmique.

Cette relative stabilité ne signifie pas l'absence d'enjeux d'équité, mais plutôt que les problèmes prennent la forme d'ajustements fins (par exemple, traitement des emprunteurs aux revenus atypiques ou des zones en transition) plutôt que de ruptures majeures entre quartiers riches et quartiers pauvres. Les politiques publiques y peuvent donc viser davantage la prévention de dérives futures que la correction d'inégalités massives

déjà constituées.

5.3.4 Connecticut : polarisation socio-économique extrême

Connecticut combine certaines des zones les plus riches des États-Unis (Fairfield County, Greenwich) et certaines parmi les plus pauvres (Bridgeport, Hartford, New Haven). Cette polarisation bimodale crée un terrain idéal pour que les modèles IA amplifient les gradients socio-économiques.

Dans les zones riches, les variables ACS signalent des environnements socio-économiques extrêmement favorables : revenu élevé, faible pauvreté, forte stabilité résidentielle. Dans les zones pauvres, les signaux sont exactement inverses.

Les modèles IA, entraînés sur ces données, apprennent ces contrastes de manière quasi parfaite. Comme le soulignent Fuster et al. (2019), ce type de polarisation renforce la dépendance des modèles aux signaux géographiques. Cette dynamique crée un risque élevé de proxy discrimination. Les zones en difficulté, déjà historiquement mal desservies, peuvent faire l'objet d'une pénalisation statistique accrue. L'IA ne crée pas ces écarts : elle les amplifie.

Du point de vue de la gouvernance régionale, Connecticut apparaît ainsi comme un cas d'école où l'introduction de modèles IA sans correctifs explicites risque d'enraciner encore davantage des géographies de l'exclusion. La combinaison d'une forte dépendance aux variables de quartier et d'un tissu socio-économique polarisé conduit à une cartographie algorithmique du risque qui épouse de près les frontières historiques de la ségrégation.

5.3.5 Synthèse régionale

Les trois États du Tri-State Area illustrent trois contextes structurels différents. Lorsque l'IA interagit avec ces contextes, ses effets redistributifs varient fortement :

- À New York, l'IA amplifie les contrastes extrêmes et pénalise fortement l'incomplétude des données.
- Au New Jersey, l'IA bénéficie d'un environnement homogène et produit des décisions plus calibrées et moins inégalitaires.
- Au Connecticut, la polarisation extrême conduit à une amplification notable des gradients socio-économiques.

En somme, l'IA ne constitue pas un système universel et neutre : elle est profondément conditionnée par la structure socio-économique et territoriale dans laquelle elle opère. Cette conclusion rejoint les analyses sociotechniques qui insistent sur l'inscription locale des technologies : un même modèle peut avoir des effets plus ou moins inévitables selon le degré de ségrégation, la qualité des données et l'intensité de la régulation régionale.

Pour les décideurs publics et les institutions financières, cette lecture régionale implique qu'aucune politique « uniforme » de gouvernance algorithmique ne peut suffire. Les garde-fous, les audits et les correctifs redistributifs doivent être adaptés à la géographie fine des inégalités, en tenant compte des configurations spécifiques de chaque État, voire de chaque aire métropolitaine.

5.4 Analyse méthodologique et comparaison avec la littérature HMDA

Au-delà des résultats substantifs, cette étude se distingue par un ensemble de choix méthodologiques qui la situent dans la continuité, mais aussi en rupture partielle, avec les travaux empiriques précédents fondés sur HMDA.

5.4.1 Positionnement par rapport aux études HMDA existantes

Les travaux de Fuster et al. (2019) et Bartlett et al. (2022) ont marqué une première génération d’analyses de l’impact du machine learning sur les marchés du crédit, en s’appuyant sur des données HMDA enrichies par diverses sources complémentaires. Ces études se concentrent principalement sur la performance relative des prêteurs « Fin-Tech » par rapport aux banques traditionnelles, et sur la mesure de discriminations raciales et géographiques dans les taux d’acceptation et les conditions de prêt.

La présente dissertation partage avec ces travaux un socle commun : l’utilisation de HMDA comme base d’observation systématique, une attention forte aux variables de quartier et une mobilisation de modèles prédictifs avancés. Cependant, elle s’en distingue sur plusieurs dimensions :

- le focus sur le Tri-State Area plutôt que sur l’ensemble du territoire national, ce qui permet une analyse plus fine des dynamiques intra-régionales ;
- la construction explicite d’une « ère IA » (2018–2023) opposée à une période pré-algorithmique, là où de nombreuses études considèrent implicitement l’IA comme déjà omniprésente ;

- l’intégration systématique de métriques de fairness (Equal Opportunity, Disparate Impact, calibration) dans l’évaluation des modèles, au-delà des seuls indicateurs de performance globale.

5.4.2 Forces et limites des modèles employés

Sur le plan technique, la combinaison de modèles logistiques, de forêts aléatoires et de XGBoost permet de couvrir un continuum allant des approches paramétriques interprétables aux méthodes non linéaires plus performantes. Ce choix est cohérent avec la littérature en économétrie appliquée (Mullainathan & Spiess, 2017 ; Friedman et al., 2001), qui recommande de comparer plusieurs familles de modèles pour distinguer ce qui relève de la structure des données de ce qui relève du choix d’algorithme.

Par rapport à d’autres études HMDA, la présente recherche apporte trois éléments méthodologiques importants :

- une évaluation systématique de la robustesse des résultats à des spécifications alternatives (par exemple, exclusion des observations extrêmes, variations dans la définition des groupes de revenu) ;
- un traitement explicite des données manquantes comme objet d’analyse à part entière, plutôt que comme simple nuisance statistique ;
- une articulation étroite entre métriques techniques et questions normatives, de manière à éviter l’illusion d’une « neutralité » des indicateurs.

Ces choix renforcent la crédibilité interne des résultats, tout en soulignant certaines limites incontournables des analyses fondées sur HMDA : absence de variables de score interne, manque d’information sur les conditions exactes des prêts, et impossibilité d’observer directement le défaut.

5.4.3 Apports spécifiques de la fusion HMDA+ACS

L’un des apports méthodologiques centraux de cette dissertation réside dans la fusion systématique de HMDA avec des données ACS au niveau des Census tracts. Cette opération permet d’ancrer les décisions de crédit dans un environnement socio-économique riche, incluant le revenu médian, la pauvreté, le chômage et la composition raciale.

Comparée à d’autres travaux qui utilisent des proxies macro (par État ou par comté), cette granularité accrue offre deux avantages majeurs :

- une meilleure capacité à détecter les gradients intra-métropolitains, souvent invisibles dans des agrégations trop larges ;
- une compréhension plus fine des mécanismes de proxy discrimination, en montrant comment des variables de quartier peuvent servir de substituts à des caractéristiques sensibles.

Ce choix méthodologique va cependant de pair avec un coût : les résultats sont fortement contextualisés et n’ont pas vocation à être généralisés « tels quels » à d’autres régions sans précaution. La portée de l’étude est donc à la fois plus profonde (sur le Tri-State Area) et plus délimitée (en termes de généralisation externe) que celle des analyses nationales.

5.5 Implications pour la théorie

Les résultats contribuent à trois débats majeurs dans la littérature, en croisant les apports de l’économie, de l’IA, du droit et des études sociotechniques.

5.5.1 L’IA comme mécanisme de standardisation

Le passage d’un régime de décision humaine — fortement hétérogène — à un régime algorithmique plus systématisé confirme les analyses de Kleinberg et al. (2018) sur la réduction du bruit décisionnel. Dans la mesure où les modèles entraînés sur la période IA produisent des décisions plus prévisibles et plus homogènes entre institutions, ils renforcent l’idée que l’IA agit comme un dispositif de normalisation des pratiques.

Dans une perspective plus large, cette dynamique fait écho aux réflexions contemporaines sur la « gouvernance par les nombres », où la standardisation des métriques et des procédures devient un instrument central de gouvernement. Loin d’être purement technique, cette standardisation redéfinit les marges de négociation entre emprunteurs et prêteurs, en encadrant plus strictement les possibilités de dérogation ou de prise en compte de situations atypiques.

5.5.2 La discrimination par proxy

Les gradients socio-économiques renforcés, observés dans plusieurs États, corroborent la thèse de Barocas et Selbst (2016) selon laquelle les algorithmes peuvent reconstruire des inégalités raciales à partir de variables en apparence neutres. Les résultats montrent que les modèles IA apprennent des schémas fortement corrélés aux structures de ségrégation résidentielle, même en l’absence de variables raciales explicites.

Les travaux de Kearns et Roth insistent sur le fait que cette reconstruction est une conséquence presque inévitable dès lors que les corrélations entre variables sensibles et variables de quartier sont fortes. De ce point de vue, la présente dissertation

offre une démonstration empirique détaillée de ces mécanismes, en révélant comment les gradients de revenu et les variables ACS deviennent des vecteurs de traduction institutionnelle des inégalités raciales et de classe.

5.5.3 Le rôle central des données manquantes

Les pénalités extrêmes associées à la catégorie « Race Not Available » démontrent que la qualité de la fairness dépend directement de la qualité des données. Ce résultat est en ligne avec les analyses récentes de Selbst, Boyd, Friedler, Venkatasubramanian, et Vertesi (2019) sur le rôle structurant des lacunes informationnelles.

En pratique, cela signifie que les débats théoriques sur la justice algorithmique ne peuvent se limiter aux propriétés formelles des modèles (contraintes de fairness, régularisation, etc.). Ils doivent aussi intégrer une réflexion sur la chaîne complète de production des données : qui est incité à fournir quelles informations, dans quelles conditions, et avec quelles garanties de protection contre les usages secondaires abusifs. Sur ce point, les travaux juridiques sur la protection des données et les droits civiques (par exemple, ceux de Citron) offrent un cadre analytique complémentaire pour penser les obligations de transparence, de consentement et de recours.

5.6 Implications pour les politiques publiques

Cinq implications majeures émergent des résultats :

5.6.1 Audits algorithmiques obligatoires

L’importance des écarts régionaux justifie des audits plus sophistiqués : calibration par groupe, Disparate Impact géographiquement contextualisé, et étude des fausses négations, dans la continuité des recommandations méthodologiques issues de la littérature sur l’IA responsable (Kearns & Roth, 2020).

Concrètement, les régulateurs devraient exiger des institutions financières qu’elles produisent, à intervalles réguliers, des rapports détaillés de performance et de fairness ventilés par race, revenu, et territoire. Ces audits devraient inclure des analyses de sensibilité (effet d’une modification des variables ACS), ainsi que des simulations de scénarios alternatifs (par exemple, suppression de certaines variables de quartier) afin d’évaluer la dépendance des modèles à des signaux potentiellement problématiques.

5.6.2 Amélioration de la complétude des données

La CFPB (Consumer Financial Protection Bureau, 2019) recommande des mécanismes robustes de collecte et de validation des données. Les résultats montrent que cette recommandation est non seulement pertinente, mais indispensable pour éviter une pénalisation systématique des profils dont les caractéristiques démographiques sont incomplètes.

Dans la pratique, cela suppose :

- de renforcer les obligations de déclaration pour les prêteurs, en sanctionnant les taux anormalement élevés de « Race Not Available » ;
- de simplifier et sécuriser les procédures de collecte des données sensibles, afin de réduire la réticence des emprunteurs à se déclarer ;

- d’expérimenter des mécanismes de « complétion assistée » des données, sous contrôle strict des régulateurs, pour éviter que l’incomplétude ne se traduise mécaniquement par une exclusion.

5.6.3 Transparence des modèles IA

Les modèles non linéaires nécessitent :

- une documentation rigoureuse ;
- des explications via SHAP/LIME ;
- un contrôle institutionnel renforcé.

Au-delà de la simple documentation interne, il s’agit de mettre en place des obligations de transparence « graduée » :

- vis-à-vis des régulateurs, qui doivent pouvoir auditer le code, les données d’entraînement et les métriques ;
- vis-à-vis des emprunteurs, qui doivent disposer d’explications compréhensibles de leur décision individuelle (refus ou conditions défavorables), sans pour autant révéler l’intégralité des modèles propriétaires.

Ces exigences s’inscrivent dans un mouvement plus large de « due process algorithmique », qui vise à aligner l’usage de l’IA sur les garanties procédurales offertes dans d’autres domaines du droit.

5.6.4 Politiques ciblées pour zones à faible revenu

Les gradients observés justifient des politiques correctrices ciblées (subventions locales, encadrement géographique, programmes de soutien aux primo-accédants) dans les territoires les plus exposés au risque de proxy discrimination.

Des dispositifs spécifiques pourraient être envisagés :

- bonifications d'intérêt ou garanties publiques pour les emprunteurs situés dans des quartiers cumulant pauvreté élevée et exclusion historique du crédit ;
- programmes pilotes combinant conseil financier, accompagnement juridique et suivi post-octroi, afin de réduire les risques de défaut tout en élargissant l'accès ;
- coordination renforcée entre autorités locales du logement et régulateurs financiers, de manière à aligner les objectifs de revitalisation urbaine et de fairness algorithmique.

5.6.5 Renforcement des capacités institutionnelles et participation des communautés

Enfin, les résultats soulignent la nécessité de renforcer les capacités d'analyse des régulateurs et des organisations de la société civile, afin qu'ils puissent dialoguer d'égal à égal avec les institutions financières sur les enjeux techniques de l'IA.

Deux axes apparaissent prioritaires :

- le développement de compétences internes en science des données et en audit algorithmique au sein des agences publiques ;
- la mise en place de mécanismes de participation des communautés affectées (par exemple, panels d'utilisateurs, consultations locales) pour discuter des effets territoriaux des modèles et co-construire des réponses adaptées.

En associant plus étroitement les populations concernées, les politiques de gouvernance de l'IA peuvent gagner en légitimité démocratique et en pertinence empirique, en particulier dans des régions comme le Tri-State Area où les contrastes socio-économiques sont extrêmes.

5.7 Contributions spécifiques de l'étude

Cette dissertation apporte quatre contributions majeures, qui combinent innovations empiriques, méthodologiques et théoriques :

- **Une analyse intégrée de performance IA et fairness dans le Tri-State.**

L'étude propose, pour la première fois à cette échelle, une évaluation conjointe de la performance prédictive et des propriétés de fairness des modèles IA d'octroi de crédit dans le Tri-State Area. En articulant métriques de risque (AUC, pseudo- R^2 , courbes ROC) et métriques d'équité (Equal Opportunity, Disparate Impact, calibration), elle montre que l'amélioration de la prédictivité n'est pas mécaniquement synonyme de progrès en matière de justice distributive, mais qu'elle peut coexister avec des gradients socio-économiques renforcés.

- **Une fusion HMDA+ACS inédite pour capter les dynamiques socio-économiques fines.**

En fusionnant systématiquement les données HMDA avec des indicateurs ACS au niveau des Census tracts, la recherche parvient à relier les décisions de crédit aux conditions socio-économiques locales de manière beaucoup plus fine que dans les études fondées sur des agrégations au niveau de l'État ou du comté. Cette approche permet de mettre en évidence des gradients intra-métropolitains et de documenter empiriquement le rôle des variables de quartier comme vecteurs de proxy discrimination.

- **Une démonstration empirique du rôle central de l'incomplétude des données.**

En isolant la catégorie « Race Not Available » et en mesurant systématiquement ses performances en termes de fairness, l'étude montre que les données manquantes ne sont pas un simple artefact technique, mais un déterminant substantiel de la distribution des opportunités de crédit. Cette démonstration

éclaire sous un jour nouveau les débats théoriques sur la production des données et leur rôle dans les systèmes sociotechniques, en montrant que l'incomplétude peut être elle-même une source d'injustice.

- **Une documentation rigoureuse de la transition institutionnelle vers l'IA.** En structurant l'analyse autour de deux régimes temporels (pré-IA vs ère IA) et en suivant l'évolution des gradients de revenu, des déterminants structurels et des métriques de fairness, la dissertation offre une chronologie précise de la transition algorithmique dans le Tri-State Area. Elle met en lumière la nature graduelle et partiellement cumulative de cette transition : l'IA ne remplace pas du jour au lendemain les régimes antérieurs, mais s'imbrique dans des institutions et des structures de marché déjà inégalitaires.

Collectivement, ces contributions positionnent la dissertation à l'interface de plusieurs champs — économie du crédit, science des données, études sur l'IA responsable et sociologie des inégalités — et en font un point de référence pour les futures recherches sur l'impact territorial de l'IA dans les services financiers.

5.8 Limites

Comme toute étude empirique utilisant HMDA, plusieurs limites subsistent :

- absence de variables de risque essentielles ;
- impossibilité d'observer les modèles propriétaires ;
- proxy du défaut imparfaite ;
- hétérogénéités institutionnelles non observées.

5.9 Pistes de recherche futures

Les futures recherches pourront explorer :

- l'accès aux modèles internes ;
- des analyses nationales multi-États ;
- des modèles hiérarchiques multi-niveaux ;
- les effets structurels des données manquantes ;
- les architectures IA modernes (Transformers, GBDT avancés).

5.10 Conclusion

Le passage à l'IA transforme profondément l'octroi du crédit dans le Tri-State Area. L'IA améliore la performance prédictive, stabilise les déterminants structurels et réduit certaines disparités conditionnelles. Cependant, elle amplifie les effets du revenu, redistribue les inégalités géographiques et pénalise les emprunteurs aux données incomplètes. L'IA ne supprime pas les inégalités : elle les reconfigure en fonction de la structure territoriale et de la qualité des données.

Une régulation attentive, une amélioration de la qualité des données, et une vigilance accrue quant aux mécanismes de *proxy discrimination* sont nécessaires pour garantir un accès équitable au crédit dans un environnement algorithmique.

Références

- Agarwal, S., Ben-David, I., & Seru, A. (2015). The effects of credit competition on lenders' risk taking. *Journal of Financial Economics*.
- Barocas, S., & Selbst, A. (2016). Big data's disparate impact. *California Law Review*, *104*(3), 671–732.
- Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the fintech era. *Journal of Finance*.
- Bhutta, N., Blair, J., & Dettling, L. (2021). The decline of access to credit. *Federal Reserve Board Working Paper*.
- Buolamwini, J., & Gebru, T. (2018). Gender shades : Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the conference on fairness, accountability, and transparency (fat*)*.
- Consumer Financial Protection Bureau. (2019). *Mortgage market trends* (Rapport technique). CFPB.
- Cowgill, B., & Tucker, C. (2020). Economics of artificial intelligence. *Journal of Economic Perspectives*.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*. Springer.
- Fuster, A., Goldsmith-Pinkham, P., Ramcharan, R., & Vickery, J. (2019). Predictably

- unequal? the effects of machine learning on credit markets. *Review of Financial Studies*.
- Gyourko, J., & Tracy, J. (2014). A regional analysis of mortgage lending. *Regional Science and Urban Economics*.
- Hardy, B., Logan, T., & Pustejovsky, J. (2023). Ai, segregation, and financial inequality. *Economic Policy Review*.
- Kearns, M., & Roth, A. (2020). *The ethical algorithm*. Oxford University Press.
- Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. (2018). Human decisions and machine predictions. *Quarterly Journal of Economics*.
- Massey, D. S., & Denton, N. A. (1993). *American apartheid : Segregation and the making of the underclass*. Harvard University Press.
- Mullainathan, S., & Spiess, J. (2017). Machine learning : An applied econometrics approach. *Journal of Economic Perspectives*, 31(2), 87–106.
- Rothstein, R. (2018). *The color of law : A forgotten history of how our government segregated america*. Liveright Publishing.
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*.
- Squires, G. (2017). *The fight for fair housing*. Routledge.