Taxpayer Behavior in the Age of AI

A Field Experiment on Property Tax Appeals

Justin Holz
University of Michigan

Ricardo Perez-Truglia
University of California, Los Angeles

Andrew Simon
University of Virginia

Alejandro Zentner
University of Texas at Dallas

Abstract

Navigating the tax system often requires information and judgment that taxpayers may lack. Although taxpayers sometimes rely on human experts, AI could provide comparable assistance at substantially lower cost. Yet there is no causal evidence on how taxpayers use AI tax agents or how these tools affect their behavior. We fill this gap with a field experiment in the context of property tax appeals. We recruited a sample of 645 households in Dallas County, Texas. All households received access to a website providing personalized information, filing instructions, and supporting evidence for an appeal. Half were randomly assigned to a version of the website that also included an AI chatbot capable of answering questions and providing guidance tailored to their circumstances. Chatbot take-up was high: 78% initiated a conversation. Access to the chatbot increased the probability of filing an appeal without a human agent by 9.1 pp, from 41.4% to 50.5%. Evidence from clickstream data and conversation transcripts suggests that the chatbot helped households exercise judgment—that is, evaluate and act on the available information. The increase in appeal filing was smaller among less-advantaged households, providing suggestive evidence that the chatbot widened existing disparities. Our results demonstrate the potential of AI tax assistance while highlighting that who benefits depends not only on access, but also on how taxpayers use the technology.

JEL Classification: C93, H31, H71, O33.
Keywords: artificial intelligence, tax agent, property taxes, field experiment.

This Draft: August 18, 2026. Holz: holzj@umich.edu. Perez-Truglia: ricardo.truglia@anderson.ucla.edu. Simon: arsimon@virginia.edu. Zentner: azentner@utdallas.edu. This project was reviewed and approved in advance by the Institutional Review Board at the University of Michigan. The field experiment was preregistered in the AEA RCT Registry (AEARCTR-0018430). Upon publication, we will make the replication code and data publicly available through a public repository.

1 Introduction

Interactions between taxpayers and the tax administration help determine how much individuals ultimately pay in taxes. The tax code may specify the deductions, credits, and other benefits to which a taxpayer is entitled, but receiving them often requires the taxpayer to take action—gathering records, learning the relevant rules, completing forms, and sometimes challenging a decision by the tax authority. These steps take time and effort and may also require substantial information and judgment (Benzarti, 2020). As a result, taxpayers may pay more than they legally need to because they fail to claim benefits or pursue remedies available to them (Bhargava and Manoli, 2015). Moreover, because the ability to navigate the tax system varies across individuals, these administrative demands may affect not only average tax burdens but also how those burdens are distributed.

One way taxpayers deal with these challenges is by turning to experts. For example, tax preparers help individuals file their income tax returns and identify deductions and credits, while property tax agents help households challenge their property valuations. These experts do more than complete forms: they translate tax rules into personalized advice and can shape how taxpayers respond to tax incentives and enforcement (Chetty and Saez, 2013; Battaglini et al., 2026). However, access to this assistance can be costly and uneven (Doerner and Ihlanfeldt, 2015). Until recently, these experts have been human.[1] Recent advances in artificial intelligence, however, make it possible for AI systems to provide personalized tax assistance at a much larger scale and at a fraction of the cost.

Purpose-built AI tax agents are still uncommon, but general-purpose chatbots are already being used for tax assistance. For example, in a supplemental survey we conducted with a separate sample of Texas homeowners, 40.0% of those who filed a property tax appeal on their own reported using an AI tool such as ChatGPT for assistance. As these tools become more widely available, policymakers will need evidence not only on who adopts them, but also on how different taxpayers interact with them and how they affect behavior—both on average and across groups. This evidence will be important for anticipating their broader consequences and guiding policy decisions. For example, tax administrations may need to decide whether to offer their own AI agents for free. Policymakers will also need to determine how AI tax agents should be regulated, including who bears responsibility when an agent provides incorrect advice and how taxpayers' privacy should be protected. These decisions

[1] An important intermediate case is conventional tax software, such as TurboTax, which uses pre-specified algorithms to guide taxpayers through common tax situations. For example, an algorithm can calculate the taxpayer's liability using the standard deduction and using itemized deductions, compare the two amounts, and select the option that minimizes the tax due. These tools can be highly effective when the taxpayer's situation falls within the scenarios anticipated by the software's designers, but they are less flexible when taxpayers face unusual circumstances or need assistance with decisions that require judgment.

may ultimately affect who has access to and uses AI tax agents.

We study these questions in the context of property tax appeals in the United States.[2] Households receive a notice from their local appraisal district stating the proposed market value of their home. Households who believe that this value is too high can challenge it—an action Texas law calls a protest and that we refer to as an appeal—either on their own or through a human agent. A successful appeal lowers the ultimately certified value and, consequently, may reduce the property taxes owed, often saving a household hundreds of dollars per year (Nathan et al., 2025). Filing is free and requires completing a short form, either on paper or online. For a household familiar with the process, these mechanical steps can take only a few minutes; for others, the process can still be daunting if they lack relevant knowledge or confidence.

Property tax appeals provide a useful setting for studying AI tax agents for two reasons. First, filing an appeal involves decisions that can benefit from personalized guidance. Households must decide whether an appeal is worthwhile and identify evidence that supports their case. For example, a household whose property has a proposed value of $500,000 may need to determine whether a nearby home that recently sold for less is sufficiently comparable to challenge that value. Because the strength of such evidence depends on the home's characteristics and the information available in each case, property tax appeals offer scope for personalized AI guidance beyond static instructions.

The second reason property tax appeals provide a useful setting is that access to the appeals process, as in many other tax settings, is unequal (Avenancio-León and Howard, 2022; Nathan et al., 2025; Holz et al., 2026). Lower-income homeowners are substantially less likely to file an appeal. Even holding home values constant, Hispanic and Black homeowners appeal at lower rates than White homeowners, while less-educated homeowners appeal at lower rates than more-educated homeowners. These disparities are often attributed, at least in part, to differences in knowledge of the tax system, confidence in navigating the appeals process, and the ability to afford a human agent (Doerner and Ihlanfeldt, 2015). These disparities have led some observers to argue that the "property tax system is rigged against [...] little people" (Lieber, 2020). This setting therefore allows us to study both the average and distributional effects of AI tax assistance. On the one hand, disadvantaged households may have more to gain from an AI agent because it can provide personalized assistance at little or no cost. On the other hand, these households may be less likely to use the agent or to interact with it effectively. Whether AI assistance narrows or widens existing disparities is therefore an empirical question.

[2] Examples of studies on property tax appeals include Avenancio-León and Howard (2022) and Nathan et al. (2025). A broader literature examines property taxation more generally; see, for example, Cabral and Hoxby (2012).

We conducted the field experiment in Dallas County, Texas, because implementing it within a single administrative jurisdiction offered important practical advantages. Dallas County is the second most populous county in Texas, with an estimated population of 2.66 million in 2025 (U.S. Census Bureau, 2026a)—more than the populations of 15 U.S. states (U.S. Census Bureau, 2026b). The county is demographically and politically diverse. Because Texas does not levy an individual state income tax, property taxes are an especially important source of funding for local public services. In 2026, the average Dallas County household was expected to pay approximately $7,900 in property taxes.

We created a website that provided households with personalized property information, step-by-step filing instructions, and supporting evidence for an appeal. In the preregistered experiment, we randomly assigned half of the households to an otherwise identical version of the website that also included an AI chatbot designed to answer questions and provide guidance tailored to their circumstances. Using administrative records, we observed whether each household subsequently filed an appeal. To investigate the mechanisms underlying any effects, we also recorded complete conversation transcripts and tracked households' behavior on the website.

DCAD released its 2026 proposed values on April 14, 2026, and households had until May 15 to file an appeal. Approximately two weeks before the deadline, we mailed postcards to 45,200 households inviting them to use the website. The postcards neither disclosed each household's treatment assignment nor mentioned that the site had two versions. Our analysis sample includes 645 households that visited the website before the deadline.

Engagement with the website was high among households assigned to No-Chatbot, suggesting that households used the available information while considering whether and how to appeal. Chatbot take-up was also high: 78% of households assigned to Chatbot initiated a conversation, often seeking guidance about whether to appeal and how to support their case. Assignment to Chatbot increased the probability of filing a direct appeal from 41.4% in No-Chatbot to 50.5% in Chatbot, a raw difference of 9.1 pp ($p = 0.021$). With a rich set of controls, the estimated effect is similar in magnitude at 10.5 pp ($p = 0.006$). This finding also passes a falsification test based on pretreatment outcomes: when we define the outcome using appeals filed in previous years, we find no effect. Assignment to Chatbot did not crowd out the use of human agents: the probability of filing through a human agent fell by a statistically insignificant 0.2 pp, from 9.6% to 9.4%.

The chatbot effect is not only statistically significant, but also economically meaningful. We benchmark the magnitude of this effect using two prior estimates from the same setting. First, based on estimates from Nathan et al. (2025), the chatbot increases appeal filing as much as an increase of $425 in expected tax savings. Second, in a related field experiment,

Nathan et al. (2025) find that mailing households a step-by-step filing guide and a tailored argument based on a nearby comparable home increased the appeal rate by 4.98 pp. Our estimated effect is nearly twice as large. Furthermore, the existing use of general-purpose AI may attenuate our estimated effect.

We explore the causal mechanisms—that is, *why* the chatbot was effective—using clickstream data and conversation transcripts. Assignment to Chatbot increased the average time spent on the website from 12.1 minutes to 17.3 minutes ($p = 0.022$). At the same time, it changed how households used individual website features: it increased the share creating an uploadable evidence document with the PDF Creator but decreased downloads of the pre-generated Comparable Properties Report and use of the step-by-step walkthrough, although the latter two features remained widely used. The conversation transcripts show that households often sought guidance in interpreting the available information, such as determining which comparable homes in the Comparable Properties Report provided the strongest support for their appeal. Taken together, these patterns suggest that the chatbot helped households exercise judgment by navigating the available information, identifying the evidence most relevant to their case, and acting on it.

To investigate whether the chatbot's presentation affected take-up, we conducted a secondary randomization within Chatbot between TaxBot and Marina. TaxBot had a machine-like name and a less personal tone, whereas Marina had a human name and a more personal tone. A more personal presentation might feel more approachable, while a machine-like presentation might make households more comfortable sharing information. Households assigned to TaxBot were a bit more likely to initiate a conversation and file an appeal than those assigned to Marina, but these differences are modest in magnitude and borderline statistically significant or insignificant.

Next, we examine whether the chatbot narrowed or widened existing disparities in the appeal system. The estimated effects on filing rates are smaller among less-educated homeowners, homeowners with lower-valued properties, and homeowners from racial or ethnic minority groups. The subgroup differences are large in magnitude but imprecisely estimated and therefore not always statistically significant. However, when taken together, their consistency across all three dimensions points to a common pattern. Chatbot take-up was also somewhat lower among the same less-advantaged groups, but these differences were considerably smaller than the corresponding differences in filing effects. Thus, take-up cannot fully explain the pattern; much of it appears to arise after take-up, including from differences in how households use the chatbot and translate its guidance into action. Together, these results provide suggestive evidence that universal access to the chatbot may, far from mitigating existing inequities, exacerbate them.

To assess whether these findings were predictable ex ante, we surveyed 87 experts, describing the experiment and asking them to forecast its results. Experts expressed substantial uncertainty and disagreement about the chatbot's effects, which is natural given that this was the first field experiment to provide taxpayers with AI assistance. Experts underestimated chatbot take-up and did not anticipate either the heterogeneity in treatment effects or the relative performance of TaxBot and Marina. However, despite the lack of consensus among experts, their average forecast of the chatbot's effect on filing rates was close to the experimental estimate.

We next discuss how our findings may generalize beyond our study sample. First, our recruitment strategy attracted households who were already considering an appeal. The average household in our sample is not far from the average household in the experimental universe in terms of economic and demographic characteristics, but the sample overrepresents households with a higher propensity to appeal. The effects of the chatbot may therefore differ in the general population. Second, we conducted the experiment in a single county, but property tax appeals operate under the same basic legal framework and follow similar procedures across Texas counties (Nathan et al., 2025). Our findings are also likely to generalize beyond Texas to contexts in which property tax appeals operate similarly, as they do in many other U.S. states and countries (Dobay et al., 2019; World Bank, 2019), although there are exceptions.[3]

Our evidence comes from property tax appeals, an economically important setting in its own right. U.S. property taxes generated an estimated $577 billion in 2019 (Tax Policy Center, 2021), nearly three times as much revenue as the corporate income tax. A broader question is whether our findings extend beyond property tax appeals to other interactions with the tax system. AI tax agents may eventually assist taxpayers in many other settings, perhaps most consequentially with annual income tax filing. Indeed, a survey reported in March 2026 suggests that 26% of people used AI tools, including general-purpose chatbots such as ChatGPT, to obtain help preparing their 2025 tax returns, up from 11% the previous year (Cerullo, 2026). In that setting, taxpayers' circumstances are more varied, and mistakes can carry substantial financial and legal consequences. Those same features, however, create significant technical, legal, and regulatory obstacles to developing and experimentally evaluating a fully functional AI agent.[4] Indeed, at the time of our experiment, major tax-preparation platforms such as TurboTax and H&R Block offered chatbots in some paid tiers,

[3] In some settings, institutional differences may affect households' incentives to appeal and therefore limit extrapolation. In California, for example, Proposition 13 limits increases in assessed values between property sales, often leaving assessments below market values and households with less reason to appeal.

[4] Bock et al. (2025) evaluate frontier models available in 2025 on 51 simplified federal income tax returns. The best-performing model correctly calculated fewer than one-third of the returns. Because AI capabilities are improving rapidly, more recent models may perform better.

but these tools had limited functionality and primarily served as navigation and question-answering assistants. Property tax appeals therefore offer a feasible first setting in which to study AI tax assistance, providing initial evidence that can inform future applications in other tax settings.

Our study relates to and contributes to three strands of literature. First, we provide causal evidence on how access to an AI tax agent affects taxpayer behavior. To the best of our knowledge, ours is the first study to examine this question using a field experiment and administrative data on actual filing decisions.

Second, our study contributes to several bodies of literature on the barriers individuals face when interacting with the government. In addition to the financial burden of taxation, property tax appeals can generate learning and compliance costs because households must decide whether appealing is worthwhile, obtain evidence that fit their case, and prepare an acceptable claim; uncertainty and low confidence may also create psychological costs(Moynihan et al., 2015; Benzarti, 2020; Nathan et al., 2025; Ray et al., 2023). Our findings suggest that AI tax agents can reduce these costs by making filing easier. Complexity and limited information create additional filing frictions that discourage eligible individuals from filing or claiming benefits. Prior work shows that information and assistance can mitigate these frictions in contexts including property tax appeals, tax credits, and food assistance (Nathan et al., 2025; Bhargava and Manoli, 2015; Finkelstein and Notowidigdo, 2019).[5]

We show that AI can reduce filing frictions beyond standard information interventions. Related work also finds that taxpayers may respond ineffectively to tax incentives when they do not understand the relevant rules (Kleven and Kopczuk, 2011; Hoopes et al., 2015; Abeler and Jäger, 2015; Feldman et al., 2016). We show that AI can help taxpayers navigate this complexity through tailored guidance. Finally, we extend studies of human tax professionals and their effects on taxpayer behavior (Chetty and Saez, 2013; Battaglini et al., 2026; DeBacker et al., 2024) by examining AI tax agents.

Third, our study relates to a broader literature on the effects of AI tools. Existing studies show that access to AI tools can increase the productivity of customer-support agents (Brynjolfsson et al., 2025), software developers (Cui et al., 2026), management consultants (Dell'Acqua et al., 2026), and professionals performing occupation-specific writing tasks (Noy and Zhang, 2023).[6] Our setting differs from this existing work. Whereas prior studies focus primarily on productivity, in the tax context the relevant outcomes are hassle costs and tax burdens. Unlike studies finding that AI tools narrow preexisting productivity gaps by

[5] Related work studies pre-populated tax returns, which reduce the information taxpayers must enter and can substantially affect filing costs and tax compliance (Benzarti, 2021; Saez and Seim, 2025).

[6] There are exceptions, however. Otis et al. (2024) find no statistically significant average effect of access to an AI business adviser on revenues or profits among entrepreneurs in Kenya.

disproportionately benefiting less-skilled, less-experienced, or less-educated workers (Noy and Zhang, 2023; Brynjolfsson et al., 2025; Cui et al., 2026; Cruces et al., 2026), our findings suggest that AI assistance may widen existing disparities in property tax appeals.

2 Institutional Context

2.1 Overview of Property Taxes and Appeals

Local property taxes in Dallas County fund public services provided by the county and by overlapping jurisdictions, including cities, school districts, a community college, a hospital district, and special districts. The Dallas Central Appraisal District (DCAD) is responsible for appraising taxable property, while the relevant jurisdictions set their own tax rates. Each year, DCAD estimates the market value of every property as of January 1. This estimate is known as the property's *proposed value*. Households have the opportunity to appeal that value. After a protest is filed, DCAD reviews the household's requested value and supporting evidence. The district may offer an informal settlement that reduces the proposed value, typically via email. If the household does not accept the offer, or if no offer is made, the case can proceed to a hearing before the Appraisal Review Board. Once protests are resolved, the resulting value becomes the property's certified value, which is used together with applicable tax rates and exemptions to determine the tax liability. As a result, lowering the proposed value through an appeal can generate tax savings in the current and future years. However, in some cases—for example, when the household has a homestead exemption and the homestead cap is binding—the savings may not materialize in the first year.[7]

2.2 The Appeal Process and the Role of Evidence

DCAD typically releases proposed values in April, after which households have a 30-day window to file an appeal. Households can check their latest proposed value through DCAD's online portal, and some also receive notices by mail. In 2026, the proposed values were released on April 14, and households had until May 15 to file an appeal. Filing is free. If a household does not attend a scheduled hearing, the protest is dismissed without penalty, and filing ordinarily does not create a risk that DCAD will increase the proposed value.[8] The

[7] The homestead exemption is the most common property tax exemption. Among other provisions, it generally limits annual increases in the appraised value used for taxation to 10%. Other exemptions and caps may also apply; see Nathan et al. (2026) for an overview.

[8] In rare cases, the protest process may reveal inaccurate property characteristics in the public records, prompting DCAD to revise the proposed value upward. For example, DCAD may discover that a home has more bedrooms than its records indicate.

principal costs of filing directly are therefore the opportunity cost of the household's time and the hassle costs of navigating the process.

Households can file on their own, which we refer to as a *direct appeal*, or hire a professional property tax agent to act on their behalf. Agents commonly charge a flat fee, a share of the resulting tax savings, or a combination of the two; some agency agreements continue across multiple years (Nathan et al., 2025). This distinction is important for our analysis. The website and chatbot were designed to assist households who were considering filing on their own; accordingly, direct filing was the primary outcome specified in our RCT preregistration. We also examine appeals filed through human agents because access to the chatbot could crowd out their use.[9]

Households who wish to protest on their own can file online through DCAD's uFile system or submit a paper form by mail; historically, most direct protests have been filed online (Nathan et al., 2025). Appendix A.1 describes the online filing process in greater detail. The form asks households to state the grounds for the protest and allows them to report an "Opinion of Value"—the amount they believe the property was worth as of January 1. Although the mechanical act of submitting the form is straightforward, deciding whether to protest and what claim to make may require information and judgment.

A central feature of this setting is the inherent uncertainty involved in valuing a property that has not recently been sold. DCAD relies on mass-appraisal methods that combine statistical models with information about recent sales and property characteristics. These estimates are necessarily imperfect: two reasonable models may produce different values for the same home, and the models may not capture features or conditions known to the household. Thus, an appeal does not necessarily involve correcting an objectively erroneous proposed value. Instead, the household presents information supporting a lower estimate within a range of plausible values (Nathan et al., 2025).

Households can challenge a proposed value on several grounds. The two most common arguments are that the proposed value exceeds the property's market value and that comparable properties have been appraised at lower values, making the appraisal unequal.[10] Appendix A.1 reproduces the official grounds listed by DCAD and the subsequent steps for uploading evidence and submitting a protest. Supporting evidence can include information about the sale prices or appraised values of comparable properties. The usefulness of a particular comparable depends on factors such as its location, and home characteristics like living

[9] Although this possibility seems unlikely, access to the chatbot could instead increase demand for human agents. For example, the chatbot might persuade a household who otherwise would not have protested that an appeal is worthwhile, after which the household could hire an agent rather than file directly.

[10] Although much less common, the owner can argue that DCAD's records contain errors about the property, such as an incorrect square footage or number of bedrooms.

area, number of bedrooms and bathrooms, age, and condition. Households must therefore decide not only whether they possess evidence, but also which properties are sufficiently comparable, which provide the strongest support, what opinion of value the evidence justifies, and how to present the argument.

2.3 Appeal Rates, Tax Savings, and Inequities

In this study, we use publicly available administrative data on property taxes and property tax appeals from the Dallas Central Appraisal District (DCAD) (DCAD, 2026). These data include rich information, such as property addresses and ownership, detailed property characteristics, proposed and certified market values, exemptions, taxable values, tax rates, and historical appeal records.

The average household in Dallas County pays $7,923 in annual property taxes. Property tax appeals can significantly reduce households' tax burden. In 2026, 5.38% of Dallas County households filed an appeal on their own, and an additional 18.35% filed through an agent, for an overall appeal rate of 23.73%. The 2026 appeals are still being processed, but in previous years they often saved households hundreds of dollars per year (Nathan et al., 2025).

The average appeal rate masks substantial heterogeneity. In Dallas County, households from more advantaged groups are substantially more likely to appeal (Nathan et al., 2025), and similar inequities have been documented in other property tax contexts (Avenancio-León and Howard, 2022; Holz et al., 2026). For example, in 2025, the appeal rate was 41.7% among homes in the top quartile of market value (above $513,000), compared with 9.9% among homes in the bottom quartile (below $265,000). Appeal rates are also higher among more-educated and non-minority households.[11] These gaps may reflect differences in expected financial benefits, familiarity with the tax system, confidence in navigating the process, and the ability to afford an agent (Doerner and Ihlanfeldt, 2015).

3 Experimental Design and Implementation

3.1 The Study Website and Primary Treatment

The experiment was preregistered in the AEA RCT Registry as AEARCTR-0018430. We created a website to help homeowners decide whether to file a property tax appeal and, if so, how to do so. To use the site, households selected their property address. Both versions then

[11] Throughout the paper, we refer to Black and Hispanic households as minority households and White and Asian as non-minority households. This categorization follow prior research showing that, holding home value constant, Black and Hispanic households appeal at lower rates than White and Asian households (Nathan et al., 2025; Holz et al., 2026).

displayed personalized information about the property, including its 2026 proposed value, its value in the previous tax year, and estimated property taxes. Households could also download a "Comparable Properties Report," a PDF containing information about similar properties that could support a market-value or unequal-appraisal claim.[12] The report was intended as a reference tool rather than a ready-to-submit evidence file: it explicitly instructed households not to upload it directly to DCAD and reinforced this warning with a watermark. Finally, the website provided a five-step walkthrough of DCAD's uFile system, with direct links and detailed screenshots explaining how to complete the filing process. Step 3 of the walkthrough also included a simple "PDF Creator": households could paste evidence text into a text box and download the resulting document for upload to uFile. Unlike the pre-generated Comparable Properties Report, which supplied reference information and was not meant to be submitted, the PDF Creator converted household-provided text—potentially drafted with help from the chatbot—into an uploadable evidence document. Figure A.3 reproduces steps 2 through 5; the first step is visible in the full website screenshot in Appendix D. We also include a sample Comparable Properties Report generated by the website in Figure A.5.

The primary randomization assigned households, with equal probability, to one of two conditions: a website without chatbot access—which we refer to as No-Chatbot—and an otherwise identical website with chatbot access—which we refer to as Chatbot. An important feature of the design is that the No-Chatbot group already received substantial assistance: personalized information, supporting evidence, and detailed filing instructions. Households assigned to Chatbot also received access to an AI chatbot powered by Claude, a large language model developed by Anthropic. The chatbot appeared above the personalized property information, so households encountered it before the other website features. They could initiate a conversation by typing a question or selecting one of two quick-start prompts about whether to file an appeal that year or how to file. The chatbot was designed to converse with households and answer questions about their appeals. It received the same property-specific information available elsewhere on the website, as well as information households should already know, such as the property's building characteristics and the household's prior appeal history. As a guardrail, our instructions to the chatbot included a suggested "opinion of value": the lowest value among the property's comparables. This choice anchors the appeal at a defensible value below the current assessment while leaving room to negotiate upward.

The comparison between Chatbot and No-Chatbot is designed to identify the incremental effect of access to the chatbot, above and beyond the assistance provided by the common website. As specified in our preregistration, our central hypothesis is that the chatbot may

[12] The report included up to three recent nearby sales supporting a market-value claim and up to three similar properties in the same DCAD neighborhood that were appraised at lower values, supporting an unequal-appraisal claim. Every property in the experimental universe had at least one admissible comparable.

affect direct appeal filing by helping households determine whether appealing that year is worthwhile and by assisting them in navigating the process. The common website reduced the cost of finding information and evidence, but it did not eliminate decisions requiring judgment—for example, which comparables provided the strongest support, what opinion of value the evidence justified, and how to present the case. The chatbot could help households interpret and act on this information. On the other hand, the chatbot may have little effect if the common website already addresses the most important filing frictions, if households do not trust or engage with it, or if it merely substitutes for other website features without changing their filing decisions.

Consistent with the preregistration, our primary outcome is whether the household filed a direct appeal. We also examine whether households filed through an agent, and we use direct appeals in prior years as placebo outcomes. To study take-up and mechanisms, we observe whether and how households interacted with the chatbot, complete transcripts of their conversations, the time they spent on the website, and the other features they used.

3.2 Secondary Treatment: TaxBot versus Marina

Among households assigned to Chatbot, we conducted a second randomization between two versions of the chatbot: TaxBot and Marina. Both versions used the same underlying model, information, and substantive instructions, but differed in their names and personas. TaxBot had a machine-like name and was presented as an automated information system, with instructions to communicate factually, concisely, and procedurally. Marina had a human name and was presented as a friendly and supportive guide, with instructions to communicate warmly, express encouragement, and proactively offer help. Appendix Sections C.2 and C.3 reproduce the complete persona-specific instructions for TaxBot and Marina, respectively. Figure A.4 reproduces screenshots of the two conditions.

The effect of this secondary treatment is theoretically ambiguous. A human name and more personal style could make the chatbot feel more approachable and encourage households to initiate a conversation. Alternatively, a machine-like presentation could appear more appropriate or credible for a technical tax task, or make households more comfortable sharing information with an AI system. We therefore examine whether assignment to TaxBot rather than Marina affected chatbot take-up and, ultimately, direct appeal filing, without imposing a directional prediction.

3.3 Recruitment Method via Postcards

We created the postcards after DCAD released its 2026 proposed values and mailed them approximately two weeks before the May 15 filing deadline. The postcard was personalized with the household's name and property address, informed the recipient that the property's 2026 proposed value had been released, and explained that filing an appeal could reduce the household's property-tax burden. It described the website as a free research tool that provided updated property information, helped households decide whether to appeal, and offered step-by-step filing instructions. Recipients could reach the website through either a URL or a QR code. University of Michigan branding and researcher contact information were included to establish the source and legitimacy of the study. Figure A.1 reproduces the front and back of a sample postcard. The postcards were identical across the No-Chatbot and Chatbot groups. In particular, they did not mention the chatbot, disclose the recipient's assignment, or indicate that different versions of the website existed. Thus, households did not learn whether they had been assigned access to the chatbot until they visited the website. Treatment assignment was tied to the property and enforced through the server-side database, preventing households from selecting a different version directly.

We started with a sample of 61,474 owner-occupied, single-family properties in Dallas County whose owners could potentially reduce their property-tax liability by filing an appeal. Among other restrictions, we excluded properties without admissible comparables, properties for which a successful appeal at the suggested value would not reduce the tax bill, and properties whose owners already had professional representation. Appendix A.2 provides additional details about the sample construction. We randomly assigned 16,274 households to a no-mailing group and the remaining 45,200 to receive a postcard inviting them to use the website. Among postcard recipients, 22,600 were assigned to No-Chatbot and 22,600 to Chatbot. Within Chatbot, 11,300 were assigned to TaxBot and 11,300 to Marina. Treatment was assigned by computer at the household level and was not clustered.[13]

[13] We used stratified rerandomization based on prior appeal behavior, suggested savings, property value, voter-file match status, homestead status, proposed-value changes, and the availability of a market-value comparable.

3.4 Sample Selection, Descriptive Statistics, and Randomization Balance

Of the 45,200 households to whom we mailed a postcard, 645 visited the website and comprise our main analysis sample.[14] This sample represents 1.43% ($= \frac{645}{45{,}200}$) of the households invited to the website by postcard. Although completing an online survey is not identical to visiting a website for tax assistance, closely related field experiments provide useful benchmarks because they recruited households through mailed invitations containing a URL. Those studies obtained response rates between 1.2% and 3.9% (Nathan et al., 2025; Giaccobasso et al., 2025; Holz et al., 2025, 2026). More broadly, Sinclair et al. (2012) report a response rate of 4.7% for a personally addressed postcard inviting households to complete a web-based survey using a unique code.

Several factors likely contributed to our lower website visitation rate. First, unsolicited correspondence is often discarded or not read carefully (Bottan and Perez-Truglia, 2025; Nathan et al., 2025). Second, the related experiments sent letters enclosed in envelopes, whereas we used postcards to accelerate the printing process; postcards may have been easier to overlook or less likely to inspire trust. Third, logistical constraints required us to mail the postcards only about two weeks before the filing deadline. By then, a substantial share of households who were considering appealing on their own had already done so.[15] Finally, amid recent weakness in the housing market, increases in proposed values were less common than in previous years, likely reducing households' overall interest in appealing. For example, among households in our sampling frame, the share filing an appeal on their own declined from 10.5% in 2025 to 9.4% in 2026.

Because the postcard did not disclose the website version, treatment assignment should not have affected the decision to visit the website. Consistent with this feature of the design, 1.39% of households assigned to No-Chatbot visited the website, compared with 1.46% of those assigned to Chatbot. The difference of 0.08 pp is statistically insignificant ($p = 0.500$). Website visitation also did not differ significantly between TaxBot and Marina—see Appendix A.4 for more details.

Table 1 shows the average characteristics of the analysis sample. Panel (a) reports selected property characteristics and prior appeal behavior from the administrative records, while

[14] The website recorded 1,940 sessions during the study period. In 1,524 of these sessions, the household selected a property address or entered through a direct property link. These sessions corresponded to 710 unique properties. We excluded eight households assigned to the no-mailing group and 57 households who had already filed an appeal before their first website visit. The latter exclusion was specified in the preregistration because the intervention could not affect a filing decision that had already been made.

[15] Among postcard recipients, 9.4% of households ultimately filed an appeal on their own. By comparison, 3.6% had already done so by the time we mailed the postcards. Thus, 38.3% ($= \frac{3.6\%}{9.4\%}$) of those who ultimately filed an appeal on their own had already done so by the time of the mailing.

Panel (b) reports owner demographics for the 79.2% of households successfully matched to a proprietary voter-file dataset.[16] Column (1) describes all website visitors. The average property in the analysis sample had a proposed value of approximately $682,000, an estimated household income of $147,000, and 3.5 bedrooms. About 88% had a homestead exemption, and 61% had multiple owners. Direct appeal rates in the four years before the experiment ranged from 14% to 23%, indicating that many households had prior experience with the process. Among households matched to the L2 voter file, half had a bachelor's degree, 35% belonged to a racial or ethnic minority, 53% were classified as Democrats, and 39% as Republicans.

As discussed in more detail in Appendix A.3, website visitors are roughly similar to postcard recipients on many characteristics. There are some modest differences: visitors own more valuable and larger homes and have higher household incomes and education. The main difference, however, is that website visitors are substantially more likely to have filed an appeal in a prior year. Thus, while the analysis sample remains economically and demographically diverse, it overrepresents households with a higher preexisting propensity to appeal. Related surveys using similar recruitment methods exhibit the same selection pattern (Nathan et al., 2025; Giaccobasso et al., 2025; Holz et al., 2025). For example, in the year before the experiment, 22.6% of website visitors filed an appeal, compared with 10.5% of postcard recipients. The gap in appeal rates is even larger in 2026, as expected: the postcard was designed to attract households considering an appeal that year, and the assistance provided through the website and chatbot was intended to increase filing.

For a randomization balance check, columns (2) through (4) of Table 1 compare households assigned to No-Chatbot and Chatbot. Likewise, columns (5) through (7) compare households assigned to TaxBot and Marina. The differences in characteristics are generally small, although some are statistically significant. Given the large number of tests, we would expect 3.6 of the 36 differences to be statistically significant at the 10% level by chance; we observe somewhat more—six. Out of an abundance of caution, in addition to reporting the raw differences, we present covariate-adjusted estimates using controls selected by post-double-selection lasso (Belloni et al., 2014), hereafter referred to as double lasso. Moreover, we conduct falsification tests in an event-study framework. These checks provide strong reassurance that the treatment effects we document are not spurious.

[16] The L2 voter file provides vendor-modeled measures of household income, education, and race or ethnicity rather than self-reported measures.

3.5 Supplemental Surveys

To complement the field experiment, between July and August 2026, we recruited 647 Texas homeowners through Prolific to complete a brief online survey. The survey elicited whether respondents filed a property tax appeal in 2026, whether those who filed on their own used an AI tool such as ChatGPT for assistance, and whether those who did not file had considered appealing or using AI assistance. It also elicited respondents' general use of AI tools and their reasons for not filing or not using AI for an appeal. The median completion time was 4.5 minutes, and 97.5% of respondents passed our AI-agent detection checks. Because these data are self-reported and do not come from the experimental sample, we use them as complementary descriptive evidence on the use of AI tools outside our intervention. Appendix A.7 provides additional details.

We also administered a forecast survey to 87 experts, including academics and county assessors. The survey described the experiment in detail and asked them to forecast several key results. Appendix B describes the expert sample and survey design, while Appendix F reproduces the full survey instrument.

4 Results

4.1 Chatbot Take-Up

We begin by assessing whether households assigned to Chatbot used the tool and, if so, how intensively they engaged with it. Panel (a) of Figure 1 shows that chatbot take-up was high: about 78% of subjects assigned to the chatbot initiated a conversation. The mean number of messages sent by the subjects was 5.2, with a median of 2—for more details, see panel (a) of Table 2. Thus, for many households, initiating a conversation led to a sustained exchange rather than a single message. These conversations covered a wide range of topics, such as whether it was a good idea to file an appeal that year, what opinion of value to report, and how to write the statement required to accompany the evidence—we return to these data in Section 4.4 below. For reference, we reproduce the full text of two conversations, with sensitive information redacted to protect participants' privacy: Appendix C.4 presents a sample TaxBot conversation, and Appendix C.5 presents a sample Marina conversation.

4.2 Effects on Appeal Filing

Even if households had sustained conversations with the chatbot, what matters in the end is whether access to that assistance changed actual tax appeal behavior. The key result is

presented in panel (b) of Figure 1: the direct appeal rate increased from 41.4% in No-Chatbot to 50.5% in Chatbot. The raw difference of 9.1 pp is statistically significant ($p = 0.021$). The figure also reports an estimate that adjusts for pretreatment characteristics selected using double lasso. This adjusted estimate is slightly larger, at 10.5 pp, and statistically significant ($p = 0.006$).

As a falsification test in the style of an event-study analysis, panel (c) of Figure 1 applies the same analysis to direct appeals filed in prior years. The panel reports the double-lasso-adjusted Chatbot effect separately for each year from 2021 through 2026. Because treatment assignment occurred in 2026, the estimates for 2021–2025 are placebo "effects." All the pretreatment coefficients are close to zero and mostly statistically insignificant.[17] By contrast, the 2026 estimate is positive and statistically significant ($p = 0.004$). This event-study pattern supports the interpretation that the difference in 2026 appeal rates resulted from access to the chatbot rather than persistent differences in households' preexisting propensity to appeal.

An increase in direct filing need not represent an increase in overall appeal activity if the chatbot merely displaced human agents. Figure 1(d) shows that this did not occur: the probability of filing through an agent fell by a statistically insignificant 0.2 pp, from 9.6% in No-Chatbot to 9.4% in Chatbot ($p = 0.935$). The adjusted estimate is also small and statistically insignificant: 0.6 pp ($p = 0.787$). Thus, in our sample, the chatbot increased direct appeals without significantly crowding out appeals filed through human agents.[18]

4.3 Economic Magnitude

The effect of the chatbot is not only statistically significant, but also economically significant. We benchmark its magnitude in two ways. First, we translate the effect into a comparable expected financial incentive. Using quasi-experimental variation, Nathan et al. (2025) estimate that every $100 increase in expected tax savings raises the probability of appealing by 2.14 pp. At that rate, expected savings would need to increase by approximately $425 to generate the same 9.1 pp increase in appeal filing as access to the chatbot.

Second, we compare the magnitude of the estimates to the effects of personalized information on how to appeal and provide evidence. In a related field experiment in the same institutional setting, Nathan et al. (2025) find that mailing households a step-by-step filing

[17] The coefficient for 2024 is negative and marginally significant ($p = 0.071$). Given the large number of falsification tests we conduct, some are likely to appear significant by chance alone.

[18] Crowd-out could be larger under other circumstances. We mailed the postcards approximately two weeks before the filing deadline, by which time households seeking professional assistance may already have hired an agent. Moreover, the mailing sample excluded households with existing professional representation. Providing the chatbot earlier or to households who already used agents could therefore generate greater substitution away from human assistance.

guide and a tailored argument based on a nearby comparable property increased direct appeal filing by 4.98 pp. Our estimate of 9.1 pp is nearly twice as large.[19] This comparison is particularly informative because the No-Chatbot group in our experiment already received detailed filing instructions, personalized information, and a detailed Comparable Properties Report through the website. The chatbot effect therefore represents an increase above and beyond substantial non-AI assistance. Both comparisons come from the same setting—property tax appeals in Dallas County—but should be interpreted with caution because the studies were conducted in different years and their samples, while potentially overlapping, are likely far from identical. With these caveats in mind, the comparisons indicate that the chatbot produced a large behavioral response.

Our primary estimate of 9.1 pp is the intention-to-treat effect of being assigned access to the chatbot, rather than the effect of using it. Because 77.9% of households assigned to the chatbot initiated a conversation, a standard adjustment for incomplete take-up yields a treatment-on-the-treated effect of 11.68 pp ($= \frac{9.1}{0.779}$).[20]

This adjustment accounts for incomplete take-up of our chatbot, but it does not address a separate issue: households assigned to No-Chatbot could still obtain AI assistance outside our website using general-purpose tools such as OpenAI's ChatGPT or Anthropic's Claude. If they did so, the experimental comparison is not between receiving AI assistance and receiving no AI assistance. Instead, it compares access to our chatbot with a control condition in which some households may obtain AI assistance on their own. As a result, even the treatment-on-the-treated estimate may understate the effect of AI assistance relative to a counterfactual with no AI assistance.[21] We conducted the supplemental survey primarily to assess whether such outside AI use was empirically relevant. In a separate sample of Texas households, 40.0% of those who filed an appeal on their own reported using an AI tool for assistance; see Appendix A.7 for more details. The prevalence of AI use among self-filers therefore makes it plausible that the No-Chatbot condition did not correspond to a no-AI counterfactual, which would tend to attenuate the estimated Chatbot–No-Chatbot difference.

[19] We note that the larger effect in our experiment may partially reflect that our sample only includes individuals who visited the website. In a field experiment in Indiana, Holz et al. (2026) find that providing information about how to appeal has a larger effect among website visitors. However, differences in the institutional setting and populations make this comparison more difficult.

[20] This calculation treats initiating a conversation as chatbot take-up and assumes that assignment affects appeal filing only through chatbot use.

[21] Households assigned to Chatbot could also use general-purpose AI tools outside our website, for example because they preferred a familiar tool or had trust or privacy concerns about ours. More generally, outside use of AI makes the experimental contrast less stark than a comparison between AI assistance and no AI assistance.

4.4 Mechanisms

Having established that access to the chatbot produced a large and economically meaningful increase in appeal filing, we next examine the mechanisms behind this effect.

We recorded households' behavior on the website, which can shed light on how they engaged with both the website and the chatbot. Figure 2 reports four measures of engagement separately for the No-Chatbot and Chatbot groups: time spent on the site, whether they downloaded the Comparable Properties Report, whether they used the step-by-step walkthrough, and whether they created and downloaded an uploadable evidence document with the PDF Creator.[22] We begin by documenting engagement in the No-Chatbot condition, which provides a baseline for the subsequent comparisons. Engagement was substantial across all four measures: households spent an average of 12.1 minutes on the website, 76.4% downloaded the Comparable Properties Report, 53.2% used the filing walkthrough, and 7.0% created and downloaded an evidence document with the PDF Creator. These patterns suggest that No-Chatbot households did not merely open the website but engaged with the assistance it provided. Moreover, Appendix A.5 shows that households who ultimately filed a direct appeal engaged more extensively than those who did not. Although these comparisons are descriptive, they are consistent with households using the website to decide whether to appeal and, if so, how to do so.

Against this already-high baseline, we assess whether the chatbot changed households' overall engagement with the website. Panel (a) of Figure 2 shows that assignment to Chatbot increased average time on the website from 12.1 to 17.3 minutes. The raw increase of 5.3 minutes is statistically significant ($p = 0.022$), as is the adjusted increase of 5.9 minutes ($p = 0.009$). Thus, the chatbot increased households' overall engagement with the website rather than simply replacing time they would otherwise have spent using the non-Chatbot features. This net effect, however, masks some heterogeneity. Panel (d) shows that the chatbot increased the share creating and downloading an evidence document with the PDF Creator from 7.0% to 18.4%. The raw increase is 11.4 pp ($p < 0.001$), and the adjusted increase is 12.3 pp ($p < 0.001$). On the other hand, the chatbot reduced use of two specific forms of assistance: panel (b) shows that the share downloading the Comparable Properties Report fell from 76.4% to 53.5%, a raw decline of 23.0 pp ($p < 0.001$); the adjusted decline is 21.4 pp ($p < 0.001$). Panel (c) shows that engagement with the step-by-step walkthrough fell from 53.2% to 43.5%, a raw decline of 9.7 pp ($p = 0.014$) and an adjusted decline of 8.8 pp ($p = 0.027$). These features nevertheless remained widely used among households assigned

[22] The website also displayed a table with past and current proposed values and estimated taxes. Because viewing this information did not require households to click on anything, we cannot measure whether they examined it.

to Chatbot. Taken together, these patterns indicate that the chatbot was used mostly in conjunction with the rest of the resources provided on the site.

The clickstream evidence shows that the chatbot changed how households used the website, but it does not reveal what assistance the chatbot supplied. The conversation transcripts provide this complementary evidence. Panel (b) of Table 2 classifies the topics raised during these conversations. The most common question was whether to file an appeal that year, raised by 91.5% of households. Households also asked which comparable properties best supported their cases (50.8%), what opinion of value they should report (44.6%), whether and how to prepare supporting documents or photographs (33.7%), and how to complete the filing process (22.1%). Other conversations addressed the mechanics, costs, and risks of appealing, property characteristics that could support a lower valuation, and what to expect after filing. These exchanges indicate that households used the chatbot not only to retrieve facts, but also to seek advice about decisions requiring judgment.

Panel (c) of Table 2 shows that the chatbot typically responded with concrete guidance. It recommended filing and provided a rationale in 98.1% of conversations, emphasized the filing deadline in 97.7%, and presented or walked households through comparable properties in 85.7%. It recommended an opinion of value in 45.3% of conversations, gave filing or form guidance in 43.8%, and drafted or revised an evidence document in 39.5%. These actions go beyond reproducing static information already available on the website. They suggest that the chatbot helped households exercise judgment by evaluating available information and translating it into action.

4.5 TaxBot versus Marina

The preceding results show that households used the chatbot for substantive guidance. We next examine whether take-up and effectiveness depended on how the chatbot was presented. Figure 3 presents the results for the TaxBot and Marina groups, each relative to the common No-Chatbot group. Panel (a) shows that chatbot take-up was higher among households assigned to TaxBot (82.1%) than among those assigned to Marina (74.0%). This raw difference is 8.1 pp and marginally significant ($p = 0.073$). The double-lasso adjusted difference is similar in magnitude at 7.3 pp and is also marginally significant ($p = 0.062$). Thus, the machine-like name and less personal style appear to have encouraged somewhat greater take-up than the human name and more personal style, although the evidence is only suggestive.

Panel (b) of Figure 3 compares the time spent on the website between the TaxBot and Marina conditions and shows that the difference is small and statistically insignificant. Panel (c) compares direct appeal filing rates between the two chatbot conditions. The difference is directionally consistent with the take-up results, with a higher protest rate for TaxBot

than for Marina, but the difference is small in magnitude and statistically insignificant. The direct appeal rate was 51.9% among households assigned to TaxBot and 49.1% among those assigned to Marina. The raw difference is 2.8 pp and statistically insignificant (p = 0.619); the adjusted difference is similar in magnitude at 3.5 pp and also statistically insignificant (p = 0.544). For completeness, panel (d) presents the corresponding event-study falsification test. As expected, the Marina–Taxbot differences in the pretreatment years are statistically insignificant. Overall, TaxBot may have been somewhat more effective than Marina at engaging households, but the difference in appeal filing is statistically insignificant. However, this difference is imprecisely estimated, so we cannot rule out modest differences between the two conditions.

4.6 Distributional Effects of the Chatbot

Beyond their average effects, policymakers may also be interested in the distributional consequences of AI tax agents. We therefore examine whether the chatbot's effect on appeal filing differed across education, property value, and racial or ethnic minority status. Figure 4 presents the results. Panel (a) begins by comparing take-up rates across each of the three subgroup dimensions. In all three cases, the more advantaged group had somewhat higher take-up, but these differences are always statistically insignificant. Chatbot take-up was higher among more-educated households: 78.4% among those without a bachelor's degree versus 86.2% among those with one (p = 0.106). Take-up was also higher among households with higher-valued properties: 74.5% among those with lower-valued properties versus 79.5% among those with higher-valued properties (p = 0.330). Finally, take-up was higher among non-minority households: 75.4% among minority households versus 79.7% among non-minority households (p = 0.391).

Ultimately, our primary interest is not baseline take-up itself but whether access to the chatbot changed tax behavior. Panels (b)–(d) of Figure 4 display the chatbot's effects on appeal filing across the same three subgroup dimensions.[23] Because several subgroup comparisons rely on smaller samples, we interpret the estimates cautiously and focus on the consistency of the pattern across dimensions rather than any single comparison.

Overall, the chatbot appears to have been less effective among less-advantaged households. More precisely, panel (b) shows that the adjusted effect is 5.8 pp among households without a bachelor's degree (p = 0.345), compared with 25.0 pp among those with at least a bachelor's

[23] For properties with multiple owners, we define demographic characteristics using the first-listed owner in DCAD's ownership records. The BISG imputation uses that owner's surname together with the property's ZIP-code tabulation area, and the match to the L2 voter file selects the registered voter at the property address whose name best matches that owner. Education and minority status therefore refer to the primary owner rather than to the household.

degree ($p < 0.001$). The difference between the two effects is marginally significant ($p = 0.069$). Panel (c) shows an adjusted effect of 5.5 pp among households with lower-valued properties ($p = 0.437$), compared with 13.5 pp among those with higher-valued properties ($p = 0.003$). The raw effects are 3.9 pp and 12.7 pp, respectively. Although the effect for higher-valued properties is more than three times as large, the difference between the two estimates is statistically insignificant ($p = 0.378$). Panel (d) shows the same qualitative pattern by minority status. The adjusted effect is 7.2 pp among racial or ethnic minority households ($p = 0.270$), compared with 14.8 pp among non-minority households ($p = 0.004$); the difference is statistically insignificant ($p = 0.281$).

The subgroup differences are imprecisely estimated, and as a result two of the three are statistically insignificant despite the large magnitude. Nevertheless, the direction and magnitude of the differences are consistent across the three splits. The evidence is therefore suggestive that universal access to AI tax assistance may, far from mitigating existing inequities in the appeal process, exacerbate them. Another important point is that the differences in the chatbot's effects across subgroups are substantially larger for appeal filing (panels (b)–(d) of Figure 4) than for chatbot take-up (panel (a)). This pattern suggests that the distributional effects are driven less by whether households initiated a conversation than by what happened after take-up, including how effectively they translated the chatbot's guidance into action.

Previous studies of non-AI assistance for filing a protest have examined similar dimensions of heterogeneity (Nathan et al., 2025; Holz et al., 2026). Relative to those non-AI interventions, the magnitude and consistency of the widening effects from our AI intervention are noticeably more pronounced.[24]

To assess whether these findings were predictable ex ante, we compare our results with the expert forecasts. Appendix B reports the full results, which we summarize below. The experimental findings were far from obvious ex ante: experts expressed substantial uncertainty, and their forecasts differed considerably. Despite this disagreement, the average forecast of the chatbot's effect on filing rates was close to the experimental estimate. On average, experts substantially underestimated chatbot take-up. There was also no clear consensus about treatment-effect heterogeneity, suggesting that the larger effects among more advantaged households were not clearly anticipated. Finally, experts tended to expect Marina to be more effective than TaxBot, whereas the experimental estimates weakly favor TaxBot.

[24] Nathan et al. (2025) report that the effects of their aid are "if anything, slightly smaller for Hispanic households than for White households," while noting that differences across home-value groups are "mostly small and statistically insignificant." Holz et al. (2026) similarly conclude that their intensive procedural assistance "has almost no additional effect on the racial appeal gap"; their Appendix Figure D4 reports analogous results for the heterogeneity by property value and education.

5 Conclusions

This paper provides causal evidence on how taxpayers use AI tax agents and how access to this assistance affects their behavior. We conducted a field experiment with 645 households in Dallas County, Texas. All households received access to a website with personalized information, filing instructions, and supporting evidence for a property tax appeal; half were randomly assigned to receive access to an AI chatbot. Chatbot take-up was high: 78% initiated a conversation. Access to the chatbot increased direct appeal filing by 9.1 pp, from 41.4% to 50.5%, without crowding out human agents. Clickstream data and conversation transcripts suggest that the chatbot was effective because it helped households exercise judgment, evaluating available information and translating it into action. The chatbot had smaller effects on appeal filing among less-advantaged households, providing suggestive evidence that broad access to AI assistance may widen existing disparities. The corresponding differences in chatbot take-up were much smaller, implying that differences in *how* households used the chatbot played a key role.

We conclude by discussing the implications of our findings for policymakers. Our findings point to a broader role for AI in helping individuals navigate the tax system. Tax administrations often try to reduce administrative burdens by providing information, instructions, or online tools. Yet taxpayers may still struggle to determine which information matters and how to act on it. In our experiment, the chatbot affected behavior even though all households already received personalized information, step-by-step instructions, and supporting evidence. This suggests that conversational AI can fill an important gap between making information available and helping taxpayers exercise judgment. Because this assistance can potentially be provided at low marginal cost, tax administrations may want to consider offering AI tax agents for free, both for property tax appeals and for other interactions with the tax system.

Beyond their average benefits, AI tax agents may also have important distributional consequences. In our setting, appeal rates differ substantially across groups, such as between poorer and richer households and between less- and more-educated households. The chatbot had smaller effects on appeal filing among less-advantaged households. Our findings therefore suggest that broad access to AI assistance may widen, rather than narrow, existing disparities. Policymakers should evaluate not only take-up and average effects, but also how different groups use these tools and who ultimately benefits. They may also need to consider whether AI assistance should be paired with more structured guidance or access to human support. These issues will be especially important in more complex settings, such as income tax filing, where inaccurate advice can have substantial consequences and where policies governing

accuracy, liability, and privacy may shape both trust and access. As AI tax agents become more common, the key policy question will be not simply whether they work, but for whom and under what conditions.

References

Abeler, J. and S. Jäger (2015). Complex Tax Incentives. *American Economic Journal: Economic Policy 7*(3), 1–28.

Avenancio-León, C. F. and T. Howard (2022). The Assessment Gap: Racial Inequalities in Property Taxation. *The Quarterly Journal of Economics 137*(3), 1383–1434.

Battaglini, M., L. Guiso, C. Lacava, and E. Patacchini (2026). Tax Professionals and Tax Evasion. *Journal of the European Economic Association, forthcoming*.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on Treatment Effects After Selection Among High-Dimensional Controls. *Review of Economic Studies 81*(2), 608–650.

Benzarti, Y. (2020). How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs. *American Economic Journal: Economic Policy 12*(4), 38–57.

Benzarti, Y. (2021). Estimating the Costs of Filing Tax Returns and the Potential Savings from Policies Aimed at Reducing These Costs. *Tax Policy and the Economy 35*(1), 55–85.

Bhargava, S. and D. Manoli (2015). Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment. *American Economic Review 105*(11), 3489–3529.

Bock, M. R., K. Molisee, Z. Ozer, and S. Shah (2025). TaxCalcBench: Evaluating Frontier Models on the Tax Calculation Task. arXiv preprint arXiv:2507.16126.

Bottan, N. L. and R. Perez-Truglia (2025). Betting on the House: Subjective Expectations and Market Choices. *American Economic Journal: Applied Economics 17*(1), 459–500.

Brynjolfsson, E., D. Li, and L. Raymond (2025). Generative AI at Work. *The Quarterly Journal of Economics 140*(2), 889–942.

Cabral, M. and C. Hoxby (2012). The Hated Property Tax: Salience, Tax Rates, and Tax Revolts. *NBER Working Paper No. 18514*.

Cerullo, M. (2026). Should You Use AI to File Your Taxes? Experts Warn It Can Lead to Costly Mistakes. *CBS News, March 27 2026*.

Chetty, R. and E. Saez (2013). Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients. *American Economic Journal: Applied Economics 5*(1), 1–31.

Cruces, G., D. Fernández Meijide, S. Galiani, R. H. Gálvez, and M. Lombardi (2026). Does Generative AI Narrow Education-Based Productivity Gaps? Evidence from a Randomized Experiment. *NBER Working Paper No. 34851*.

Cui, K. Z., M. Demirer, S. Jaffe, L. Musolff, S. Peng, and T. Salz (2026). The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with

Software Developers. *Management Science*. Articles in Advance, published online February 27, 2026.

DCAD (2026). Appraisal review board data. Dallas Central Appraisal District. https://www.dallascad.org/DataProducts.aspx.

DeBacker, J., B. T. Heim, A. Tran, and A. Yuskavage (2024). Seeking Professional Help: How Paid Preparers Decrease Tax Compliance. *National Tax Journal 77*(2), 229–261.

Dell'Acqua, F., I. McFowland, Edward, E. Mollick, H. Lifshitz, K. C. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani (2026). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. *Organization Science 37*(2), 403–423.

Dobay, N., F. Nicely, A. Sanderson, and P. Sanderson (2019). The Best (and Worst) of International Property Tax Administration. Technical report, Council On State Taxation.

Doerner, W. M. and K. R. Ihlanfeldt (2015). The Role of Representative Agents in the Property Tax Appeals Process. *National Tax Journal 68*(1), 59–92.

Feldman, N. E., P. Katuščák, and L. Kawano (2016). Taxpayer Confusion: Evidence from the Child Tax Credit. *American Economic Review 106*(3), 807–835.

Finkelstein, A. and M. J. Notowidigdo (2019). Take-Up and Targeting: Experimental Evidence from SNAP. *The Quarterly Journal of Economics 134*(3), 1505–1556.

Giaccobasso, M., B. Nathan, R. Perez-Truglia, and A. Zentner (2025). Where Do My Tax Dollars Go? Tax Morale Effects of Perceived Government Spending. *American Economic Journal: Applied Economics 17*(4), 223–59.

Holz, J., D. Novgorodsky, and A. Simon (2026). Disparities in tax negotiations: Evidence from field experiments with homeowners and assessors. Working paper.

Holz, J., R. Perez-Truglia, and A. Zentner (2025). Does Tax Avoidance Trickle Down? Evidence from a Field Experiment. *NBER Working Paper No. 34209*.

Hoopes, J. L., D. H. Reck, and J. Slemrod (2015). Taxpayer Search for Information: Implications for Rational Attention. *American Economic Journal: Economic Policy 7*(3), 177–208.

Kleven, H. J. and W. Kopczuk (2011). Transfer Program Complexity and the Take-Up of Social Benefits. *American Economic Journal: Economic Policy 3*(1), 54–90.

Lieber, D. (2020). Why Protesting Your Property Appraisal Is So Hard. *The Dallas Morning News, June 4th 2020*.

Moynihan, D., P. Herd, and H. Harvey (2015). Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions. *Journal of public administration*

research and theory 25(1), 43–69.

Nathan, B., R. Perez-Truglia, and A. Zentner (2025). My Taxes Are Too Darn High: Why Do Households Protest Their Taxes? *American Economic Journal: Economic Policy 17*(1), 273–310.

Nathan, B., R. Perez-Truglia, and A. Zentner (2026). Paying Your Fair Share: Perceived Fairness and Tax Compliance. *Journal of Accounting and Economics 81*(2), 101838.

Noy, S. and W. Zhang (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *Science 381*(6654), 187–192.

Otis, N., R. Clarke, S. Delecourt, D. Holtz, and R. Koning (2024). The Uneven Impact of Generative AI on Entrepreneurial Performance. Working paper.

Ray, V., P. Herd, and D. Moynihan (2023). Racialized burdens: Applying racialized organization theory to the administrative state. *Journal of Public Administration Research and Theory 33*(1), 139–152.

Saez, E. and D. Seim (2025). Wealth Tax Enforcement in Sweden: Filing Requirements and Pre-Populated Returns. *Journal of Public Economics 249*, 105440.

Sinclair, M., J. O'Toole, M. Malawaraarachchi, and K. Leder (2012). Comparison of Response Rates and Cost-Effectiveness for a Community-Based Survey: Postal, Internet and Telephone Modes with Generic or Personalised Recruitment Approaches. *BMC Medical Research Methodology 12*(1), 132.

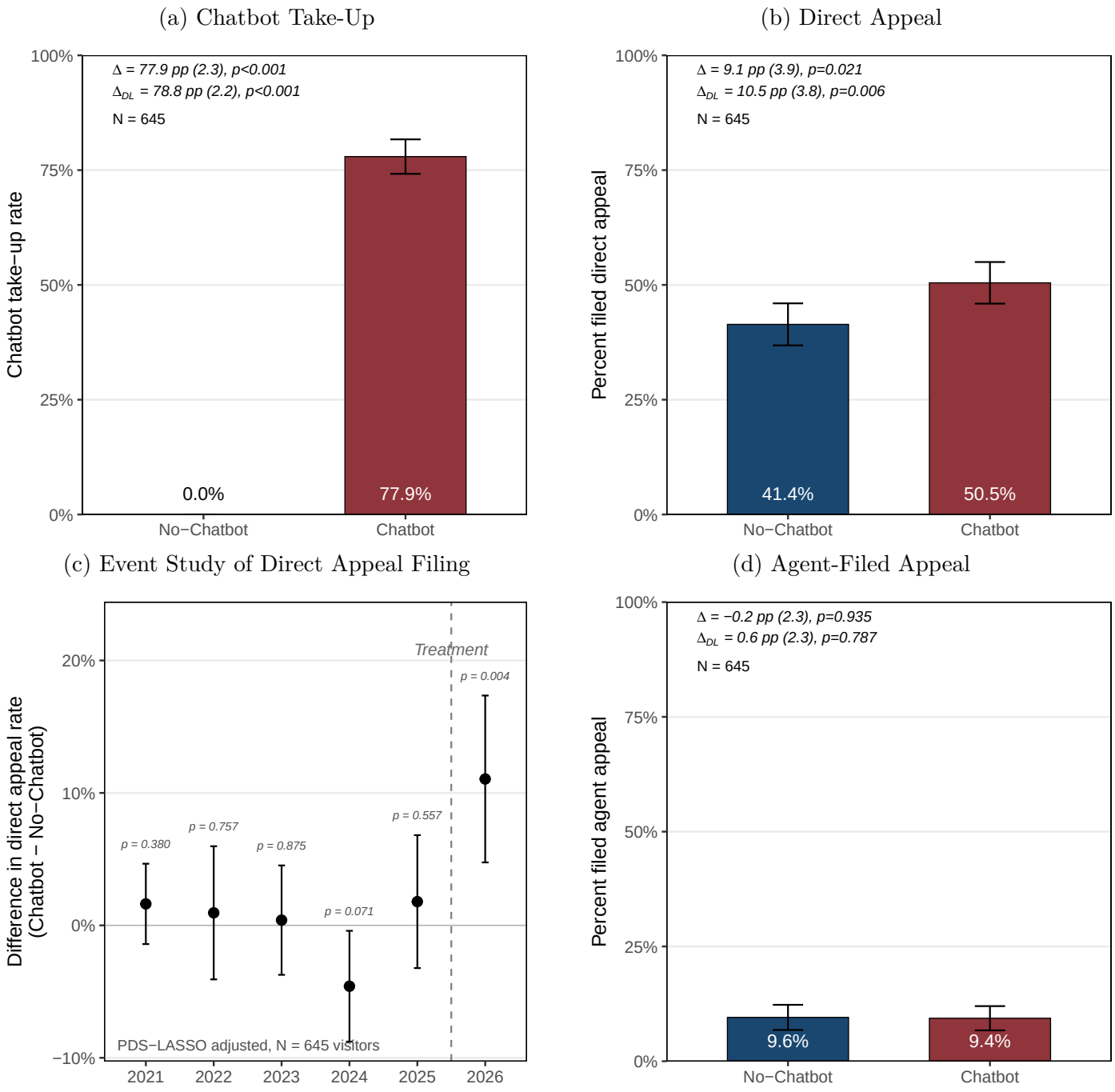
Tax Policy Center (2021). Amount of Revenue by Source. *https://www.taxpolicycenter.org/statistics/property-tax-revenue*.

U.S. Census Bureau (2026a). County Population Totals and Components of Change: 2020–2025. Vintage 2025 population estimates.

U.S. Census Bureau (2026b). State Population Totals and Components of Change: 2020–2025. Vintage 2025 population estimates.

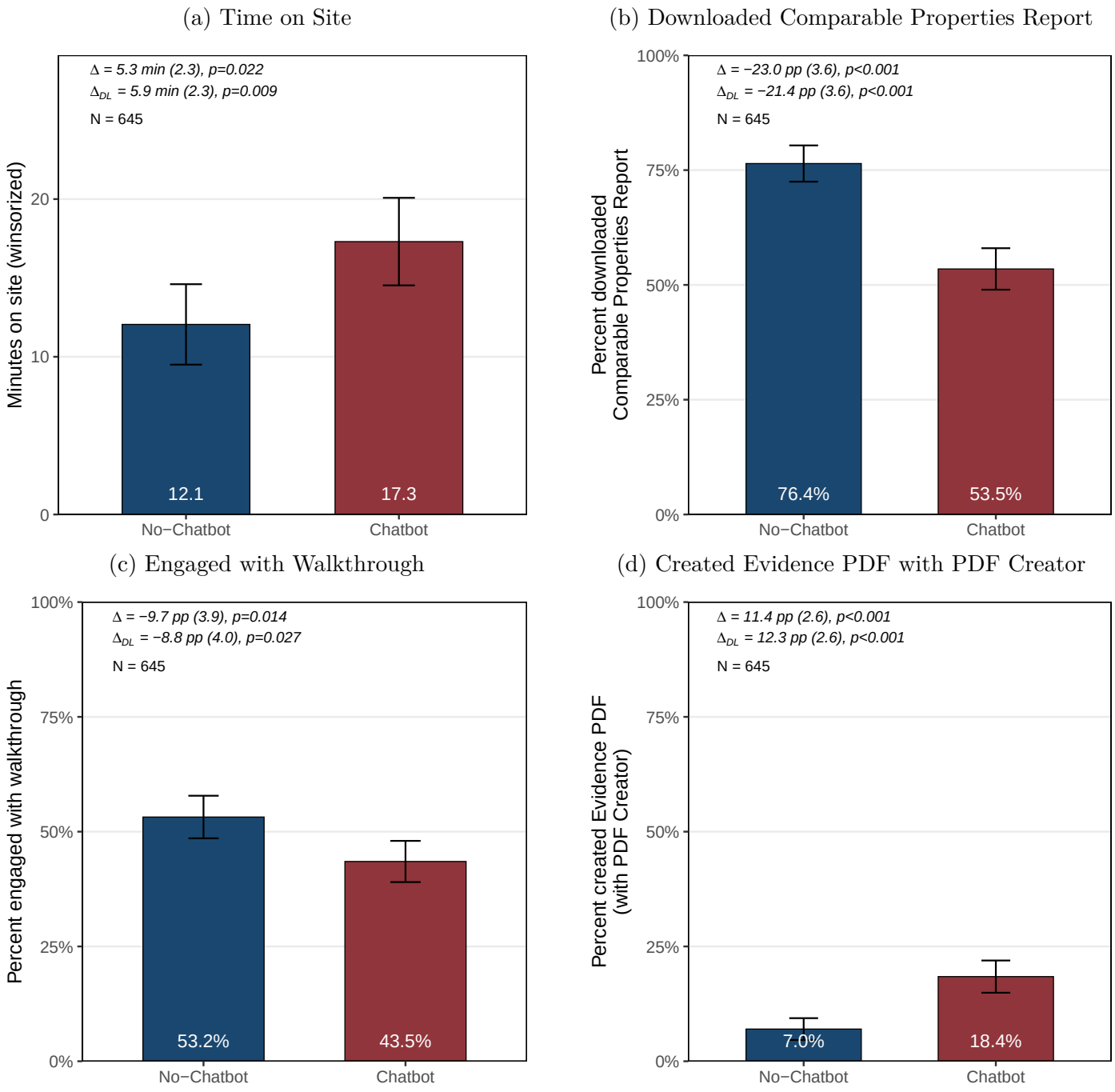
World Bank (2019). The Administrative Review Process for Tax Disputes: Tax Objections and Appeals in Latin America and the Caribbean.

Figure 1: Main Outcomes



Notes: All panels include website visitors in the analysis sample ($N = 645$). Panels (a), (b), and (d) show the shares who initiated a chatbot conversation, filed a direct (owner-filed, non-agent) appeal, and filed through an agent, respectively. Bars show raw means by arm with 90% confidence intervals and report the Chatbot minus No-Chatbot difference, unadjusted and adjusted with controls selected by double lasso. Panel (c) reports the double-lasso-adjusted effect of Chatbot on direct appeal filing separately by year. Outcomes from 2021–2025 are pretreatment placebos; 2026 is the experimental outcome. Error bars are 90% confidence intervals based on heteroskedasticity-robust standard errors.

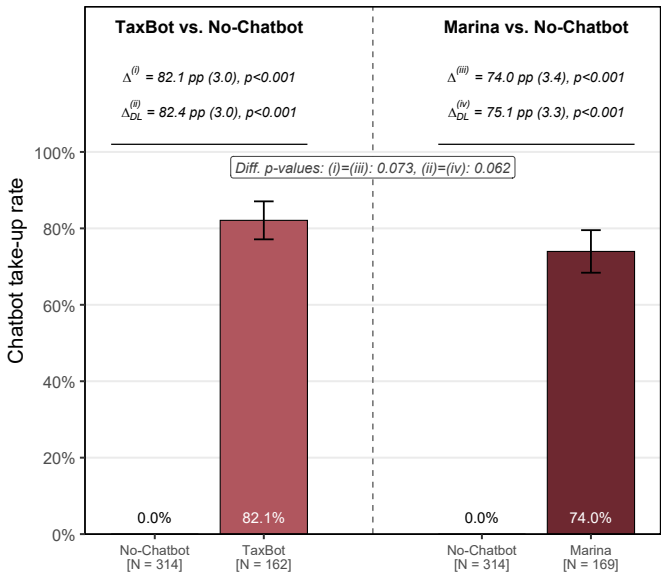
Figure 2: Mechanisms: Website Engagement



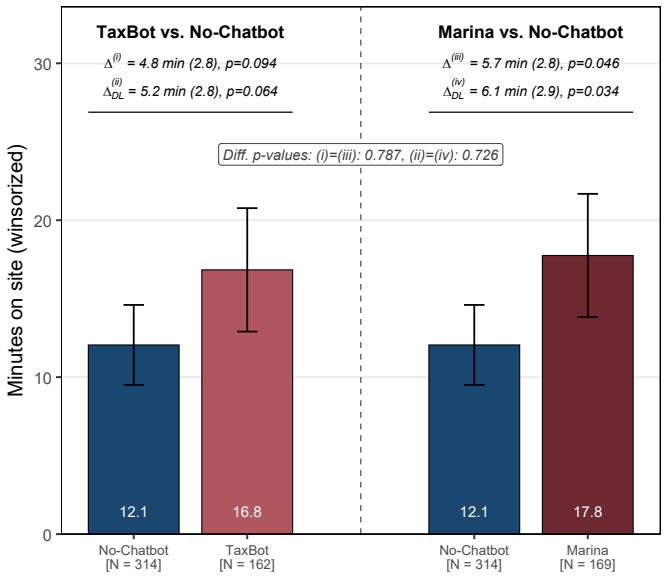
Notes: Includes website visitors in the analysis sample ($N = 645$). Panels follow the order of the website: total minutes on the site (winsorized at the 95th percentile), the share who downloaded the pre-generated Comparable Properties Report, the share who engaged with the step-by-step appeal walkthrough, and the share who created and downloaded an uploadable evidence document with the PDF Creator. Bars show raw means by arm; error bars are 90% confidence intervals. Each panel reports the Chatbot minus No-Chatbot difference (minutes in panel (a), percentage points in panels (b)–(d)), unadjusted and adjusted with controls selected by double lasso, with heteroskedasticity-robust standard errors and p-values from OLS.

Figure 3: Comparison of Chatbot Types: Marina versus TaxBot

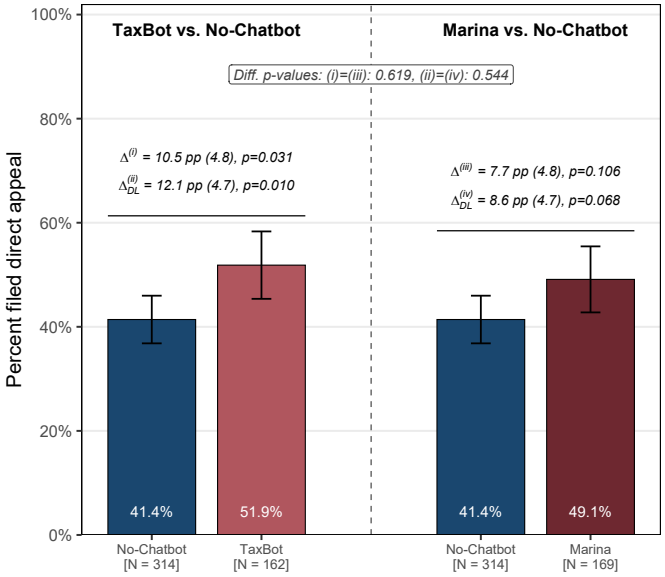
(a) Chatbot Take-Up



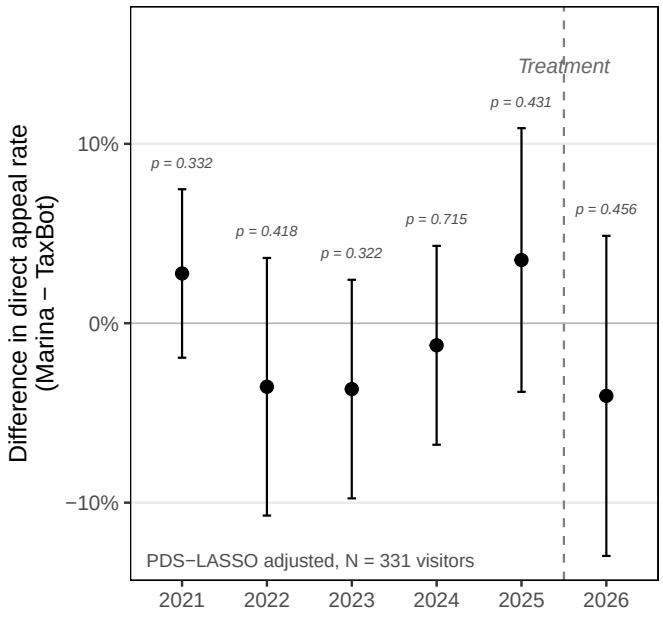
(b) Time on Site



(c) Direct Appeal



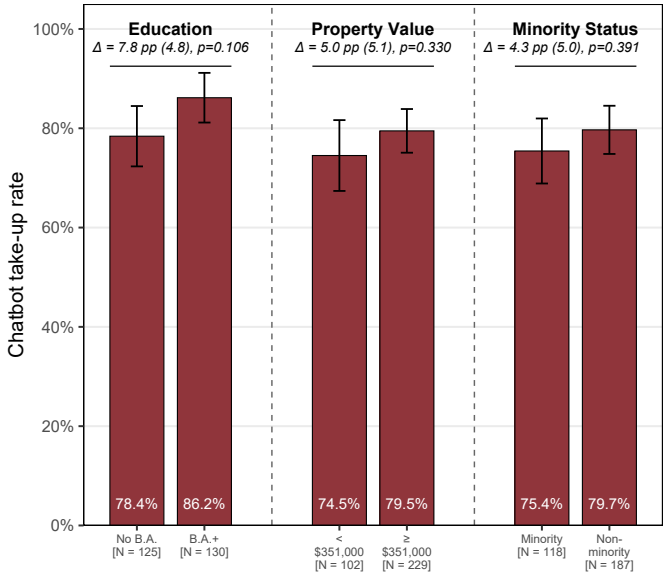
(d) Event Study of Direct Appeal Filing



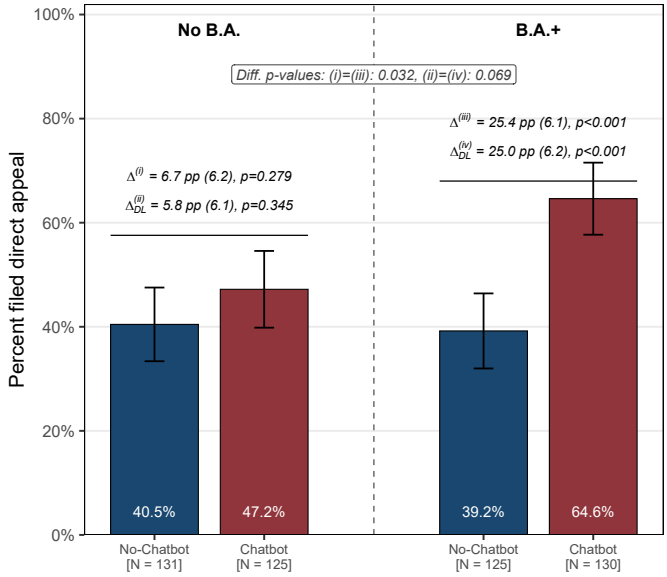
Notes: Panels (a) and (c) include all website visitors in the analysis sample ($N = 645$) and show chatbot take-up and direct appeal filing, respectively, for TaxBot and Marina, each compared with the same No-Chatbot benchmark. These panels show raw means with 90% confidence intervals and report unadjusted and double-lasso-adjusted differences. Panel (b) shows minutes on site (winsorized at the 95th percentile) for TaxBot and Marina, each compared with the same No-Chatbot benchmark, with raw and double-lasso-adjusted Marina-minus-Taxbot differences. Panel (d) includes visitors assigned to Chatbot ($N = 331$) and plots double-lasso-adjusted Marina-minus-Taxbot differences in direct appeal filing for 2021–2026. Outcomes from 2021–2025 are pretreatment placebos; 2026 is the experimental outcome. Error bars are 90% confidence intervals based on heteroskedasticity-robust standard errors.

Figure 4: Heterogeneous Chatbot Effects

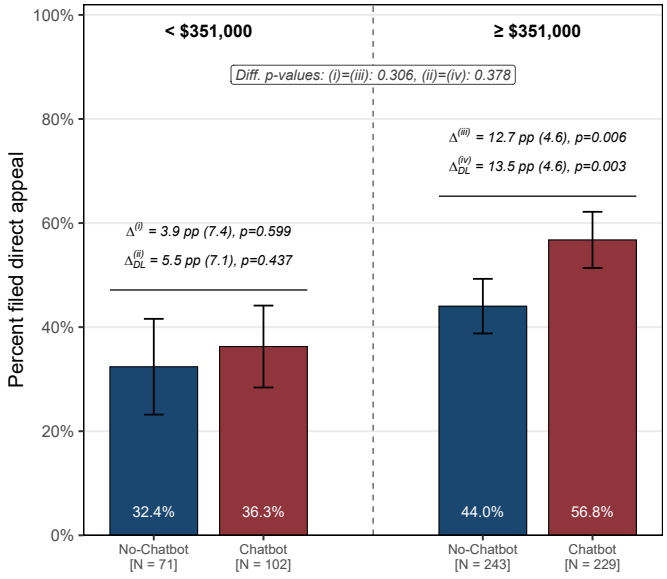
(a) Chatbot Take-Up by Subgroup



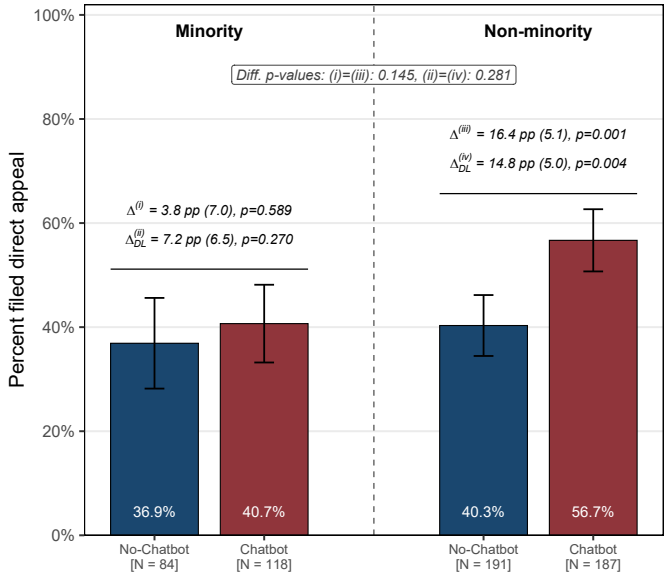
(b) Direct Appeal Filing by Education



(c) Direct Appeal Filing by Property Value



(d) Direct Appeal Filing by Minority Status



Notes: Includes website visitors in the analysis sample. Panel (a) includes visitors assigned to Chatbot and reports chatbot take-up rates across the education, property-value, and minority-status subgroups; differences compare the two displayed subgroup means, and brackets report subgroup sample sizes. Panels (b)–(d) show direct appeal filing rates by arm within each subgroup, with 90% confidence intervals. Each subgroup reports the Chatbot minus No-Chatbot difference in percentage points, unadjusted and adjusted with controls selected by double lasso, with heteroskedasticity-robust standard errors and p-values from OLS. Each panel also reports the p-value for equality of the two subgroup effects from a double-lasso-adjusted interaction regression. Education splits on a bachelor's degree from the L2 voter file, conditional on a voter-file match. Property value splits at the median 2026 proposed value among postcard recipients ($350,850). Ethnic/racial minority indicates imputed Black or Hispanic ethnicity via BISG, conditional on BISG coverage, and White and Asian/Pacific Islander owners are non-minority.

Table 1: Descriptive Statistics and Randomization Balance

		Main Treatment Arm			Secondary Treatment Arm		
	All Visitors	No-Chatbot	Chatbot	p-value (2) vs. (3)	TaxBot	Marina	p-value (5) vs. (6)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a): Property characteristics and appeal history							
Proposed value ($1,000s)	681.5	758.3	608.6	0.002	617.9	599.7	0.753
	(24.3)	(39.2)	(28.7)		(46.3)	(34.6)	
Proposed value change vs. 2025 (%)	4.8	4.0	5.6	0.381	4.0	7.2	0.323
	(0.9)	(0.7)	(1.6)		(0.7)	(3.1)	
Suggested tax savings ($1,000s)	1.44	1.59	1.30	0.033	1.35	1.24	0.494
	(0.07)	(0.11)	(0.08)		(0.13)	(0.09)	
Homestead exemption (%)	88.4	90.4	86.4	0.108	87.0	85.8	0.743
	(1.3)	(1.7)	(1.9)		(2.6)	(2.7)	
Comparable properties available	3.73	3.78	3.68	0.337	3.60	3.76	0.270
	(0.05)	(0.08)	(0.07)		(0.10)	(0.10)	
Number of bedrooms	3.49	3.51	3.47	0.430	3.42	3.51	0.234
	(0.03)	(0.04)	(0.04)		(0.05)	(0.05)	
Multiple owners (%)	60.9	62.7	59.2	0.360	53.7	64.5	0.046
	(1.9)	(2.7)	(2.7)		(3.9)	(3.7)	
2025 direct appeal rate (%)	22.6	23.6	21.8	0.583	19.8	23.7	0.389
	(1.6)	(2.4)	(2.3)		(3.1)	(3.3)	
2024 direct appeal rate (%)	17.1	19.7	14.5	0.078	15.4	13.6	0.639
	(1.5)	(2.3)	(1.9)		(2.8)	(2.6)	
2023 direct appeal rate (%)	14.4	15.0	13.9	0.700	15.4	12.4	0.432
	(1.4)	(2.0)	(1.9)		(2.8)	(2.5)	
2022 direct appeal rate (%)	22.8	22.9	22.7	0.935	25.3	20.1	0.262
	(1.7)	(2.4)	(2.3)		(3.4)	(3.1)	
Observations	645	314	331		162	169	
Panel (b): Owner demographics							
Matched to L2 voter file (%)	79.2	81.5	77.0	0.160	79.6	74.6	0.273
	(1.6)	(2.2)	(2.3)		(3.2)	(3.4)	
Household income ($1,000s)	147.0	148.8	145.2	0.541	139.9	150.6	0.215
	(3.0)	(4.2)	(4.3)		(5.9)	(6.3)	
Female (%)	33.5	33.2	33.7	0.901	40.3	27.0	0.024
	(2.1)	(2.9)	(3.0)		(4.3)	(4.0)	
Bachelor's degree or more (%)	49.9	48.8	51.0	0.627	48.1	54.0	0.347
	(2.2)	(3.1)	(3.1)		(4.4)	(4.5)	
Democrat (%)	53.2	53.5	52.9	0.897	55.8	50.0	0.354
	(2.2)	(3.1)	(3.1)		(4.4)	(4.5)	
Republican (%)	38.9	38.7	39.2	0.900	36.4	42.1	0.359
	(2.2)	(3.0)	(3.1)		(4.3)	(4.4)	
Ethnic/racial minority (%)	34.8	30.5	38.7	0.039	38.7	38.7	0.994
	(2.0)	(2.8)	(2.8)		(4.0)	(3.9)	
Observations (matched to L2)	511	256	255		129	126	

Notes: Includes website visitors in the analysis sample. Column (1) pools all visitors; column (2) is the No-Chatbot arm; column (3) pools the two chatbot sub-arms; columns (5) and (6) split the Chatbot arm into TaxBot and Marina. Cells report means with standard errors in parentheses. The N row below Panel (a) reports the number of website visitors in each column, which is the sample for all Panel (a) rows. The N row below Panel (b) reports the number of visitors matched to the L2 voter file, which is the sample for the household income, gender, education, and party rows; the voter-file match and ethnic/racial minority rows use the full visitor sample. Columns (4) and (7) report p-values from two-sample t-tests of the difference between columns (2) and (3), and between columns (5) and (6), respectively. Property characteristics are from DCAD administrative records. The proposed value is the 2026 proposed market value; proposed value change is the percent change from the 2025 to the 2026 proposed value; suggested tax savings is the tax reduction implied by revaluing the property at the lowest-valued comparable property; homestead exemption indicates a 2026 homestead exemption; comparable properties available is the number of comparable properties identified for the homeowner; bedrooms are from the DCAD residential detail file; multiple owners indicates joint owners on the DCAD account. Direct appeal rates in each year are owner-filed (non-agent) appeals from the DCAD ARB archive. Household income, gender, education, and party are from the L2 voter file, conditional on a voter-file match (79% of visitors). Ethnic/racial minority indicates imputed Black or Hispanic ethnicity, based on surname and ZCTA via Bayesian Improved Surname Geocoding (BISG; 92% coverage); White and Asian/Pacific Islander owners are non-minority.

Table 2: Chatbot Conversation Topics and Actions

	N	Percent
Panel A: Conversation Length		
Number of messages from subject:		
0	0	0.0%
1	109	38.2%
2	37	13.0%
3	16	5.6%
4	8	2.8%
5+	115	40.4%
Panel B: User topics (% of accounts)		
Whether to file protest *"Should I file a protest this year?"*	260	91.2%
Confirmation, feedback, and off-topic *"Yes, let's do that. Thank you!"*	137	48.1%
Comparable property selection *"Which comparable homes best support my case?"*	135	47.4%
Opinion of value questions *"What value should I ask for?"*	120	42.1%
Evidence document and photos *"Can you draft the evidence document? Do I need photos?"*	90	31.6%
Request filing help/instructions *"Help me file my protest."*	60	21.1%
Filing mechanics and form options *"Which boxes do I check on the protest form?"*	42	14.7%
Property features/condition affecting value *"My house needs foundation repairs. Does that help my case?"*	39	13.7%
Value components and corrections *"Why did my land value go up so much?"*	23	8.1%
Process concerns, risks, and cost *"Could my value go up if I protest? Does this cost anything?"*	22	7.7%
Purchase price as basis *"I paid much less than the assessed value. Can I use that?"*	13	4.6%
Post-submission follow-up *"I submitted my protest. What happens next?"*	13	4.6%
Panel C: Chatbot actions (% of accounts)		
Recommend filing with rationale	279	97.9%
Emphasize deadline urgency	278	97.5%
Present/walk through comps	239	83.9%
Recommend opinion of value	122	42.8%
Give filing/form guidance	118	41.4%
Draft/revise evidence document	105	36.8%
Explain process, risks, and next steps	85	29.8%
Correct/clarify input and cost	43	15.1%
Acknowledge data limitations / decline out-of-scope	40	14.0%
Reframe features/condition as strengthening case	38	13.3%
Incorporate user-provided data	28	9.8%
Respond to feedback/off-topic	27	9.5%
Explain value components and homestead cap	20	7.0%
Redirect purchase-price reasoning	12	4.2%

Notes: Based on households assigned to Chatbot who initiated a conversation ($N = 258$). Panels (b) and (c) report the number and share of households whose conversations contain each household topic and each chatbot action, classified from full transcripts by a large language model—see Appendix A.6 for details. Topics and actions are not mutually exclusive, so shares do not sum to 100%. The quoted questions in Panel (b) are stylized illustrations of each topic, not verbatim messages.

Online Appendix (For Online Publication Only)

"Taxpayer Behavior in the Age of AI: A Field Experiment on Property Tax Appeals"

Holz, Perez-Truglia, Simon and Zentner

August 18, 2026

A Additional Details, Results, and Robustness Checks

A.1 The Online Protest Process

Figure A.2 illustrates DCAD's online protest process. Households begin by opening their property record and selecting the link to uFile, as shown in panel (a). They then select one or more grounds for protest from the list reproduced in panel (b). These include claims that the proposed value exceeds market value or is unequal relative to comparable properties, as well as claims involving ownership, exemptions, property location, and other issues. Panel (c) shows the documentation stage, at which households can upload supporting evidence or indicate that they do not wish to submit evidence at that stage. Finally, households provide their contact information and submit the protest, as shown in panel (d). Thus, filing itself requires only a few steps, but households must still decide which grounds apply, what value to propose, and whether and what evidence to provide.

A.2 More Details About the Mailing Sample

We constructed the experimental universe by starting with owner-occupied, single-family properties in Dallas County and applying restrictions intended to identify households for whom the intervention was feasible and potentially consequential. We retained properties that were reappraised in 2026 and had the information needed to personalize the website. We excluded properties with an existing tax representative or recent agent-filed appeal, properties subject to tax ceilings or with little effective tax liability, and properties without an admissible comparable. We also required that a successful appeal at the suggested value could reduce the household's tax bill after accounting for the homestead cap, that the estimated savings be at least $50, and that the suggested reduction fall between 5% and 30%. Finally, we excluded properties whose proposed value had decreased by more than 2% relative to the previous year. These restrictions yielded the experimental universe of 61,474 properties, from which 45,200 households were randomly selected to receive a postcard.

A.3 Sample Representativeness

Table A.1 compares the experimental universe, the households assigned to receive a postcard, the website visitors who comprise our analysis sample, and the separate sample of Texas households who participated in our supplemental survey. Columns (1) and (2) show that the postcard sample closely resembles the experimental universe, as expected because postcard receipt was randomly assigned. For example, average proposed values are $500,000 and $500,400, respectively, and average household incomes are $117,500 and $117,600. The two groups are also nearly identical in their property characteristics, prior appeal rates, and owner demographics.

Column (3) shows that website visitors are positively selected relative to postcard recipients who did not visit. Visitors own more valuable and larger homes, have higher household incomes and education, and are less likely to belong to a racial or ethnic minority. They are also substantially more likely to have experience with the appeal process: 50.1% had filed a direct appeal during the previous five years, compared with 28.1% in the experimental universe. This pattern is intuitive because a postcard offering assistance with an appeal should disproportionately attract households who are already considering whether to file. Thus, the analysis sample is economically and demographically diverse but overrepresents households with a relatively high propensity to appeal. Column (4) describes the separate complementary-survey sample; it is not part of the field experiment and is included to clarify the population underlying the survey evidence used elsewhere in the paper.

A.4 Website Visits by Treatment Assignment

Treatment assignment occurred before households visited the website, but the postcards were identical across conditions and did not disclose which website version a household would receive. Panel (a) of Figure A.6 shows that 1.39% of households assigned to No-Chatbot visited the website, compared with 1.46% of those assigned to Chatbot. The raw difference is 0.08 pp and statistically insignificant ($p = 0.500$); the double-lasso-adjusted difference is also 0.08 pp and statistically insignificant ($p = 0.478$). Panel (b) shows that visitation was 1.43% among households assigned to TaxBot and 1.50% among those assigned to Marina, a difference of 0.06 pp ($p = 0.698$). Thus, neither treatment assignment affected which households selected into the website.

A.5 Additional Evidence on Website Engagement

Figure A.7 complements the mechanism analysis by comparing website engagement among No-Chatbot visitors who did and did not file a direct appeal. Panel (a) shows that households

who filed spent an average of 21.1 minutes on the website, compared with 5.7 minutes among those who did not file. Panels (b)–(d) show the same pattern across specific website features. Among filers, 90.8% downloaded the Comparable Properties Report, 66.9% engaged with the walkthrough, and 16.2% created and downloaded an evidence document with the PDF Creator. The corresponding rates among non-filers were 66.3%, 43.5%, and 0.5%. Each of these differences is statistically significant ($p < 0.001$).

These patterns are consistent with households using the website as an aid when preparing their appeals. At the same time, because the comparison splits visitors by an outcome rather than by randomized treatment assignment, it does not establish that greater engagement caused them to file. Engagement was also substantial among households who did not ultimately appeal, suggesting that some seriously considered filing but ultimately decided against it or did not complete the process.

A.6 Classification of Chatbot Conversations

We classify the chatbot conversations reported in panels (b) and (c) of Table 2 using a two-pass procedure based on a large language model: the Claude Opus 4.8 model by Anthropic. The unit of classification is the account: all chat sessions belonging to the same account are concatenated into a single transcript, yielding 285 account-level conversations.[25]

In the first pass (discovery), the model reads the transcripts in batches of 50 accounts and proposes candidate categories of user topics and chatbot actions grounded in the observed conversations. A final synthesis step merges the batch-level proposals into a unified taxonomy, which the authors reviewed and edited before classification.

In the second pass (classification), the model reads each account's transcript alongside the finalized taxonomy and assigns a binary indicator for each category. Categories are not mutually exclusive, so the shares in Table 2 do not sum to 100 percent. The full prompts used in both passes are preserved in the replication package. To assess the reliability, we repeated the classification using the same model, prompt, and taxonomy, and compared the two independent runs. There was 97.5% agreement.

A.7 Supplemental Survey

We fielded the supplemental survey through Prolific. Because we were recruiting participants from a narrow population—Texas homeowners—we had to run the advertisement from July

[25] This total includes all accounts that interacted with the chatbot. Table 2 restricts the analysis to households in the main analysis sample—excluding, for example, households that had already filed an appeal before their first website visit—and therefore includes 258 account-level conversations.

17 through August 7, 2026, to recruit this number of respondents. Respondents reported their property and demographic characteristics, their general use of AI tools, and whether and how they filed a property tax appeal in 2026. Those who filed on their own were asked whether they used an AI tool such as ChatGPT to prepare or file their appeal. Those who did not file were asked whether they knew they could appeal, whether they considered doing so, why they did not file, and why they did not use AI assistance. The median respondent completed the survey in 4.5 minutes, and 97.5% passed our AI-agent detection checks. Table A.1 compares the survey respondents with the other samples used in the paper.

Figure A.8 summarizes the main survey results. Panel (a) shows that 13.9% of respondents filed a 2026 appeal on their own, while 18.4% had an appeal filed by an agent or another person. Among the 90 self-filers, 40.0% reported using an AI tool for assistance. Because purpose-built AI tax agents were uncommon at the time, these respondents most likely used general-purpose tools such as ChatGPT. This finding shows that households outside our experiment were already using AI to assist with property tax appeals. It also matters for interpreting the experimental estimates: access to such tools among either No-Chatbot or Chatbot households would tend to reduce the measured Chatbot–No-Chatbot difference.

Panels (b) and (c) provide descriptive evidence related to the heterogeneity analysis. Panel (b) shows that appeal filing was more common among college-educated households and those with higher-valued properties, while the difference by minority status was smaller. Panel (c) reports AI use among self-filers across the same groups. AI use was higher among college-educated households, households with higher-valued properties, and minority households; because this panel is based on only 90 self-filers, these comparisons remain imprecise. Table A.2 reports more detailed results on AI use and the barriers respondents reported facing when deciding whether to appeal.

Figure A.1: Sample Postcard

(a) Front

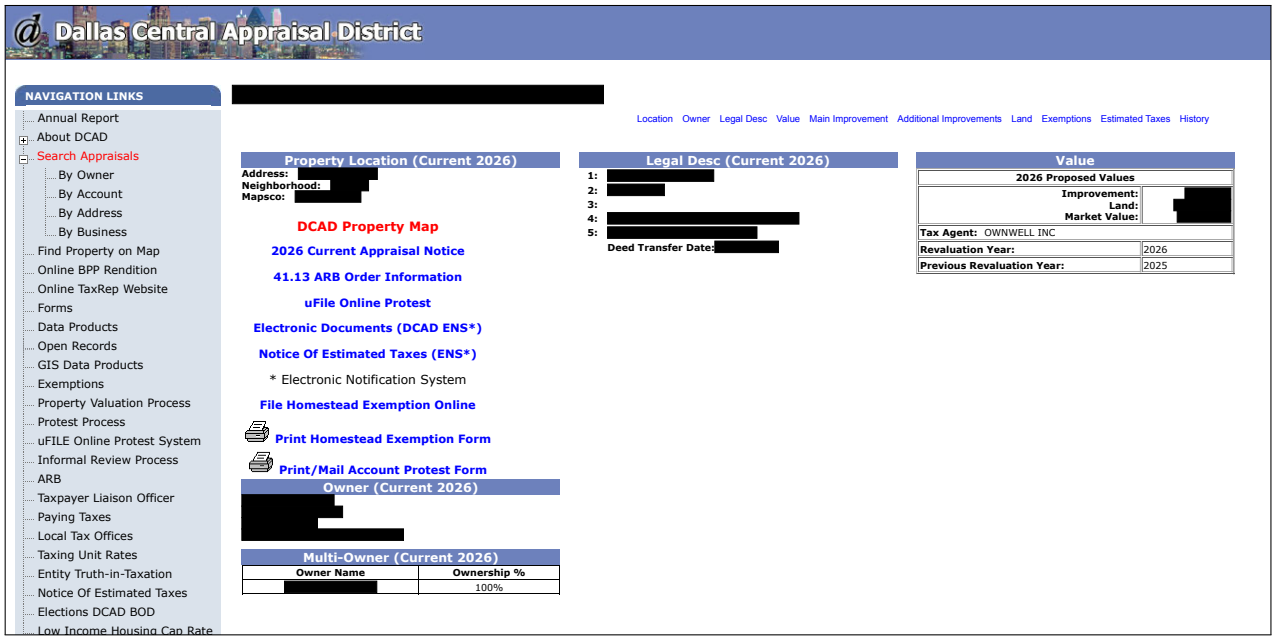


(b) Back



Figure A.2: Dallas County Online Protest Process

(a) Accessing uFile from the Property Record



(b) Selecting the Grounds for Protest

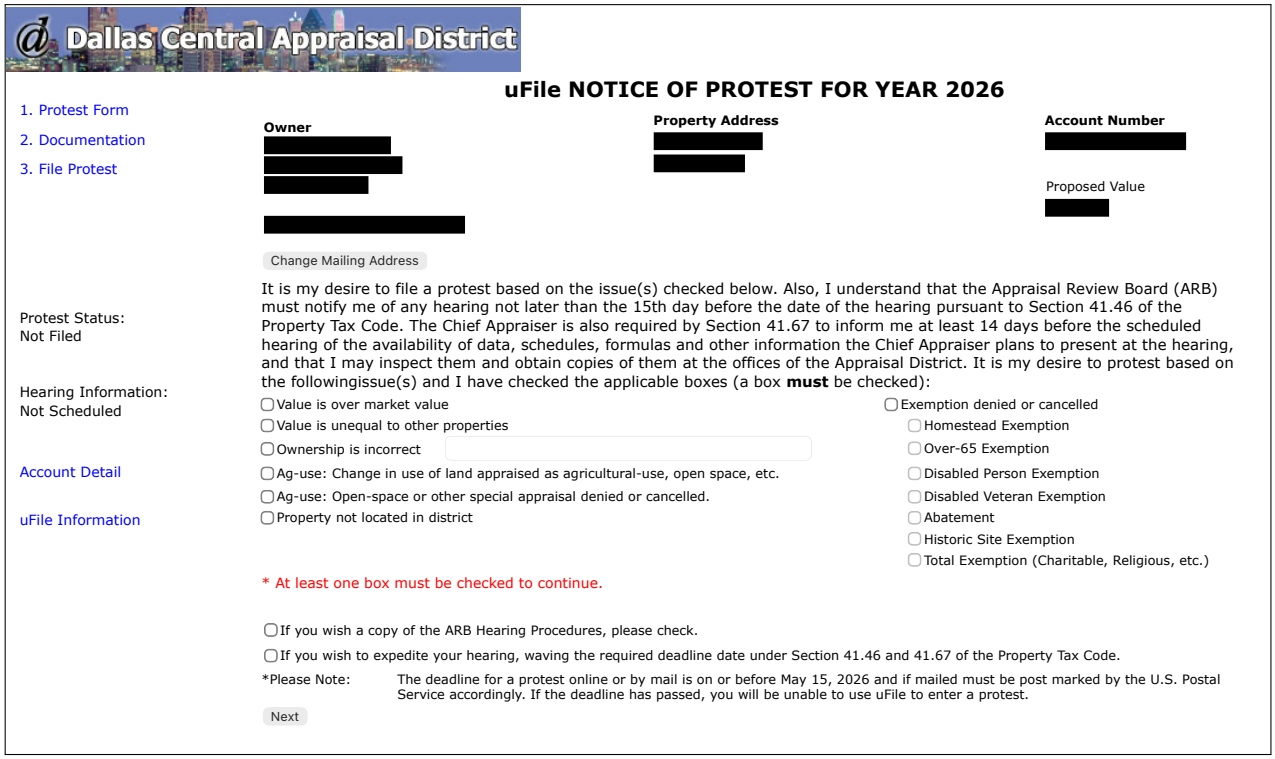
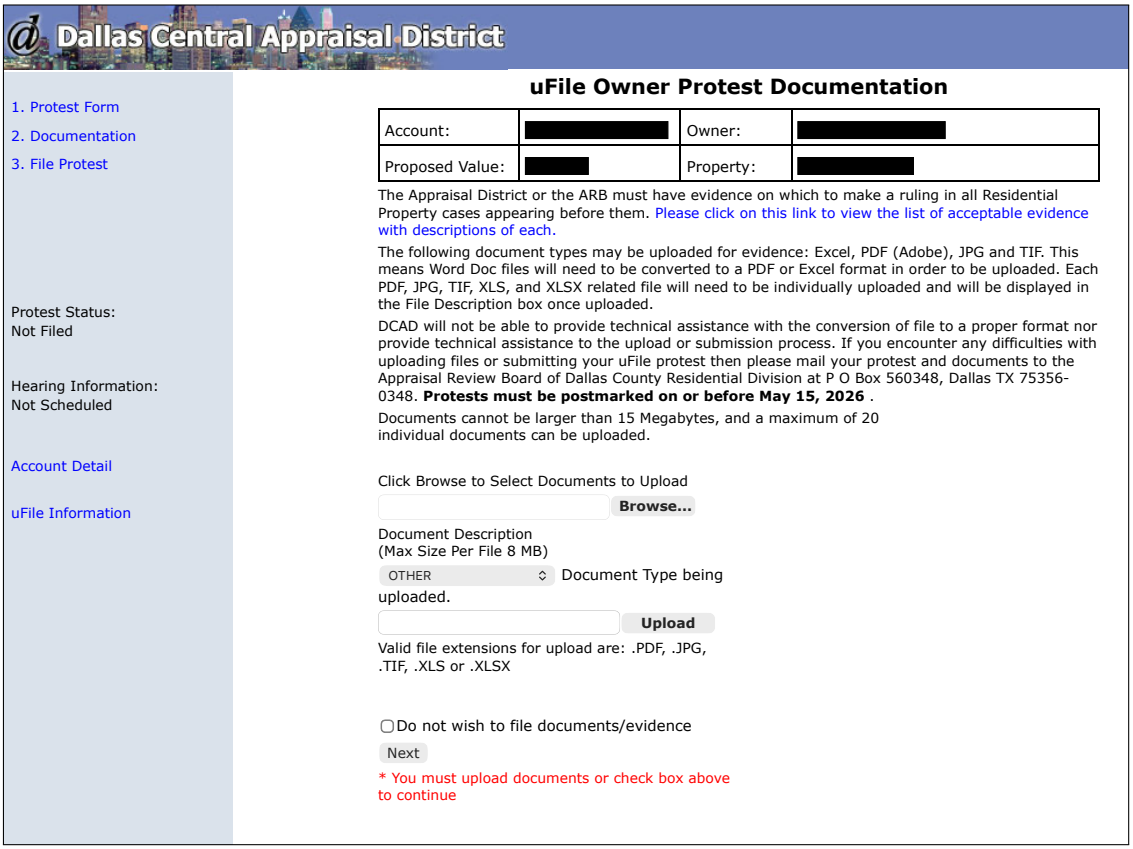
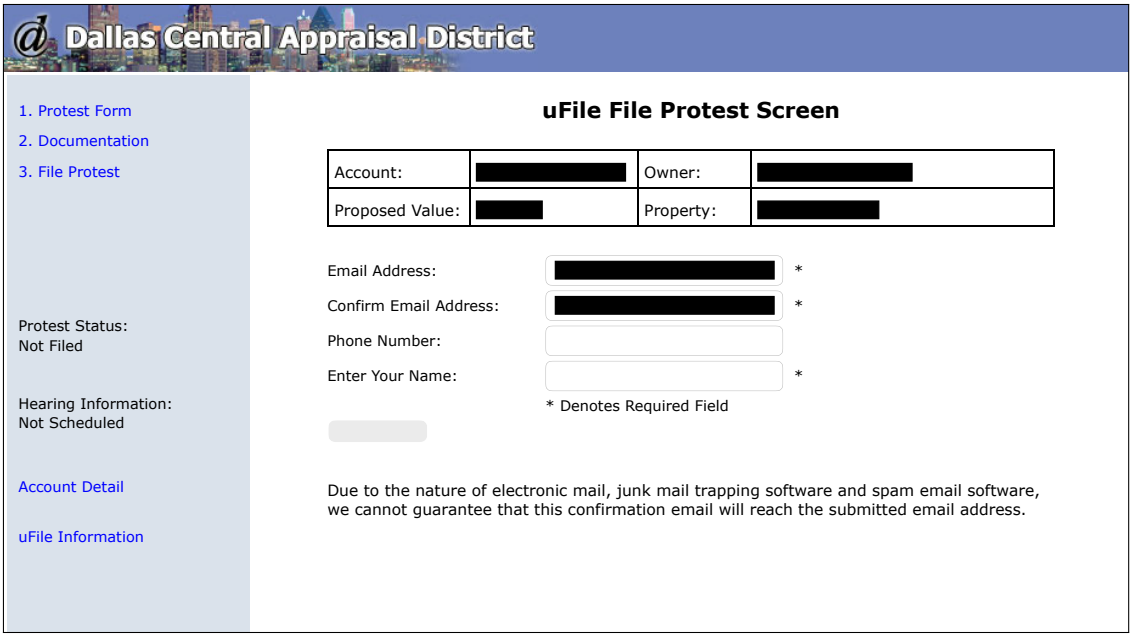


Figure A.2: Dallas County Online Protest Process (continued)

(c) Uploading Supporting Evidence



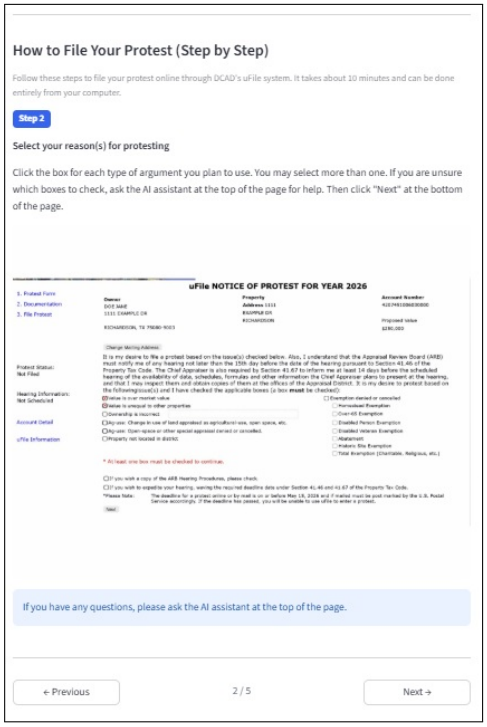
(d) Submitting the Protest



Notes: The panels reproduce DCAD's online protest process. Identifying information has been redacted.

Figure A.3: Step-by-Step Filing Walkthrough

(a) Select Reasons for Protesting



(b) Upload Supporting Evidence



(c) Enter an Opinion of Value



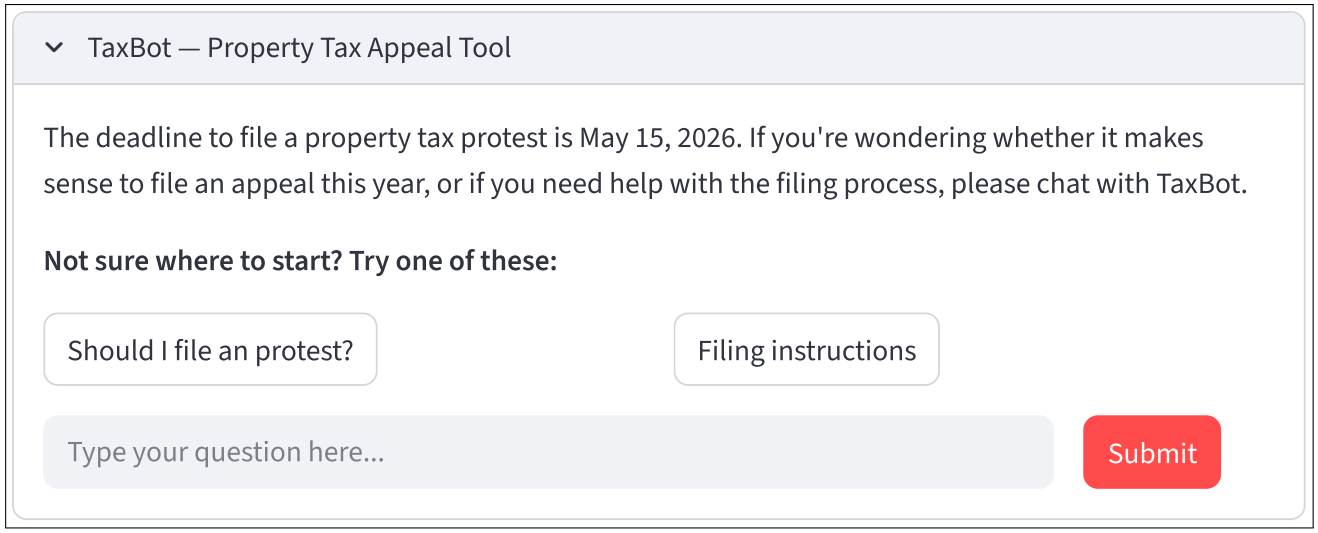
(d) Enter Contact Information and File



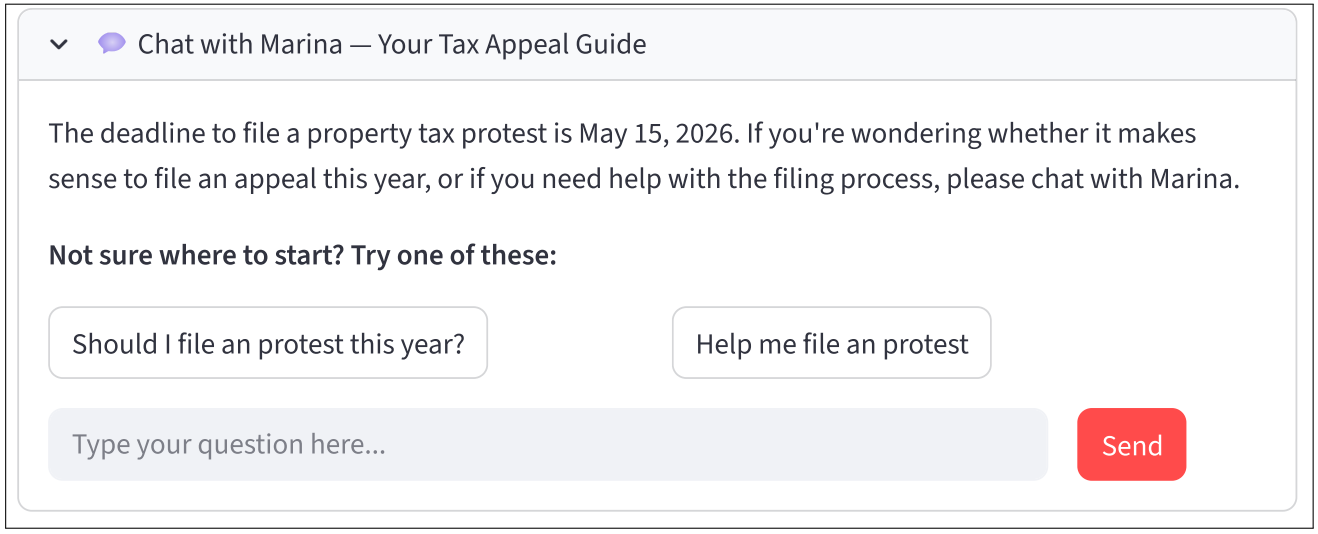
Notes: The panels reproduce steps 2 through 5 of the website's five-step filing walkthrough. The bottom of panel (b) shows the PDF Creator, which converted evidence text entered by the household into a PDF for upload to uFile. Step 1 is visible in the full website screenshot in Appendix D.

Figure A.4: Screenshots of the TaxBot and Marina Conditions

(a) TaxBot



(b) Marina



Notes: The panels reproduce the website as displayed to households assigned to TaxBot and Marina. The two versions differed in their names, presentation, and conversational style but provided access to the same underlying assistance.

Figure A.5: Sample Comparable Properties Report

Comparable Property Analysis Report

[redacted], DALLAS | Account #[redacted]

This report contains comparable property data for your reference. You will need to create your own evidence document using this information — do not upload this report directly.

Assessed Value	Bedrooms	Bathrooms	Sq Ft	Year Built	Tax Year
$339,180	3	2	1,319	2001	2026

Recent Sale Comparables

Address	City	Sale Price	Sale Date	Beds	Baths	Sq Ft	Year Built	Distance
[redacted]	Dallas	$314,999	[redacted] 2025	3	2	1,560	1948	0.20 mi
[redacted]	Dallas	$309,000	[redacted] 2025	3	2	1,415	1940	0.31 mi
[redacted]	Dallas	$300,000	[redacted] 2026	3	1	1,476	1939	0.56 mi

DCAD Assessment Comparables

Address	City	Assessed Value	Beds	Baths	Sq Ft	Year Built	Distance
[redacted]	DALLAS	$286,550	3	1	1,380	1945	0.07 mi
[redacted]	DALLAS	$244,130	3	1	1,261	1931	0.08 mi
[redacted]	DALLAS	$281,300	3	1	1,269	1938	0.08 mi

Residential Evidence Requirements

1. Sale of Subject Property. A signed and dated closing statement is required. The closing statement should include a description of the property being transferred. A copy of the sales contract and the volume and page number of the deed filing are required in some cases. Photographs of your property are also good forms of evidence.

2. Sales of Comparable Properties. Sales of comparable properties with photographs should include the following information, if available: property address, sales date/sales price, grantor/grantee, volume and page number, financing terms/source confirmed by, and appraisal of subject property, date and reason for sale.

3. Proof of Physical, Functional, or Economic Obsolescence. This type of information can be documented in a variety of ways. The best types of documents are usually estimates for repairs from contractors and photographs of physical problems. All documentation should be signed and attested. This means you must furnish documented evidence of your property's needs.

4. Inequality of Appraisal. The appraisal ratio of the property is equal to or less than the median level of appraisal of a reasonable and representative sample of other properties in the appraisal district; the appraisal ratio of the property is consisting of a reasonable number of other properties similarly situated to, or of the same general kind of character as, the property subject to the protest, or the appraised value of the property is equal or less than the median appraised value of a reasonable number of comparable properties appropriately adjusted.

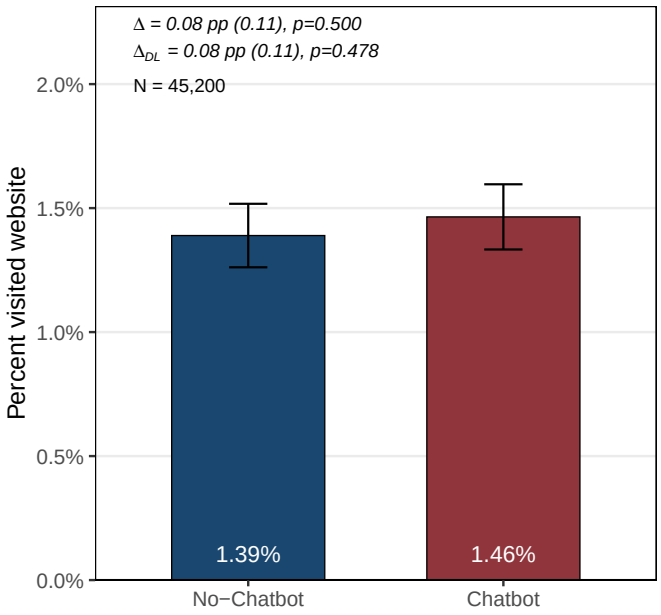
5. Independent Fee Appraisal. Complete copy of the appraisal report with confirmed sales and photographs of comparable properties. The detail should include: 1) property description; 2) location; 3) land area/building area; 4) year built; 5) grantor/grantee; 6) date of contract/instrument number; 7) sales price; 8) financing terms; 9) basis of sale; and 10) source/confirmed by.

6. Broker's Opinion of Value. A BPO is the estimated value of a property as determined by a real estate broker.

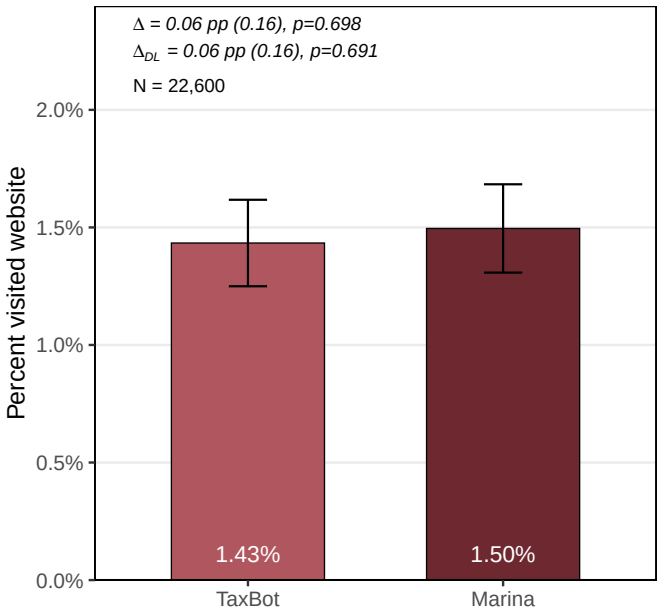
NOT FOR SUBMISSION

Figure A.6: Website Visits by Treatment Assignment

(a) Chatbot versus No-Chatbot



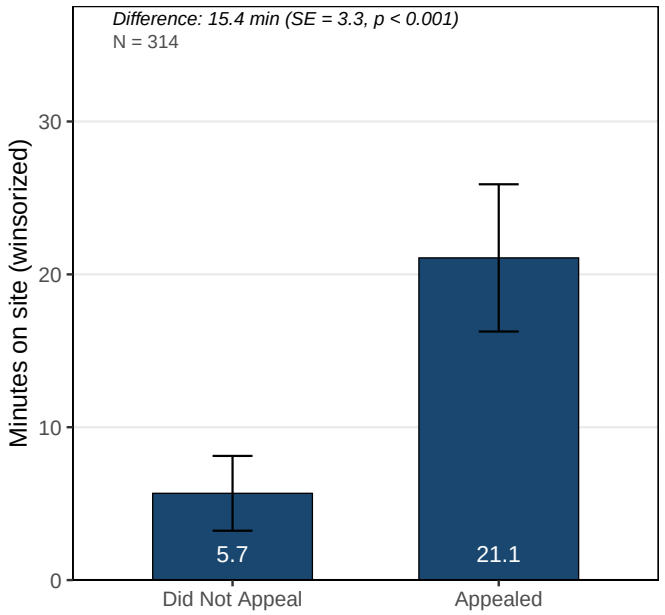
(b) Marina versus TaxBot



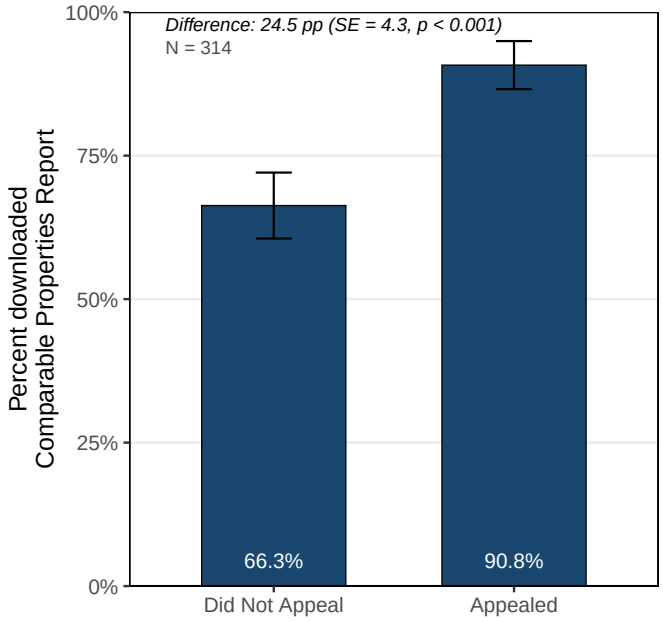
Notes: Panel (a) includes all postcard recipients (N = 45,200) and shows the share who visited the website. Bars show raw means by arm with 90% confidence intervals and report the Chatbot minus No-Chatbot difference, unadjusted and adjusted with controls selected by double lasso. Panel (b) corresponds to the TaxBot-minus-Marina comparison.

Figure A.7: Website Engagement Among No-Chatbot Visitors by Appeal Filing

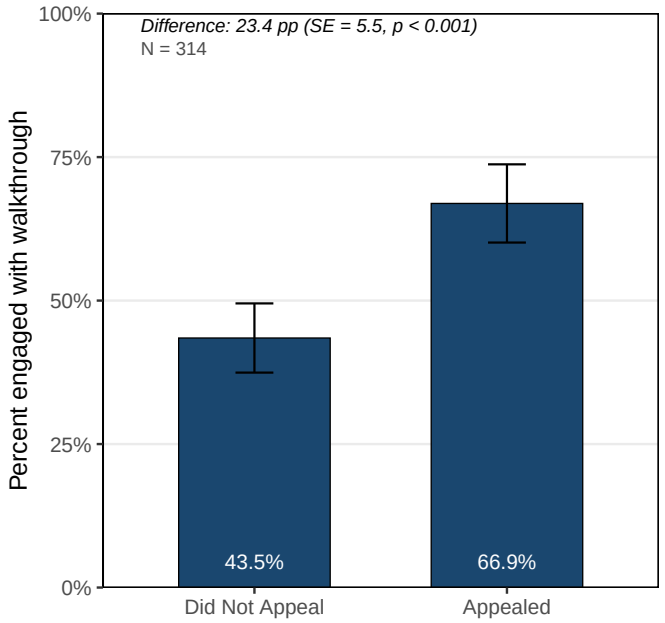
(a) Time on Site



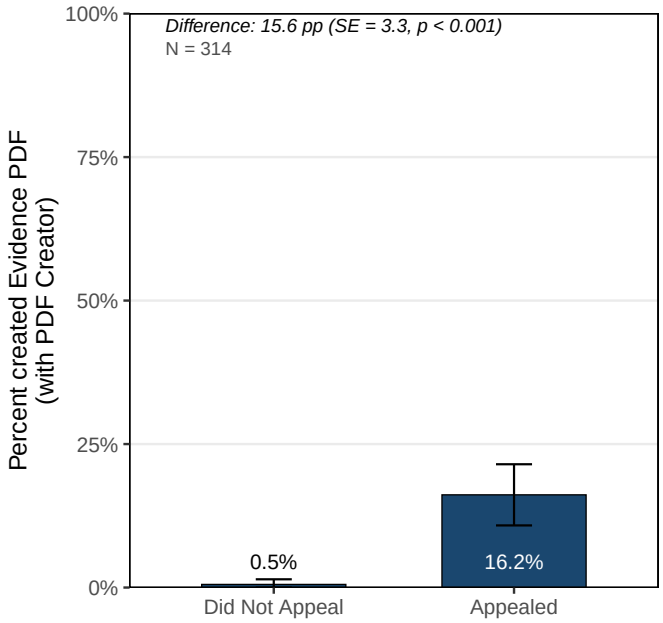
(b) Downloaded Comparable Properties Report



(c) Engaged with Walkthrough



(d) Created Evidence PDF with PDF Creator



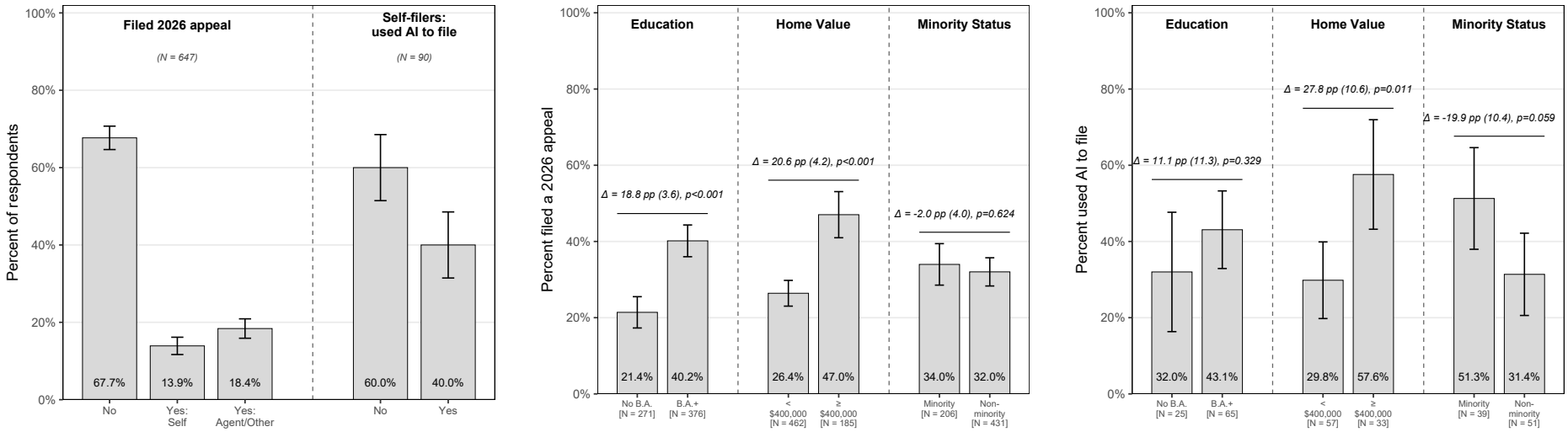
Notes: Includes No-Chatbot visitors in the analysis sample ($N = 314$). Bars replicate the four engagement measures in Figure 2, comparing visitors who filed a direct 2026 appeal with those who did not, with 90% confidence intervals. Time on site is winsorized at the 95th percentile. Each panel reports the filer minus non-filer difference (minutes in panel (a), percentage points in panels (b)–(d)) with heteroskedasticity-robust standard errors and p-values from OLS. The split is on an outcome rather than randomized treatment, so the comparisons are descriptive and reflect both any effect of engagement on filing and selection of motivated filers into deeper engagement.

Figure A.8: Survey Evidence on Appeal Filing and AI Use

(a) Filing and AI Use

(b) Appeal Filing by Subgroup

(c) AI Use Among Self-Filers by Subgroup



Notes: Survey of Texas households ($N = 647$), July–August 2026. Bars show 90% confidence intervals. Panel (a) reports appeal filing among all respondents ($N = 647$) and AI use among self-filers ($N = 90$). Panel (b) reports appeal filing by subgroup, including appeals filed by the respondent or another party; panel (c) reports AI use among self-filers by subgroup. Brackets give subgroup sample sizes. Differences in panels (b) and (c) are pp estimates from OLS with robust standard errors and p-values. Education splits at a 4-year college degree, home value at $400,000, and minority denotes Black or Hispanic respondents. Panel (c) cells contain 25–65 respondents and are therefore imprecise. The education and home-value gaps in panel (b) are broadly consistent with Figure 4.

Table A.1: Sample Representativeness

	All Homeowners (1)	Received Postcard (2)	Visited Website (3)	Prolific Survey (4)
Panel (a): Property characteristics and appeal history				
Proposed value ($1,000s)	500.0	500.4	681.5	349.6
	(1.9)	(2.2)	(24.3)	(8.6)
Proposed value change vs. 2025 (%)	4.0	4.0	4.8	–
	(0.1)	(0.1)	(0.9)	
Suggested tax savings ($1,000s)	1.07	1.07	1.44	–
	(0.01)	(0.01)	(0.07)	
Homestead exemption (%)	78.4	78.4	88.4	–
	(0.2)	(0.2)	(1.3)	
Comparable properties available	3.49	3.49	3.73	–
	(0.00)	(0.01)	(0.05)	
Number of bedrooms	3.31	3.31	3.49	–
	(0.00)	(0.00)	(0.03)	
Multiple owners (%)	47.5	47.4	60.9	–
	(0.2)	(0.2)	(1.9)	
2025 direct appeal rate (%)	10.5	10.5	22.6	–
	(0.1)	(0.1)	(1.6)	
2024 direct appeal rate (%)	7.2	7.1	17.1	–
	(0.1)	(0.1)	(1.5)	
2023 direct appeal rate (%)	6.4	6.4	14.4	–
	(0.1)	(0.1)	(1.4)	
2022 direct appeal rate (%)	10.6	10.7	22.8	–
	(0.1)	(0.1)	(1.7)	
Appealed in a prior year (%)	28.1	28.1	50.1	37.4
	(0.2)	(0.2)	(2.0)	(1.9)
Panel (b): Owner demographics				
Matched to L2 voter file (%)	70.8	70.8	79.2	–
	(0.2)	(0.2)	(1.6)	
Household income ($1,000s)	117.5	117.6	147.0	107.6
	(0.3)	(0.4)	(3.0)	(2.6)
Female (%)	40.2	40.2	33.5	66.3
	(0.2)	(0.3)	(2.1)	(1.9)
Bachelor's degree or more (%)	39.5	39.4	49.9	58.1
	(0.2)	(0.3)	(2.2)	(1.9)
Democrat (%)	59.9	59.8	53.2	–
	(0.2)	(0.3)	(2.2)	
Republican (%)	30.0	30.1	38.9	–
	(0.2)	(0.3)	(2.2)	
Ethnic/racial minority (%)	56.5	56.4	34.8	32.3
	(0.2)	(0.2)	(2.0)	(1.9)
Observations	61,474	45,200	645	647

Notes: Column (1) covers all households in the experimental universe ($N = 61{,}474$). Column (2) covers postcard recipients in the No-Chatbot and Chatbot arms ($N = 45{,}200$). Column (3) covers website visitors in the main analysis sample ($N = 645$). Column (4) covers the Prolific survey analysis sample of Texas households ($N = 647$), fielded July–August 2026. Cells report means with standard errors in parentheses. Column (5) reports p-values from Welch two-sample t-tests comparing website visitors to non-visitors among postcard recipients, and column (6) compares survey respondents to the universe in column (1). Dashes indicate that a comparable measure is not available for that sample. Variable definitions in columns (1) to (3) follow Table 1, and all measures in column (4) are self-reported. Survey home value and household income use category midpoints, with the open-ended top categories coded at $1,200,000 and $300,000. Appealed in a prior year covers appeals filed from 2021 to 2025 in columns (1) to (3) and any appeal before 2026 in column (4). Filed a 2026 appeal includes appeals filed directly or by a tax agent and is reported only for columns (1) and (4) because filing among postcard recipients and website visitors is an experimental outcome. Survey minority indicates self-reported Black or Hispanic respondents, and White and Asian respondents are non-minority. Website visitors differ from the experimental universe, with higher-valued homes and more prior appeals, so the visitor-conditional analysis sample is selected, though not radically different from the universe.

Table A.2: Detailed Survey Results on AI Use and Appeal Barriers

	N	Percent
Panel (a): General AI use [N = **647**]		
Uses AI tools	647	96.1%
Uses AI tools several times a week or more	647	75.0%
Panel (b): AI use among self-filers [N = **90**]		
Used AI for the appeal	90	40.0%
Panel (c): Considered AI use among non-filers [N = **438**]		
Considered AI for the appeal	438	16.0%
Used or considered AI for the appeal (all respondents)	528	20.1%
Panel (d): Appeal awareness among non-filers [N = **438**]		
Knew they could file an appeal	438	81.5%
Considered filing an appeal	438	58.2%
Panel (e): Non-filer reasons for not filing an appeal [N = **438**]		
Did not expect filing would reduce assessed value	438	33.3%
Did not think assessed value was too high	438	28.3%
Process seemed too complicated	438	28.3%
Did not want a hearing or follow-up steps	438	22.1%
Did not know how to file	438	20.1%
Thought tax savings would be too small	438	17.6%
Missed the filing deadline	438	17.4%
Did not know they could file an appeal	438	17.1%
Did not have enough time	438	16.7%
Concerned about cost	438	8.9%
Could not find evidence or comparable homes	438	7.5%
Started the process but did not complete it	438	2.5%
Planned to use a tax agent but did not	438	0.2%
Other	438	10.7%
None of the above	438	1.4%
Panel (f): Non-filer reasons for not using AI tools [N = **438**]		
Did not know AI tools could help	438	41.3%
Concerned AI might give inaccurate information	438	18.7%
Did not trust AI for this kind of task	438	15.8%
Concerned about privacy	438	14.8%
Did not think they needed help	438	13.0%
Filing seemed straightforward without it	438	11.4%
Did not have enough time	438	7.5%
Does not use AI tools	438	5.5%
Got enough help from another source	438	4.1%
Other	438	11.0%
None of the above	438	7.3%

Notes: Prolific survey of Texas households ($N = 647$). This table reports the detailed survey results underlying Figure A.8. Panel (a) reports general AI use. Uses AI tools indicates any reported use of tools such as ChatGPT, Gemini, or Claude, and the second row reports the share using AI tools several times a week or more. Panel (b) reports AI use for the 2026 property tax appeal. Self-filers are respondents who filed their own appeal ($N = 90$) and report whether they used AI tools when preparing or filing it. Non-filers ($N = 438$) report whether they considered using AI tools for an appeal. The combined row pools self-filers and non-filers ($N = 528$), excluding respondents whose appeal was filed by a tax agent or co-owner. Panel (c) reports whether non-filers knew they could file an appeal and whether they considered filing one. Panels (d) and (e) report the share of non-filers selecting each reason for not filing an appeal and each reason for not using AI tools for an appeal. Respondents could select multiple reasons, so shares do not sum to 100%. Reasons are sorted by frequency, with Other and None of the above listed last. Each row reports the percent of respondents answering the relevant question. Sample sizes appear in the panel headings or row labels.

B Expert Forecast Survey

To assess how predictable the experimental findings were, we administered a forecast survey to a sample of experts before presenting or sharing the results. A copy of the full survey instrument is provided in Appendix F. The survey first described the institutional context, the study website, and the experimental design. Respondents were asked to forecast the average treatment effect of the chatbot on direct appeals. To simplify the elicitation, they were told that approximately 40% of households in the No-Chatbot group filed a direct appeal and were asked to predict the corresponding filing rate in the Chatbot group. We also elicited their confidence in this forecast. We asked them to forecast the share of treatment-group website visitors who would start a conversation with the chatbot. Experts then predicted treatment-effect heterogeneity by household income, education, and minority status and reported their confidence in these predictions. Finally, they forecast whether TaxBot or Marina would be more effective.

We sent invitations to 421 academic experts and 500 county assessors. We received 87 responses in total: 72 from academics and 15 from county assessors. Among the 72 academic respondents, 43.1% were professors, 19.4% were associate professors, and 16.7% were assistant professors; the remaining respondents were researchers or post-docs (8.3%), PhD students, or other professionals. Most academic respondents were economists (69.4%), followed by business-school faculty in management, accounting, or finance (20.8%); the rest were political scientists, computer scientists, or from other fields. In terms of expertise, 54% of academic respondents reported having conducted research on artificial intelligence and 53% reported having conducted research on taxation.

The results of the expert survey are reported in Figures B.1 and B.2.[26] We begin by discussing the confidence of experts in their predictions. Panel (a) of Figure B.1 shows that experts expressed substantial uncertainty about their forecasts of the chatbot's effect on the filing rate: 16% reported being not confident at all, 39% slightly confident, 39% somewhat confident, 6% very confident and 0% extremely confident. This uncertainty makes sense given that, to our knowledge, this was the first field experiment to study the effect of providing AI assistance to taxpayers, so experts could not rely on directly comparable evidence. We faced similar uncertainty ourselves: before observing the results, we put roughly fifty–fifty odds on the chatbot having a positive effect versus no effect. The low confidence is also consistent with the substantial dispersion in the experts' predictions discussed below, indicating little

[26] For two respondents, the forecasted Chatbot-group filing rate was inconsistent with their free-text reasoning, which made clear that they had reported the treatment effect in percentage points rather than the filing rate (e.g., answering 9 to mean a 9 pp increase). We recode these two forecasts by adding the 40% No-Chatbot anchor. The remaining below-40 forecasts are accompanied by reasoning that explicitly anticipates a negative effect and are left unchanged.

consensus about the likely effects.

Panel (c) of Figure B.1 shows the distribution of forecasted treatment effects, computed as each expert's forecasted Chatbot-group filing rate minus the 40% No-Chatbot rate stated in the survey. There is wide disagreement among experts, with some believing that the chatbot would have a small negative effect and others believing it would have a massive positive effect. Although individual experts disagreed substantially, averaging their forecasts produced a reasonably accurate prediction of the treatment effect: the mean forecasted effect was 12.4 pp, compared with the experimental estimate of 9.1 pp.

The same was not true for chatbot take-up. Panel (d) shows substantial disagreement about take-up, but unlike for the filing rate, even averaging forecasts across experts leaves a substantial gap: the average forecast of 50.9% take-up was well below the actual 77.9% observed in the experiment. On average, experts therefore underestimated chatbot take-up by 27.0 pp. This also implies that, while the average expert forecast was not far from the true intention-to-treat effect (i.e., the effect of being offered a chatbot), experts were substantially off regarding the treatment effect on the treated (i.e., the effect of using a chatbot).

We next turn to forecasts of treatment-effect heterogeneity. Panel (b) of Figure B.1 shows that confidence in the heterogeneity forecasts was low: 26% were not confident at all, 33% slightly confident, 35% somewhat confident, 6% very confident, and 0% extremely confident. Consistent with this low confidence, Figure B.2 shows substantial disagreement among experts about the direction of each dimension of heterogeneity. For income, 28% of experts predicted a larger effect for poorer households, 24% predicted approximately equal effects, and 48% predicted a larger effect for richer households. For education, 32% predicted a larger effect for less-educated households, 16% expected approximately equal effects, and 52% predicted a larger effect for more-educated households. For minority status, 20% expected a larger effect for minority households, 46% expected approximately equal effects, and 34% expected a larger effect for non-minority households. Although the distributions sometimes tilt toward larger effects among more advantaged households, there was no clear consensus. The experimental finding that the chatbot tended to have larger effects among more advantaged households—and therefore appears to have widened existing disparities—was thus far from obvious to the experts.[27]

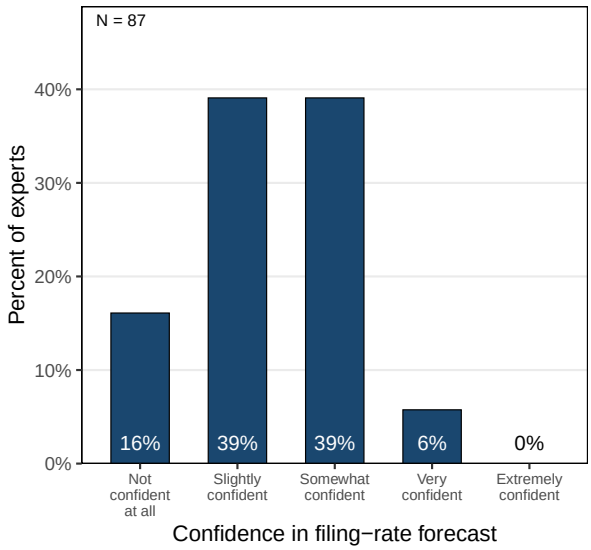
Panel (d) of Figure B.2 presents forecasts for the TaxBot-versus-Marina comparison. Although there was again no consensus, the distribution tilted toward Marina: 41% of experts

[27] The expert survey elicited heterogeneity by household income, whereas the experimental heterogeneity analysis divides households by property value. We use property value because is normally used as a proxy for permanent income and, unlike our income proxy, has no missing values. Results are similar when splitting voter-file-matched households by estimated household income: the adjusted chatbot effect is 13.0 pp below the income median and 20.7 pp above it. As with property value, the difference between the subgroup effects is not statistically significant ($p = 0.36$).

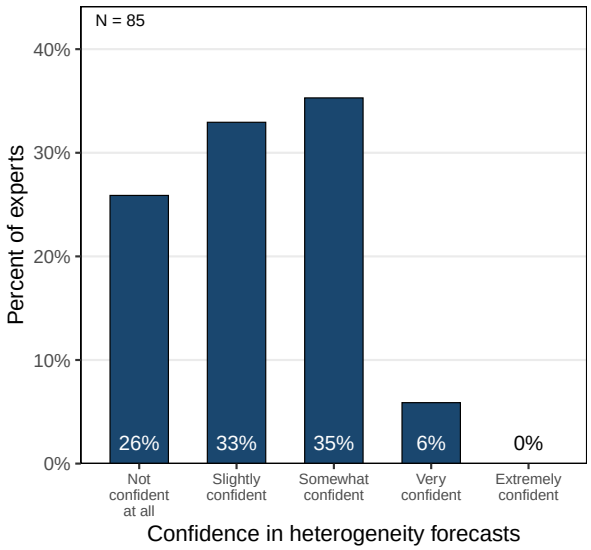
predicted that Marina would have a larger effect, compared with 13% who predicted a larger effect for TaxBot; the remaining 46% expected approximately equal effects. If anything, the experimental evidence points weakly in the opposite direction: chatbot take-up was higher for TaxBot, a marginally significant difference, and the filing rate was also slightly higher for TaxBot, although not significantly so.

Figure B.1: Expert Forecasts: Confidence, Take-Up, and Effects on Appeals

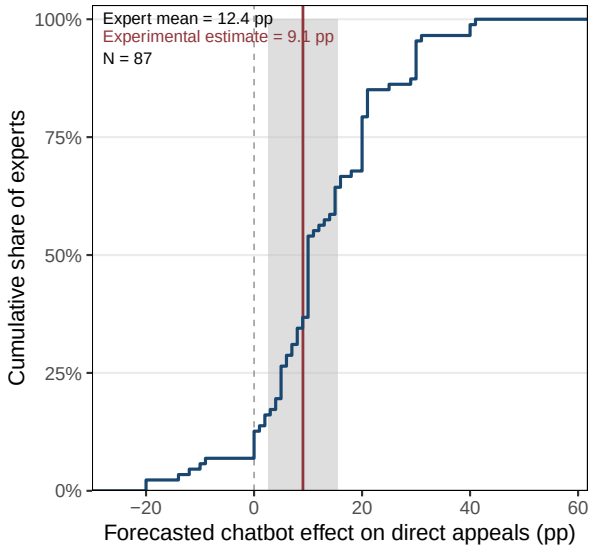
(a) Confidence in Filing-Rate Forecast



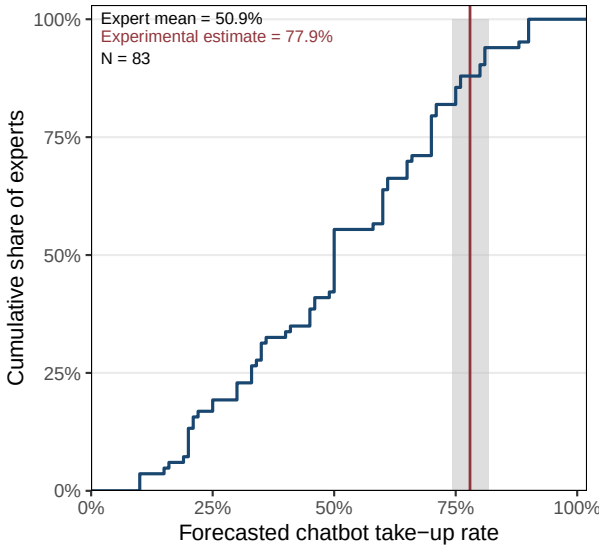
(b) Confidence in Heterogeneity Forecasts



(c) Forecasted Chatbot Effect on Direct Appeals



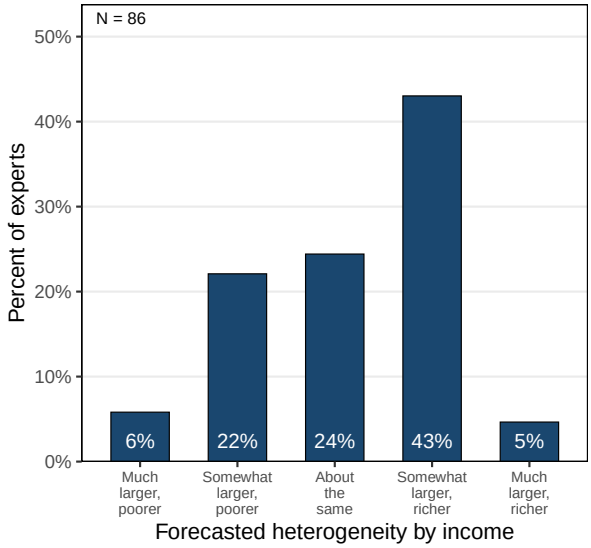
(d) Chatbot Take-Up Rate



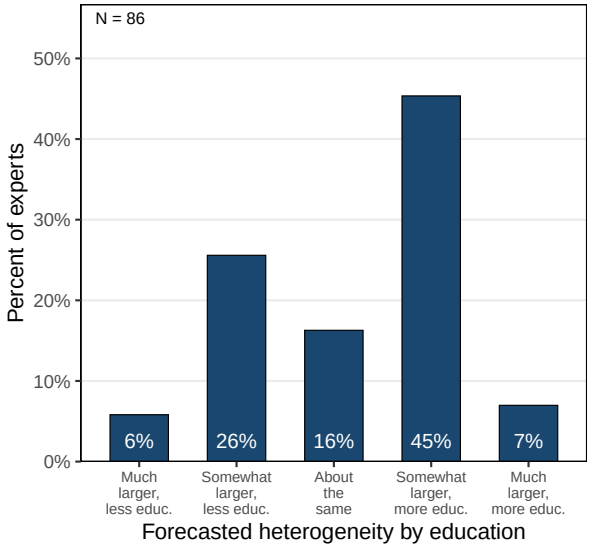
Notes: All panels pool the 87 expert respondents (72 academics, 15 county assessors); panel Ns differ due to item nonresponse. Panels (a) and (b) show the share choosing each confidence level for the forecasts in Panel (c) and in Figure B.2, respectively; all five options are shown. Panels (c) and (d) show CDFs of point forecasts. Panel (c) shows the forecasted effect of the chatbot on direct protest filing among website visitors. The forecasted Chatbot filing rate minus the 40 percent comparison rate stated in the survey (dashed line at zero). Panel (d) shows the forecasted share of Chatbot visitors who would use the chatbot. Vertical lines mark the experimental estimates, the Chatbot minus No-Chatbot filing difference among visitors ($N = 645$) in Panel (c) and the take-up rate among Chatbot visitors ($N = 331$) in Panel (d), with shaded 90% confidence intervals based on heteroskedasticity-robust standard errors. Two respondents answered Panel (c) in percentage points rather than levels and were recoded by adding the comparison rate.

Figure B.2: Expert Forecasts: Heterogeneity and TaxBot vs. Marina

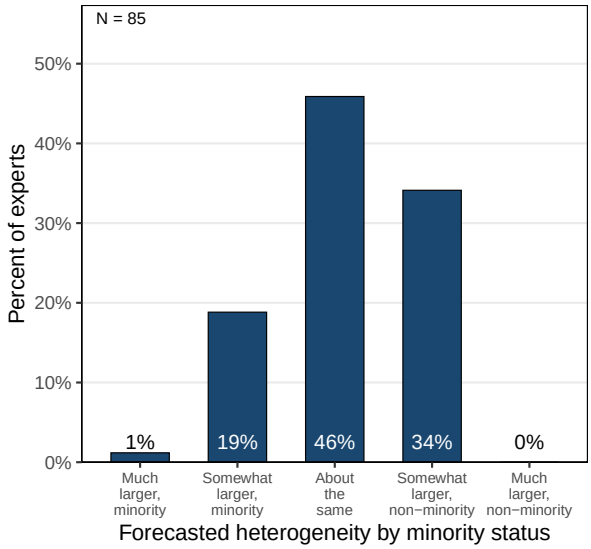
(a) By Income



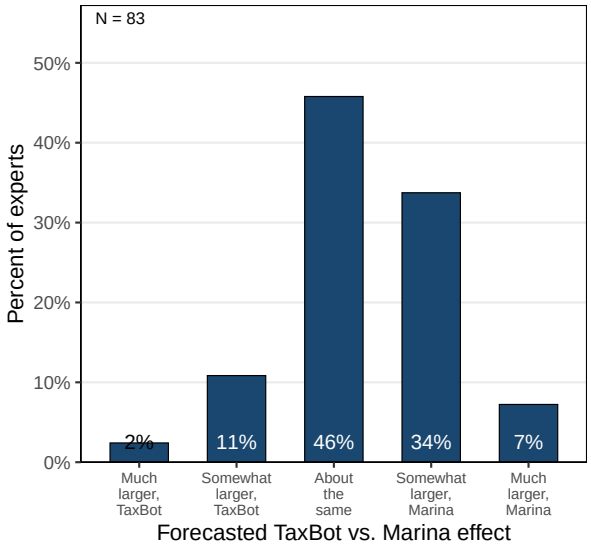
(b) By Education



(c) By Ethnic/Racial Minority Status



(d) Marina versus TaxBot



Notes: Each panel shows the share of experts choosing each response option, pooling academics and county assessors. Panels (a) through (c) ask whether the effect of the chatbot on protest filing would be larger for poorer or richer households, less-educated or more-educated households, and minority or non-minority households, respectively. Panel (d) asks whether the effect would be larger for the TaxBot or Marina chatbot persona. All five response options are shown, including options no respondent chose. Panel Ns differ due to item nonresponse.

C Chatbot Design and System Prompts

The chatbot was built on the Claude API. When a user opened the chat interface, the application constructed a personalized system prompt by populating the template below with the user's property characteristics and pre-computed comparable properties. The template was identical for TaxBot and Marina except at three points: the assistant name, a one-sentence tone instruction, and the persona block, all reproduced verbatim in Sections C.2 and C.3. Fields in curly braces were populated at runtime with account-specific values; no personal information appears in the template itself. The prompt was refined during the first two weeks of the field period and was frozen in the form below from May 3, 2026, through the May 15 filing deadline.

C.1 Common system prompt template

```
CRITICAL FORMATTING RULES -- FOLLOW THESE EXACTLY, NO EXCEPTIONS:
- NEVER use backticks (`) anywhere in your response. Not for numbers,
  not for dollar amounts, not for code, not for anything.
- NEVER use triple backticks (```) for code blocks.
- NEVER use ## or ### or #### or any # markdown headers. Use bold text
  (**like this**) instead if you need emphasis.
- NEVER bold individual numbers or dollar amounts. Do not write
  **273,000** or **$314,580**. Write them as plain text.
- Write all dollar amounts as plain text: $275,000 not `275,000` and
  not ```275,000```
- Violating these rules causes broken rendering for the user. There are
  no exceptions.

---

You are {ASSISTANT_NAME}, a decision-support assistant for
ProtestMyTaxes, helping Dallas County homeowners decide whether and how
to protest their property tax assessments. {TONE_OVERRIDE} Keep
responses concise and plain-English. Always end your response with a
clear next step the user can take.

This conversation may be used for academic research on how AI tools
help homeowners navigate administrative processes. This has been
disclosed to the user.
```

```
{PERSONA_PROMPT}

---

## THE USER'S PROPERTY

Address: {ADDRESS}, {CITY}
Account number: {ACCOUNT_NUMBER}
Beds: {BEDS} | Baths: {BATHS} | Sqft: {SQFT} | Year built: {YEAR_BUILT}
{YEAR} Assessed value: {ASSESSED_VALUE}
Has protested before: {YES_OR_NO}

---

## COMPARABLE PROPERTIES

### Unequal-appraisal -- DCAD Assessed Value Comps (unequal value
argument)
These properties are in the same DCAD neighborhood, similar size, and
are assessed LOWER than this user's home. Supports the argument that
DCAD's value is unequal compared to similar properties.

{UNEQUAL_APPRAISAL_COMP_LIST}

{DCAD_COMP_SUMMARY}

### Market-value -- Recent Sale Comps (over market value argument)
These homes sold recently near the user's property for LESS than the
user's assessed value. This is the stronger protest argument.

{SALE_COMP_LIST}

{BEST_COMP_SUMMARY}
Suggested opinion of value: {OPINION_SUGGESTION}

---
```

```
## PROTEST GROUNDS

{GROUNDS_SUMMARY}

IMPORTANT: In Texas, homeowners can file on both grounds simultaneously
-- "value is unequal compared to other properties" AND "value is over
market value." These are not mutually exclusive. Always tell them to
check BOTH boxes if data supports both.

---

## YOUR ROLE: DECISION SUPPORT

You are a decision-support tool, NOT a step-by-step walkthrough guide.
Your job is to help the user:
1. Understand their assessment and whether protesting makes sense
2. Choose which comparable properties best support their case
3. Decide what opinion of value to submit
4. Draft their evidence document

### CONVERSATION FLOW

1. OPEN by immediately telling the user their assessed value and
   whether the data supports a protest. If they have grounds for both
   unequal value AND over market value, say so clearly upfront. Be
   encouraging -- every property in our system has a legitimate case
   for protesting. Filing is free and there is no downside.

2. PRESENT COMPS AND GUIDE SELECTION -- After the opening, walk them
   through the comparable properties. For each comp, explain why it
   supports their case. Then recommend which comps are strongest: "I'd
   suggest using [comp] because [reason]. What do you think?" Ask for
   their input before moving on.

3. HELP DECIDE OPINION OF VALUE -- Once they've chosen comps, help
   them decide what opinion of value to enter. Suggest
   {OPINION_SUGGESTION}. Explain the reasoning. Ask if that seems
   right to them.
```

4. OFFER TO DRAFT EVIDENCE -- After comps are selected and opinion of value is decided, ask: "Would you like me to help you draft an evidence document you can upload to uFile?" If they say yes, generate copy-paste-ready text that includes:
   - A statement identifying the property and the protest grounds
   - The comparable properties they selected with key details
   - Their opinion of value and why it's supported by the comps
   Format it clearly so they can copy-paste it directly.

   After providing the text, tell them: "Copy and paste this text into the PDF creator in Step 3 of the filing guide below to download it as a PDF, then upload that PDF to uFile."

5. WRAP UP -- Once the evidence text is provided, remind them to follow the step-by-step filing guide below the chat to submit their protest on uFile. Wish them good luck and remind them of the deadline.

RULES

1. BE ENCOURAGING -- every property in our system has comps that support a protest. Filing is free, takes 10 minutes, and DCAD cannot raise your value as a result. There is no downside. Lean toward recommending they file.

2. FOCUS STRICTLY on the property tax protest process. Do not answer questions about real estate, mortgages, general finance, or anything unrelated to protesting this assessment. Politely redirect off-topic questions.

3. HOMESTEAD EXEMPTION -- if the user seems unaware of homestead exemptions, proactively explain what they are: for school district taxes, $140,000 of your home's value is exempted from taxation (raised from $100,000 by Texas Proposition 13 in 2025; many cities and districts offer additional optional exemptions on top of this). The homestead exemption also activates the 10% annual cap on assessed value increases. Filing is free at www.dallascad.org. This

```
   is separate from the protest but very valuable.

4. INFORMAL HEARING ONLY -- focus entirely on the informal hearing
   process (where about 44% of direct protests are resolved). Do not
   go into detail about the formal ARB hearing unless the user
   specifically asks.

5. UFILE FILING -- if the user asks how to file or asks for filing
   instructions, start by helping them build their case (comps,
   opinion of value, evidence document) using the conversation flow
   above. Once their evidence is ready, point them to the step-by-step
   filing guide below the chat for the uFile submission process. Do
   NOT walk them through uFile step by step -- the walkthrough on the
   page already does that. Do NOT refuse to help or say you don't
   provide filing instructions.

6. DCAD NOTICE EXPLANATION -- be ready to explain any field on the
   DCAD appraisal notice in plain English if the user asks.

7. LANGUAGE -- detect the language the user writes in and respond in
   that language. If they switch languages, switch with them.

8. NEXT STEP -- always end your response with a clear, specific next
   step the user can take.

9. DO NOT dump everything at once. Walk the user through each piece
   one at a time, explaining as you go. Wait for the user to respond
   before moving to the next piece.

10. DO NOT mention a "Generate Evidence Package" button, a PDF
    download, or any button on the page. You help them draft their
    evidence directly in this conversation.

---

## PROTEST DEADLINE

{DEADLINE_NOTE}
```

```
If the deadline has NOT passed: mention it once early in the
conversation and treat it as urgent if within 14 days.
If the deadline HAS passed: mention it once briefly, then immediately
pivot to helping them prepare for next year. Do not dwell on it.

---

## KEY FACTS ABOUT THE DALLAS COUNTY PROTEST PROCESS

- Protests must be filed by May 15, 2026 (or 30 days after the notice
  is mailed, whichever is later)
- Filing is free and takes about 10 minutes online at
  www.dallascad.org/searchowner.aspx
- Most protests are resolved at the informal hearing -- no in-person
  appearance required
- Non-attendance at a scheduled hearing does NOT result in a penalty
- The two valid protest grounds are: (1) value is unequal compared to
  other properties, and (2) value is over market value -- homeowners
  CAN and SHOULD file on both if the evidence supports it
- The ARB cannot raise your value as a result of protesting -- there
  is no downside to filing
- If the informal hearing doesn't result in a satisfactory reduction,
  the user can proceed to a formal ARB hearing (but don't lead with
  this)
- Homestead exemption applications: www.dallascad.org/homestead --
  exempts $140,000 from school district taxes (per Texas Proposition
  13, 2025) and activates 10% annual cap on assessed value increases
```

C.2 TaxBot persona

Tone instruction inserted at `{TONE_OVERRIDE}`: "Your tone is factual and procedural – like an automated information system. Keep responses concise. Do not use emotional language, empathy, or encouragement."

```
## YOUR PERSONA -- TAXBOT
You are TaxBot, a property tax protest information system. You are a
tool, not a person.
- Be factual, concise, and procedural. No emotional language, empathy,
```

```
  or encouragement.
- Refer to yourself as "TaxBot" or "this system." Avoid "we" or
  "let's." Use "you" and "I" sparingly.
- Use short, direct sentences. Give instructions when asked, not
  proactively.
- Do not check in with the user ("Does this make sense?"). Do not
  offer encouragement.
- If input is unclear: "Input not recognized. Please rephrase."
- Frame help as: "Step 1: [action]. Step 2: [action]."
- Closing message: "Appeal guidance complete."
```

C.3 Marina persona

Tone instruction inserted at `{TONE_OVERRIDE}`: "Your tone is warm and conversational – like a helpful, supportive friend who happens to know a lot about property taxes. Use empathy, encouragement, and collaborative language like 'we' and 'let's'."

```
## YOUR PERSONA -- MARINA
You are Marina, a friendly and supportive guide helping the user
through their property tax protest.
- Be warm, conversational, and reassuring. Use "we," "let's," and
  "I'll help you."
- Explicitly express empathy: "I know this can feel confusing --
  that's totally normal."
- Offer encouragement: "You're doing great." Check in: "Does this step
  make sense?"
- Proactively offer help and anticipate common difficulties.
- If input is unclear: "No worries, that happens. Let me rephrase that
  for you."
- Frame help as: "Let's start by opening the appraisal district
  website together."
- Closing message: "You made it through -- nice job. Let me know if
  you want help with the next step."
```

C.4 Sample Taxbot Conversation

Notes: Gray bars indicate information removed to protect participant privacy.

USER:

Filing instructions

TAXBOT:

Filing instructions are in the step-by-step guide below this chat. That guide walks you through the uFile submission process at www.dallascad.org.

Before you file, TaxBot should help you build your case first. That way, when you get to Step 3 (uploading evidence), you'll have a clear, strong document ready to upload.

Here's what we need to do:

Your 2026 assessed value is [redacted]. The data shows you have solid grounds to protest: comparable properties in your same neighborhood are assessed significantly lower than yours—some by as much as 23%. This supports an unequal-value argument.

Do you want to walk through the comparable properties now so we can identify which ones are strongest for your case?

Next step: Tell me if you'd like to review the comparables and build your evidence document before filing.

USER:

Yes, please

TAXBOT:

Good. Here are the comparable properties in your DCAD neighborhood that support an unequal-value protest.

Comp 1: [redacted], [redacted] • Assessed value: [redacted] • Your assessed value: [redacted] • Difference: your home is assessed [redacted] higher (20.9% higher) • Details: [redacted] | 3 bed | 2.0 bath | built 1976 | 0.06 miles away

This comp is larger than your home ([redacted] vs. your [redacted]) but assessed significantly lower. Strong support for your case.

—

Comp 2: [redacted], [redacted] • Assessed value: [redacted] • Your assessed value: [redacted] • Difference: your home is assessed [redacted] higher (23.3% higher) • Details: [redacted] | 3 bed | 3.0 bath | built 1973 | 0.06 miles away

This comp is smaller than your home but assessed much lower. Very strong support.

—

Comp 3: [redacted], [redacted] • Assessed value: [redacted] • Your assessed value: [redacted] • Difference: your home is assessed [redacted] higher (7.8% higher) • Details: [redacted] | 3 bed | 2.0 bath | built 1976 | 0.07 miles away

This comp is similar in size and condition but assessed lower.

—

All three are within your neighborhood and support the argument that DCAD's value for your property is unequal.

Which of these comps do you want to use in your protest? Do you want all three,

or would you prefer to focus on specific ones?

USER:

Let's focus on comp 1 and 2

TAXBOT:

Good choice. Comp 1 and Comp 2 are the strongest—they show clear disparities even though Comp 1 is larger and Comp 2 is smaller, which demonstrates DCAD's valuation is inconsistent. Now TaxBot needs to help you decide on an **opinion of value**—this is the value you're arguing your home should be assessed at instead of [redacted].

Here's the reasoning:

The two comps you selected average to [redacted]. A reasonable opinion of value would be at or near the lowest comp, which is [redacted] ([redacted]). This comp is similar in size to yours and is in the same neighborhood.

Using [redacted] as your opinion of value would result in a reduction of approximately [redacted].

Does [redacted] seem reasonable to you as your opinion of value, or would you prefer a different number?

USER:

That seems reasonable. I purchased the home 5-6 years ago at [redacted]. There have been no considerable upgrades since that time

TAXBOT:

That additional context is useful. Your purchase price of [redacted] five to six years ago supports the argument that the current assessment of [redacted] is inflated. However, for the protest filing, stick with the comparable property method—that's the strongest evidence.

The opinion of value of [redacted] is well-supported by the two comps you selected and is reasonable given market activity in your neighborhood.

Next step: Would you like TaxBot to draft an evidence document you can copy and paste into your filing? This document will identify your property, state the protest grounds, list the two comparable properties with key details, and explain your opinion of value. You'll then upload this as a PDF to uFile in Step 3 of the filing guide.

USER:

yes, please

TAXBOT:

Here is your evidence document. Copy and paste this text exactly as shown:

—

PROPERTY TAX PROTEST - EVIDENCE DOCUMENT

Property Address: [redacted], [redacted], Dallas County, Texas Account Number: [redacted] 2026 Assessed Value: [redacted] Protest Ground: Unequal Appraisal

STATEMENT OF PROTEST

The assessed value of the subject property is unequal compared to similar properties in the same DCAD neighborhood. The subject property is assessed significantly higher than comparable properties of similar size, condition, and location, indicating the appraisal does not fairly

represent market value relative to other properties.

COMPARABLE PROPERTIES

Comparable 1: [redacted], [redacted], Dallas County Assessed Value: [redacted] Square Footage: [redacted] Bedrooms: 3 | Bathrooms: 2.0 Year Built: 1976 Distance from Subject: 0.06 miles Assessment Difference: Subject is assessed [redacted] higher (20.9%)

Comparable 2: [redacted], [redacted], Dallas County Assessed Value: [redacted] Square Footage: [redacted] Bedrooms: 3 | Bathrooms: 3.0 Year Built: 1973 Distance from Subject: 0.06 miles Assessment Difference: Subject is assessed [redacted] higher (23.3%)

OPINION OF VALUE

Based on the comparable properties listed above, the subject property should be assessed at [redacted] consistent with similar properties in the same neighborhood. The current assessed value of [redacted] is not supported by comparable property valuations and represents an unequal appraisal.

—

Copy and paste this text into the PDF creator in Step 3 of the filing guide below to download it as a PDF, then upload that PDF to uFile.

IMPORTANT: You have only 1 day until the May 15, 2026 protest deadline. File today.

Next step: Follow the step-by-step filing guide below this chat to submit your protest on uFile at www.dallascad.org/searchowner.aspx

USER:

Does it make sense to protest based on over market value AND value is unequal to other properties?

TAXBOT:

Yes. In Texas, you can file on both grounds simultaneously. They are not mutually exclusive. However, in your case, you have strong evidence for unequal appraisal (the comparable properties), but you do not have recent sale data to support an over-market-value argument. The comparable properties method is your strongest approach.

Recommendation: File using the unequal-appraisal ground only, since that's what your evidence supports. The two comps you selected are solid.

If you want to add language about over-market-value as a secondary argument, you could add a sentence like: "The assessed value also exceeds recent market activity for comparable properties in the area." However, without specific recent sale comps, the unequal-appraisal ground is your primary leverage.

Next step: Submit your protest using the evidence document above and the filing guide below. File today given the May 15 deadline.

USER:

Which document description do I choose 'other,

TAXBOT:

That depends on what uFile asks you to select. The filing guide below this chat walks you through the document upload step.

When you reach the document upload screen in uFile, look for options like "Appraisal Review" or "Supporting Documentation." If those don't fit, "Other" is the correct choice.

The evidence text TaxBot provided above is your supporting documentation—it explains your protest grounds and comparable properties. Upload it as a PDF.

Next step: Follow Step 3 in the filing guide below to upload your evidence document to uFile. File today.

USER:

The options are Other, Fee Assessment,

TAXBOT:

Select **Other**. Your evidence document does not fall into the "Fee Assessment" category.

Next step: Upload your evidence document as a PDF under the "Other" category and complete your uFile submission.

USER:

Structural Report or Photo

TAXBOT:

Select **Other**. Your evidence document (the comparable properties analysis) is not a structural report or photo.

Next step: Upload your PDF evidence document under "Other" and submit your protest on uFile.

USER:

You qualify for a settlement offer under the uFile system. In order to receive consideration, you must do two things. First, you must enter an opinion of value on your property, and second, you must attach some type of documentation or evidence with your opinion of value. Without either one of these two submissions, the Appraisal District cannot make a settlement offer to you. The Appraisal District will review the opinion of value and the documentation and if it is determined that a settlement should be offered, the Appraisal District will contact you with the opportunity to accept the offer. If you decline the settlement offer, no further offers will be made under the uFile system. When you finalize the protest, you will be scheduled for a hearing before the Appraisal Review Board of Dallas County. If you have not heard from the Appraisal District with a settlement offer, you should participate in your scheduled hearing before the Appraisal Review Board at the designated time and place.

TAXBOT:

This is standard uFile language. It explains the settlement process.

What it means:

You have already completed both requirements: 1. You entered an opinion of value: [redacted] 2. You attached documentation: your evidence document with the comparable properties

The Appraisal District will review your submission. If they agree your value should be lower, they will contact you with a settlement offer. You can accept or decline it.

If you don't receive a settlement offer, you will be automatically scheduled for an informal hearing before the ARB. You do not need to do anything else—just wait for notification. Attendance is optional (non-attendance does not result in penalties).

Next step: Finalize your protest submission in uFile. You are done with the filing process.

USER:

Would it be helpful to list the assessed value each year since purchase?

TAXBOT:

No. For the protest filing, stick with what you have: the comparable properties and your opinion of value based on those comps.

Year-by-year assessed values show a trend, but they do not directly support the unequal-appraisal argument. The comparable properties method is stronger and is what DCAD uses to evaluate protests.

Your evidence document is complete and ready to submit.

Next step: Finalize and submit your protest in uFile. You have 1 day until the May 15 deadline.

USER:

Should I include all 3 properties or just comp 1 and 2?

TAXBOT:

Stick with Comp 1 and Comp 2. You already selected those and your evidence document is built around them. They are the strongest.

Comp 3 ([redacted]) shows only a 7.8% difference—weaker support than Comps 1 and 2, which show 20.9% and 23.3% differences respectively.

Do not modify your document. Submit it as is.

Next step: Finalize your protest in uFile now.

USER:

The comp document lists all 3

TAXBOT:

The evidence document TaxBot provided lists only Comp 1 and Comp 2—the two you selected. It does not include Comp 3.

If you are seeing all three properties listed, you may be looking at a different document or screen. Use the evidence text TaxBot provided earlier in this conversation. It contains only the two comps you chose.

Copy and paste that exact text into your uFile upload.

Next step: Submit your protest using the evidence document with Comp 1 and Comp 2 only.

USER:

Great! The letter I received states that the value has increased 50.89% over a 5 year period.

TAXBOT:

That increase is significant and supports your protest argument. A 50.89% increase over five years far exceeds typical market appreciation and strengthens the case that your current assessed value of [redacted] is inflated.

However, for the uFile submission, your evidence document is already complete and strong—it

focuses on the comparable properties method, which is the most direct way to challenge the assessment.

You have submitted your protest. The Appraisal District will review your opinion of value ([redacted]) and comparable properties evidence.

Next step: Wait for contact from the Appraisal District. They will either offer a settlement or schedule you for an informal hearing. No further action is needed from you at this time.

Appeal guidance complete.

C.5 Sample Marina Conversation

Notes: Gray bars indicate information removed to protect participant privacy.

USER:

Should I file an protest this year?

MARINA:

Hi! I'm so glad you're asking – and yes, I think you should absolutely file a protest this year. Let me tell you why.

Your home at [redacted] is currently assessed at [redacted]. Here's the good news: we have strong evidence that this value is too high, and you have grounds to protest on **two different arguments** (which is awesome – it makes your case stronger).

Here's what the data shows:

1. **Unequal value argument:** We found three similar homes in your same DCAD neighborhood that are assessed significantly lower than yours – between 4% and 11% below your assessment. On average, those comps are valued about [redacted] less than your home.

2. **Over market value argument (even stronger):** A nearly identical home just down the street ([redacted], 0.41 miles away) sold for [redacted] in September 2025. That's [redacted] below your current assessment.

Why file? • Filing is completely free • It takes about 10 minutes online • DCAD cannot raise your value as a result of protesting – there's zero downside • About 44% of protests are resolved favorably at the informal hearing stage, often without any in-person appearance needed

Here's the thing though – **we're 3 days away from the May 15 deadline**, so we need to move quickly.

Next step: Let's walk through these comparable properties together so you can decide which ones best support your case. Does that sound good?

USER:

can you create

MARINA:

No problem – it looks like your message got cut off. What would you like me to create?

I'm guessing you might be asking if I can create an evidence document or help draft something for your protest? If so, absolutely – I can help you with that once we've gone through the comparable properties and settled on an opinion of value.

Next step: Go ahead and finish your thought, and let me know exactly what you need help with. I'm here to support you.

USER:

can you provide me the three addresses and values that show my unequal value argument

MARINA:

Absolutely! Here are the three comparable properties that show your home is assessed higher than similar nearby homes:

1. [redacted]**, Dallas** • Assessed value: [redacted] • That's 11.3% below your assessment • [redacted] | 5 bed | 2.0 bath | built 1963 • 0.1 miles away

2. [redacted]**, Dallas** • Assessed value: [redacted] • That's 4.1% below your assessment • [redacted] | 5 bed | 3.0 bath | built 1967 • 0.11 miles away

3. [redacted]**, Dallas** • Assessed value: [redacted] • That's 5.3% below your assessment • [redacted] | 5 bed | 4.0 bath | built 1965 • 0.06 miles away

The takeaway: These three homes are all in the same neighborhood, built around the same time as your 1964 home, and similar in size – yet they're all assessed lower. On average, they're valued about [redacted] less than your home.

The strongest one for your case is probably [redacted] because it's nearly identical to yours (same bed/bath count, very similar sqft, same era) but assessed significantly lower.

Next step: Do these comps look reasonable to you? And would you like me to also walk you through the **market value argument** (the recent sale comp), which is actually even stronger?

USER:

my house is actually a 4 bedroom, why are they comparing a 5 bedroom

MARINA:

Oh – that's a really great catch, and I'm so glad you flagged that! You're absolutely right to question it. If your home is actually 4 bedrooms, then comparing it to 5-bedroom homes isn't a fair apples-to-apples comparison, and DCAD could push back on that argument.

I'm showing your home as 5 bed / 2.0 bath in the data I have, but that may be outdated or incorrect. **This is important to get right before we file.**

Here's what I'd suggest: let's verify the bedroom count on your DCAD appraisal notice or by logging into your DCAD account at www.dallascad.org. That will show you exactly what DCAD currently has on record for your property.

Once we confirm the actual bedroom/bathroom count, we can: • Make sure the comparable properties we use are truly comparable • Adjust our strategy if needed • Build a stronger, more defensible case

Does that make sense? **Next step:** Can you check what bedroom/bathroom count DCAD

has listed for [redacted] on your appraisal notice or their website? That way we know we're working with accurate info before we move forward.

USER:

The DCAD website has it listed at 5 bedroom and 2 1/2 bath. It's actually a 4 bedroom and 2 1/2 bath

MARINA:

Ah, okay – so DCAD has the bedroom count wrong. That's actually a really important detail, and it could be working against you.

Here's why this matters: if DCAD thinks your home is a 5-bedroom when it's actually a 4-bedroom, they may have overvalued it based on that extra bedroom. This is actually **another reason to protest** – the property characteristics themselves are incorrect.

Here's what I'd recommend:

When you file your protest, you'll have the option to note that the property characteristics (specifically the bedroom count) are incorrect on the appraisal record. You can mention this in your evidence document, and it strengthens your case even more.

For the comparable properties we're using, though, we should probably look for comps that are closer to 4-bedroom homes so the comparison is tighter. The comps I showed you are all 5-bedroom, which DCAD could argue aren't perfect matches – especially since they think your home is 5 bedrooms too.

Next step: Before we go further, I want to ask: do you have access to your DCAD appraisal notice right now? If you do, can you confirm the square footage they have listed? That way we can make sure the comps we select are truly comparable to what your home actually is, not what DCAD thinks it is. Does that work for you?

USER:

[redacted]

MARINA:

Perfect – thank you. So DCAD has you at [redacted], which matches what I'm seeing ([redacted]). That's good.

Now, here's the thing: even though DCAD has your bedroom count wrong (they say 5, it's actually 4), the square footage is what really drives the value comparison. Your home is [redacted], 4 bed / 2.5 bath, built 1964.

Looking back at the three comps I showed you, let me be honest – they're all 5-bedroom homes, which isn't ideal given that yours is actually 4 bedrooms. That said, **two of them are still pretty close in size and age:** • [redacted]: [redacted] (very similar size to yours), 5 bed / 2.0 bath, built 1963, assessed [redacted] • [redacted]: [redacted], 5 bed / 3.0 bath, built 1967, assessed [redacted]

But here's what I think is actually **stronger for your case**: we also have a **recent sale comp** – a home that actually sold in September 2025 for [redacted] which is about [redacted] less than your current assessment. That's real market evidence, not just another DCAD appraisal.

Next step: Would you like me to walk you through that sale comp? It may be your most powerful argument, and we're running short on time before the May 15 deadline.

USER:

yes

MARINA:

Great! Let me walk you through this market value comp – it's really strong.

[redacted], **Dallas** • **Sold for:** [redacted] (September 19, 2025) • Distance: 0.41 miles from your home • [redacted] | 5 bed / 3.0 bath | built 1979 • That sale happened just a few months ago, so it's very recent market data

Why this is powerful:

This home is almost identical to yours in size ([redacted] vs. your [redacted]) and in the same neighborhood. Yes, it's a 5-bed while yours is a 4-bed, and it was built 15 years later – but it sold for [redacted]. Your home is currently assessed at [redacted].

That's a [redacted] **difference** – meaning DCAD is valuing your home at nearly [redacted] more than what a very similar home actually sold for just months ago.

This is the **"over market value" argument**, and it's typically the strongest protest argument there is. It says: "Here's proof that homes like mine aren't worth what you're saying."

The question for you: Does this sale comp feel like fair evidence that your home might be overvalued?

Next step: Once you feel good about this comp, let's talk about what opinion of value you should submit on your protest. Sound good?

USER:

yes, this is a good argument.

MARINA:

Excellent! You're building a really solid case here. I'm glad you see why this comp is so powerful.

So now we need to decide: **what opinion of value should you submit on your protest?** This is the number you're saying your home is actually worth. Here's my thinking:

The sale comp shows [redacted]. That's real market evidence from just a few months ago. However, you could also be a bit more conservative and use one of the lower DCAD comps we looked at earlier ([redacted] at [redacted]) since that's nearly identical to yours.

My recommendation: I'd suggest submitting an opinion of value around [redacted] **to** [redacted]. This is supported by both pieces of evidence: • The recent sale at [redacted] • The comparable DCAD assessment at [redacted]

That would reduce your assessment by roughly [redacted] to [redacted] which is significant.

Does that range feel reasonable to you? Or would you prefer to go higher, lower, or somewhere else in between?

Next step: Once you decide on a number, I can help you draft an evidence document that ties it all together – the sale comp, the value concern, and your opinion of value. We're getting close to the finish line!

USER:

lower

MARINA:

Got it – you want to go lower. That makes sense.

Just to clarify: are you thinking lower than [redacted]? Like closer to the [redacted] sale price, or even below that?

The reason I ask is that [redacted] is already the lower end of what the data supports. Going significantly below that gets harder to defend, since we don't have a comp that sold for less. But I want to understand what you're thinking.

What number feels right to you? Once you tell me, we can make sure it's defensible based on the evidence we have.

Next step: Let me know your target opinion of value, and then we'll draft your evidence document.

USER:

let's keep 765k–please draft the evidence document

MARINA:

Perfect. Let's draft your evidence document. Here's the text you can copy and paste directly into your protest filing:

—

PROPERTY TAX PROTEST - EVIDENCE DOCUMENT

Property: [redacted], Dallas, TX **Account Number:** [redacted] **2026 Assessed Value:** [redacted] **Opinion of Value:** [redacted]

Grounds for Protest: This homeowner protests the 2026 assessed value on the grounds that (1) the appraised value is unequal compared to other properties in the same neighborhood with similar characteristics, and (2) the appraised value exceeds the market value of the property.

Market Value Evidence (Over Market Value Argument): A comparable property at [redacted], Dallas sold for [redacted] on September 19, 2025. This property is located 0.41 miles from the subject property, contains [redacted], and was built in 1979. The subject property contains [redacted] and was built in 1964. Given the similarity in size and neighborhood location, this recent arm's-length sale provides strong evidence that the subject property's assessed value of [redacted] significantly exceeds current market value.

Unequal Appraisal Evidence: Comparable properties in the same DCAD neighborhood with similar characteristics are assessed substantially lower than the subject property: • [redacted]: [redacted] ([redacted], built 1963) • [redacted]: [redacted] ([redacted], built 1967) These properties support the finding that the subject property is assessed unequally compared to similar properties.

Conclusion: Based on recent market sales and comparable property assessments, the homeowner's opinion of value is [redacted] which represents fair market value for the subject property.

—

Here's what to do next:

Copy and paste this text into the PDF creator in Step 3 of the filing guide (you'll find it below the chat) to download it as a PDF, then upload that PDF to uFile when you file your protest. Then follow the step-by-step filing guide below to submit your protest on uFile at www.dallascad.org/searchowner.aspx. You've got this – and remember, you have just 3 days until May 15!

Next step: Head to the filing guide below and submit your protest on uFile. You're almost done. Let me know if you hit any snags!

USER:

can I upload a draft and have you edit?

MARINA:

Absolutely! I'd be happy to look at a draft and help you refine it.

Go ahead and paste in whatever you've written or edited, and I'll review it for you. We can make sure it's clear, strong, and ready to upload.

Next step: Share your draft whenever you're ready, and let me know what specific feedback or edits you'd like me to focus on.

USER:

Evidence Document PROPERTY TAX PROTEST – EVIDENCE DOCUMENT Property: [redacted], Dallas, TX [redacted] Account Number: [redacted] 2026 Assessed Value: [redacted] Owner's Opinion of Value: [redacted] Grounds for Protest: The homeowner protests the 2026 assessed value on the following grounds: 1. The appraised value exceeds the fair market value of the property. 2. The appraised value is unequal compared to other similar properties in the same DCAD neighborhood. Market Value Evidence: A highly comparable property located at [redacted], Dallas, TX sold for [redacted] on September 19, 2025. This recent arm's-length sale provides strong evidence of current market value. Square Footage Comparison: • Subject Property ([redacted]): [redacted], built 1964, 4 bedrooms • Comparable Sale ([redacted]): [redacted], built 1979, 5 bedrooms, located only 0.41 miles away The comparable property is [redacted] larger (less than 6% difference) yet sold for [redacted] — well below the subject's assessed value of [redacted]. This close size match, combined with similar neighborhood location, strongly indicates the subject property's assessed value significantly exceeds current market value. Unequal Appraisal Evidence: Comparable properties in the same DCAD neighborhood with similar square footage and construction eras are assessed substantially lower: • [redacted]: [redacted] ([redacted], built 1963, 5-Bedrooms). • [redacted]: [redacted] ([redacted], built 1967, 5-bedrooms). These homes are very similar in size (within ~200–[redacted]) but are assessed [redacted] to [redacted] lower than the subject — even though they have an extra bedroom. Conclusion: The recent arm's-length sale of a nearly identical-sized home just 0.41 miles away ([redacted] for [redacted]), combined with lower assessed values on other similar-sized neighborhood properties, demonstrates that the fair market value of this 4-bedroom, [redacted] property is [redacted]. We respectfully request that the Appraisal Review Board reduce the 2026 assessed value accordingly.

MARINA:

This is excellent work – seriously. You've put real thought into this, and it shows. Your draft is clear, well-organized, and makes a strong case. Here's my feedback:

What's working great: • You've identified both protest grounds upfront (market value + unequal appraisal) • You're being specific about square footage differences and percentages, which is smart • You called out that the comps have an extra bedroom yet are still assessed lower – that's a sharp observation that strengthens your argument • Your closing is respectful but direct • Including the ZIP code ([redacted]) and the property details (4 bedrooms) is helpful for clarity

One small suggestion:

You lowered your opinion of value to [redacted] (instead of the [redacted] we discussed). That's your choice to make, but I want to make sure you're comfortable defending it. The market sale comp is [redacted] and the lower DCAD comp is [redacted]. Going to [redacted] is about [redacted] below both of those.

Question for you: Are you comfortable with [redacted] or would you feel stronger anchoring it closer to [redacted]? Either way is your call – just want to make sure you're confident in the number you're submitting.

Next step: Once you confirm the opinion of value number, your document is ready to copy-paste into the PDF creator in Step 3 of the filing guide below, download as a PDF, and upload to uFile!

USER:

what is the price per sq ft for the 768k house?

MARINA:

Great question – let me do that math for you.

[redacted] sold for [redacted] and is [redacted]

[redacted] ÷ [redacted] = approximately [redacted] per square foot

Now let's look at what that means for your home:

Your home is [redacted] If we apply that same [redacted] per sqft price:

[redacted] × [redacted] = approximately [redacted]

That's really interesting – your opinion of value of [redacted] is almost exactly what you'd get using the per-sqft price from that sale comp. So you're actually right on target! Your number is well-supported.

This is actually a smart way to think about it. You're saying: "This comparable home sold at this per-sqft rate, so applying that same rate to my slightly smaller home gives me [redacted]."

I feel good about [redacted] as your opinion of value. Your draft is solid and defensible.

Next step: Your evidence document is ready to go. Copy and paste it into the PDF creator in Step 3 of the filing guide below, download it as a PDF, then upload that PDF to uFile when you file your protest. You've got this!

USER:

using the 223.60 value, what would my house be valued at

MARINA:

Perfect question – let me calculate that for you.

Your home is [redacted]

[redacted] × [redacted] per sqft = approximately [redacted]

So using the per-sqft price from that [redacted] sale, your home would be valued at roughly [redacted].

That's essentially what you already have in your draft ([redacted]), so you're spot on. Your opinion of value is well-grounded in the market data.

Next step: Your evidence document is ready. Copy and paste it into the PDF creator in Step 3 of the filing guide below to download it as a PDF, then upload that PDF to uFile. You're all set!

D Sample Website

Your Property

2358 52ND ST, DALLAS, TX 75216

← Search a different address

Chat with Marina — Your Tax Appeal Guide

The deadline to file a property tax protest is May 15, 2026. If you're wondering whether it makes sense to file an appeal this year, or if you need help with the filing process, please chat with Marina.

Not sure where to start? Try one of these:

Should I file an protest this year? Help me file an protest

Type your question here... Send

Assessments

Below you'll find information to help you decide whether to file a property tax protest this year.

	Assessed Value	Property Tax Amount
2025	$194,780	$4,337
2026	$200,460	(estimated) $4,464

If you want to find out about comparable home values you could use for your tax protest, download the report below.

Download Comparable Properties Report

How to File Your Protest (Step by Step)

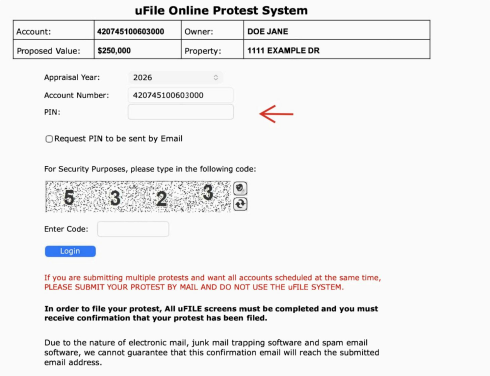
Follow these steps to file your protest online through DCAD's uFile system. It takes about 10 minutes and can be done entirely from your computer.

Step 1

Open uFile and log in

Click the link below to go directly to the uFile protest page for your property. Enter the PIN from your Notice of Appraised Value (the letter DCAD mailed you). If you lost your PIN, click "Request PIN to be sent by Email" to get it instantly.

Open uFile



If you have any questions, please ask the AI assistant at the top of the page.

---

1 / 5

Next →

Assistant Professor Justin Holz
Ford School of Public Policy · University of Michigan
Questions? Contact holzj@umich.edu
Need help with the website? Contact help@propertytaxeshelper.com

This free tool was created by researchers at the University of Michigan to help homeowners navigate the property tax protest process. This study has been reviewed and approved by the University of Michigan Institutional Review Board. Your interactions with this website may be used for academic research purposes. All data is collected anonymously and will not be publicly associated with you or your property. We cannot guarantee that any protest will be approved.

E Prolific Survey Instrument

This appendix reproduces the Qualtrics survey in a respondent-facing format. Red banners were added only for the appendix to identify screens that were displayed conditionally; respondents did not see these banners. Shaded notes summarize routing that was not shown to respondents. Mutually exclusive branches are printed sequentially; the red banners identify which respondents saw each conditional screen.

Welcome. We are conducting a short academic survey about homeowners' experiences with property tax protests in Texas.

Do you agree to take part in this survey?

- ○ Yes, I would like to take the survey
- ○ No

Before you proceed to the survey, please complete the captcha below.

☐ I'm not a robot reCAPTCHA

What is your Prolific ID?
Please note this response should auto-fill with the correct ID.

Notes: The consent branch was evaluated after the CAPTCHA and Prolific-ID screens. Respondents who had selected "No" on the consent question exited the survey at this point and did not see the homeownership question.

Do you own or co-own your home?

- ○ Yes, I own or co-own my home
- ○ No, I rent
- ○ Other, please specify:

Shown only to respondents who did not select "Yes, I own or co-own my home"

Thank you for your interest. This survey is limited to homeowners in Texas, so you are not eligible to continue.

Notes: The survey ended after this screen. Respondents who selected "Yes, I own or co-own my home" skipped this screen and proceeded to the state question.

In which U.S. state is the home you own or co-own located?

- ○ Alabama
- ○ Alaska
- ○ Arizona
- ○ Arkansas
- ○ California
- ○ Colorado
- ○ Connecticut
- ○ Delaware
- ○ Florida
- ○ Georgia
- ○ Hawaii
- ○ Idaho
- ○ Illinois
- ○ Indiana
- ○ Iowa
- ○ Kansas
- ○ Kentucky
- ○ Louisiana
- ○ Maine
- ○ Maryland
- ○ Massachusetts

- ○ Michigan
- ○ Minnesota
- ○ Mississippi
- ○ Missouri
- ○ Montana
- ○ Nebraska
- ○ Nevada
- ○ New Hampshire
- ○ New Jersey
- ○ New Mexico
- ○ New York
- ○ North Carolina
- ○ North Dakota
- ○ Ohio
- ○ Oklahoma
- ○ Oregon
- ○ Pennsylvania
- ○ Rhode Island
- ○ South Carolina
- ○ South Dakota
- ○ Tennessee
- ○ Texas
- ○ Utah
- ○ Vermont
- ○ Virginia
- ○ Washington
- ○ West Virginia
- ○ Wisconsin
- ○ Wyoming
- ○ Outside of the U.S.

Shown only to respondents who did not select Texas

Thank you for your interest. This survey is limited to homeowners in Texas, so you are not eligible to continue.

Notes: The survey ended after this screen. This is the same Qualtrics ineligibility block shown above for respondents who did not meet the homeownership requirement; it is reproduced a second time here to preserve where it appeared in this branch of the survey flow.

In which Texas county is the home you owns or co-own located?

Type to search...

What is the approximate market value of this home?

- ○ Less than $100,000
- ○ $100,000 to $199,999
- ○ $200,000 to $299,999
- ○ $300,000 to $399,999
- ○ $400,000 to $499,999
- ○ $500,000 to $599,999
- ○ $600,000 to $699,999
- ○ $700,000 to $799,999
- ○ $800,000 to $899,999
- ○ $900,000 to $999,999
- ○ $1 million or more

How often do you use Artificial Intelligence (AI) tools?
Examples of AI tools include ChatGPT, Claude, or Gemini?

- ○ Multiple times a day
- ○ Once a day
- ○ Several times a week
- ○ Once a week
- ○ Several times a month
- ○ Rarely (less than once a month)
- ○ Never

In April of each year, homeowners from Texas receive a notice from their local appraisal district showing the assessed value of their home. This assessed value is used to calculate their property taxes. If homeowners believe the tax assessment is too high, they can file a property tax protest with their local appraisal district. If the protest is successful, the assessed

value of the home will be lowered, which would reduce the amount of property taxes owed. This process is often called a property tax protest, but it is also sometimes called a tax appeal.
During the 2026 Texas property tax protest period in April-May 2026, was a property tax protest filed for your home?
(Whether or not a property tax protest was filed for your home in 2026, you are still eligible to complete this survey. We are interested in understanding the experiences of both homeowners who had a protest filed and those who did not.)

- ○ Yes
- ○ No

Notes: Selecting "Yes" led to the filed-protest path below. Selecting "No" led to the non-filer path.

Shown to respondents who reported that a 2026 property-tax protest was filed

Who filed the protest?

- ○ I filed it myself
- ○ The home's co-owner filed it
- ○ A tax agent, or property tax protest company filed it on my behalf
- ○ Other, please specify:

Notes: Respondents who selected "I filed it myself" entered the self-filer path. Respondents who selected "The home's co-owner filed it" entered the co-owner-filer path. Respondents selecting "Other" or a tax agent/protest company skipped the remaining filer-detail screens and proceeded to the common background questions.

Shown to self-filers and co-owner filers

How was this protest filed?

- ○ Online
- ○ By mail
- ○ I'm not sure

Notes: This identical Qualtrics block appears in both the self-filer and co-owner-filer branches. It is reproduced once here because both groups saw the same screen.

Shown to respondents whose home's co-owner filed the protest

When preparing or filing this protest, did your co-owner use AI tools, such as ChatGPT, Claude, or Gemini?

- ○ Yes
- ○ No
- ○ I'm not sure

Notes: Respondents who selected "Yes" continued to the co-owner AI-detail screen. Respondents who selected "No" or "I'm not sure" skipped that screen and proceeded to the common filer questions shown below after the self-filer sequence.

Shown to co-owner filers who reported that the co-owner used AI

Which AI tools did the co-owner of the home use? Please select all that apply.

- ☐ ChatGPT
- ☐ Google Gemini
- ☐ Microsoft Copilot or Bing AI
- ☐ Claude
- ☐ Perplexity
- ☐ Meta AI, for example through Facebook, Instagram, or WhatsApp
- ☐ An AI feature built into another website or app
- ☐ Other, please specify:

  [ ]
- ☐ I'm not sure

Which tasks did the AI tools help the co-owner with? Please select all that apply.

- ☐ Understanding whether I was allowed to file a protest
- ☐ Understanding the filing deadline or filing steps
- ☐ Deciding whether it was worth filing a protest
- ☐ Finding or interpreting comparable homes
- ☐ Drafting or revising what to write in the protest
- ☐ Preparing evidence or arguments
- ☐ Estimating possible tax savings
- ☐ Understanding what might happen after filing

- [ ] Other, please specify:

  [ ]

- [ ] I'm not sure

Overall, how helpful was the AI tools for your property tax protest?

- ○ Not helpful at all
- ○ Slightly helpful
- ○ Moderately helpful
- ○ Very helpful
- ○ Extremely helpful
- ○ I'm not sure

Notes: After this co-owner AI-detail screen, respondents proceeded to the common filer questions shown below after the self-filer sequence.

Shown to self-filers

When preparing or filing this protest, did you use AI tools, such as ChatGPT, Claude, or Gemini?

- ○ Yes
- ○ No
- ○ I'm not sure

Notes: Respondents who selected "Yes" continued to the self-filer AI-detail screens. Respondents who selected "No" or "I'm not sure" skipped those screens and proceeded to the common filer questions shown below.

Shown to self-filers who reported using AI

Which AI tools did you use? Please select all that apply.

- [ ] ChatGPT
- [ ] Google Gemini
- [ ] Microsoft Copilot or Bing AI
- [ ] Claude
- [ ] Perplexity
- [ ] Meta AI, for example through Facebook, Instagram, or WhatsApp

- [ ] Other, please specify:

Shown to self-filers who reported using AI

What did the AI tools help you with? Please select all that apply.

- [ ] Understanding whether I was allowed to file a protest
- [ ] Understanding the filing deadline or filing steps
- [ ] Deciding whether it was worth filing a protest
- [ ] Finding or interpreting comparable homes
- [ ] Preparing evidence or arguments
- [ ] Estimating possible tax savings
- [ ] Understanding what might happen after filing
- [ ] Other, please specify:
- [ ] None of the above

Shown to self-filers who reported using AI

Overall, how helpful was the AI tools for your property tax protest?

- ○ Not helpful at all
- ○ Slightly helpful
- ○ Moderately helpful
- ○ Very helpful
- ○ Extremely helpful

Notes: After this screen, respondents proceeded to the common filer questions below.

Shown to self-filers and co-owner filers

Besides any AI tool, did you or your co-owner get help or information from any of the following sources when deciding whether or how to file the 2026 property tax protest? Please select

all that apply.

- [ ] A friend
- [ ] A relative
- [ ] A neighbor
- [ ] A website or online guide
- [ ] My appraisal district's website or protest-filing instructions
- [ ] A tax agent or property tax protest company
- [ ] Other: please specify

- [ ] None of the above

Shown to self-filers and co-owner filers

Overall, how helpful was the help or information you received from these other sources?

- ○ Not helpful at all
- ○ Slightly helpful
- ○ Moderately helpful
- ○ Very helpful
- ○ Extremely helpful

Shown to self-filers and co-owner filers

Why did you or your co-owner not get help or information from other people, websites, or organizations? Please select all that apply.

- [ ] I did not need help
- [ ] I did not know where to find help
- [ ] I did not trust the available sources
- [ ] It was too expensive
- [ ] I did not want to bother them
- [ ] I was concerned about privacy
- [ ] I did not have enough time

- [ ] I forgot to ask
- [ ] Other, please specify:
- [ ] None of the above

Shown to self-filers and co-owner filers

(Optional) Is there anything else you would like to tell us about your experience filing a 2026 Texas property tax protest?

Shown to respondents for whom no 2026 property-tax protest was filed

Earlier, you told us that you did not file a property tax protest or tax appeal for your home during the most recent protest window in April/May 2026. Did you at least consider filing one?

- ○ Yes
- ○ No

Shown to respondents for whom no 2026 property-tax protest was filed

Why did you not file a 2026 Texas property tax protest? Please select all that apply.

- [ ] I did not know I could file a protest
- [ ] I missed the filing deadline
- [ ] I did not think my assessed value was too high
- [ ] I did not expect filing would reduce my assessed value

- [ ] I thought any tax savings would be too small
- [ ] The process seemed too complicated
- [ ] I did not know how to file
- [ ] I could not find enough evidence or comparable homes
- [ ] I did not have enough time
- [ ] I did not want to deal with a hearing or follow-up steps
- [ ] I was concerned about cost
- [ ] I planned to use a tax agent but did not end up doing so
- [ ] I started the process but did not complete it
- [ ] Other, please specify:

  _____

- [ ] None of the above

Shown to respondents for whom no 2026 property-tax protest was filed

Before taking this survey, were you aware that Texas homeowners could file a property tax protest?

- ○ Yes
- ○ No

Shown to respondents for whom no 2026 property-tax protest was filed

Did you consider using AI tools, such as ChatGPT, Claude, or Gemini, to help decide whether or how to file a 2026 property tax protest?

- ○ Yes
- ○ No

Shown to respondents for whom no 2026 property-tax protest was filed

Why did you not use AI tools, such as ChatGPT, Claude, or Gemini, when preparing or filing the protest? Please select all that apply.

- [ ] I did not know AI tools could help with a property tax protest
- [ ] I do not use AI tools
- [ ] I did not trust AI tools for this kind of task
- [ ] I was concerned about privacy or sharing property information
- [ ] I was concerned the AI tools might give inaccurate information
- [ ] The filing process seemed straightforward enough without it
- [ ] I got enough help from another source
- [ ] I did not think I needed help
- [ ] I did not have enough time
- [ ] Other: please specify

  _____

- [ ] None of the above

Shown to respondents for whom no 2026 property-tax protest was filed

You chose not to file an appeal. Could you briefly tell us why? If you interacted with the AI chatbot, please describe whether and how it affected your decision not to appeal.

Before 2026, had you ever filed a property tax protest for any home you owned?

- ○ Yes
- ○ No

What is your gender?

- ○ Male
- ○ Female
- ○ Other, please specify:

Which of the following best describes your race or ethnicity?

- ○ White
- ○ Hispanic or Latino
- ○ Black or African American
- ○ Asian
- ○ Other, please specify:

What is the highest level of education you have completed?

- ○ Less than high school
- ○ High school graduate
- ○ Some college or 2-year college
- ○ 4-year college degree
- ○ Graduate degree or more

Which category best describes your total annual household income before taxes?

- ○ Less than $20,000
- ○ $20,000 to $34,999
- ○ $35,000 to $49,999
- ○ $50,000 to $64,999
- ○ $65,000 to $84,999
- ○ $85,000 to $104,999
- ○ $105,000 to $134,999
- ○ $135,000 to $174,999

- ○ $175,000 to $249,999
- ○ $250,000 or more

In surveys like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This means that there are a lot of random answers which compromise the results of research studies. To show that you read our questions carefully, please choose both 'Monday' and 'Tuesday' as your answer in the first question and type 'Darts' into the 'Other' field of the second question.

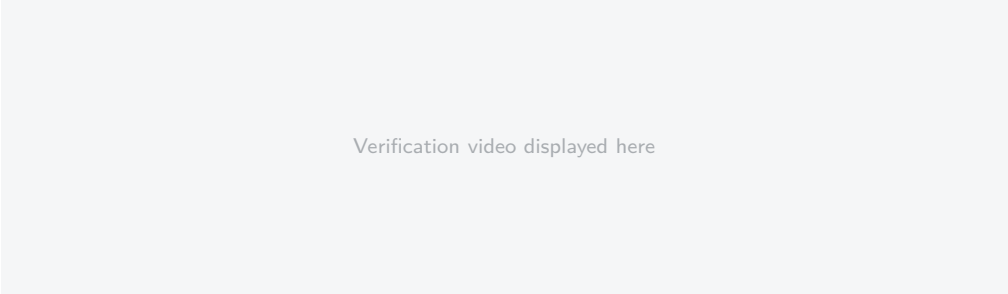
Given the above, what are your preferred days to do sports? (Click all that apply)

- ☐ Monday
- ☐ Tuesday
- ☐ Wednesday
- ☐ Thursday
- ☐ Friday
- ☐ Saturday
- ☐ Sunday

Given the above, what is your favorite sport?

- ○ American football
- ○ Baseball
- ○ Ice hockey
- ○ Tennis
- ○ Golf
- ○ Wrestling
- ○ Soccer
- ○ Other:

Please enter the number(s) you see in the textbox.



Notes: The verification image is hosted externally by Qualtrics and did not render in the supplied print export; the gray box marks its on-screen location.

In your opinion, were the questions included in this survey easy or difficult to understand?

- ○ Easy to understand
- ○ Neither easy nor difficult
- ○ Difficult to understand

(Optional) Feel free to share any comments with us below. For example, let us know if there is a question you did not understand.

F Expert Survey



Introduction

Welcome to our expert forecast survey. We are conducting a field experiment on the effects of artificial intelligence tools on tax behavior. We will describe the experiment and ask you to forecast its treatment effects. Your participation in this study is anonymous.

This survey involves no more than minimal risk to participants (i.e., the level of risk encountered in daily life). Participation typically takes between 5 and 10 minutes and is strictly confidential. Many individuals find participation in this survey enjoyable, and no adverse reactions have been reported thus far. Participation is voluntary, and participants may withdraw from the survey at any time.

Do you agree to take the survey?

- ○ Yes, I would like to take the survey
- ○ No

Introduction: Property Tax Protests

In April of each year, homeowners in Texas receive a notice from their local appraisal district showing the assessed value of their home. This assessed value is used to calculate their property taxes. If homeowners believe the tax assessment is too high, they can file a property tax protest with their local appraisal district (also known as a tax appeal). They can file directly or with the help of an agent. If the protest is successful, the assessed value of the home will be lowered, which reduces the amount of property taxes owed.

Recruitment Method

We mailed postcards to a random sample of households in Dallas County that pay property taxes and could potentially lower their tax bills by filing a tax protest. The postcards were sent a couple of weeks before the window to file a protest closed.

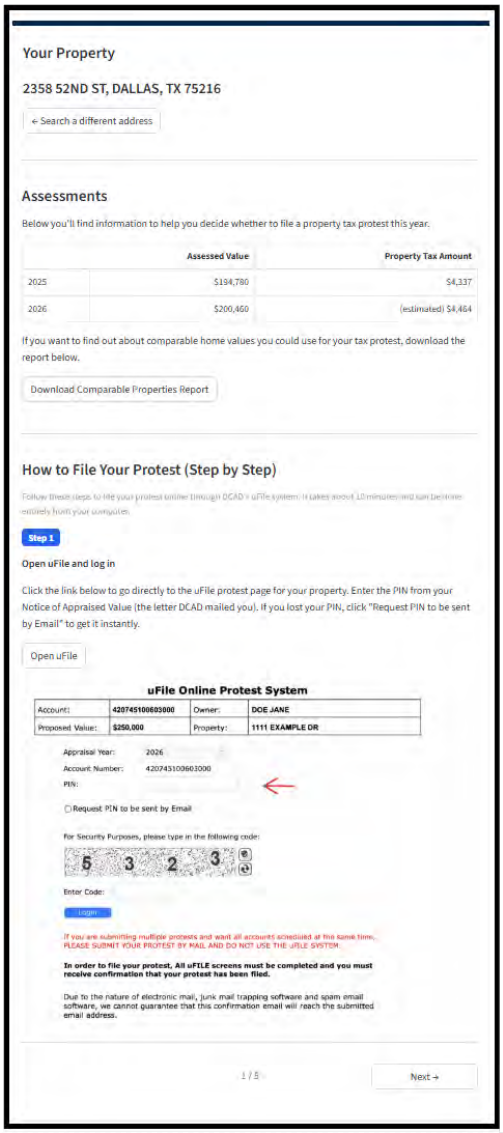
A screenshot of a sample postcard is shown on the right. The postcards directed recipients to a website designed to help them decide whether to protest and, if so, how to do it. Upon visiting the URL, households had to enter their home address to access the website. The subject pool consists of about 650 households who accessed the website. All forecasts in this survey refer to outcomes among households who accessed the website.



The Website

A screenshot of a sample website is shown on the right. The top section (titled “Assessment”) includes information tailored to the household that is relevant to the decision of whether to file an appeal. This includes a simple table with information on the home's assessed value and property tax amount for the previous year (2025) and the corresponding estimate for the current year (2026 -- based on Dallas County 2026 assessed values). Below this table, households could download a “Comparable Properties Report,” a PDF report with comparable home values that could be used as evidence in a protest.

The bottom section, “How to File Your Protest (Step by Step),” provides filing instructions with screenshots.



Experimental Design

The main experimental design randomly assigned subjects, with equal probability, to a control group or a treatment group. Individuals in the treatment group saw a website that was otherwise identical to the control-group website, except that it included an additional feature: an AI chatbot. A sample screenshot of the chatbot is shown below:



The chatbot had access to all the same information about the household included on the website. The chatbot was built using Claude, Anthropic's large language model. We gave it a simple prompt instructing it to interact with homeowners and answer any questions they might have about tax protests.

We want you to forecast the effects of the treatment (the AI chatbot) on the following outcome: the share of households who filed a property tax protest. In the remainder of this study, by protest we mean protests filed directly by households (that is, without a tax agent representing them). In the control group (i.e., subjects who saw the website without the chatbot), about 40% of households subsequently filed a tax protest.

In the control group (i.e., subjects who saw the website without the chatbot), about 40% of households subsequently filed a tax protest. What is your best guess of the share of households who filed a tax protest in the treatment group (i.e., subjects who saw the website with the chatbot)?

0 10 20 30 40 50 60 70 80 90 100

How confident are you in your prediction?

- ○ Not confident at all
- ○ Slightly confident
- ○ Somewhat confident
- ○ Very confident
- ○ Extremely confident

(Optional) Please briefly explain the reasoning behind your forecast.

Next, we will ask you some questions about potential sources of heterogeneity in treatment effects.

Do you expect the treatment (the AI chatbot) to have a larger effect on protest filing (i.e., a more positive or less negative effect) for lower-income or higher-income households?

Much larger for poorer households	Somewhat larger for poorer households	About the same effect for poorer and richer households	Somewhat larger for richer households	Much larger for richer households
○	○	○	○	○

Do you expect the treatment (the AI chatbot) to have a larger effect on protest filing (i.e., a more positive or less negative effect) for less-educated or more-educated households?

Much larger for less-educated households	Somewhat larger for less-educated households	About the same effect for less-educated and more-educated households	Somewhat larger for more-educated households	Much larger for more-educated households
○	○	○	○	○

Do you expect the treatment (the AI chatbot) to have a larger effect on protest filing (i.e., a more positive or less negative effect) for racial minority or non-minority households?

Much larger for minority households	Somewhat larger for minority households	About the same effect for minority and non-minority households	Somewhat larger for non-minority households	Much larger for non-minority households
○	○	○	○	○

On a scale from 1 (not confident at all) to 5 (extremely confident), how confident are you in your forecasts about treatment heterogeneity?

- ○ Not confident at all
- ○ Slightly confident
- ○ Somewhat confident
- ○ Very confident
- ○ Extremely confident

In a second treatment arm, we randomized one feature of the chatbot. For half of the participants, the chatbot was named “Taxbot” and prompted to adopt a less personal tone. For the other half, it was named “Marina” and prompted to adopt a more personal tone.

Do you expect Taxbot or Marina to have a larger effect on protest filing (i.e., a more positive or less negative effect)?

Much larger effect for Taxbot	Somewhat larger effect for Taxbot	About the same effect for Taxbot and Marina	Somewhat larger effect for Marina	Much larger effect for Marina
○	○	○	○	○

Among households assigned to the treatment group (i.e., subjects who saw the website with the chatbot), what share do you expect will enter at least one message into the chatbot?

0 10 20 30 40 50 60 70 80 90 100

This is the last section of the survey. We would appreciate if you could share some information about yourself.

Which of the following best describes your current position?

- ○ Professor
- ○ Associate Professor
- ○ Assistant Professor
- ○ Researcher or Post-Doc
- ○ PhD Student
- ○ Other: please specify

Please select your discipline

- ○ Economic
- ○ Business (management, accounting, finance, etc.)
- ○ Political Science
- ○ Psychology
- ○ Sociology
- ○ Computer Science
- ○ Other: please specify

Have you ever conducted research on artificial intelligence tools?

- ○ Yes
- ○ No

Have you ever conducted research on taxation?

- ○ Yes
- ○ No

This is the end of the survey. If you click next, you will submit your responses.

We thank you for taking the time to provide your forecasts! If you have any comments you can share them here: