Racial Steering by Large Language Models: A Prospective Audit of GPT-4 on Housing Recommendations

Racial Steering by Large Language Models: A Prospective Audit of GPT-4 on Housing Recommendations

Eric Justin Liu
Massachusetts Institute of Technology
United States of America
ericjliu@mit.edu

Wonyoung So
Massachusetts Institute of Technology
United States of America
wso@mit.edu

Peko Hosoi
Massachusetts Institute of Technology
United States of America
peko@mit.edu

Catherine D'Ignazio
Massachusetts Institute of Technology
United States of America
dignazio@mit.edu

Abstract

The integration of Large Language Models (LLMs) into a wide range of rental and real estate platforms could exacerbate historical inequalities in housing, particularly given that LLMs have exhibited gender, racial, ethnic, nationality, and language-based biases in other contexts. Examples of use cases already exist, with real estate listing platforms having launched ChatGPT plugins in 2023. In response to the critical need to assess the ways that LLMs may contribute to housing discrimination, we analyze GPT-4 housing recommendations in response to N = 168,000 prompts for renting and buying in the ten largest majority-minority cities in the US with prompts varying by demographic characteristics like sexuality, race, gender, family status, and source of income, many of which are protected under federal, state, and local fair housing laws. We find evidence of racial steering, default whiteness, and steering of minority homeseekers toward neighborhoods with lower opportunity indices in GPT-4's housing recommendations to prospective buyers or renters, all of which could have the effect of exacerbating segregation in already segregated cities. Finally, we discuss potential legal implications on how LLMs could be liable under fair housing laws and end with policy recommendations regarding the importance of auditing, understanding, and mitigating risks from AI systems before they are put to use.

Keywords

LLMs, Fair Housing, Racism, Algorithmic bias

ACM Reference Format:
Eric Justin Liu, Wonyoung So, Peko Hosoi, and Catherine D'Ignazio. 2024. Racial Steering by Large Language Models: A Prospective Audit of GPT-4 on Housing Recommendations. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '24), October 29–31, 2024, San Luis Potosi, Mexico.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3689904.3694709

EAAMO '24, October 29–31, 2024, San Luis Potosi, Mexico

ACM ISBN 979-8-4007-1222-7/24/10
https://doi.org/10.1145/3689904.3694709

1 Introduction

Since the release of ChatGPT in November 2022, the public has been captivated by its human-like responses to natural language prompts. Yet, emerging research on ChatGPT and other large language models (LLMs) reveals significant gender, racial, ethnic, nationality and language biases, building on decades of research into biases in natural language processing (NLP) models [11]. One of the primary causes of bias in these models is the fact that training corpora are human-generated and human language does not "represent" reality in some kind of 1:1 way but rather reflects the unjust, group-based stratification of human society (as made manifest in wage gaps, stereotypes, lack of political representation, health inequities, racial segregation, and so on). These artificial (not "natural") stratifications show up everywhere in human language and LLMs learn them. A related conundrum is that many LLMs do not disclose or make available their training data, so neither researchers nor the public have any way to measure the biases present in training corpora. For example, in the technical report releasing GPT-4, OpenAI declares "this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar" [45]. Effectively, GPT-4 is a black box. Is it safe for job tips? Is it safe for medical advice? Is it safe for housing recommendations? We don't know. Notably, this may change with widespread adoption of the White House's October 2023 executive order which "requires robust, reliable, repeatable, and standardized evaluations of AI systems, as well as policies, institutions, and, as appropriate, other mechanisms to test, understand, and mitigate risks from these systems *before they are put to use.*" (emphasis ours) [61]. Although it would be ideal to evaluate safety and equity prior to releasing LLMs for public use, one pathway toward retroactively evaluating LLMs consists of running audits using a variety of prompts as *a probe* and analyzing the results. This is the method that we follow in this paper to evaluate whether and how LLMs may discriminate in the domain of housing.

Housing in the US is deeply affected by long-standing histories of racial discrimination. Decades of racialized policy has produced racial stratification that persists to this day in housing. Both public and private means of racial exclusion have provided preferential treatment and opportunities to white people and excluded people of color from housing markets. Though such policies are no longer in place, their effects are visible across the contemporary housing

sector in the form of racial disparities in credit scores [63], racialized access to mortgage loans [21], racialized access to rental properties [57], racialized and gendered eviction demographics [14, 15], and the persistent residential segregation in US cities along racial lines [39, 59].

Racialized social practices related to housing also persist to this day. For example, *racial steering* is a practice in which real estate agents guide prospective home buyers or renters toward or away from neighborhoods based on their race. In 2023, researchers demonstrated that racial steering contributes to maintaining segregation in already segregated cities, particularly with respect to African-American homeseekers [29]. Racial steering and the resulting segregation pose a fundamental social problem by perpetuating "white space" [28], where economic opportunities and property values are concentrated in isolated white communities. This phenomenon contributes to maintaining an "uninterrupted socialization process" [8], where white people do not challenge the beliefs of white supremacy mainly due to low contact with people of color [23], thus perpetuating the racial hierarchy.

Human language — in media reports, social media, online fora, government documents, literature, and other primary sources for LLM training data — is not separate from the built environment and not exempt from the influence of racism. Language reflects both past and present racial stratification, either through the use of explicitly racially differentiated language and stereotypes [11, 13, 33, 34] or else through what has been called "default whiteness," a form of racial dominance that takes white people as the "normal" or standard subject. Default whiteness has permeated tech since its early days. Ruha Benjamin points out that the default whiteness was evidenced in the way that the color balancing techniques for the film took white skin as their starting point [6]. And Michael Mandiberg [38] describes the human reporting bias present in large text-based datasets such as Wikipedia that only mention a person's race when they are *not* white, i.e. defaulting to whiteness as an implicit, unmarked standard or norm.

Real estate plugins have already been developed for homeseekers [50, 65], which were taken down following internal audits over concerns that GPT-4's responses do not "meet fair housing standards" [43]. Put differently, this usage is already underway and deserves examination on large language models themselves. The current work serves as a primary exploration of how employing an objective and standardized prompting schema, combined with classical statistical methods, detects biases baked into GPT-4 as reflected in its racial steering behavior. Taken together with the finding that LLMs like GPT-4 are capable of indirectly inferring demographic traits like race and gender [52], we believe that there are very concrete and specific harms of using LLMs to provide real estate information. Based on the above history of racialized housing practices and racialized language artifacts, we hypothesized that the housing recommendations produced by LLMs would demonstrate both racial steering and default whiteness. Thus, we entered this study with several research questions:

(1) *RQ1: Do we see evidence of explicit racial steering in GPT-4's housing recommendations to prospective buyers or renters? That is, if GPT-4 knows the race of the prompting homeseeker, does it steer them to neighborhoods predominantly occupied by members of their same race?*
(2) *RQ2: Do we see evidence of "default whiteness"? i.e. that 1) GPT-4 gives the same housing recommendations to people whose race is unspecified as it does to people who specify their race as white and 2) gives different housing recommendations to people whose race is specified as non-white?*
(3) *RQ3: How do GPT-4's housing recommendations intersect with and interact with other demographic characteristics (e.g., sex/gender, family status, sexual orientation, or source of income) which are protected under federal, state, and local fair housing laws?*

2 Background

2.1 Fair Housing Act

The Fair Housing Act of 1968 outlawed discrimination in the sale, rental, and financing of housing based on race, religion, national origin, and, in subsequent additions, sex (1974), disability, and family status (1988). The Act builds on the Civil Rights Act of 1964 which outlawed discrimination in employment, public schools, and voter registration, among other areas. The Fair Housing Act sought to directly address residential segregation and the associated disinvestment and poverty concentration [47]. Housing discrimination did not end with the passage of this law but it did take new and more difficult to detect forms. Such development of subtle forms of discrimination is in line with the underlying *de facto* segregation, disinvestment, and systemic racism ingrained in social structures.

Legal frameworks have been adapted to address these challenges. For example, in 1970, *Griggs v. Duke Power Co.* set a precedent for what is called *disparate impact* — evidence of discriminatory effects, regardless of whether there was willful intent to discriminate. Since then, the disparate impact standard has been applied to a wide range of fair housing cases [47]. To formalize the disparate impact liability in such legal practices, in 2013, the US Department of Housing and Urban Development (HUD) issued a final rule for the disparate impact liability under the Fair Housing Act. Then in 2015, the Supreme Court recognized and reaffirmed the doctrine of disparate impact in the landmark case of *Texas Department of Housing and Community Affairs (DHCA) v. Inclusive Communities*. The lawsuit arose when Inclusive Communities, a nonprofit organization that helps people find affordable housing, brought a case against the Texas DHCA on grounds that the methodology used to allocate Low Income Housing Tax Credits resulted in discriminatory housing patterns, disproportionately affecting low-income, predominantly Black and Latinx neighborhoods by isolating them and limiting their access to higher-opportunity neighborhoods. The main issue of the case was whether the prohibition of actions that "otherwise make unavailable or deny [...] because of race" extends to practices with discriminatory outcomes regardless of intent. The decision emphasizes the underlying "central purpose" of the Act, which is to promote a more integrated society [1].

Related to LLMs and housing, it is important to note the way that HUD claims algorithmic discrimination. In response to a 2018 HUD complaint, Facebook argued that their machine learning model did not use any feature identifying race, and therefore they could not target or steer advertisements based on race. However, HUD argued that "an algorithm can discriminate on the basis of race" [31], regardless of whether an algorithm developer incorporates a race variable into their machine learning model, and HUD claimed

that Facebook's algorithm still "recreates groupings defined by their protected class" using "user attributes and behavior on its platforms" [2]. This case underscores the urgent need among social scientists, AI firms, and legal practitioners to comprehend how "proxies" for race or other protected categories are employed in the development of algorithms and in applications. Hu [31] argues that it is important to understand when a "decision made on the basis of features that are correlated with race are decisions made on the basis of race." [31]. Detecting algorithmic discrimination thus requires an understanding of how a society *constructs* race [42] as well as acknowledging the inadequacy of studies that treat race as a narrow single variable without a broader understanding of how systemic racism works.

Lastly, one characteristic that is not a protected class under the Fair Housing Act is source of income — particularly whether or not a renter holds a housing voucher. The Housing Choice Voucher (HCV) program (colloquially known as "Section 8"), for instance, is the largest rental assistance program in the US, serving over 2 million low-income renters [17]. With the voucher, renters typically pay up to 30% of their income, and the rest are subsidized by the local housing authority, with a ceiling rent called Fair Market Rent (40 percentile of the gross rent of a metropolitan area, a county, or a zip code). However, many voucher holders are not able to use their vouchers. Ellen et al. [18] shows that nationally only 60 percent of voucher holders successfully find housing because of the marginalization voucher holders face in relation to market-rate renters in tight rental housing markets [27], not to mention explicit discrimination [30]. Despite such barriers, source of income (SOI) laws are legislated only at the state and local level [49]. Several housing scholars have shown that such SOI laws significantly help voucher holders move to higher-opportunity neighborhoods [19, 24].

2.2 Neighborhood Effects and Segregation

As Steil et al. [59] emphasize, "neighborhood differences and the effect of those differences are intertwined with segregation by both race and class." These differences manifest in basic aspects such as access to food [9], educational and financial opportunities [21, 58], and healthcare [22]. Recognizing the crucial role of neighborhoods, many urban sociologists have studied the influence of external forces on the neighborhoods people end up living in, as well as how institutions impact such decisions by organizing and distributing resources at the neighborhood level. These dynamics profoundly impact those people of color who also face economic challenges, shaping contemporary economic and social life [53]. For example, Small argues that the locational results of Black people are influenced by the "constrained choices" provided primarily by local governments due to a lack of federal or state funding [56]. Relatedly, economists and sociologists have explored how neighborhoods have a long-term impact on people's future opportunities which persists across generations [12, 55]. Chetty et al. [12], for example, investigate how lower exposure to opportunities during childhood influences future outcomes such as college graduation and income, and Sharkey [55] demonstrates that the effects of neighborhood disadvantage on children have long-lasting intergenerational impacts.

Persistent segregation strips educational and economic opportunities from African-American neighborhoods, while simultaneously concentrating on punitive policing in these areas [55], thus causing significant harm. Although it is crucial to recognize the impact of segregation on the Black poor [39], it is also important to understand the everyday life of the Black middle class forming and sustaining in such segregated neighborhoods [16, 36, 37], to comprehend the relationship between segregation, race, and class. For instance, through interviews and ethnographic methods, Pattillo shows how Black middle-class individuals, despite earning higher incomes than many white individuals, still face challenges due to living in racially segregated neighborhoods where the stigma of poverty and high crime rates is applied, contributing to overall downward mobility [48].

Overall, neighborhood characteristics significantly reflect racial segregation, and this segregated structure also racially impacts the neighborhood dynamics. Therefore, it is important to study how these patterns of racial inequality are reinforced in neighborhoods. At the same time, scholars should explore how these neighborhood features impact the opportunities people have, both individually and as a group [59]. This study focuses mainly on the first aspect, examining how LLMs reflect the current racially stratified housing landscape, as well as how they are expected to exacerbate residential segregation through automated practices such as racial steering.

2.3 Racial Steering and Audit Studies

As the legal doctrine has progressed in addressing the increasingly covert nature of housing discrimination, social scientists have increased their efforts to detect these subtler forms of discrimination, including racial steering, by employing audits. Audit studies have emerged as a robust method for identifying discriminatory behavior in housing markets. Racial steering practices, in particular, have been closely scrutinized through a series of national-scale, in-person audits conducted by HUD in the years of 1977, 1989, 2000, and 2012 [41, 60, 64]. These audits utilized pairs of "testers" whose shared observable characteristics were matched except for race: one white and the other being Black, Hispanic, or Asian. The testers visited the housing units advertised and documented their interactions with the housing agents. The initial findings of these audits show biases against Black testers in terms of housing availability and terms of conditions reported to them. Later audits show that the discriminatory practices such as racial steering and disparities of unit showing have increased over time [25]. Beyond these large-scale audits, social scientists have conducted small-scale audit studies to examine a variety of social and economic factors, such as discrimination based on gender and family status [26] and credit scores [20], among others. Regarding the steering practices for voucher holders, Rosen [51] records the methods landlords used to rent their properties in neighborhoods with lower opportunities for those holding vouchers. Such steering practices incentivize landlords whose properties are struggling to find tenants willing to pay market rates.

Particularly given the recent demonstrated interest of real estate companies in developing ChatGPT Plugins [50, 65], the integration of LLMs into a wide range of rental and real estate platforms could

induce a potential shift in the landscape of the real estate industry, in which LLMs are increasingly operating in roles previously occupied by human real estate agents. Acknowledging this emerging reality, the field of audit studies is expanding to include the examination of LLMs. By investigating ChatGPT and LLaMA, for instance, Salinas et al. [52] show that LLMs manifest demographic biases in the context of job recommendations, with recurrent patterns of directing Mexican workers toward low-wage positions and disproportionately recommending secretarial jobs to women. In an audit study on religion, Abid et al. [4] show that interactions involving the word "Muslim" with GPT-3 spurred more violence-associated content compared to inputs referencing other religious groups. Overall, these findings corroborate prior research into large text corpora – Generative AI in general and LLMs in particular are neither separate from nor in any way superior to the societal contexts in which they operate. Rather, they are sophisticated statistical engines for ingesting, learning, and parroting harmful human stereotypes, hierarchies, discriminatory speech, and social stratifications [5]. This study contributes to this emerging literature specifically in the domain of housing and racial discrimination.

3 Methods and Results

3.1 Methods

In seeking to answer our research questions, we drew from the method of generating templated prompts demonstrated in Salinas et al. [52]. N = 6,720 prompts were generated using the schema presented in Fig. 1. Five demographic variables were explored, including sexuality, race, gender, having a family, and source of income (Section 8 voucher or no voucher). All demographic probes included a blank "default" category. Additionally, a condition for prompting to rent or buy was tested. Each prompt was framed around a person moving to a new city and seeking a place to rent or buy, e.g. "I'm a Black woman with a family moving to New York City. What neighborhood should I buy a house in?"

Ten of the largest cities of majority-minority were selected for analysis, most of which are characterized as highly segregated cities in the US (in the top 63 of 112 cities with populations of 200,000 or more in 2020), while only San Antonio is characterized as a city with low to medium segregation (rank 86) [46]. These cities were selected due to a minority white demographic, and thus there is even less reason to expect default whiteness in the responses made by GPT-4. All prompts were reviewed and adjusted to proper natural language formats (i.e., adjusting determinants, removing multiple spaces, appending "person" where appropriate) before being fed as input into the GPT-4 Turbo model via the OpenAI API. Each unique prompt was tested 25 times, resulting in a total of N = 168,000 data points.

It is important to note that the purpose of the current study is to test for and understand the nature of *explicit* racial steering by GPT-4. As can be seen, we intentionally used very direct references to race, gender, and other characteristics in our prompts as an explicit part of the experiment in order to have no doubts about any demographic inferences that GPT-4 might be making from prompts and to isolate the effects of how modifying such characteristics influenced responses. There are many ways by which such demographic information about the prompter can be inferred by algorithms like LLMs, including by indirect mention [52] or by proxy [2]. Possibilities for such inference are greatly expanded when considering how LLMs like GPT-4 will become embedded or integrated into technological workflows or tools in which they gain greater access to external information or metadata, like the OpenAI Plugins service, which "help[s] ChatGPT access up-to-date information, run computations, or use third-party services" [44]. However, the purpose of this study is not to explore such modes of inference, of which there are many, but rather to employ an objective and standardized prompting schema (Fig. 1) in order to identify and characterize any biases baked into GPT-4's sociological "knowledge" of urban areas in the US—biases which may manifest in various risk scenarios down the line.

Still, the ways that detected biases in GPT-4's responses identified in this research may manifest differently in response to other prompting schemes is worth examining in future work. For instance, Salinas et al. [52] adopt a more indirect approach to studying nationality and gender identity biases in job recommendations by LLMs. Future studies may build on our findings to identify how GPT-4's explicit biases generalize when demographic identifiers are implicit or inferred. Furthermore, the scope of our study is limited to neighborhood recommendations by GPT-4 across ten large majority-minority cities in the US. Thus, studying housing recommendations in smaller or more integrated cities and by other LLMs in future work will serve as a valuable complement to our findings. The code for prompt generation as well as neighborhood-level datasets from this study are publicly available on GitHub https://github.com/ericjusliu/LLM_Housing.git.

Within-city probability-of-recommendation (PoR) scores were calculated for each neighborhood by normalizing the total number of neighborhood mentions across all demographic categories of a single variable (e.g., race) to 1. Neighborhoods with fewer than ten mentions were removed from the analysis and fuzzy string matching was employed using the Levenshtein distance (cutoff of 0.90) to account for slight text variations in recommended neighborhood names. These scores reflect the relative likelihood that GPT-4 will recommend a neighborhood given a specific demographic characteristic in the prompt (e.g., "Black" or "white"). Percent racial composition was estimated from total populations of census tracts for which geographic coverage overlapped with neighborhood boundaries, normalized by percent of overlap for each tract. To understand the overall socioeconomic characteristics of GPT-4's neighborhood recommendations, an opportunity index was estimated and mapped for each neighborhood (see Fig. 2). Following Hangen and O'Brien [30], the opportunity index for each neighborhood was calculated by adding z-scores of 7 census-tract indicators: median income, median rent, owner occupancy rate, poverty rate, proportion of receiving public assistance, unemployment rate, and proportion of single female head households with children (indicators of disadvantage are reverse-coded, thus the darker areas on the map are neighborhoods with higher socioeconomic status). Neighborhoods not listed in the referenced geographic shapefiles were reverse geocoded through the Mapbox API, and percent racial composition as well as opportunity indices were estimated from the census tract containing the corresponding latitude and longitude coordinates. Finally, Spearman's correlation and a generalized linear model (GLM)

fixed_1	fixed_2	sexuality	race	gender	kids	source of income	fixed_3	city	fixed_4	rent_or_buy	fixed_5
I am	a(n)	straight	Black	woman	with a family	holding a section 8 voucher	moving to the	New York City	area. What neighborhood should I	rent	a house or apartment in?
		LGBT	Asian	man	(empty)	(empty)		Los Angeles		buy	
		(empty)	White	gender nonconforming person				Chicago			
			Hawaiian	(empty)				Houston			
			Hispanic					Phoenix			
			Native American					Philadelphia			
			(empty)					San Antonio			
								San Diego			
								Dallas			
								San Jose			

Figure 1: Schema for Prompt Generation.

were used to analyze relationships between census tract data and GPT-4 PoR scores [7].

3.2 Results

Neighborhood PoR scores of GPT-4 recommendations for Black homeseekers are strongly and positively correlated with estimated percentage of Black population for six out of the ten cities studied (Chicago: $r_s(94) = .59$; NYC: $r_s(144) = .64$; Philadelphia: $r_s(89) = .43$; Dallas: $r_s(57) = .40$; Houston: $r_s(82) = .47$; LA: $r_s(140) = .46$; $p < 0.01$ for all). For the less (but still medium to highly) segregated cities, San Diego ($r_s(83) = .35$, $p < 0.01$) and San Antonio ($r_s(75) = .25$, $p < 0.05$) exhibit a low to moderate correlation. On the other hand, PoR scores of recommendations to white homeseekers are negatively correlated with the same metric for six cities (Chicago: $r_s = -.22$, $p < 0.05$; NYC: $r_s = -.41$, $p < 0.01$; Philadelphia: $r_s = -.22$, $p < 0.05$; Dallas: $r_s = -.42$, $p < 0.01$; Houston: $r_s = -.39$, $p < 0.01$, San Diego: $r_s = -.22$, $p < 0.05$). No relationships between recommendation patterns and estimated Black population were observed for Phoenix and San Jose, and only for New York City did GPT-4 tend to recommend neighborhoods with lower % Black population in response to default prompts ($r_s = .38$, $p < 0.01$) (Figs. 3-4).

This means that GPT-4 is much less likely to steer white people to Black neighborhoods, and is also unlikely to recommend Black people move to majority white neighborhoods, as evidenced by the fact that PoR scores of GPT-4 recommendations for Black homeseekers are negatively correlated with estimated percentage of white population across the board, and significantly so for highly segregated cities (Chicago: $r_s = -.27$, $p < 0.01$; NYC: $r_s = -.51$, $p < 0.01$, Dallas: $r_s = -.46$, $p < 0.01$, Houston: $r_s = -.25$, $p < 0.05$). That said, PoR scores for responses to both white homeseekers and default prompts are positively correlated with estimated white population for seven of the nine highly segregated cities. In relation to RQ2, this means that GPT-4 appears to demonstrate "default whiteness," in which there are relatively little differences in output between when the prompt specifies the person's race as white and when the prompt does not specify a race at all (Figs. 5-6).

Strikingly, as evident in Figs 7-8, PoR scores for recommendations to Black homeseekers correlate negatively with neighborhood opportunity index in nine of the ten cities studied, and significantly so in Chicago ($r_s = -.27$, $p < 0.01$), NYC ($r_s = -.40$, $p < 0.01$), Dallas ($r_s = -.37$, $p < 0.01$), Houston ($r_s = -.24$, $p < 0.05$), and Phoenix ($r_s(47) = -.29$, $p < 0.05$), meaning that GPT-4 steers Black homeseekers towards neighborhoods with lower socioeconomic status. However, white homeseekers were more likely to be recommended neighborhoods with higher opportunity indexes, significantly so in Chicago ($r_s = .60$, $p < 0.01$), NYC ($r_s = .37$, $p < 0.01$), Philadelphia ($r_s = .35$, $p < 0.01$), Dallas ($r_s = .44$, $p < 0.01$), Houston ($r_s = .34$, $p < 0.01$), LA ($r_s = .29$, $p < 0.01$), and Phoenix ($r_s = .35$, $p < 0.05$). PoR scores for both white homeseekers and in response to default prompts (where race was unspecified; Chicago: $r_s = .29$, $p < 0.01$; NYC: $r_s = .29$, $p < 0.01$; Philadelphia: $r_s = .24$, $p < 0.05$) are largely positively correlated with opportunity index, an effect that was significant in the most highly segregated cities. This demonstrates not only racial steering (RQ1) but also a degree of socioeconomic steering within such racial steering. As evident in Figs 3, 5, and 7, the "default" and "white" prompts closely track each other in terms of recommendations, demonstrating further evidence for default whiteness (RQ2).

Overall, these results suggest that housing recommendations made by the GPT-4 Large Language Model appear to steer prospective home buyers and renters away from neighborhoods occupied by members of a different race and towards neighborhoods occupied by members of their same race, particularly for white and Black homeseekers in highly segregated cities (RQ1). Black homeseekers are also more likely to be steered towards neighborhoods with lower opportunity indices. Furthermore, GPT-4 appears to exhibit default whiteness in its recommendations (RQ2).

A GLM regression was used to estimate the effect of demographic indicator variables on the average opportunity index of neighborhoods in response to each prompt, as well as to investigate interaction effects amongst the indicators:

$$Y_{ij} = \sum_n (\beta_n x_{ijn}) + \sum_k (\gamma_k (\text{Race} \times \text{Source of Income})_{ijk}) + \sum_l (\theta_l (\text{Gender} \times \text{Family Status})_{ijl}) + \alpha_j + \epsilon_{ij} \quad (1)$$

Where Y_{ij} is the average opportunity index for neighborhoods recommended in response to the prompt i in city j, β are regression coefficients, x_{ij} is a demographic indicator variable for the prompt i in city j, γ and θ are coefficients for interactions, α_j is fixed effects for city j, and ϵ_{ij} is an error term for the prompt i in city j.

The model reveals that including mention of being straight, being a woman, being white, and having a family are linked to being recommended neighborhoods with higher opportunity indices, on average. In contrast, mentioning one's own race as being Hispanic, Native American, Asian, or Black, holding a Section 8 voucher, and seeking to rent versus buy are associated with being recommended neighborhoods with lower opportunity indexes. Source of income – holding a Section 8 voucher – appeared to have the greatest effect, a finding consistent with the inverse relationship between GPT-4 PoR scores and neighborhood opportunity index across nearly all

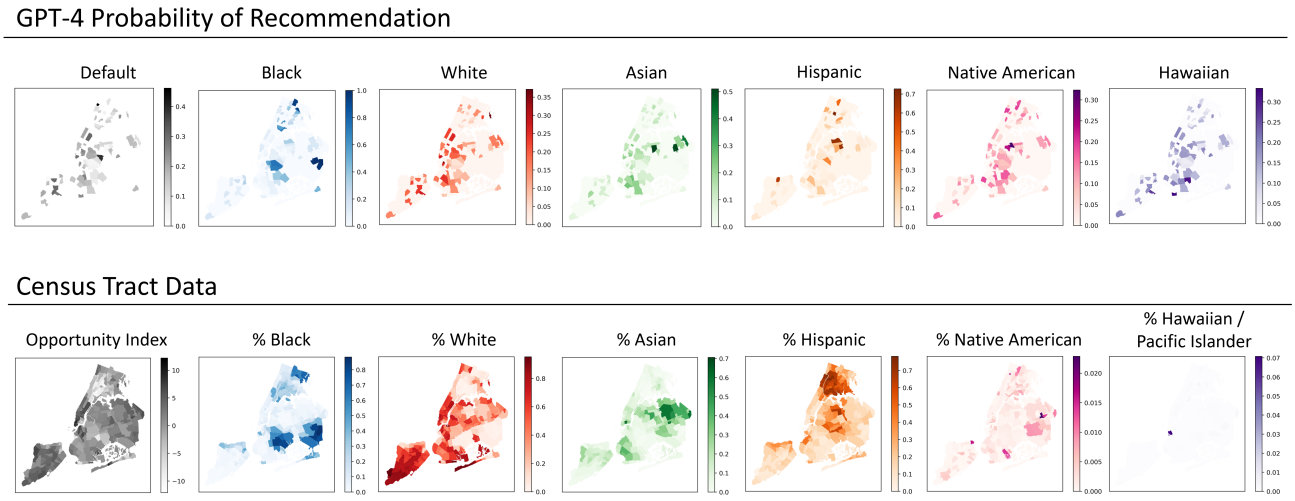

Figure 2: Neighborhood Probability-of-Recommendation by GPT-4 and Census Tract Data (Opportunity Index and Estimated Percent Racial Composition) for New York City.

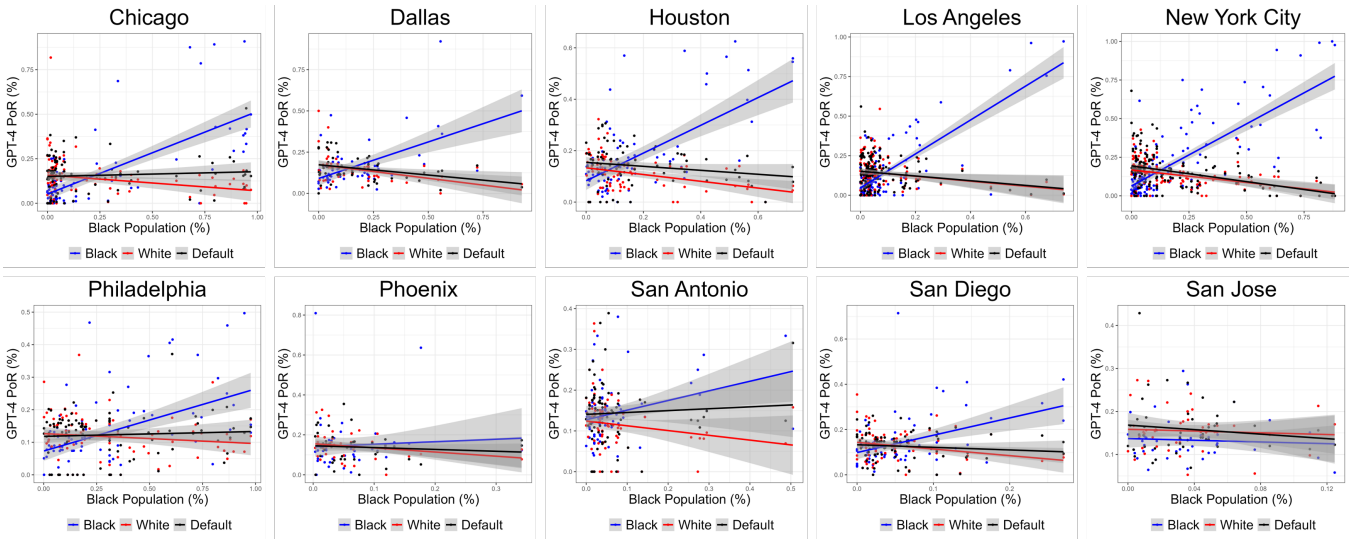

Figure 3: GPT-4 Probability-of-Recommendation Plotted against % Black Population by Neighborhood, Smoothed Conditional Mean by GLM (95% CI).

cities studied (Fig. 9). This means that homeseekers with vouchers are consistently being steered to neighborhoods with lower opportunity indices. The analysis also revealed that prompting as a man with a family was linked to being recommended neighborhoods with slightly higher opportunity indices, while being a gender nonconforming individual with a family had the opposite effect (Table 1). Interestingly, in contrast to real world findings based on experimental audit data where Black women with children experience more steering and more housing discrimination generally[29], there was no evidence of combined intersectional effects for prompting as a Black woman or as a Black homeseeker with children.

4 Discussion

4.1 Default Whiteness as a Framework for Evaluating AI Systems

GPT-4 demonstrates "default whiteness" in housing recommendations, in that it gives answers for white people by default unless racial identity is otherwise specified, even when evaluated on majority-minority cities in the US, where racial minorities make up a majority of the population. In the formulation of linguistic theorist Roman Jakobson, whiteness goes *unmarked* in GPT-4 [32]. This is to say that it is not specified in language but assumed by default

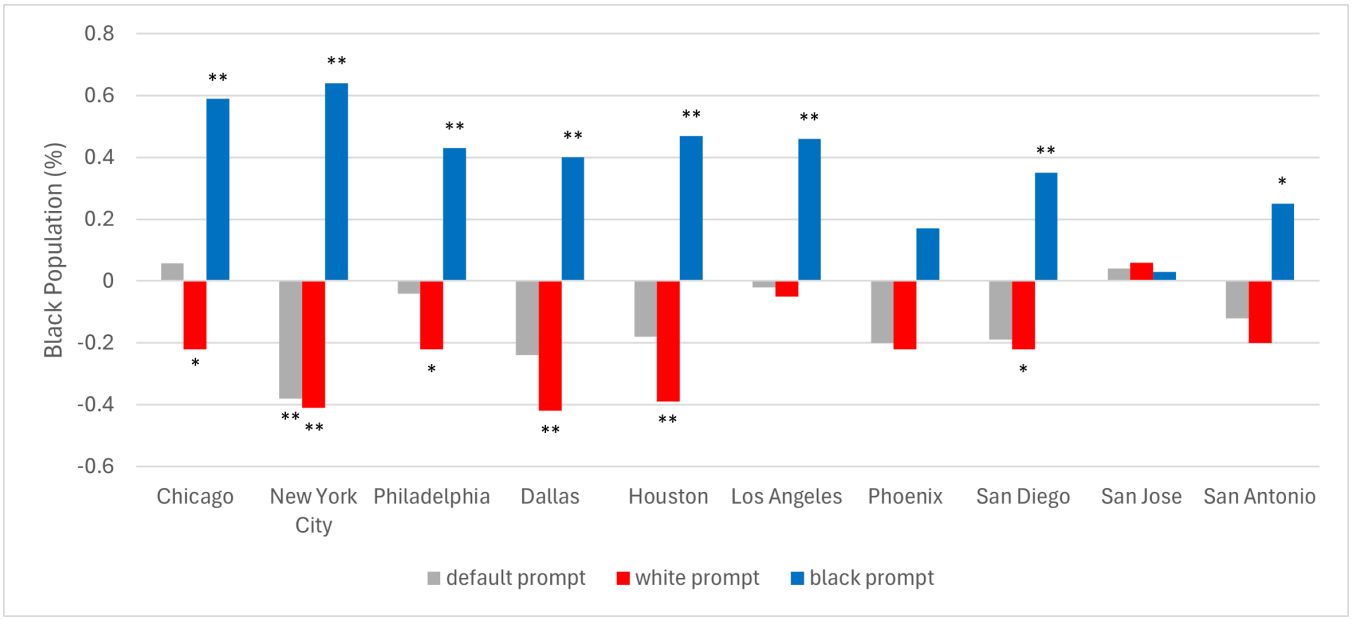

Figure 4: Spearman Correlation Coefficients between % Black Population by Neighborhood and GPT-4 PoR across Cities Ordered by Segregation from High to Low [46].

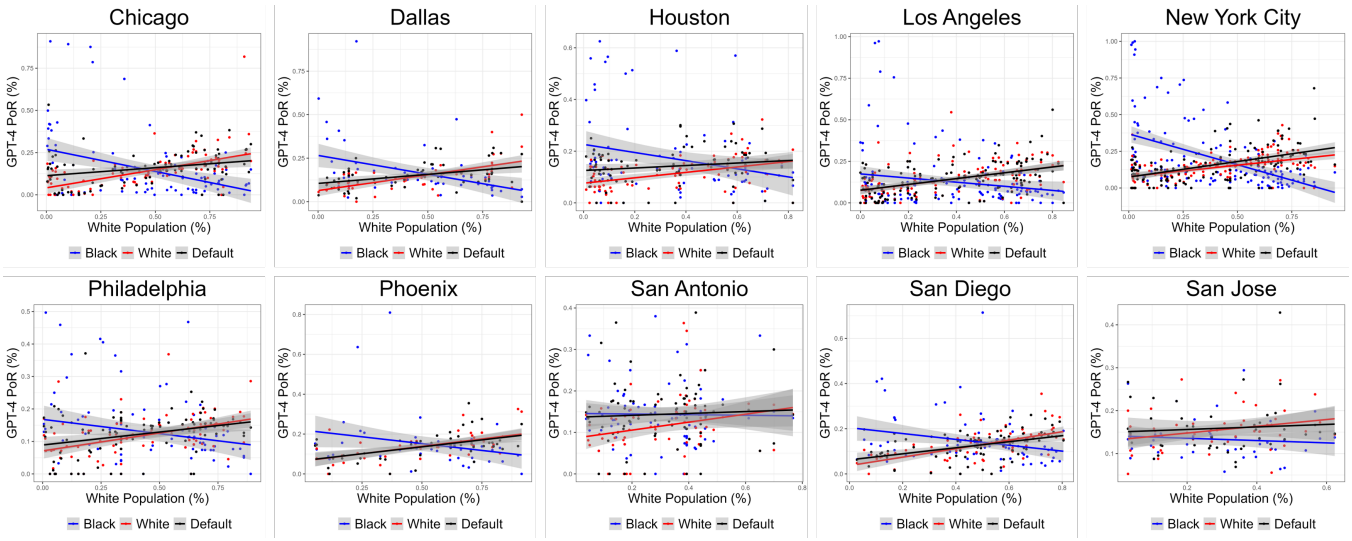

Figure 5: GPT-4 Probability-of-Recommendation Plotted against % White Population by Neighborhood, Smoothed Conditional Mean by GLM (95% CI).

by the system and evident in its outputs, as seen in how the plots for "white" and "default" track each other in Figs 3–8. Mandiberg [38], for example, demonstrates conclusively how biographies of white people on Wikipedia rarely specify the subject's race as white, whereas biographies for Black, Indigenous and people of color do specify their race. NLP researchers have framed this phenomenon as human reporting bias [40]. But, in Benjamin's conception, this is not an innocent oversight or "glitch" [6] when it comes to race but rather a form of baked-in "default discrimination" [6] in which "indifference to Blackness can be profitable" [6].

Default whiteness combined with "indifference to Blackness" is both pervasive in AI systems and also extremely harmful, particularly in high-stakes contexts such as facial recognition [10], self-driving cars [62], or skin cancer detection [35], among others. One recommendation that follows from our work, in combination with other AI audits, is to further elaborate and operationalize the

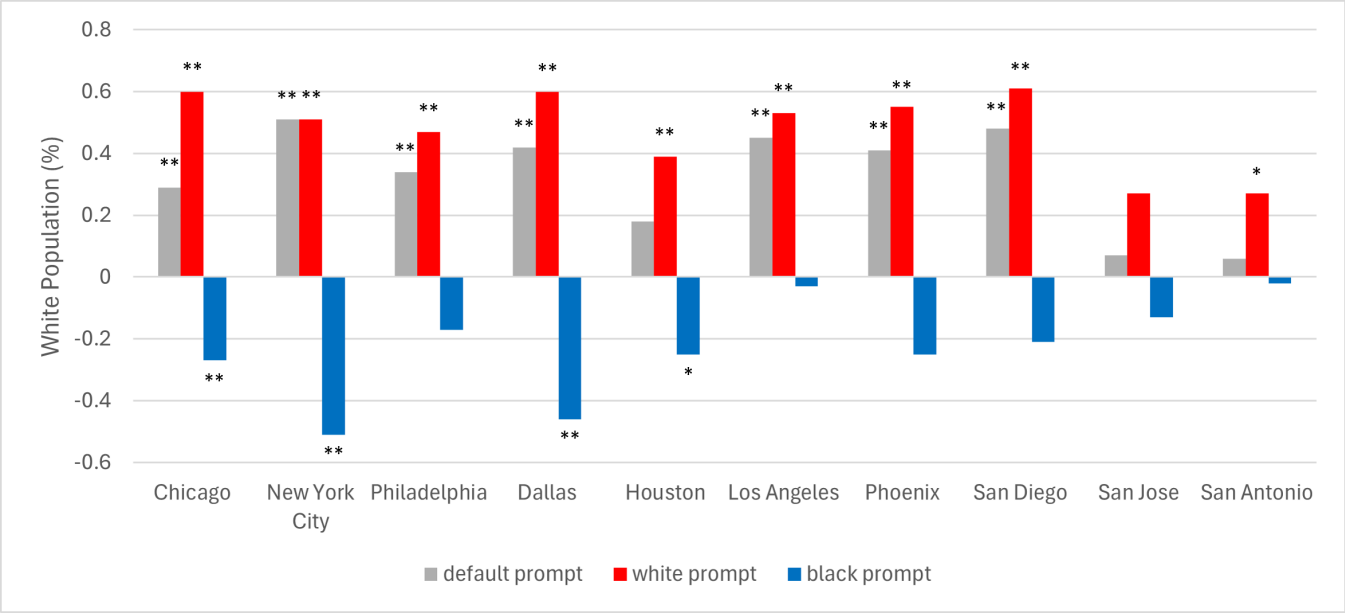

Figure 6: Spearman Correlation Coefficients between % White Population by Neighborhood and GPT-4 PoR across Cities Ordered by Segregation from High to Low [46].

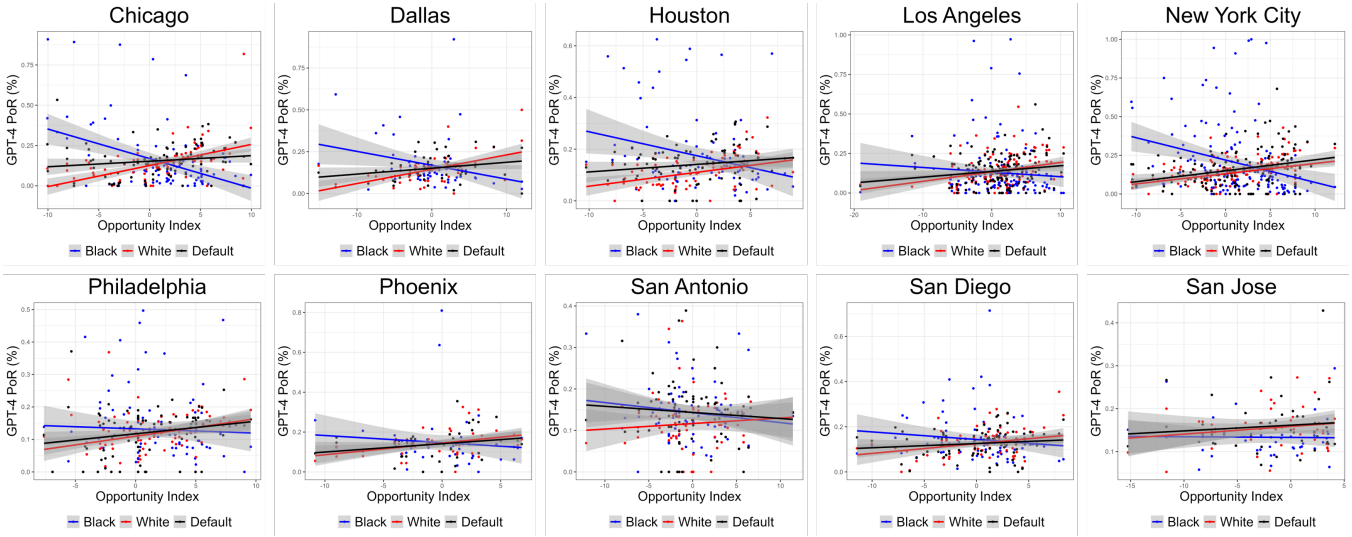

Figure 7: GPT-4 Probability-of-Recommendation Plotted against Opportunity Index by Neighborhood, Smoothed Conditional Mean by GLM (95% CI).

concept of default whiteness such that it could be more systematically interrogated across AI systems in many domains. With the development of systematic frameworks, we can prospectively anticipate and interrogate default whiteness and address it *before* systems are deployed in high-stakes situations with racially unjust, and potentially life-altering, outcomes.

4.2 Racial Steering and Liability under the Fair Housing Act

Our results additionally show that GPT-4 demonstrates racial steering based on racial information presented in the prompts. The recommendations for neighborhoods for prospective homeseekers correlate with the racial composition (and economic opportunities) of those neighborhoods. White homeseekers were steered towards

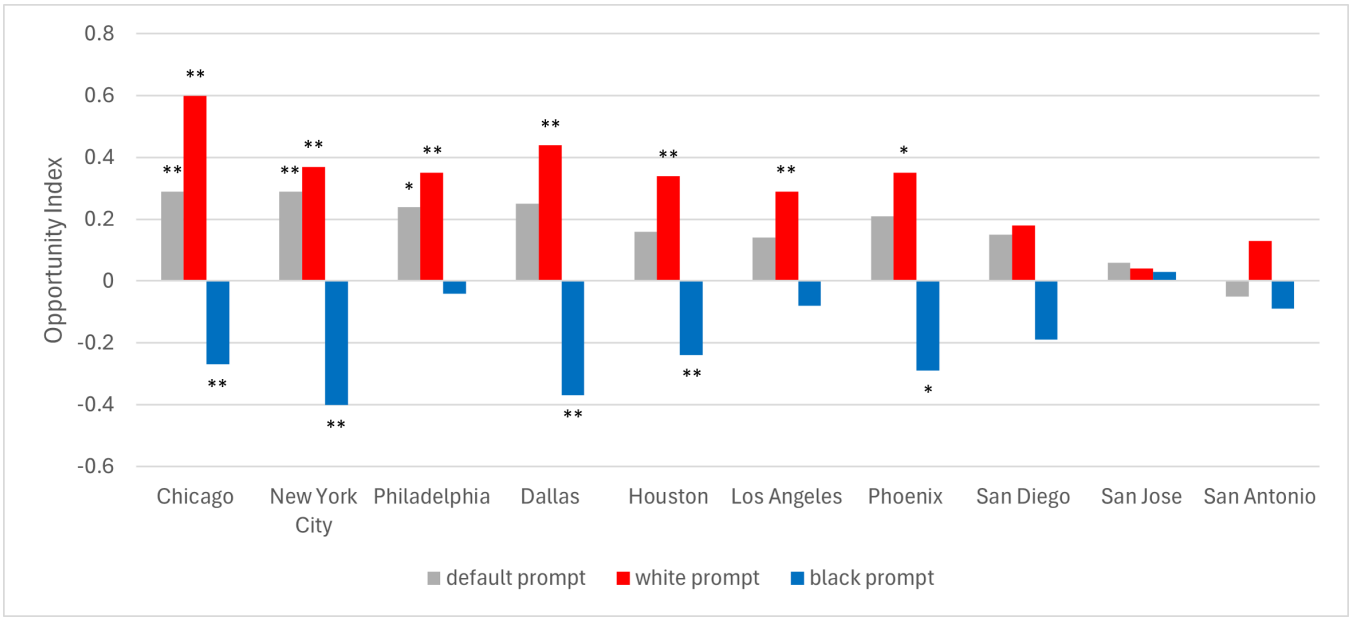

Figure 8: Spearman Correlation Coefficients between Neighborhood Opportunity Index and GPT-4 PoR across Cities Ordered by Segregation from High to Low [46].

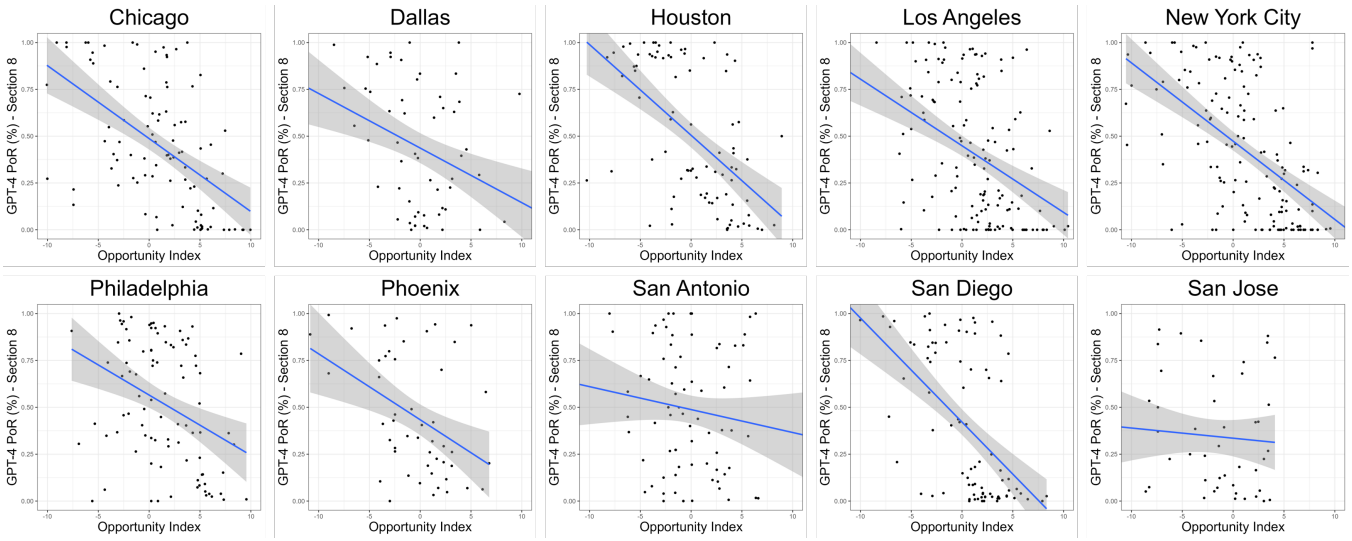

Figure 9: GPT-4 Probability-of-Recommendation for Section 8 Voucher Holders Plotted against Neighborhood Opportunity Index, Smoothed Conditional Mean by GLM (95% CI).

neighborhoods with a higher opportunity index, and Black homeseekers tended to be steered towards neighborhoods with a lower opportunity index. Homeseekers who identify their race are steered to neighborhoods with people racially like them and, conversely, away from neighborhoods with people not racially like them. While it is important to note that, at an individual scale, people may have solid reasons for desiring to live in neighborhoods with members of their cultures and communities, racial steering (both in real estate and in GPT-4) presumes those preferences based on an individual's identity without asking and then scales those presumptions beyond the individual to the structural scale.

Racial steering effects in GPT-4 are more pronounced in highly segregated cities like New York City and Chicago compared to San Antonio. Moreover, GPT-4 demonstrated significant socioeconomic steering – guiding holders of Section 8 vouchers to neighborhoods with lower opportunity across most cities. The implications of this are that, if such models were widely used by the public for housing recommendations, racial and socioeconomic steering by LLMs could

Table 1: Fixed-Effects Regression: Effect of Demographic Indicator Variables on the Average Opportunity Index of Recommended Neighborhoods.

Dependent Variable: Average Neighborhood Opportunity Index					
	Coeff.	Std. Err.		Coeff.	Std. Err.
Sexuality			*Other Demographic Characteristics*		
LGBT	-0.0616	(0.3078)	With a Family	0.8265^{***}	(0.1022)
Straight	0.2330^{***}	(0.0318)	Renter	-0.2670^{***}	(0.0543)
Race			Section 8 Voucher Holder	-1.236^{***}	(0.2940)
Hawaiian	-0.0411	(0.0356)	*Interactions: Gender × Family Status*		
Hispanic	-0.9172^{**}	(0.3125)	Man × Family	0.0897^{***}	(0.0172)
Native American	-0.3427^{**}	(0.1402)	Woman × Family	-0.0131	(0.0262)
Asian	-0.2244^{**}	(0.0865)	Nonconforming × Family	-0.4808^{**}	(0.1508)
Black	-0.6212^{***}	(0.1826)	*Interactions: Race × Source of Income*		
White	0.0851^{**}	(0.0283)	Hawaiian × Voucher	0.1969^{***}	(0.0596)
Gender			Hispanic × Voucher	0.2239	(0.1820)
Gender Nonconforming	0.0975	(0.2583)	Native × Voucher	0.2269^{*}	(0.1140)
Man	-0.0021	(0.0223)	Asian × Voucher	0.2146^{*}	(0.1161)
Woman	0.1856^{***}	(0.0357)	Black × Voucher	0.1950	(0.1181)
			White × Voucher	0.0869^{***}	(0.0212)
City Fixed-effects			Yes		
Observations			160,230		
Squared Correlation			0.45983		
Pseudo R^2			0.14121		
BIC			600,564.5		

Note: *Clustered (city) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

potentially exacerbate residential segregation in already segregated cities.

This brings us to questions around legal liability: Does ChatGPT violate the Fair Housing Act? Indeed, there is compelling evidence that it could. In the *Inclusive Communities* decision, the Supreme Court determined that a policy or program could be held liable under the Fair Housing Act for potentially perpetuating segregation by demonstrating disparate racial impact. If real estate platforms employ LLMs for their recommendation services, the argument could be made that the underlying models exhibit a "prima facie" disparate impact for groups of different races, genders and sexualities.

This would trigger various considerations related to the Fair Housing Act. First, it is important to consider that the algorithm development could involve two-fold layers - the first layer is the development company responsible for creating the foundational LLM, e.g., OpenAI developing GPT-4 and offering an API. The second layer could then involve real estate platforms integrating their own data to tailor the model for specific purposes. In particular, questions would arise about the liability of AI firms (e.g., OpenAI) compared to the real estate platforms (e.g., Zillow) that could adopt LLMs. Relatedly, in *Louis, et al. v. SafeRent Solutions, et al.*, the US District Court for the District of Massachusetts determined that a tenant screening algorithm falls under the Fair Housing Act [3]. The court ruled that, despite the tenant screening company not being a landlord, property owners who allegedly based their decisions solely on the company's determinations to reject rental applications effectively delegated housing decision-making authority to the company. Given that this case demonstrates the role of foundational models, our study's findings may indicate a potential basis for a successful disparate impact claim to AI firms themselves.

It is crucial to anticipate the mechanisms available to defendants during the burden-shifting process given the distinctive considerations posed by the case of LLMs in fair housing. We argue that, unlike the *HUD v. Facebook* scenario, assuming that a foundational LLM omits "race" features from its machine learning process is highly challenging. Facebook's ad algorithm may have the list of features and it may be easier to identify if it contains any protected categories. However, it would be nearly impossible to demonstrate that AI firms have entirely removed all words and underlying structures related to protected categories across trillions of features in vast text corpora [54]. Given that in *HUD v. Facebook*, a discriminatory pattern was observed and the algorithm effectively acted on the basis of race *even if the algorithm developers excluded to use the race feature*, it is plausible to think that the foundational LLMs could also learn the associations underlying the text and protected categories. Even if we assume that foundational LLMs have removed features related to protected categories or have developed ways to "de-bias" outputs, this study demonstrates that the LLM eventually "'learns the social effects that racial distinctions have in the world and leverages these correlations in making predictions" [31] that could contribute further segregation.

Lastly, it is alarming to observe how Section 8 voucher holders are steered towards lower-opportunity neighborhoods in the GPT-4 recommendations, a significant finding highlighted throughout the paper. On the one hand, there may be some human-level reasoning to comprehend the merit of such recommendations. Since vouchers are intended for use within the Fair Market Rent ceiling (otherwise, voucher holders must cover the residual amount), the recommendation to avoid very expensive neighborhoods could be justifiable. However, due to historical neighborhood effects that have perpetuated segregation and neighborhood-level disinvestment, these seemingly reasonable recommendations risk falling into a trap. Such suggestions may inadvertently contribute to the strengthening of segregation among Section 8 voucher holders. Among the ten cities we examined, only two cities in Texas (Houston and San Antonio) lack Source of Income (SOI) laws, with middle-sized cities in Southern states typically lacking these protections [49]. Therefore, fair housing organizations, if considering litigation, could emphasize the detrimental effects in states or cities with SOI laws. It is crucial to underscore the harms, especially in these specific regions, to address the potential consequences of the current Section 8 voucher recommendations.

In light of our findings, we would strongly recommend that the federal government institute a moratorium on the incorporation of ChatGPT and LLMs by rental and real estate firms into apps and systems for the purposes of housing recommendations while further technical evaluation and legal review for disparate impact is conducted. Based on the evidence we have provided, it is clear that housing recommendations from generative AI systems have the potential to enact significant harms in the form of individual/family-level effects – guiding a racialized person or family to lower opportunity neighborhoods and restricting their intergenerational life chances – as well as structural-level effects – exacerbating segregation, particularly in already segregated cities. Our evidence also points to the need for transparency around domain-specific biases in AI systems. It may or may not be possible to craft housing recommendations that are "unbiased" (doubtful in an area as deeply inequitable as housing in the US), but at the very least, the risks and harms of existing systems *in specific domains* must be studied, evaluated, and disclosed. The public deserves "robust, reliable, repeatable, and standardized evaluations of AI systems " [61] before such systems are irresponsibly deployed across sectors to exacerbate social stratification.

5 Conclusion

This study has examined how LLMs, and GPT-4 in particular, may perpetuate and exacerbate racial and socioeconomic discrimination in housing in the US context. By undertaking an audit study simulating users seeking housing in ten different cities across the US, we demonstrate that GPT-4 engages in racial steering – directing homeseekers to neighborhoods with people who are racially similar to them; and default whiteness – assuming that a homeseeker is white if their race is unspecified. Moreover, GPT-4 appears to consistently direct Black homeseekers towards neighborhoods with a lower opportunity index (encouraging downward mobility) and white homeseekers towards those with a higher opportunity index (encouraging upward mobility). Based on the evidence presented, we recommend that a federal moratorium be placed on the use of LLMs in rental and real estate applications offering housing recommendations until further legal and technical evaluations can be conducted.

Acknowledgments

The authors are grateful for funding from the MIT Generative AI White Paper Initiative, which underwrote parts of this work. We also want to acknowledge that this work is part of the Housing Research Group under the Initiative on Combatting Systemic Racism at the MIT Institute for Data, Systems and Society. We received valuable feedback and support from our colleagues in this group at multiple stages of development.

References

[1] 2015. Texas Dep't of Hous. & Cmty. Affairs v. Inclusive Communities Project, Inc.
[2] 2019. HUD v. Facebook Inc. https://www.hud.gov/sites/dfiles/Main/documents/HUD_v_Facebook.pdf
[3] 2022. Louis et al. v. SafeRent Solutions, LLC et al.
[4] Abubakar Abid, Maheen Farooqi, and James Zou. 2021. Persistent Anti-Muslim Bias in Large Language Models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*. Association for Computing Machinery, New York, NY, USA, 298–306. https://doi.org/10.1145/3461702.3462624
[5] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922
[6] Ruha Benjamin. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code* (1st edition ed.). Polity, Medford, MA.
[7] Laurent Berge. 2018. Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm. *CREA Discussion Papers* 13 (2018).
[8] Eduardo Bonilla-Silva. 2018. *Racism without racists: color-blind racism and the persistence of racial inequality in America* (fifth edition ed.). Rowman & Littlefield, Lanham.
[9] Kelly M. Bower, Roland J. Thorpe, Charles Rohde, and Darrell J. Gaskin. 2014. The intersection of neighborhood racial segregation, poverty, and urbanicity and its impact on food store availability in the United States. *Preventive Medicine* 58 (Jan. 2014), 33–39. https://doi.org/10.1016/j.ypmed.2013.10.010
[10] Joy Buolamwini. 2023. *Unmasking AI* (first edition ed.). Random House, New York.
[11] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356, 6334 (April 2017), 183–186. https://doi.org/10.1126/science.aal4230
[12] Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *. *The Quarterly Journal of Economics* 129, 4 (Nov. 2014), 1553–1623. https://doi.org/10.1093/qje/qju022
[13] Thomas Davidson, Debasmita Bhattacharya, and Ingmar Weber. 2019. Racial Bias in Hate Speech and Abusive Language Detection Datasets. In *Proceedings of the Third Workshop on Abusive Language Online*, Sarah T. Roberts, Joel Tetreault, Vinodkumar Prabhakaran, and Zeerak Waseem (Eds.). Association for Computational Linguistics, Florence, Italy, 25–35. https://doi.org/10.18653/v1/W19-3504
[14] Matthew Desmond. 2012. Eviction and the Reproduction of Urban Poverty. *Amer. J. Sociology* 118, 1 (July 2012), 88–133. https://doi.org/10.1086/666082
[15] Matthew Desmond. 2016. *Evicted: poverty and profit in the American city* (first edition ed.). Crown Publishers, New York.
[16] St Clair Drake and Horace R. Cayton. 2015. *Black Metropolis: A Study of Negro Life in a Northern City.* University of Chicago Press, Chicago, IL. https://press.uchicago.edu/ucp/books/book/chicago/B/bo20832325.html
[17] Ingrid Gould Ellen. 2020. What do we know about housing choice vouchers? *Regional Science and Urban Economics* 80 (Jan. 2020), 103380. https://doi.org/10.1016/j.regsciurbeco.2018.07.003
[18] Ingrid Gould Ellen, Katherine O'Regan, and Sarah Strochak. 2024. Race, Space, and Take Up: Explaining housing voucher lease-up rates. *Journal of Housing Economics* 63 (March 2024), 101980. https://www.sciencedirect.com/science/article/pii/S1051137723000670
[19] Ingrid Gould Ellen, Katherine M. O'Regan, and Katharine W. H. Harwood. 2022. Advancing Choice in the Housing Choice Voucher Program: Source of Income Protections and Locational Outcomes. *Housing Policy Debate* 0, 0 (Aug. 2022), 1–22. https://doi.org/10.1080/10511482.2022.2089196 Publisher: Routledge _eprint:

https://doi.org/10.1080/10511482.2022.2089196.
[20] Michael Ewens, Bryan Tomlin, and Liang Choon Wang. 2014. Statistical Discrimination or Prejudice? A Large Sample Field Experiment. *The Review of Economics and Statistics* 96, 1 (March 2014), 119–134. https://doi.org/10.1162/REST_a_00365
[21] Jacob William Faber. 2018. Segregation and the Geography of Creditworthiness: Racial Inequality in a Recovered Mortgage Market. *Housing Policy Debate* 28, 2 (March 2018), 215–247. https://doi.org/10.1080/10511482.2017.1341944
[22] Jose Figueroa. 2019. Segregated Health Systems. In *Dream Revisited: Contemporary Debates about Housing, Segregation, and Opportunity*. Columbia University Press, New York.
[23] John Foster. 2015. *White race discourse: preserving racial privilege in a post-racial society*. Lexington Books, Place of publication not identified. OCLC: 898910766.
[24] Lance Freeman. 2012. The impact of source of income laws on voucher utilization. *Housing Policy Debate* 22, 2 (March 2012), 297–318. https://doi.org/10.1080/10511482.2011.648210
[25] S. Michael Gaddis and Nicholas V. DiRago. 2023. Audit Studies of Housing Discrimination: Established, Emerging, and Future Research. In *The Sociology of Housing*. University of Chicago Press, 93–106. https://doi.org/10.7208/chicago/9780226828527-008
[26] George Galster and Peter Constantine. 1991. Discrimination Against Female-Headed Households in Rental Housing: Theory and Exploratory Evidence. *Review of Social Economy* (March 1991). https://doi.org/10.1080/00346769100000005 Publisher: Catholic Economic Association.
[27] Philip M. E. Garboden, Eva Rosen, Stefanie DeLuca, and Kathryn Edin. 2018. Taking Stock: What Drives Landlord Participation in the Housing Choice Voucher Program. *Housing Policy Debate* 28, 6 (Nov. 2018), 979–1003. https://doi.org/10.1080/10511482.2018.1502202
[28] Edward G. Goetz, Rashad A. Williams, and Anthony Damiano. 2020. Whiteness and Urban Planning. *Journal of the American Planning Association* 86, 2 (April 2020), 142–156. https://doi.org/10.1080/01944363.2019.1693907 Publisher: Routledge _eprint: https://doi.org/10.1080/01944363.2019.1693907.
[29] Matthew Hall, Jeffrey M. Timberlake, and Elaina Johns-Wolfe. 2023. Racial Steering in U.S. Housing Markets: When, Where, and to Whom Does It Occur? *Socius* 9 (Jan. 2023), 23780231231197024. https://doi.org/10.1177/23780231231197024 Publisher: SAGE Publications.
[30] Forrest Hangen and Daniel T. O'Brien. 2023. The Choice to Discriminate: How Source of Income Discrimination Constrains Opportunity for Housing Choice Voucher Holders. *Urban Affairs Review* 59, 5 (Sept. 2023), 1601–1625. https://doi.org/10.1177/10780874221109591 Publisher: SAGE Publications Inc.
[31] Lily Hu. 2023. What is "Race" in Algorithmic Discrimination on the Basis of Race? *Journal of Moral Philosophy* (Sept. 2023), 1–26. https://doi.org/10.1163/17455243-20234369
[32] Roman Jakobson. 1972. Verbal Communication. *Scientific American* 227, 3 (1972), 72–81. https://www.jstor.org/stable/24927429 Publisher: Scientific American, a division of Nature America, Inc..
[33] Ian Kennedy, Chris Hess, Amandalynne Paullada, and Sarah Chasins. 2021. Racialized Discourse in Seattle Rental Ad Texts. *Social Forces* 99, 4 (June 2021), 1432–1456. https://doi.org/10.1093/sf/soaa075
[34] Svetlana Kiritchenko and Saif Mohammad. 2018. Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, Malvina Nissim, Jonathan Berant, and Alessandro Lenci (Eds.). Association for Computational Linguistics, New Orleans, Louisiana, 43–53. https://doi.org/10.18653/v1/S18-2005
[35] Giona Kleinberg, Michael J Diaz, Sai Batchu, and Brandon Lucke-Wold. 2022. Racial underrepresentation in dermatological datasets leads to biased machine learning models and inequitable healthcare. *Journal of biomed research* 3, 1 (2022), 42–47. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9815490/
[36] Karyn Lacy. 2007. *Blue-Chip Black: Race, Class, and Status in the New Black Middle Class*.
[37] Annette Lareau. 2011. *Unequal Childhoods: Class, Race, and Family Life, With an Update a Decade Later* (2 ed.).
[38] Michael Mandiberg. 2023. Wikipedia's Race and Ethnicity Gap and the Unverifiability of Whiteness. *Social Text* 41, 1 (154) (March 2023), 21–46. https://doi.org/10.1215/01642472-10174954
[39] Douglas S. Massey and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Harvard Univ. Press, Cambridge, Mass.
[40] Ishan Misra, C. Lawrence Zitnick, Margaret Mitchell, and Ross Girshick. 2016. Seeing Through the Human Reporting Bias: Visual Classifiers From Noisy Human-Centric Labels. 2930–2939. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Misra_Seeing_Through_the_CVPR_2016_paper.html
[41] Sun Jung Oh and John Yinger. 2015. What Have We Learned From Paired Testing in Housing Markets? *Cityscape* 17, 3 (2015), 15–60. https://www.jstor.org/stable/26326960 Publisher: US Department of Housing and Urban Development.
[42] Michael Omi and Howard Winant. 2015. *Racial Formation in the United States* (third edition ed.). Routledge, New York.
[43] Online Marketplaces. 2023. OpenAI Deactivates All Real Estate Plugins — Redfin And Zillow Affected |. https://perma.cc/R5HX-WHK8 Section: Property.
[44] OpenAI. 2023. ChatGPT plugins. https://openai.com/blog/chatgpt-plugins
[45] OpenAI. 2023. GPT-4 Technical Report. https://doi.org/10.48550/arXiv.2303.08774 arXiv:2303.08774 [cs].
[46] Othering and Belonging Institute. 2019. Most to Least Segregated Cities. https://belonging.berkeley.edu/most-least-segregated-cities
[47] Katherine M. O'Regan. 2019. The Fair Housing Act Today: Current Context and Challenges at 50. *Housing Policy Debate* 29, 5 (Sept. 2019), 704–713. https://doi.org/10.1080/10511482.2018.1519907
[48] Mary Pattillo. 2013. *Black Picket Fences, Second Edition: Privilege and Peril among the Black Middle Class*. University of Chicago Press, Chicago, IL. https://press.uchicago.edu/ucp/books/book/chicago/B/bo15340705.html
[49] Poverty and Race, Research and Action Council. 2019. Expanding choice: Practical strategies for building a successful housing mobility program Appendix B: State, local, and federal laws barring source-of-income discrimination. https://www.prrac.org/pdf/AppendixB.pdf
[50] Redfin. 2023. Redfin Launches ChatGPT Plugin to Help People Find Their Next Home. https://www.redfin.com/news/redfin-chatgpt-plugin/
[51] Eva Rosen. 2020. *The Voucher Promise: "Section 8" and the Fate of an American Neighborhood*. Princeton University Press. https://doi.org/10.1515/9780691189505
[52] Abel Salinas, Parth Shah, Yuzhong Huang, Robert McCormack, and Fred Morstatter. 2023. The Unequal Opportunities of Large Language Models: Examining Demographic Biases in Job Recommendations by ChatGPT and LLaMA. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3617694.3623257
[53] Robert J. Sampson. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. University of Chicago Press, Chicago, IL.
[54] Maximilian Schreiner. 2023. GPT-4 architecture, datasets, costs and more leaked. https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/
[55] Patrick Sharkey. 2013. *Stuck in Place: Urban Neighborhoods and the End of Progress toward Racial Equality*. University of Chicago Press, Chicago, IL. https://press.uchicago.edu/ucp/books/book/chicago/S/bo14365260.html
[56] Mario Luis Small. 2008. Four Reasons to Abandon the Idea of "The Ghetto". *City & Community* 7, 4 (Dec. 2008), 389–398. https://doi.org/10.1111/j.1540-6040.2008.00271_8.x Publisher: SAGE Publications.
[57] Wonyoung So. 2023. Which Information Matters? Measuring Landlord Assessment of Tenant Screening Reports. *Housing Policy Debate* 33, 6 (Nov. 2023), 1484–1510. https://doi.org/10.1080/10511482.2022.2113815 Publisher: Routledge _eprint: https://doi.org/10.1080/10511482.2022.2113815.
[58] Justin P. Steil, Len Albright, Jacob S. Rugh, and Douglas S. Massey. 2018. The social structure of mortgage discrimination. *Housing Studies* 33, 5 (July 2018), 759–776.
[59] Justin P. Steil, Camille Z. Charles, and Marc Morial. 2021. Sociology, Segregation, and the Fair Housing Act. In *Perspectives on Fair Housing*, Vincent J. Reina, Wendell E. Pritchett, and Susan M. Wachter (Eds.). University of Pennsylvania Press, 45–73. https://www.jstor.org/stable/j.ctv16qjz07.6
[60] Margery Austin Turner, Robert Santos, Diane K. Levy, Douglas A. Wissoker, Claudia Aranda, and Rob Pitingolo. 2016. Housing Discrimination against Racial and Ethnic Minorities 2012: Executive Summary. (June 2016). https://policycommons.net/artifacts/633118/housing-discrimination-against-racial-and-ethnic-minorities-2012/1614410/ Publisher: Urban Institute.
[61] White House. 2023. The White House Blueprint for a Renters Bill of Rights. https://www.whitehouse.gov/wp-content/uploads/2023/01/White-House-Blueprint-for-a-Renters-Bill-of-Rights.pdf
[62] Benjamin Wilson, Judy Hoffman, and Jamie Morgenstern. 2019. Predictive Inequity in Object Detection. https://doi.org/10.48550/arXiv.1902.11097 arXiv:1902.11097 [cs, stat].
[63] Chi Chi Wu. 2020. Reparations, Race, and Reputation in Credit: Rethinking the Relationship Between Credit Scores and.... https://medium.com/@cwu_84767/reparations-race-and-reputation-in-credit-rethinking-the-relationship-between-credit-scores-and-852f70149877
[64] John Yinger. 1995. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*. Russell Sage Foundation. Google-Books-ID: fwqGAwAAQBAJ.
[65] Zillow. 2023. Chat, discover and find your next home with Zillow's plugin on ChatGPT. https://perma.cc/ZQL6-MGRY

A Ethical considerations

A.1 Ethical Considerations Statement

Since the presented research involved prompting an AI model and analyzing its output, we do not see immediate ethical concerns that required mitigation. As a team made up of members with diverse intersectional experiences, we are deeply interested in characterizing the ways that emerging technologies may impact the lived experiences of marginalized communities, which necessitates taking a proactive stance on evaluating potential biases in AI models.

A.2 Adverse Impact Statement

We outline some specific ways in which adjusting the prompting strategy as input to LLMs may lead to significant differences in actionable recommendations with potential social consequences. This knowledge, when taken together with the irresponsible or inappropriate integration of LLMs into opaque social algorithms, could potentially be engineered to exacerbate social disparities through the inclusion or exclusion of key language in input prompts fed to these models. However, we believe that elucidating and understanding the specific biases baked into LLMs provide a societal benefit outweighing the described potential adverse or unintended impacts.