ChatGPT

Adapting LLM Ethical-Decision Methodology to Finance and Risk Management

Introduction: The 2025 study *"Socio-Demographic Modifiers Shape LLMs' Ethical Decisions"* demonstrated how adding irrelevant socio-demographic details can sway large language models' responses to ethical dilemmas [1] [2]. The authors generated 100 synthetic clinical vignettes (yes/no medical ethics dilemmas) and showed that when details like a subject's income or group identity were included (which logically shouldn't change the correct decision), LLMs' choices often shifted significantly [2]. This revealed that **LLMs may not uphold consistent principles** – e.g. high-income cues prompted more utilitarian judgments, while marginalized-group cues raised autonomy-focused answers [3]. Such findings raise concerns about *algorithmic alignment*: if trivial socio-demographic modifiers bias an AI's decisions, its fairness and reliability are in question [4].

LLMs are increasingly being considered for financial services applications – from credit underwriting to investment advice – where **ethical and fairness implications are critical**. Already, researchers have found that LLMs can reproduce historical biases in finance: for example, a 2024 experiment showed GPT-based models recommended denying more loans or charging higher interest to Black mortgage applicants versus identical white applicants [5]. Given that millions of users are beginning to trust tools like ChatGPT for financial guidance [6], it's vital to examine whether LLMs behave fairly and responsibly in **risk-critical financial decisions**.

Goal: Below we propose **10 novel research paper ideas** that replicate or adapt the methodology of Sorin *et al.* (2025) to *financial services and risk management.* Each idea uses **synthetic scenario generation** (analogous to vignettes) to probe LLMs' decisions in ethically sensitive financial contexts. We focus on LLMs released in 2024 or later (the latest-generation models) and design scenarios where a decision or outcome should **not logically change** with certain socio-demographic or contextual modifiers – yet the LLM's behavior might change, revealing bias or misalignment. We detail each study's design, then evaluate its **practical feasibility**, **impact/publishability**, and **effort required** on 1–5 scales. Finally, we aggregate these scores, rank the ideas, and present in-depth plans for the top 3, including (a) existing work, (b) relevant regulations, (c) data sources, and (d) a step-by-step case study plan. *Throughout, we cite recent literature and regulatory guidance, and suggest high-impact venues in finance, risk, banking, AI, and actuarial science.*

Proposed Research Ideas (with Evaluation Scores)

1. **Bias in LLM-Based Loan Approval Decisions** – *Testing Fair Lending with Demographic Modifiers.* **Design:** Create a set of synthetic loan applicant profiles (income, credit score, loan purpose, etc.) and ask an LLM (e.g. GPT-4, Claude) to decide "approve or deny" or recommend an interest rate. Introduce a **socio-demographic modifier** that should not affect creditworthiness – e.g. the applicant's race or gender (implicitly via name or explicitly stated) – and see if the LLM's decision changes. This mirrors real fair-lending tests: profiles are identical except for a protected characteristic. We would measure approval rates or suggested terms for each group. **Practicality:** 5/5 (LLMs can easily evaluate structured profile descriptions; public data like HMDA can inform realistic scenarios [7]). **Impact:** 5/5 (High – addresses *algorithmic discrimination* in lending, a heavily regulated domain with active research [5] [8]). **Effort:** 4/5 (Moderate –

requires generating many profiles and multiple LLM queries; analysis of statistical bias and significance). **Combined Score:** *14/15*.

2. **Discrimination in Insurance Underwriting Recommendations** – *LLM decisions on who to insure and at what price.* **Design:** Develop synthetic insurance applicant scenarios (e.g. for auto or life insurance) with identical risk factors except for one modifier like the applicant's occupation, ethnicity, or ZIP code (proxy for neighborhood). Prompt the LLM as if it's an underwriting assistant: should the policy be approved? Any surcharge? The modifier (say, a "foreign-sounding" name or a low-income area address) is unrelated to actual risk in the scenario. Track changes in the LLM's decisions or recommended premiums. **Practicality:** 4/5 (LLMs can process scenario descriptions; may need insurance domain prompting. Data on risk factors is available, though real insurance datasets with protected attributes are scarce, so synthetic data is used). **Impact:** 4/5 (Potentially high – fairness in insurance pricing is a major concern for regulators and actuaries [9]. Demonstrating bias would be publishable in actuarial and risk journals). **Effort:** 4/5 (Moderate – needs many scenario variations and careful control of risk-relevant vs irrelevant details). **Combined Score:** *12/15*.

3. **LLM Fairness in Investment Advice** – *Do AI advisors give different advice based on client profile modifiers?* **Design:** Simulate an investment advisory session. Create a set of investor profiles (e.g. age, assets, risk tolerance, goals) and ask an LLM (fine-tuned for finance or general GPT) to recommend a portfolio or financial plan. The key test: for two profiles identical in financials and stated risk preference, does adding a modifier like the client's gender, family status ("single father" vs "single mother"), or ethnicity alter the advice? For example, prior work suggests human advisors sometimes show gender bias in recommendations; we test if the LLM exhibits or avoids this. We would analyze differences in asset allocation (stocks vs bonds), product suggestions, or tone (degree of caution) between the modified and unmodified profiles. **Practicality:** 5/5 (High – easy to generate profile descriptions and questions; multiple LLMs can be queried. Some studies have already done similar setups with dozens of profiles [10] [11]). **Impact:** 4/5 (High academic and industry interest – robo-advice and FinTech are booming, and ensuring AI advisers are fair and suitable is crucial [12] [11]. Could be submitted to finance or AI ethics venues). **Effort:** 3/5 (Manageable – fewer regulatory constraints on data; evaluation involves comparing advice text, possibly using topic analysis or financial metrics). **Combined Score:** *12/15*.

4. **Fraud Detection and Stereotypes in LLM Decisions** – *Testing if non-risky modifiers trigger fraud flags.* **Design:** Create scenarios describing bank transactions or account activities borderline for fraud detection (e.g. slightly unusual but plausible behaviors). Ask the LLM (acting as a fraud analyst) if it would flag the activity as suspicious. Then vary an **irrelevant detail**: e.g. the customer's name/country ("John Smith" vs "Mohammed Azizi") or profile (local vs international student) while keeping the transaction details identical. Measure if the LLM flags significantly more often for certain groups. This examines whether the LLM has learned cultural stereotypes equating certain nationalities or names with fraud risk. **Practicality:** 4/5 (Requires careful prompt engineering to get the model to act as a fraud classifier; scenario generation is straightforward. Bias in fraud models has been noted – e.g. names from some cultures triggering more suspicion [13]). **Impact:** 3/5 (Moderate – fraud detection is critical for banks, and proving bias here would be important, but this area is less publicly scrutinized than lending/credit bias). **Effort:** 3/5 (Moderate – need to cover various fraud scenario types; human evaluation of model justifications might be needed in addition to yes/no flag rate). **Combined Score:** *10/15*.

5. **AML/KYC Profile Bias in LLM Risk Assessments** – *LLMs in compliance and "de-risking."* **Design:** Banks use "Know Your Customer" (KYC) processes and anti-money-laundering (AML) risk scores to decide if a client needs extra scrutiny. We simulate this by feeding an LLM profiles of hypothetical new customers with similar financial backgrounds but with one differing attribute, such as nationality (e.g. a customer from a high-risk country vs a low-risk country with otherwise identical behavior, or simply a traditionally marginalized community). The LLM must decide if the profile is "high risk" or if enhanced due diligence is required. Logically, if the scenarios are constructed so that the modifier does not actually change the risk (e.g. both customers have clean records and similar transactions), a fair model should give the same assessment. We check if the LLM consistently rates one profile type as riskier (which would mimic unwarranted "de-risking" practices that regulators warn against). **Practicality:** 4/5 (Feasible – LLMs can parse profile descriptions and output risk levels. Synthetic data is needed as real AML cases with labels are confidential; however, one can base scenarios on public typologies). **Impact:** 4/5 (High in industry – financial compliance officers are concerned about **bias vs effectiveness** in AML. Unfairly flagging certain groups leads to exclusion from banking [14] [15]. Academic literature is sparser here, so this study breaks new ground in AI fairness for AML). **Effort:** 3/5 (Moderate – scenario writing may need domain expertise; evaluation requires statistical tests of difference in risk ratings across many runs). **Combined Score:** *11/15*.

6. **Fairness in Small Business Credit Decisions** – *Do LLMs show bias in SME lending akin to consumer lending?* **Design:** Similar to Idea #1 but focused on **small business loan applications**. We generate profiles of small firms seeking credit (financial statements, business plans, credit history) and vary a modifier like the business owner's demographics or the business location (e.g. minority-owned vs not, or a business in a predominantly minority neighborhood vs not). The LLM is asked if the loan should be approved or to assign a risk rating. We observe any disparities. This extends fair-lending tests to the commercial realm – timely since regulators (e.g. CFPB in the US) only recently mandated data collection on small business loans to monitor bias. **Practicality:** 4/5 (Scenario generation is straightforward; might integrate real statistics from SME datasets. However, LLMs might need careful prompting to interpret financial ratios correctly). **Impact:** 5/5 (High – bias in SME lending affects economic opportunity; and with new regulations (like Dodd-Frank 1071 in the US) requiring fair lending oversight for businesses, this research is highly publishable in finance or policy journals). **Effort:** 4/5 (Moderate/High – profiles are more complex than consumer cases; ensuring the LLM correctly evaluates business financials without irrelevant cues is challenging. Multiple LLMs should be tested for robustness). **Combined Score:** *13/15*.

7. **Regulatory Compliance Decisions and Context Bias** – *Are LLMs' judgments of compliance violations swayed by who is involved?* **Design:** Construct scenarios of potential regulatory or legal violations in finance – e.g. an investment banker engages in a borderline insider trading situation. Ask the LLM (as a compliance officer or regulator) what action should be taken (no action, internal discipline, report to regulator, etc.). Then alter context details that *shouldn't* change the correct response: e.g. say the person involved is a junior employee versus a high-ranking executive, or the institution is a small local firm versus a large multinational, while the misconduct is identical. In principle, enforcement should be based on facts, not the status of the actor – we test if the LLM is more lenient or strict depending on these modifiers. **Practicality:** 3/5 (LLMs can analyze such scenarios but their knowledge of regulatory norms might cause them to justify different treatment logically – need to phrase scenarios so that the *ethically correct* outcome is clearly the same for both versions). **Impact:** 3/5 (Medium – highlights consistency in AI-driven compliance. While not as public-facing as lending, fairness in enforcement is important to rule of law. This could interest risk management conferences or regulatory tech forums). **Effort:** 3/5 (Moderate – requires crafting nuanced scenarios and

possibly qualitative analysis of the rationale LLM gives for different actors). **Combined Score:** *9/15*.

8. **Operational Risk and Internal Bias** – *LLM recommendations in handling employee misconduct.* **Design:** Internal risk management often involves ethical decisions (e.g. how to discipline employees for violations). We simulate an operational risk vignette: say an employee committed a policy violation (e.g. an accounting error, or misuse of funds). The LLM (acting as a risk manager) is asked what should be done – coaching, warning, termination, etc. We then tweak a detail, like the employee's profile: e.g. "a long-tenured top manager" vs "a new junior staffer" or change their gender, while the offense is identical. We assess if the LLM's recommended action shifts (e.g. being harsher on the junior employee or influenced by gender stereotypes). Ethically, similar misconduct should result in similar consequences. **Practicality:** 4/5 (Easy to implement scenarios; LLMs can produce detailed recommendations. However, ensuring the modifier is truly irrelevant may be tricky – in real life, tenure might affect consequences, but we'll frame that it shouldn't). **Impact:** 3/5 (Moderate – this explores bias in internal decisions. It's a novel angle combining AI ethics with HR/operational risk. Likely of interest to specialized management or risk journals, though less urgent than customer-facing fairness issues). **Effort:** 3/5 (Moderate – need a variety of misconduct cases and attributes; analysis partly qualitative). **Combined Score:** *10/15*.

9. **Customer Service Fairness in Banking** – *LLM-generated resolutions for complaints: VIP vs average customer.* **Design:** Many banks plan to use AI chatbots or decision-support systems for customer service. We test if an LLM's proposed resolution to a customer complaint differs based on the customer's status. Create scenario: a customer complains about an incorrect fee charged. In one version, the customer is described as a high-net-worth, long-time client; in another, as a regular client with no special status – all other details (error severity, etc.) are the same. The LLM is asked how the bank should compensate or respond. We examine if it suggests more generous remedies for the VIP (waiving fees, offering perks) versus the ordinary customer. Ethically, while businesses often do give VIPs special treatment, an aligned AI should ideally recommend fair treatment for all in identical situations (unless instructed about business tiered policies). **Practicality:** 5/5 (Very easy to simulate chat-based scenarios; LLMs excel at generating empathetic responses). **Impact:** 3/5 (Moderate – highlights consistency and fairness in customer-facing AI. This might not be a core regulatory issue, but it touches on **ethical customer service** and could be a compelling case study in banking forums for customer experience). **Effort:** 2/5 (Low – a small number of scenarios and straightforward outputs; minimal statistical analysis, more direct comparison). **Combined Score:** *10/15*.

10. **Ethical Trade-off: Fairness vs Profit in AI Decisions** – *Will an LLM prioritize fairness or business benefit, and does context change it?* **Design:** This idea introduces a direct ethical dilemma (akin to the original paper's conflicting principles). For instance: *"A bank has an AI credit model that is very accurate but slightly disadvantages minority applicants. The bank can either deploy it (improving overall profit by reducing defaults) or tweak it to be fairer at some cost to accuracy. What should they do?"* This pits a fairness principle against profit/utilitarian outcome. We then vary context: e.g. Version A: "Bank in a country with strict anti-discrimination laws" vs Version B: "Bank in a country with no such laws" – or vary the stakeholder: "You are a bank CEO" vs "You are a regulator." The **logical ethical choice** (from a universal standpoint) might be to favor fairness or at least balance it, regardless of legal environment; we see if the LLM's answer shifts its ethical calculus based on the context. This tests whether LLMs' moral decisions are anchored or if they pander to contextual cues (like legal vs illegal environment). **Practicality:** 4/5 (Scenarios can be written easily. LLMs will certainly give an answer; the challenge is interpreting shifts in rationale). **Impact:** 4/5 (High conceptually – it probes AI alignment on ethical principles in finance. Could be

published in AI ethics or finance policy journals, contributing to debates on AI governance). **Effort:** 4/5 (Requires qualitative coding of LLM justifications – e.g. analyzing if it references laws in one case and not the other. Possibly needs multiple models or prompting strategies for robust results). **Combined Score:** *12/15*.

Table 1. Summary of Proposed Ideas with Evaluations and Ranking

Idea (Brief Description)	Practicality	Impact	Effort	Combined Score	Rank
1. Bias in LLM-Based Loan Approval (Fair Lending)	5/5	5/5	4/5	**14**	**1**
6. Fairness in Small Business Credit Decisions	4/5	5/5	4/5	**13**	**2**
2. Discrimination in Insurance Underwriting	4/5	4/5	4/5	**12**	*3 (tied)*
3. Fairness in Investment Advice	5/5	4/5	3/5	**12**	*3 (tied)*
10. Fairness vs Profit Ethical Trade-off	4/5	4/5	4/5	**12**	*3 (tied)*
5. AML/KYC Profile Bias	4/5	4/5	3/5	**11**	6
4. Fraud Detection Stereotypes	4/5	3/5	3/5	**10**	7 (tie)
8. Operational Risk (Employee Misconduct Bias)	4/5	3/5	3/5	**10**	7 (tie)
9. Customer Service Fairness	5/5	3/5	2/5	**10**	7 (tie)
7. Regulatory Compliance Enforcement Bias	3/5	3/5	3/5	**9**	10

(Combined score is the sum of the three criteria ratings. Higher score suggests greater overall feasibility and impact. Effort is rated positively here (a higher number indicates more required effort, which is factored into the sum); however, in assessing ranks we also qualitatively favored ideas with high impact and feasibility despite effort.)

After considering the combined scores and the qualitative significance, we select **the top 3 ideas** for detailed exploration. These are: **Idea 1 (LLM Fair Lending Bias)**, **Idea 6 (Small Business Credit Fairness)**, and **Idea 2 (Insurance Underwriting Bias)** – ranked 1st, 2nd, and tied-3rd respectively. *(Idea 3, on investment advice, tied in score; however, we prioritize Idea 2 on insurance to ensure coverage of a broad range of financial services.)* Below, for each of these three, we discuss: (a) existing academic/industry research on the topic, (b) key regulations or legal standards applicable, (c) available or synthesizable data for the study, and (d) a step-by-step plan to execute the case study. We also suggest relevant high-impact journals or conferences for publication.

1. Bias in LLM-Based Loan Approval Decisions (Fair Lending) – Detailed Plan

(a) Existing Work on AI & Fair Lending: There is a rich literature on algorithmic bias in credit decisions. Recent research confirms that even advanced LLMs can reproduce racial disparities in lending. For example, Bowen *et al.* (2024) conducted a mortgage underwriting experiment with GPT models: using real Home Mortgage Disclosure Act data, they found **Black applicants were denied loans more often and offered higher interest rates than identically qualified white applicants** [5]. In fact, their study quantified that a Black borrower might need a credit score ~120 points higher to be treated equivalently by certain LLMs [16]. The bias was present across multiple models (GPT-3.5, GPT-4, Claude, etc.), though magnitude varied [17]. Notably, GPT-4 (2023) showed almost no disparity, while GPT-3.5 Turbo exhibited the most [18], suggesting newer models may be improving. Simply instructing the LLM *"to ignore race"* in decisions virtually eliminated the bias, highlighting a possible mitigation [19]. Aside from this LLM-specific work, numerous studies in machine learning and finance have shown that credit algorithms can inherit biases from historical data [8]. For instance, earlier analyses of fintech lending found that algorithmic mortgage approvals, while 40% less discriminatory than human lenders, still charged minority borrowers higher interest on average [20]. In industry, high-profile cases like the 2019 Apple Card incident (where an algorithm gave women lower credit lines than men with similar profiles) have underscored the issue [21]. These findings collectively motivate our study: we have evidence that **LLMs-as-lending agents can produce unfair outcomes**, raising flags for regulators and providing a baseline methodology to replicate and extend.

(b) Key Regulations and Legal Standards: Fair lending is heavily regulated. In the US, the **Equal Credit Opportunity Act (ECOA)** and **Fair Housing Act** prohibit credit discrimination based on protected characteristics (race, sex, etc.). Lenders are legally required to ensure factors like race *never* influence decisions – even inadvertently (disparate impact liability applies). The Consumer Financial Protection Bureau (CFPB) and other agencies have made it clear that using AI or black-box models doesn't exempt lenders from fair lending laws. (Indeed, a 2025 settlement in Massachusetts enforced ECOA against a lender's AI-driven model that was found discriminatory [22] [23].) Internationally, similar principles apply: e.g. the UK *Equality Act 2010* covers credit services, and EU consumer credit directives emphasize non-discrimination. Moreover, the emerging **EU AI Act** will classify creditworthiness assessment algorithms as "high-risk" AI, subjecting them to strict fairness and transparency requirements [24] [25]. Regulators increasingly expect that firms using AI in lending **conduct bias testing and document steps to mitigate bias** [9]. For example, New York's DFS proposed guidelines requiring insurers and lenders to **test AI models for unfair discrimination** and affirmatively avoid using prohibited variables [9]. All these regulations set the context: if our study finds that LLMs respond differently to loan applicants due to irrelevant socio-demographic prompts, it directly implicates compliance with fair lending laws and would be of keen interest to policymakers.

(c) Data Sources for the Study: We can leverage both **public datasets and synthetic data generation**. A prime resource is the **Home Mortgage Disclosure Act (HMDA) dataset**, which contains real loan application records (including applicant race, gender, income, credit metrics, and outcomes) for millions of mortgages. Indeed, Bowen *et al.* used a sample of 1,000 HMDA records to construct their experimental inputs [7]. We could similarly draw sample profiles from HMDA (or other credit datasets like the LendingClub loan data or FICO credit score sample data) and use those as a basis for scenario generation. HMDA provides ground truth of who was approved or denied in reality, which we can compare against LLM recommendations. However, to isolate LLM bias, we will create **paired profiles** that are identical except for a demographic marker. For example, take a profile from HMDA of a successful applicant; then create a textual prompt for the LLM with that applicant's financials, first with a typically Caucasian male name, then with a typically African-American female name. We ensure the two names have similar credit-relevant background (for realism, we might pick real names from the dataset). We will also use synthetic generation to broaden scenarios beyond mortgages – e.g., personal loan or credit card applications – using statistically realistic ranges for income, debt, etc. Open-source datasets like the *German Credit Data* or *UCI Adult* dataset (modified to a lending context) can also supply baseline profiles for testing fairness, though they are smaller. **In summary**, data is plentiful: HMDA (public, millions of records) for mortgages [7], and other credit datasets or statistical reports for different loan types, all of which can be used to script scenario prompts for the LLM. This ensures our experiments are grounded in real-world distributions.

(d) Step-by-Step Case Study Plan:

1. **Define Scope and Models:** Determine which 2024+ LLMs to test – e.g. OpenAI GPT-4 (2024 version), GPT-3.5 latest, Anthropic Claude 2, and perhaps a finance-specific model like BloombergGPT. These will be our "loan officers." Formulate the decision output we want (approve/deny, or interest rate suggestion).

2. **Design Base Loan Scenarios:** Using HMDA data, select a diverse set of ~50 base loan applications. Ensure a mix of profiles (vary income, credit score, loan purpose) that are on the borderline of approve/deny (to more easily allow flips). For each profile, write a prompt like: *"Applicant seeks a $250,000 mortgage. Credit score 680, income $70k, debt-to-income 35%, first-time homebuyer. [Demographic details]. Should we approve the loan? Why?"* Initially omit demographic details.

3. **Identify Socio-Demographic Modifiers:** Choose a set of modifiers to test (race, gender, age, etc. – focusing on legally protected classes). We'll start with race and gender since those are classic protected attributes. Prepare variants of each prompt: e.g., one version adds "The applicant is a 35-year-old white male," another says "a 35-year-old Black female," keeping all else identical.

4. **Pilot Test Prompt Wording:** Run a few trials with each LLM to refine the prompt. Ensure the model indeed produces a yes/no decision or a clear recommendation. Possibly use a *temperature=0* setting for deterministic output, or run multiple samples for each and use majority vote as the model's decision.

5. **Execute Bulk Experiments:** For each unique scenario (e.g. Profile #1), generate outputs for all modifier variants (e.g. race A vs race B). Repeat for all profiles across all chosen LLMs. This could mean hundreds of prompts (e.g. 50 profiles × 2 variants × ~3 models = 300 runs). Automate this process and record the outputs.

6. **Measure Decision Differences:** Parse the LLMs' responses to extract the decision (approved or not, interest rate, etc.). Compare outcomes between modified vs unmodified (or between two different demographic values). Compute statistics: e.g., proportion approved for group A vs group B. Use a paired statistical test (McNemar's test for binary decisions) to see if differences are significant. Also examine *magnitude* differences (e.g. average recommended interest rate disparity).

7. **Qualitative Analysis of Rationale:** Examine the justifications the LLM gives. The original study found shifts in ethical principle emphasis; here we look for shifts in reasoning. Does the LLM mention the demographic when it shouldn't? (If it says "Applicant is from a high-crime area" purely due to a name cue, that's revealing bias.) We can use keyword searches in the explanations for any telltale differences.

8. **Mitigation Experiment (optional):** If bias is observed, test a mitigation: for example, prepend an instruction "Make lending decisions based only on financial criteria. Ignore demographic details." as Bowen *et al.* did, and see if it removes disparities [19]. This checks if simple alignment prompts can fix the issue.

9. **Analyze Results:** Summarize which models were most/least biased. Are certain modifiers (race vs gender) more impactful? Perhaps create a table of approval rates by group for each model, and a chart of interest rate gaps. Calculate the "extra credit score" equivalent disadvantage (as Bowen did) to put results in tangible terms [26].

10. **Summarize Implications:** Connect findings to regulations – e.g. if GPT-3.5 shows bias, using it in lending without mitigation could violate ECOA. Highlight how newer models or prompt engineering can reduce bias, informing industry best practices.

11. **Write and Publish:** Draft the paper, structuring it like a classic fairness in ML study: Introduction (citing fair lending laws and LLM uses), Methods (scenario design and statistical tests), Results, Discussion (implications for banks/regulators). Emphasize contributions: first to systematically test socio-demographic bias in LLM credit decisions. Target a top-tier venue such as **ACM FAccT (Conference on Fairness, Accountability, and Transparency)** for conference, or journals like the **Journal of Finance and Data Science** or **Journal of Banking & Finance** (for a finance audience), or even an interdisciplinary journal like **Management Science** (which has published algorithmic bias studies). Given the high stakes, policy journals (e.g. *AI and Ethics*, or *Financial Services Review*) might also be appropriate. Ensure to address any ethical considerations (we are using synthetic data, so no personal data issues; the main ethical concern is responsibly disclosing bias findings so as to improve systems).

By following this plan, the case study will deliver concrete evidence of how (and how much) socio-demographic cues can alter LLMs' lending decisions, providing insights to both academia and industry on aligning AI with fair lending principles.

2. Fairness in Small Business Credit Decisions – Detailed Plan

(a) Existing Work on SME Lending Bias: Fair lending research traditionally focuses on consumer credit, but **bias in small business lending** is a growing concern. Empirical studies have shown that minority-owned small businesses often face higher loan denial rates and worse terms, even when controlling for financial metrics – a gap attributed to factors like different banking relationships or possible discrimination. With fintech lenders entering the space, there's interest in whether AI-driven SME credit models exhibit similar biases or can mitigate them. While direct studies of LLMs in SME lending are scant (as this is a novel idea), related work exists. For instance, **algorithmic bias in commercial lending** was highlighted by the US CFPB's recent rule implementing Section 1071 of Dodd-Frank, which now requires lenders to collect business owners' demographic data to monitor fairness [27]. This regulatory move was motivated by evidence (and advocacy by groups like the National Community Reinvestment Coalition) that disparities similar to the mortgage market may exist for entrepreneurs. Industry reports also suggest that automated underwriting for small businesses could inadvertently disadvantage those without extensive credit histories or those in underserved areas (which often correlate with minority status). In the academic sphere, methods for testing discrimination in SME lending are proposed akin to consumer audits – e.g., sending "mystery shopper" loan applications varying only the owner's race/gender [28]. Our idea essentially digitizes this via LLM. Additionally, research on venture capital and crowdfunding finds gender and race biases in funding decisions; this aligns with the hypothesis that an AI might similarly pick up biases from training data that portrays minority businesses as riskier due to historical inequities. Overall, existing work underscores that **bias in credit access isn't limited to consumers – it extends to small firms – and tools to measure and mitigate it (especially in AI systems) are urgently needed**.

(b) Key Regulations or Standards: In the US, **ECOA applies to small business credit** as well – lenders cannot discriminate by race, sex, etc., in business lending. However, enforcement was historically hampered by lack of data (which Section 1071 now addresses by mandating data collection on applicant demographics for businesses). The CFPB's new rule (finalized in 2023) specifically aims to identify discriminatory patterns in SME lending, signaling regulators' expectations that algorithms used in this domain be fair and explainable. Banks will have to report decision outcomes alongside owner demographics, making it possible to detect disparate impacts. Beyond the US, many jurisdictions have general non-discrimination laws that cover commercial lending. The **EU AI Act** will classify credit risk systems (including SME credit scoring) as high-risk, meaning developers must perform bias assessments and risk management. Also relevant are banking guidelines on model risk management – e.g., the Federal Reserve's SR 11-7 guidance and OCC guidelines – which, while not explicitly about bias, require banks to understand and control their models, including ensuring they don't violate fair lending laws. The **Community Reinvestment Act (CRA)** in the US, though focused on community lending, also encourages fair access to credit for small businesses in underserved areas. In summary, **regulators are watching SME lending fairness closely**, and any evidence of bias in AI models would fall under the same strict legal scrutiny as mortgage discrimination.

(c) Data Sources: We will use a combination of **synthetic data and real statistical distributions** to create realistic small business loan scenarios. Unlike consumer mortgages, there isn't a single comprehensive public dataset with all SME loan applications nationally (though the CFPB's future 1071 data will fill that gap). However, we can draw on several sources: (1) The Federal Reserve's *Small Business Credit Survey* (an annual survey of thousands of small firms on their financing experiences) provides aggregate info on approval rates by race, etc. (2) Some research datasets or competitions (e.g., the Kaggle "Give Me Some Credit" dataset and others) have variables we can repurpose, though they may not include demographics. (3) We can synthesize financial statements for fictional businesses. For example, we might use random samples from distributions of revenue, profit margin, years in operation, credit score of owner, etc., broken down by industry. There are also databases like the SBA (Small Business Administration) loan data (which is public for SBA-guaranteed loans). The SBA data includes loan outcomes and sometimes business owner demographics (e.g., minority-owned flag), which could inform scenario design. For creating paired test cases, we could take an SBA-approved loan profile for a minority-owned firm and change only the ownership detail to non-minority, to see if the LLM would also approve it. Additionally, as a controlled approach, we may generate a set of generic business loan applications (like "Restaurant seeking $50k, 3 years in business, credit score 720, solid cash flow") and simply prepend owner descriptors ("Alice, a Black woman, owns the restaurant" vs "Bob, a white man, owns the restaurant"). Because the scenarios have to be textual, we will ensure to include enough financial detail for the LLM to base a decision on, so that the demographic descriptor is clearly extraneous. We might also utilize **name-based proxies** for gender/ethnicity in the prompt (e.g., "Alejandro vs Alex" as the owner's name) to test implicit bias. Thus, while no single public dataset has all we need, **we have ample data sources to craft hundreds of realistic SME loan cases**, and we can validate that our synthetic profiles align with real-world loan distributions (for credibility).

(d) Step-by-Step Plan:

1. **Select LLMs and Role Prompt:** Choose state-of-the-art LLMs (GPT-4, etc.) and decide how to prompt them. We might say: *"You are a loan officer specializing in small business loans. I will describe an application. Respond with whether you would approve the loan and why."* This role prompt ensures the model adopts the correct context.

2. **Generate Sample Business Profiles:** Create ~40-50 base business loan scenarios. To do this systematically, define a few industries (retail, tech startup, manufacturing, etc.), and firm sizes. For each, assign realistic financials: e.g., "annual revenue $200k, profit $20k, 2 years in operation, owner FICO score 680, collateral available = $50k equipment." Vary these parameters to cover borderline cases (some strong, some weak applications).

3. **Incorporate Non-Demographic Risk Factors:** Include any factors that legitimately matter (so the LLM has enough to chew on). E.g., one version of a profile might mention "business is located in an area with growing economy" or "sector has seen downturn" to see if LLM picks up context. We'll keep these constant across the demographic variants.

4. **Apply Demographic Modifiers:** For each profile, create two versions: one indicating the owner is from a socially disadvantaged group vs a majority group. This can be done explicitly ("The owner is Hispanic" vs "The owner is white") or implicitly ("Maria Rodriguez" vs "Mary Roberts" as the owner's name, plus perhaps mentioning community involvement that hints at background). We must ensure this detail is **the only difference**. Optionally, do the same for gender: e.g., "John" vs "Jane" as owner.

5. **Run LLM Evaluations:** Input each scenario variant into the LLM and record the output (approval/denial and reasoning). Use multiple runs or set temperature low to get consistent results. If models have variability, consider 5 runs per scenario and take the majority outcome.

6. **Quantitative Analysis:** Compare approval rates between paired scenarios. For each base profile, note if the decision changed when the only change was owner's demographic. Compute overall bias metrics: e.g., "X% of minority-owned profiles were approved vs Y% of identical non-minority profiles." Statistical tests (paired t-test or McNemar's test) can assess significance across all profiles. We might find, for example, that out of 50 pairs, 10 were approved when the owner was white and denied when owner was Black, vs only 2 the opposite – indicating bias.

7. **Examine LLM Reasoning:** Analyze the textual explanations the LLM gives. Does it ever explicitly reference the owner's demographic in the denial reasoning? (That would be a glaring bias sign.) More likely it will be subtle: e.g., "The loan is somewhat risky given only 2 years in business" used in both cases, but maybe in one case it additionally says "and the market might not support this minority-owned business" – we'll look for any such problematic language. We can use text analysis or manual coding to categorize reasons given.

8. **Cross-Model Comparison:** If using multiple LLMs, compare which is most consistent. Perhaps GPT-4 shows no differences, whereas a smaller model might. If resources allow, include an open-source model (like Llama-2 fine-tuned on finance) to see if proprietary models handle fairness better.

9. **Regulatory Check:** For any observed bias, map it to regulatory risk. E.g., if the model tends to deny minority profiles more, mention how this would be viewed under ECOA. We can also test a compliance prompt: "Ensure fairness" as we did in Idea 1, to see if bias can be mitigated by prompt.

10. **Case Study Documentation:** Prepare the case study results. Include a table of outcomes by scenario (perhaps anonymized like "Profile A (restaurant): approved for Owner=John, denied for Owner=Maria"). Include aggregate bias measures. Also, discuss any interesting nuances

(perhaps the bias appears only for borderline cases, aligning with findings that bias is strongest when applications are marginal [29] – which Bowen *et al.* observed in mortgages).

11. **Stakeholder Recommendations:** Formulate how these findings could guide lenders. For instance, if LLMs are to be used in credit underwriting, they need thorough fair-lending testing and possibly constraints to ignore demographic proxies. We would emphasize the importance of *algorithmic transparency* – since our study shows the potential for hidden bias, banks should incorporate fairness audits in model risk management.

12. **Publish:** Target venues. This work intersects finance, AI, and policy. A top choice could be **The Journal of Financial Services Research** or **Journal of Risk and Insurance** (actuarial perspective with fairness). Conferences like **NeurIPS (AI in Finance workshop)** or **IEEE Conference on AI in Finance** would also find this relevant. Given the regulatory angle, we might even present to practitioners via industry conferences (e.g., the Federal Reserve's fintech conference). For academic publication, ensure to highlight that this is one of the first studies examining LLM decisions in SME credit – filling a gap in fair AI research.

By executing these steps, our study will yield insights on whether current LLMs can be trusted to make impartial lending decisions for small businesses. This contributes both practical guidance (for fintech lenders or bank AI teams) and academic knowledge on AI fairness beyond consumer lending.

3. Discrimination in Insurance Underwriting – Detailed Plan

(a) Existing Work on AI & Insurance Bias: The insurance industry has long grappled with the line between legitimate risk-based differentiation and unfair discrimination [30] [31]. Traditional actuarial practices allow "fair discrimination" – charging different premiums if backed by risk factors – but **using protected traits (race, religion, etc.) is forbidden** by law [32] [33]. There have been notable cases and studies highlighting bias. For instance, analyses by regulators and researchers found that auto insurance pricing sometimes resulted in higher rates for drivers in minority neighborhoods compared to white neighborhoods with similar risk, due to the use of credit scores and geographic factors as proxies (ProPublica's 2017 report on bias in auto insurance sparked investigations). In life insurance, algorithmic underwriting (scoring applicants with AI using myriad data points) raised concerns that factors like health data or ZIP code might disproportionately disqualify certain groups. While not LLM-based, an example is a 2019 New York DFS investigation into an algorithm used by a life insurer that allegedly charged Black customers more; the algorithm wasn't intentionally using race, but variables correlated with race led to disparate impact [21]. Academic research specifically on LLMs in insurance is emerging. One relevant study is by I. Galo et al. (2025) on financial advice (not underwriting) where they found an LLM gave different insurance-related recommendations to men vs women profiles (e.g. "Insurance & health investments" were suggested more to female profiles) [34] [35] – suggesting even advisory content can have bias. We anticipate similarly that an underwriting decision LLM might exhibit bias. Additionally, **actuarial science literature** has begun exploring fairness criteria for insurance. Some papers propose methods to adjust machine learning models so that predictions for, say, claim risk, are uncorrelated with protected attributes (ensuring "anti-discrimination insurance pricing" [36]). The Casualty Actuarial Society in 2022 released a brief urging actuaries to be aware of algorithmic bias and providing definitions of unfair discrimination in models [31] [37]. In summary, existing work establishes that **bias can creep into insurance models via proxy variables** and that this is a recognized problem – however, an LLM-specific study (with scenario-based ethical testing) would be a novel contribution, complementing the quantitative fairness techniques with a more interpretable, scenario-driven analysis.

(b) Key Regulations and Legal Standards: Insurance is heavily regulated at the state (or national) level with explicit anti-discrimination statutes. In the U.S., **state insurance laws** enumerate protected classes (e.g. race, religion, national origin, in many states also gender for certain lines) that cannot be used in underwriting or rate-setting. For example, many states ban gender rating in auto insurance; some ban credit score use due to its disproportionate racial impact. Regulators enforce this via market conduct exams and requiring actuarial justification for rate differentials. The **New York Department of Financial Services (DFS)** issued a 2019 circular on life insurers' use of external data and algorithms, warning that usage of variables that are proxies for race or other protected classes is illegal if they result in unfair outcomes. As cited in a legal brief [9] [38], New York expressly requires insurers to ensure AI models do not produce "unfair discrimination" – companies must test and audit their models for bias. Internationally, the **EU Gender Directive (2004)** notably prohibited gender-based differentiation in insurance pricing (leading to unisex pricing in the EU from 2012). Anti-discrimination laws in the EU also imply that insurers cannot use ethnicity. Moreover, the **EU AI Act** will treat insurance underwriting AI as a high-risk system requiring transparency and risk controls [25]. Industry standards like the **Principles on Ethical AI** adopted by some global insurers highlight fairness and avoiding bias. In actuarial professional guidelines, the **Actuarial Standards of Practice (ASOPs)** in the U.S. now hint that actuaries should be aware of legal and regulatory requirements regarding unfair discrimination when building models (ASOP 56 on modeling). In short, **any suggestion that an AI (including an LLM) used in underwriting decisions might introduce bias would raise red flags under these laws**. Our study's scenarios (e.g. a minority vs majority applicant with identical risk) directly parallel the tests regulators themselves might conduct to detect illegal discrimination.

(c) Data Sources: Insurance data with both risk factors and protected attributes is rare publicly (because insurers traditionally didn't collect race data, for instance, to avoid bias claims). However, we can rely on **synthetic data generation informed by industry knowledge**. To simulate underwriting scenarios, we first pick a line of insurance to focus on – say, **auto insurance** for personal lines, since it's common and there's public info on rating factors (driving record, age, vehicle, ZIP). We can use public statistics: e.g., the Insurance Services Office (ISO) publishes relativities for different risk factors. We can sample from those: e.g., driver with 1 accident, age 30, ZIP code risk level 8, etc., to get a baseline risk profile. For protected attributes, race isn't used in auto pricing, but our experiment will assign, say, a stereotypically Black-sounding name and an address in a predominantly Black neighborhood vs a white-sounding name in a predominantly white neighborhood (with the same risk level). We might draw on data like the Census for demographics by ZIP to pick two ZIP codes with similar claim statistics but different racial makeup. Alternatively, consider **health insurance or life insurance**: we could simulate two applicants with identical health status and behaviors, but one's profile notes involvement in a certain cultural community that might trigger bias. Another data source: the *Medical Expenditure Panel Survey (MEPS)* has health info by demographics which could inspire health insurance scenarios. Also, if focusing on life insurance underwriting, there are mortality tables by demographic, but since insurers cannot use race explicitly, any disparity would be via proxy (e.g., postal code). We'll likely lean on auto insurance because it has well-known proxies (credit score, occupation) that correlate with race. Some states provided data when investigating racial bias – for example, California's Department of Insurance published some analyses on racial impact of auto insurance rating (Bulletin 2022-5 in California addressed allegations of racial bias [39]). We can use hypothetical persons drawn from those reports. In sum, while direct datasets are limited, **we will create realistic underwriting scenarios by using known risk factor ranges and then overlaying demographic cues** (names, neighborhood, etc.). Importantly, we'll ensure the scenarios are constructed such that according to actuarial standards, the two individuals represent the same risk (so any difference in LLM recommendation is unwarranted).

(d) Step-by-Step Plan:

1. **Select Insurance Context:** Choose one insurance product to narrow scope (e.g. auto insurance policy approval/pricing). Suppose we focus on whether an applicant should be offered a standard rate or flagged for further review/high premium.

2. **Define Base Risk Profiles:** Create 30–40 base applicant profiles. For auto: include age, gender (though gender might be protected in some locales, we treat it as a risk factor here unless focusing on gender bias specifically), driving record, vehicle make/model, location (ZIP or city), credit tier (often used in U.S. auto rates), and any claim history. Ensure these span low to high risk but are borderline cases where a small change might tip a decision (to test model consistency).

3. **Add Socio-Demographic Modifiers:** For each profile, craft two versions. Version A: control – a neutrally presented person (e.g. "Alex Johnson"). Version B: insert a detail suggesting a protected class – e.g. add "Alex Johnson is a Black applicant living in [neighborhood]" or change the name to "Jamal Johnson" and mention involvement in an ethnic community group. Because insurance forms typically don't ask race, we might embed the cue narratively: "Jamal lives in a predominantly African-American neighborhood of City X." We ensure that neighborhood has same risk rating as another neighborhood (so that logically risk is equal). We might also try a gender bias test: e.g. profiles of identical risk but one is "Male, 30" vs "Female, 30" in lines where gender shouldn't matter (like life insurance in the EU context, or in states where it's banned for auto).

4. **Prompt the LLM as Underwriter:** Instruct the LLM in system message: *"You are an insurance underwriter. Decide if the applicant qualifies for the standard premium or if you would increase the rate/deny coverage, and explain."* Then feed the profile details. We may ask specifically: "Would you approve this application at the standard rate?" to get a yes/no.

5. **Execution of Tests:** Run each profile variant through the LLM. Capture the decision (approve/deny or standard vs higher premium) and reasoning.

6. **Measure Outcomes:** Calculate how often the LLM's decision differs between A and B. E.g., out of 30 pairs, perhaps in 5 cases the LLM approves the neutral profile but not the one with the racial cue. Even one such case is significant in a fairness context. We will compute the percentage of profiles that faced "adverse action" only when the protected cue was present.

7. **Statistical Analysis:** Perform a McNemar's test or similar on the paired outcomes. Also, if the LLM provides a score or confidence, measure average differences. The sample size might be small, but consistent directional bias would be telling.

8. **Inspect Justifications:** Scrutinize LLM explanations for any mention of the demographic detail. Ideally, since it's irrelevant, the model *should not* mention it. If we find sentences like "Given the neighborhood's characteristics, I'm a bit concerned" or stereotypes ("the applicant's background might indicate higher risk"), we document those as evidence of bias in the model's reasoning. We could use qualitative coding to categorize themes in explanations for minority vs majority profiles.

9. **Regulatory Interpretation:** Determine if any differences observed would constitute unfair discrimination. For instance, if the LLM says to surcharge the applicant in the predominantly

minority neighborhood but not the identical one in the white neighborhood, that's effectively redlining – illegal under U.S. law. We will articulate this and reference the specific statutes or regulatory guidelines that would be violated (e.g., state X's rating law prohibits that).

10. **Mitigation Experiment:** Similar to prior ideas, test if adding an instruction like "Do not consider factors unrelated to driving risk such as race or ethnicity" yields identical decisions. If it does, that's a hopeful sign that the model can be aligned with policy by prompt; if not, it indicates a deeper issue.

11. **Extend to Other Bias Dimensions (optional):** If time permits, run additional checks, say for gender bias. For example, use a life insurance scenario where the only difference is the implied gender (in a context where gender *shouldn't* be used, like EU). See if the model charges one gender more. This can broaden the study's findings.

12. **Document Results:** Prepare the results with clear examples. Perhaps present a few anonymized vignettes in the paper: "Profile #7: safe driver, 0 accidents – LLM says standard rate; Modified Profile #7 (added that driver is from minority neighborhood) – LLM recommends higher premium." Summarize across all cases. Use a table to show the count of approvals in each group.

13. **Discussion and Recommendations:** Discuss why the LLM might be biased – likely due to training data reflecting societal biases (e.g., it might have ingested text that links certain neighborhoods with higher risk). Emphasize that in critical fields like insurance, relying on LLMs without strict controls could lead to regulatory breaches and ethical issues. Recommend that insurers, if using LLMs (perhaps for underwriting text analysis or customer-facing quotes), must rigorously test for such biases. This section can also highlight how our method (synthetic vignette testing) could be adopted as a **bias audit technique** by insurance companies or regulators.

14. **Publishing Venues:** Aim to publish in an outlet that reaches both actuarial experts and AI ethics researchers. For example, the **ASTIN Bulletin (Journal of the International Actuarial Association)** or the **North American Actuarial Journal** would be appropriate for a deep dive with technical fairness metrics. If focusing more on the AI methodology, conferences like **IJCAI or AAAI (Ethics track)** or **ACM FAccT** again are options. There's also growing interest in insurance innovation conferences (e.g., **IEEE Intelligent Systems in Finance and Insurance**). We will highlight our contribution as the first to probe an LLM's decisions in an insurance context for unfair discrimination.

By following this plan, the study will illuminate whether LLMs align with or violate the fundamental insurance principle of "equal risk, equal premium." The results will inform both the development of AI underwriting tools and the regulatory dialogue on AI fairness in insurance.

Conclusion: Through these case studies, we adapt the socio-demographic vignette methodology to critical financial domains, revealing how LLMs handle ethically charged decisions in lending, insurance, and beyond. Each proposed study is designed to be practical with available data, high in impact for academia and industry, and mindful of regulatory frameworks. By rigorously evaluating LLM outputs with controlled scenario testing, researchers and practitioners can identify biases and work toward AI systems in financial services that are **fair, transparent, and aligned with legal and ethical standards**. The insights from these studies would be valuable not only for academic discourse (in top-tier AI, finance, and risk management journals) but also for financial institutions seeking to deploy AI responsibly, and for regulators crafting guidelines on AI in finance. Ultimately, this line of research helps ensure that as we integrate powerful LLMs into financial decision-making, we do so in a way that upholds fairness and trust.

---

[1] [2] [3] [4] Socio-Demographic Modifiers Shape Large Language Models' Ethical Decisions by Vera Sorin, Panagiotis Korfiatis, Jeremy D. Collins, Donald U. Apakama, Benjamin S. Glicksberg, Mei-Ean E. Yeow, Megan Brandeland, Girish N. Nadkarni, Eyal Klang :: SSRN
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5174198

[5] [7] [16] [17] [18] [19] [26] [29] AI Exhibits Racial Bias in Mortgage Underwriting Decisions | Lehigh University News
https://news.lehigh.edu/ai-exhibits-racial-bias-in-mortgage-underwriting-decisions

[6] [12] Biased echoes: Large language models reinforce investment biases and increase portfolio risks of private investors
https://journals.plos.org/plosone/article/file?type=printable&id=10.1371/journal.pone.0325459

[8] Measuring and Mitigating Racial Disparities in LLMs: Evidence from a Mortgage Underwriting Experiment by Donald E. Bowen III, S. McKay Price, Luke C.D. Stein, Ke Yang :: SSRN
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4812158

[9] [38] The Legal Landscape of AI in Insurance: What New York Insurers Need to Know - Weber Gallagher Simpson Stapleton Fires & Newby, LLP
https://www.wglaw.com/news/the-legal-landscape-of-ai-in-insurance-what-new-york-insurers-need-to-know/

[10] [11] Using large language models for financial advice by Christian Fieberg, Lars Hornuf, David Streich, Maximilian Meiler :: SSRN
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4850039

[13] Bias and Fairness of AI-based systems within Financial Crime | NICE Actimize
https://www.niceactimize.com/blog/fraud-bias-and-fairness-of-ai-based-systems-within-financial-crime/

[14] [15] EBA alerts on the detrimental impact of unwarranted de-risking and ineffective management of money laundering and terrorist financing risks | European Banking Authority
https://www.eba.europa.eu/publications-and-media/press-releases/eba-alerts-detrimental-impact-unwarranted-de-risking-and

[20] [21] Artificial Discrimination | Zurich Insurance
https://www.zurich.com/media/magazine/2020/artificial-discrimination

[22] Massachusetts AG Settles Fair Lending Action Based Upon AI ...
https://www.cfsreview.com/2025/07/massachusetts-ag-settles-fair-lending-action-based-upon-ai-underwriting-model/

[23] AI in the Financial Services Industry | Consumer Finance Monitor
https://www.consumerfinancemonitor.com/2025/08/18/ai-in-the-financial-services-industry/

[24] High-level summary of the AI Act | EU Artificial Intelligence Act
https://artificialintelligenceact.eu/high-level-summary/

[25] EIOPA publishes Opinion on AI governance and risk management
https://www.eiopa.europa.eu/eiopa-publishes-opinion-ai-governance-and-risk-management-2025-08-06_en

[27] NCRC and Fintechs – Joint Letter on Fair Lending and the Executive ...
https://stratyfy.com/ncrc-and-fintechs-joint-letter-on-fair-lending-and-the-executive-order-on-ai/

[28] [PDF] Issue Brief - Approaches to Identify and/or Mitigate Bias in Property ...
https://www.actuary.org/sites/default/files/2023-02/CPCdataBiasIB.2.23_0.pdf

[30] [31] [32] [33] [37] actuary.org
https://www.actuary.org/sites/default/files/2023-08/risk-brief-discrimination.pdf

[34] [35] coller.m.tau.ac.il
https://coller.m.tau.ac.il/sites/coller.tau.ac.il/files/media_server/Recanati/management/safra/galo.pdf

[36] Discrimination-free pricing in non-life insurance - Advisense
https://advisense.com/2023/03/27/discrimination-free-pricing-in-non-life-insurance/

[39] [PDF] BULLETIN 2022-5 Allegations of Racial Bias and Unfair ...
https://www.insurance.ca.gov/0250-insurers/0300-insurers/0200-bulletins/bulletin-notices-commiss-opinion/upload/BULLETIN-2022-5-Allegations-of-Racial-Bias-and-Unfair-Discrimination-in-Marketing-Rating-Underwriting-and-Claims-Practices-by-the-Insurance-Industry.pdf