Nava Labs

Nava Public Benefit Corporation

March 13, 2026

Helping the Helpers:

Evaluating a GenAI-powered assistive chatbot for caseworkers

Nava Labs

Better Government Lab

GEORGETOWN UNIVERSITY & UNIVERSITY OF MICHIGAN

CornellBowers

Information Science

Acknowledgements

The report authors would like to acknowledge members of the Nava Labs team: Alicia Benish, Charlie Cheng, Diana Griffin, Foad Green, Genevieve Gaudet, Kanaiza Imbuye, Kasmin Scott, Kevin Boyer, Ryan Hansz, and Yoom Lam for contributing to the development and implementation of the pilot and evaluation plan; Bob Wilkinson, Chloe Hilles, Greg Jordan-Detamore, and Noam Leead for design and communications support; members of the Amplifi team: Jill Bauman, Brit Gilmore, Karen Van Kirk, and Ryan Fendy, for partnering on the real-world pilot implementation and recruiting caseworkers for the advisory council and pilot participation; and the 12 members of the caseworker advisory council and the over 60 caseworkers who participated in the pilot from various human services agencies and nonprofit organizations in Los Angeles and who provided invaluable insights about the experience using the chatbot.

Report authors

Nava Public Benefit Corporation

Michael Chen,
PhD,
Evaluation Lead,
Nava PBC

Martelle Esposito,
MS, MPH,
Director of Partnerships and Advocacy,
Nava PBC

Better Government Lab, Georgetown University

Eric Giannella,
PhD,
Better Government Lab,
Georgetown University

Zhaowen Guo,
PhD,
Better Government Lab,
Georgetown University

Department of Information Science, Cornell Tech

Jennah Gosciak,
MUP,
Department of Information Science,
Cornell Tech

Allison Koenecke,
PhD,
Department of Information Science,
Cornell Tech

Contents

Executive summary

Project overview

Nava Labs developed and evaluated an artificial intelligence (AI)-powered chatbot designed to help caseworkers more effectively assist clients find and enroll in public benefit programs. This first-of-its-kind study addresses narrowing the gap of over $228 billion in annual unclaimed benefits by improving caseworker capabilities with generative AI (GenAI) technology.[1]

Our solution

The assistive chatbot explains program rules in plain language, uses retrieval-augmented generation to pull information from pre-vetted sources, provides direct citations, supports multilingual translation, and prevents hallucinations through careful guardrails. We built the technology to integrate with existing workflows via an application programming interface (API) and tested the chatbot in real scenarios with nonprofits and government agencies.

Evaluation methodology

Using an implementation science framework in which we sought to identify factors that affected the uptake of the assistive chatbot in addition to impact outcomes, researchers conducted a mixed-methods evaluation, including:[2,3]

A randomized controlled trial with 125 caseworkers examining accuracy effects from being shown AI-generated responses to hypothetical client questions developed from real experiences.

A 14-week, real-world pilot with Benefit Navigator, a web-based tool by Amplifi that helps caseworkers navigate benefit and tax credit programs on behalf of benefit-seeking clients, that included 61 caseworkers across six organizations in Los Angeles County and a quasi-experimental design comparing intervention and comparison groups.

A mixed-methods analysis aggregating qualitative and quantitative findings across multiple data sources, including product logs, surveys, and in-depth interviews.

Key findings

Accuracy

The chatbot is estimated to improve caseworker accuracy by an average of 40% with stronger improvements for more difficult client questions.

Acceptability

Throughout the pilot, about 65% of caseworkers with access to the chatbot used it, submitting an average of 14 prompts each. However, usage declined over time and varied by site, highlighting the value of sustained engagement strategies. The Net Promoter Score (NPS), a metric used to assess user experience, was 11, indicating moderate average satisfaction. While 40% of respondents were "promoters" meaning they would recommend the chatbot to their colleagues, others expressed less enthusiasm, highlighting mixed adoption and opportunities for improvement.

Administrative burden

Results from the pilot showed promising but inconclusive evidence of reduced learning and psychological costs on caseworkers, though these findings lack statistical significance due to sample size limitations and lower response rate on the endpoint survey.

Accessibility

Chatbot responses averaged a 10th to 12th grade reading level — higher than the recommended 8th grade standard but significantly more accessible than source policy manuals, which require college-level reading ability.

Implementation insights

We identified "super users" with high engagement and chatbot usage at sites with active participation in training, consistent organizational reinforcement, and peer support. Service delivery models significantly influenced use patterns, with rapid-intake sites showing higher activity than long-term case management programs.

Implications

This evaluation demonstrates that rigorously tested AI-powered tools can meaningfully improve benefit navigation accuracy, particularly for newer staff. However, successful implementation requires intentional support strategies, ongoing engagement, and attention to accessibility. The findings provide a roadmap for scaling AI-powered tools in public benefit contexts while highlighting critical areas for continued refinement.

About the Nava Labs AI research program

Nava Labs is the philanthropically-funded division within Nava Public Benefit Corporation focused on prototyping systems changes for government programs. We research and prototype products, practices, and policies within government programs and advocate for the adoption of what works. This interdisciplinary team leverages Nava's deep technology delivery experience to identify critical junctures where philanthropy can help accelerate public interest projects and build more trustworthy public institutions.

A first-of-its-kind exploratory project, the Nava Labs AI research program has sought to answer if and how caseworkers can use generative AI powered by large language models (LLMs) to help more eligible people get enrolled in programs like Medicaid and the Supplemental Nutrition Program (SNAP). These efforts can help distribute the over $228 billion in benefits that go unclaimed annually.[1]

Complex policies and processes can make navigating and enrolling in public benefits programs difficult. As a result, people often seek and receive guidance from caseworkers. However, caseworkers can also struggle to understand and interpret eligibility rules or help families complete lengthy applications. For nonprofits and government agencies, it can be costly to train and difficult to retain skilled caseworkers due to the challenges of the job.[4] Over the years, we have seen many examples of long call center wait times, indicating that demand is likely outpacing supply.[5–7]

The Nava Labs approach to developing and testing GenAI solutions that support caseworkers is rooted in human-centered design, iterative agile development, and rigorous program evaluation. The Nava Labs team conducted user research to understand caseworker needs and where GenAI might be an appropriate fit to address those needs, such as reducing administrative burdens, freeing up more quality time to meet with clients, and lowering the training barrier to helping clients. The Nava Labs team also developed proofs of concept for five GenAI-powered tools to address caseworker needs and started piloting some of those tools in real-world settings; Table 1 outlines each tool and its development phase as of December 2025. This report describes the evaluation methods and results for the assistive chatbot pilot.

Table 1.

GenAI tools Nava Labs is researching and developing as of October 2025

Tool	Purpose	Development Phase
Assistive chatbot	Retrieves program rules and provides plain-language explanations	Pilot complete
Referrals generator	Suggests local resources and government programs with action plans	Piloting
Application submission agent	Pulls data from a variety of sources to autocomplete benefits applications	Piloting
Document processor	Verifies that documents meet requirements	Piloting
Call notes generator	Minimizes note-taking burden and outlines next steps for the client	Proof of concept complete

About the assistive chatbot

Chatbot functionality

The chatbot aims to make it easier for caseworkers to find credible answers to questions about health and human services program eligibility and enrollment to discuss with their clients in real-time. Nava Labs sought to reduce cognitive load for caseworkers, speed up responses, and build their confidence. The chatbot solution:

- Leverages a foundational Large Language Model (LLM).
- Provides plain-language descriptions about program rules.
- Uses retrieval augmented generation (RAG) to pull information from pre-vetted sources only.
- Provides direct source citations from the pre-vetted sources with links to the original source if further exploration is needed.
- Provides multilingual translation support.
- Prevents hallucinations with clear guardrails around the chatbot's scope of knowledge; if a question is out of scope, the chatbot responds "I don't know the answer" or provides a link to the topic rather than making up a response.

The chatbot also lets caseworkers give real-time feedback on the quality of its responses to their questions.

Preparing the chatbot for real-world implementation

We intentionally built the chatbot technology to adapt to different settings and integrate with a range of different workflows through an application programming interface (API).

The team completed several rounds of prototyping and testing to prepare the assistive chatbot for the pilot with nonprofits and government agencies. Rigorous testing ensures the assistive chatbot is ready to use in real-world settings and collect data on usage and impact. The team iterated on building the solution until technical evaluations signaled high chatbot response accuracy and user testing showcased promise for addressing caseworker needs. The team also confirmed they met all security, privacy, and infrastructure requirements. Then they readied the assistive chatbot to integrate into the workflows of our pilot partner Amplifi, who integrated the chatbot with their Benefit Navigator tool and ensured a seamless user interface across the tools.

Chatbot evaluation design

Evaluation questions

Informed by principles of implementation science, we sought to identify factors that influenced the uptake and impact of the assistive chatbot.[3] In particular, we wanted to better understand the impact of the assistive chatbot on benefit navigation and considered five main research questions:

1
Accuracy
Does the chatbot help caseworkers make accurate decisions and improve the quality of information shared with clients?

2
Appropriateness
Does the chatbot enhance caseworkers' ability to help clients access the full breadth of benefits for which they are eligible?

3
Acceptability
Are caseworkers comfortable using the chatbot? To what extent do caseworkers prefer the chatbot over existing tools?

4
Administrative burden
Does the chatbot make it easier for caseworkers to help clients identify and enroll in benefits?

5
Accessibility
To what extent does the chatbot enable caseworkers of different abilities and backgrounds to use its features effectively?

Terminology

Before delving into our evaluation findings, it's helpful to clarify some terminology. For this report, we will use the term "caseworker" to mean anyone who helps people navigate and enroll in benefits, and it encompasses a range of formal role titles at organizations. Caseworkers may also have additional responsibilities in their roles beyond benefits navigation like providing ongoing support and coordination of services, assessing eligibility, and processing applications.

Evaluating the chatbot

An offline experiment

We partnered with researchers from the Better Government Lab and Cornell University's Department of Information Science to better understand how AI-powered chatbots impact the accuracy of information communicated by caseworkers.

The researchers conducted a randomized controlled trial (RCT) experiment to examine how chatbot response quality affected caseworker accuracy in practice. Do caseworkers actually benefit from helpful chatbot suggestions? Would caseworkers give up on inaccurate chatbots? To what extent would caseworkers follow incorrect chatbot suggestions?

In our randomized experiment, caseworkers answered multiple-choice questions about real-world situations. SNAP Quality Control (QC) errors and input from SNAP auditors determined the difficulty of the questions. We randomly assigned some caseworkers to receive simulated chatbot suggestions, while others saw no suggestions and served as the control group. For those who received chatbot suggestions, we systematically varied the chatbot's accuracy between 55% and 100% to examine how caseworkers' behavior changed in response to different levels of perceived chatbot quality. Institutional review boards at both Georgetown University and Cornell University determined this study as exempt from human subjects research.

Drawing on the researchers' findings, we estimated the effect of the assistive chatbot on improving caseworkers' accuracy. Further results from this offline experiment will be presented in forthcoming academic publications.

Piloting the chatbot with caseworkers

We partnered with Amplifi (formerly known as Imagine LA) — a 501(c)(3) nonprofit organization that creates tools that simplify complex systems so people can access essential benefits and resources with dignity — to implement the chatbot and study its impact on caseworkers in real-world settings.[8] Amplifi has over 18 years of experience providing direct social services to clients and and is the home of Benefit Navigator, a web-based tool that helps caseworkers and their clients navigate federal, state, and local public benefit and tax credit programs.[9] The tool includes a benefit calculator, trusted benefit information hub, personalized action plans with application tips and links, tools to help people identify and avoid benefit cliffs and dashboards to track impact.

The Nava Labs and Amplifi teams integrated the open source assistive chatbot as a new feature in Benefit Navigator, with the aim to help caseworkers find and share information with clients more quickly and easily. Our teams engineered the chatbot interface to be consistent with Benefit Navigator's overall design, and we gathered a large set of trusted sources from Benefit Navigator as well as publicly available policy manuals to serve as the source material for the assistive chatbot.

We also partnered with the Better Government Lab, a joint research center from Georgetown University's McCourt School of Public Policy and the University of Michigan's Gerald R. Ford School of Public Policy, who provided evaluation expertise and direct support throughout the pilot.

Pilot methodology

We used a mixed methods approach that included both quantitative and qualitative data to evaluate the effects of the assistive chatbot on caseworkers' experiences. By combining multiple data sources, we sought to better understand how and why the chatbot affected benefit navigation for caseworkers. During the pilot, we collected and analyzed data from three sources:

1

Product log data

We analyzed log data on a regular basis to monitor chatbot adoption and engagement to identify usage patterns and trends.

2

Survey data

Caseworkers completed baseline and endpoint surveys to help us evaluate changes in caseload, administrative burden, workflow, and attitude toward AI. In addition to structured questions, our surveys included open-ended items to invite qualitative feedback and nuances.

3

In-depth interviews

We interviewed caseworkers before and after the pilot to understand the chatbot's usability and relevance within the context of their workflows and organizations.

Pilot sample and study design

Nava Labs recruited 61 caseworkers across six human services agencies (sites) in Los Angeles County within Amplifi's network to participate in the assistive chatbot pilot. Sites included organizations that support people with chronic homelessness, underserved youth, community college students, low income families, and uninsured and publicly insured individuals, helping them connect to health and social services, employment opportunities, and financial counseling.

To study the effects of the assistive chatbot on caseworkers, we used a quasi-experimental design that included intervention and comparison groups. At each site, we gave half of the caseworkers access to the chatbot (i.e., intervention group), while the remaining caseworkers maintained their existing workflows without chatbot access (i.e., control group). A supervisor, manager, or director at each site determined the group assignment.

After setting up intervention and comparison groups, caseworkers completed a brief baseline survey. To mitigate the risk of bias, we verified that caseworkers in both groups were relatively similar at baseline, including their knowledge of benefit programs, time spent with clients, workload, how often they handled complex client situations, administrative burden, and attitude toward AI. With the exception of a few measures, we found no statistically significant differences between the two groups, indicating that the intervention and comparison groups were fairly balanced.

The assistive chatbot pilot operated from March 11 to June 11, 2025. At the end of the three-month pilot, caseworkers completed a longer endpoint survey. Since each caseworker had a unique identifier, we were able to link their baseline and endpoint responses at the individual level. Additionally, among caseworkers in the intervention group, product log data provided detailed insights about caseworkers' usage patterns and interactions with the chatbot, including the prompts they submitted and the accompanying responses generated by the chatbot's large language model.

Out of 61 caseworkers, 57 completed the baseline survey, corresponding to a 93% response rate. Forty-three out of 57 caseworkers completed the endpoint survey, corresponding to a 75% response rate. As in many real-world prospective studies, the study experienced attrition when 4 caseworkers changed jobs in the midst of the pilot, decreasing our sample size from 61 to 57 participants.

Our final data set consisted of the following subgroups:

39 caseworkers who completed both baseline and endpoint surveys

18 caseworkers who completed only the baseline survey

4 caseworkers who completed only the endpoint survey

Pilot sample and study design

We collected data exclusively for product development, quality improvement, and product evaluation purposes. We obtained informed consent from all participating caseworkers prior to data collection. We administered the surveys using Qualtrics and included caseworkers' email addresses to enable linkage between baseline and endpoint responses. To protect privacy and minimize security risks, we didn't collect client-level data. Additionally, we stored all project data on password-protected servers that were accessible to authorized project team members and partners, including staff from Amplifi and researchers at the Better Government Lab and Cornell University.

Outcomes

1

Domain 1
Measurement and instrument development

Outcome 1.1:
We effectively adopted validated measures and instrumentation to study the effects of AI-powered tools in a public benefits context.

What we did

Adapted the Administrative Burden Scale

In the pilot, we adapted a validated survey instrument to quantify the chatbot's impact on caseworkers' experience of administrative burden, which is defined broadly as any "challenge imposed on people that makes it significantly more difficult to access or maintain a benefit for which they would otherwise be eligible."[10] The Administrative Burden Scale (ABS) assesses the barriers people face when navigating government programs.[11–13] These burdens include three types:

Learning cost
The effort to understand eligibility and rules.

Compliance cost
The time and resources spent on paperwork and procedures.

Psychological cost
Stress, frustration, and stigma incurred by the process of applying for benefits.

Taken together, these burdens determine whether people access benefits and how they experience government services. We slightly modified each item in the ABS to improve fit with the study context. We also conducted informal cognitive testing to ensure that questions were clear and interpreted as intended by caseworkers.

Implemented a quasi-experimental design

We used a quasi-experimental design, which included intervention (chatbot access) and comparison (status quo) groups, to evaluate the impact of the chatbot on caseworkers' experiences. Baseline and endpoint survey data enabled us to compare pre-post changes in key measures during the pilot.

Implemented a complementary randomized controlled trial (RCT)

Although the real-world pilot did not include a random assignment of caseworkers, we conducted a separate "offline experiment" RCT to examine how chatbot accuracy affected human accuracy among caseworkers.

What we learned

Through this pilot and offline experiment, we demonstrated the feasibility and value of applying rigorous research methods, including the use of validated scales, quasi-experimental design, and randomized experiments, to evaluate the impact of AI tools on caseworkers in the public benefit context.

Why it matters

This pilot and offline experiment highlight the feasibility and value of applying rigorous research methods — including a validated scale, quasi-experimental design, and a randomized experiment — to assess the impact of AI tools on caseworkers within the public benefits context. Our findings demonstrate that AI-powered tools, including chatbots, can be systematically evaluated for its effects on caseworkers' experience, benefit navigation, and service delivery.

Outcomes

2

Domain 2
Accuracy

Outcome 2.1:
Use of the assistive chatbot is estimated to improve caseworkers' accuracy by an average of 40%.

What we did

Nava Labs partnered with researchers from the Better Government Lab and the Cornell Department of Information Science to conduct a randomized experiment, known as the "offline experiment," to assess how variation in chatbot response accuracy impacted caseworkers' ability to navigate complex benefit scenarios.

In this experiment, 125 caseworkers with working knowledge of CalFresh, California's SNAP program, completed an online questionnaire that consisted of 45 multiple-choice questions simulating complex scenarios about the program. Questions ranged in difficulty level (easy, medium, and hard) and each question included at least four response options. A panel of CalFresh Quality Control auditors with expert knowledge of this benefit program determined the best response to each question.

We randomly assigned participating caseworkers to one of four groups as listed opposite:

Control group
Caseworkers received no simulated chatbot assistance; they answered multiple-choice questions based entirely on their own knowledge and experience.

"Good" simulated chatbot group
Caseworkers in this group received some simulated chatbot assistance. For each multiple-choice question, the system displayed a suggested answer from the simulated chatbot. Caseworkers could choose to adopt the suggested answer or select a different response. Across the 45 questions in the questionnaire, we intentionally calibrated the simulated chatbot to be correct approximately 60–65% of the time.

"Better" simulated chatbot group
Similar to the experimental group described above, but we intentionally calibrated the simulated chatbot to be correct approximately 85–90% of the time.

"Best" simulated chatbot group
Similar to the two experimental groups described above, but we intentionally calibrated the simulated chatbot to be correct at least 95% of the time.

2

Domain 2
Accuracy

Caseworkers in the three treatment groups (i.e., "good," "better," and "best" simulated chatbot) received training and technical assistance on chatbot usage prior to completing the CalFresh questionnaire. At the time of testing (Fall 2025), the Nava-developed assistive chatbot performed in the "better" category, outperforming all baseline LLMs.

What we learned

Caseworkers who received assistance — in the form of suggested answers — from the "better" and "best" simulated chatbots scored, on average, 24 and 25 percentage points higher than those in the control group, who received no simulated chatbot assistance.

Across all simulated chatbot groups ("good," "better," and "best"), caseworkers scored, on average, about 20 percentage points higher than those in the control group.

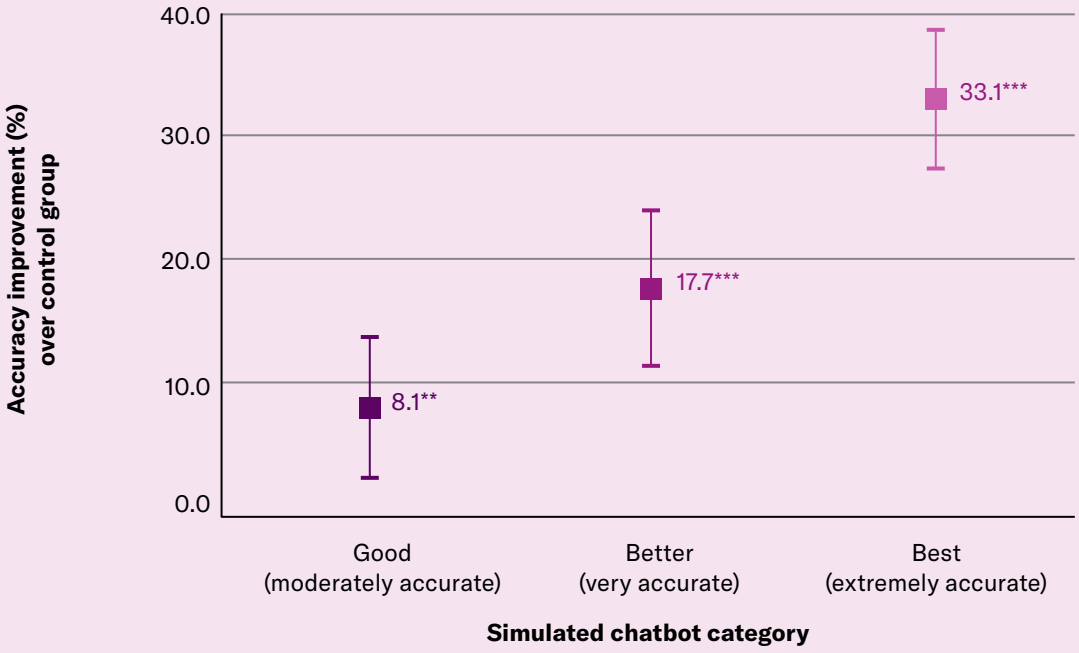
Our analyses showed statistically significant improvements, suggesting that the observed effects were unlikely to have occurred by chance, meaning that the results were likely due to the chatbot support. Overall, the assistive chatbot increased accuracy scores by approximately 24 percentage points, representing a 40% improvement relative to the control group. These results are summarized in Figure 1.

Why it matters

Overall, this randomized experiment provides strong evidence that an assistive chatbot can help caseworkers perform their work more accurately. Caseworkers showed especially clear improvements when handling complex client scenarios related to the CalFresh (SNAP) program.

Figure 1.
Average benefit navigator accuracy improvements by simulated chatbot category

** p<0.01;
*** p<0.001



2

Domain 2
Accuracy

Outcome 2.2:
Chatbot accuracy had the largest impact on more difficult client questions.

What we did

During the offline experiment, researchers from the Better Government Lab and Cornell Department of Information Science analyzed the data by comparing accuracy across questions of varying difficulty levels — easy, medium, and hard. They examined how caseworkers performed when the simulated chatbot provided either correct or incorrect suggestions. They determined question difficulty by using the control group's performance, with questions grouped into the three difficulty categories based on control group accuracy rates.

What we learned

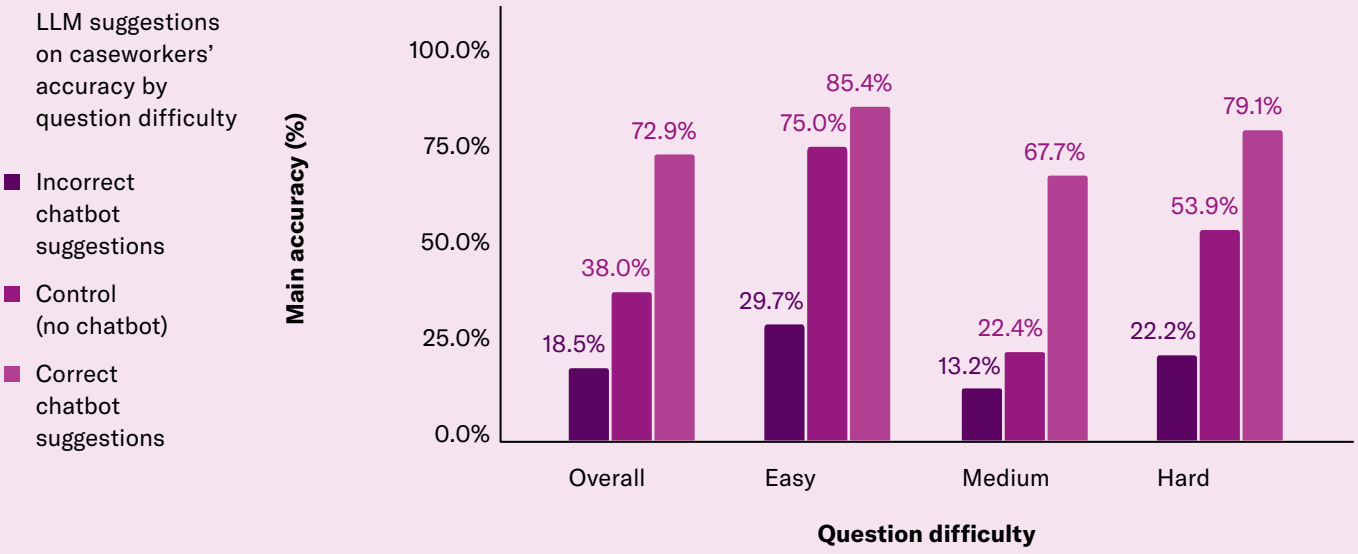
Caseworkers generally followed the simulated chatbot's suggestions, whether those suggestions were correct or not. When faced with a correct suggestion from the simulated chatbot, caseworkers' average accuracy increased by about 35 percentage points. However, when faced with an incorrect suggestion from the simulated chatbot, accuracy decreased by about 20 percentage points.

The simulated chatbot's correct suggestions proved to be most helpful on difficult questions. When the simulated chatbot provided correct suggestions, caseworkers achieved the largest accuracy gains on the most difficult questions. Conversely, when the simulated chatbot gave incorrect suggestions, caseworkers' accuracy declined noticeably on "easy" and "medium" questions. This finding implies that caseworkers could have second-guessed their otherwise correct responses. These findings are displayed in Figure 2.

Why it matters

These findings suggest that chatbots may be most valuable when caseworkers encounter complex or unfamiliar client scenarios. In contrast, chatbots may be less useful, and potentially counterproductive, for simpler questions, where chatbot suggestions could lead caseworkers to second-guess their otherwise correct responses.

Figure 2.
Impact of correct vs. incorrect LLM suggestions on caseworkers' accuracy by question difficulty



Outcomes

3

Domain 3
Appropriateness

Outcome 3.1:
The assistive chatbot offered caseworkers reliable and timely information about a wide range of public benefit programs.

What we did

To better understand how caseworkers used the assistive chatbot to help clients navigate benefit programs, we analyzed a set of 271 chatbot prompts submitted during the pilot. A team of three reviewers conducted line-by-line review and systematically coded each prompt by the benefit programs mentioned, the type of inquiry, and the underlying client needs. In addition, we interviewed caseworkers to gather further qualitative insights about their experiences using and interacting with the chatbot.

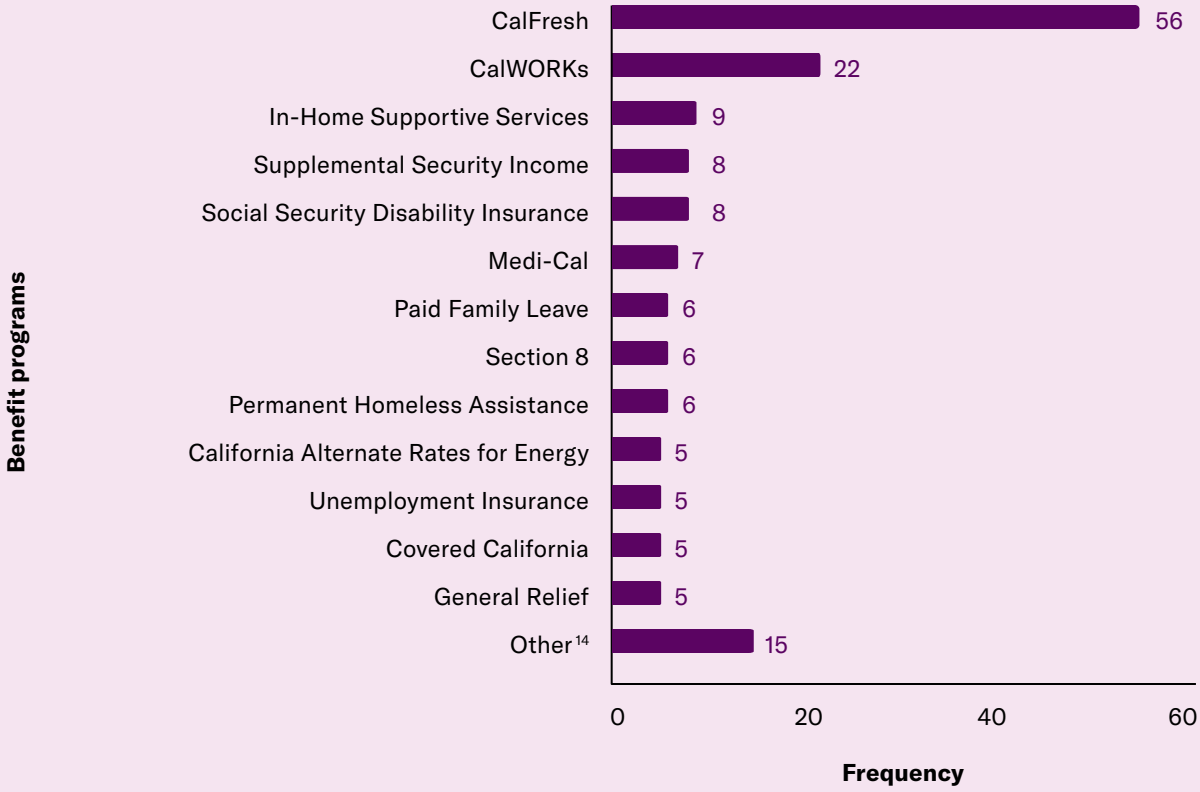
What we learned

Programs inquired

The assistive chatbot provided information about 22 distinct benefit programs to caseworkers. Of the 271 prompts, 163 (60%) referenced specific programs or included sufficient detail to infer the intended program. The most frequently mentioned programs included CalFresh (34%), CalWORKs (14%), In-Home Supportive Services (6%), Supplemental Security Income (5%), Social Security Disability Insurance (5%), Medi-Cal (4%), Paid Family Leave (4%), Section 8 housing (4%), and Permanent Homeless Assistance (4%). Figure 3 shows the frequency distribution of these programs inquired.

Domain 3
Appropriateness

Figure 3.
Frequency of benefit programs inquired by caseworkers

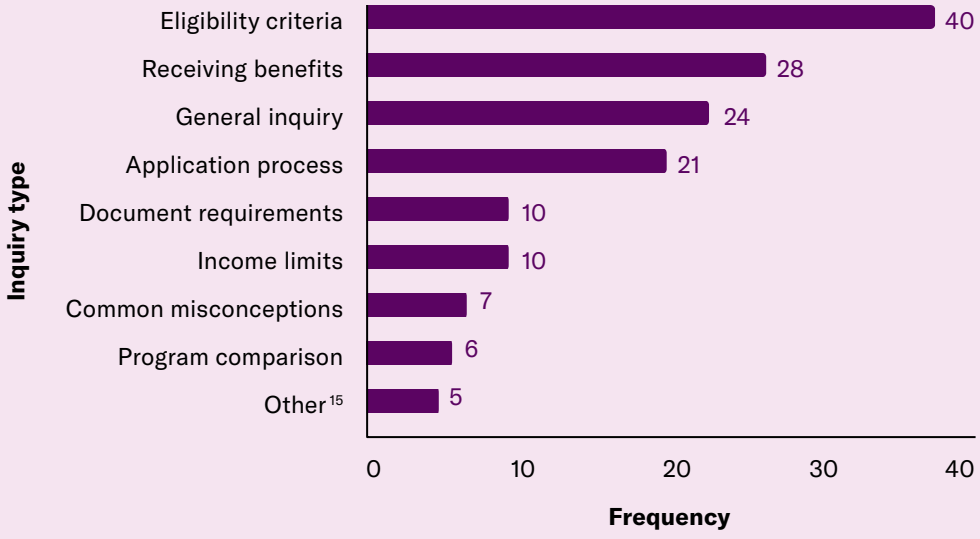


Domain 3
Appropriateness

Types of inquiries

Prompts submitted by caseworkers covered 11 types of inquiries. We were able to categorize 151 (56%) of the 271 prompts by type. The most common inquiry types included program eligibility criteria (27%), benefit access post-enrollment (19%), general program inquiries (16%), and application processes (14%). Other additional inquiries included document requirements (7%), income limits (7%), clarifications on common misconceptions (5%), and cross-program comparisons (4%). Figure 4 displays the frequency distribution of inquiry types submitted.

Figure 4.
Types of inquiry submitted by caseworkers

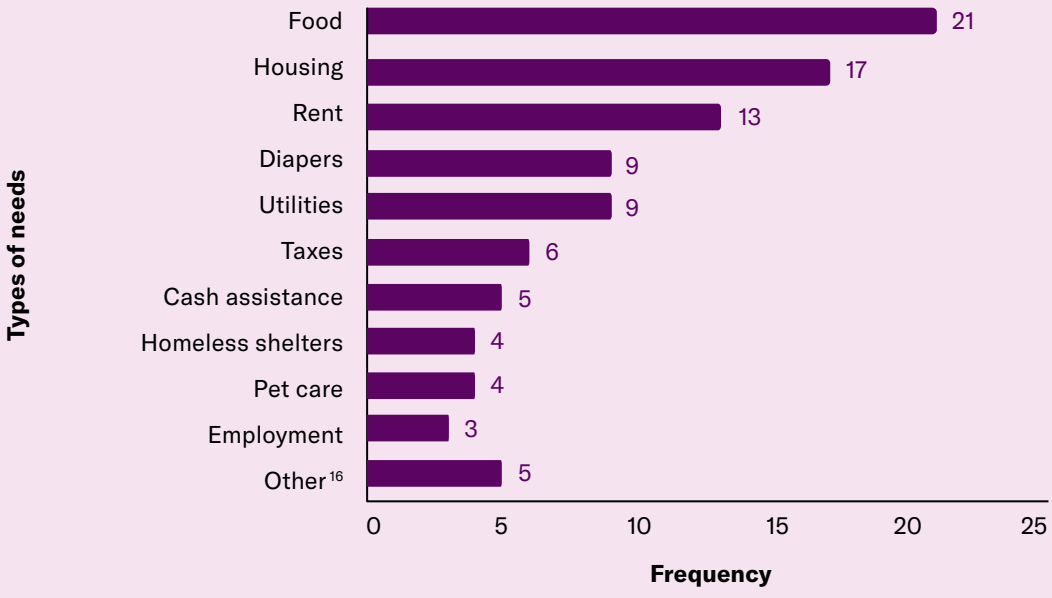


Domain 3
Appropriateness

Client needs addressed

Of the 271 prompts, 96 (35%) contained sufficient detail to identify or infer the specific client need that caseworkers aimed to address. The most common client needs included food access (22%), housing support (18%), rent assistance (14%), diapers and child care items (9%), utilities assistance (9%), tax filing and financial literacy support (6%), temporary housing shelters (4%), and pet care (4%). Figure 5 shows the frequency distribution of client needs identified in prompts.

Figure 5.
Client needs identified based on caseworker prompts



Domain 3
Appropriateness

Time optimization

Qualitative interviews with caseworkers revealed that the assistive chatbot supported benefit navigation by reducing the time and effort spent researching and interpreting complex policy materials. Instead of pausing to take notes, review documents, and conduct follow-up communication, caseworkers used the assistive chatbot to quickly access accurate, up-to-date, and plain language information that can be readily shared with clients. This tool enabled caseworkers to spend more time engaging directly with clients — including listening to their needs, identifying suitable solutions, explaining eligibility and certification requirements, and building client trust.

"I don't have to go outside of the chatbot to go and search for [answers]. So, being able to just have that response on hand within seconds for our [clients] was amazing, instead of 'let me write down all the questions you have and I'll go and search for themand then I'll read them back to you later on.'"

Caseworker

Why it matters

Taken together, these descriptive findings reveal that caseworkers used the assistive chatbot in a variety of ways to support clients and optimize service delivery. The wide range of programs, inquiry types, and client needs reflected in the prompts illustrates the chatbot's ability to adapt to the complex demands of benefit navigation. The breadth of usage also highlights the chatbot's appropriateness, compatibility, and relevance for caseworkers seeking accurate and timely information about benefit programs. Finally, these usage patterns indicate that the chatbot not only met its technical requirements but also effectively supported caseworkers in guiding benefit-seeking clients.

Outcomes

4

Domain 4
Acceptability

Outcome 4.1:
65% of caseworkers with chatbot access reported real-world use.

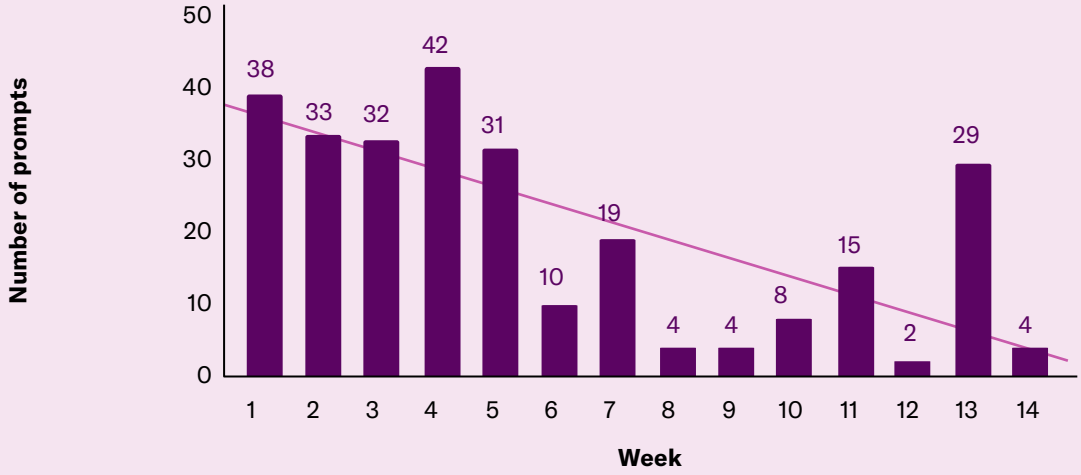
What we did

We collected chatbot use data throughout the pilot and linked it to individual caseworkers. This enabled us to examine how each caseworker interacted with the assistive chatbot, including who used it, when they used it, and what prompts they submitted.

What we learned

During the 14-week pilot, 20 out of 31 caseworkers (65%) with chatbot access used the tool actively, generating a total of 271 prompts. Each caseworker submitted an average of 14 prompts, though individual usage varied from as few as 1 prompt to as many as 40 prompts. The median number of prompts per caseworker was 11. Figure 6 shows chatbot use by pilot week and the downward-sloping trendline reflects declining usage over time.

Figure 6.
Chatbot use over time



Domain 4
Acceptability

To explore this usage trend, we disaggregated the chatbot product data and compared usage by site, week, and caseworkers, as shown in Figure 7. Our analyses surfaced these potential explanations and insights:

1

Use varied by site

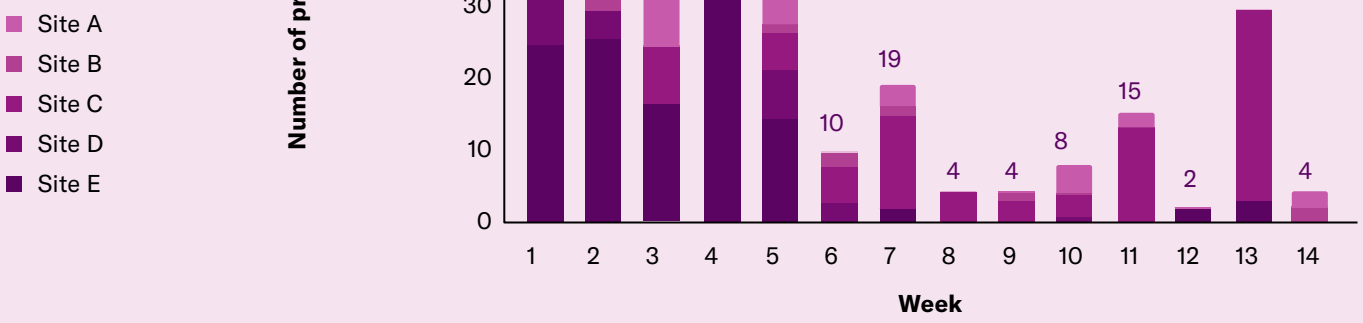
Caseworkers at site E (in blue) used the chatbot frequently from the very beginning, and this activity level persisted for about five weeks. Interestingly, just as their usage began to decline, caseworkers at site C (in yellow) began to use the chatbot more regularly. The remaining sites — including A (in orange), B (in green), and D (in red) — demonstrated occasional use of the assistive chatbot.

2

Use depended on service delivery models

The ways in which sites delivered client services also shaped how caseworkers used the chatbot. For example, at site E (blue), caseworkers conducted client intake, triage, and referrals through one-time interactions that typically did not require follow-up visits. In contrast, at site C (yellow) caseworkers maintained longer-term relationships with clients, many of whom participated in follow-up visits.
These differing service delivery models likely shaped chatbot usage — whereas a rapid triage approach may have created more opportunities to use the chatbot, ongoing case management may have needed less frequent chatbot interactions.

Figure 7.
Use by site measured by number of prompts submitted to the chatbot per week



4

Domain 4
Acceptability

Why it matters

Usage variation reflects how caseworkers viewed the assistive chatbot's acceptability and usefulness with respect to their existing workflow. The high initial uptake (weeks 1–5) indicates that caseworkers found the chatbot sufficiently acceptable to try. However, the subsequent decline in use (beginning in week 6) suggests challenges to sustain engagement over time. In the absence of sending reminders and providing training support, the chatbot may have lost some relevance and became deprioritized in caseworkers' workflow.

Differences in use across sites further indicate that acceptability is affected by multiple factors, including caseworkers' needs and workflows, organizational capacity and buy-in, and other external factors (e.g., large-scale wildfire that coincided with the pilot period). These findings underscore that maintaining acceptability requires not only a well-designed product but also consistent implementation support over time. For future iterations, we identified two key insights:

1

Review product data early, consistently, and collaboratively with partners

Analyzing usage data and feedback from caseworkers enabled emerging patterns and gaps to be identified. These data points informed strategies to refine the chatbot and improve service delivery. For example, upon observing a growing number of prompts inquiring about diapers and childcare items, we quickly added a trusted resource (via the Special Supplemental Nutrition Program for Women, Infants, and Children) to the chatbot, which subsequently enabled caseworkers to provide timely and accurate guidance to parents and caregivers.

2

Plan for sustained engagement

Maintaining product use requires more than a successful launch. Strategies such as sharing regular usage updates, hosting informal feedback sessions, and highlighting innovative prompting by caseworkers can help keep the product top-of-mind and encourage continued use.

4

Domain 4
Acceptability

Outcome 4.2:
Several "super users" of the assistive chatbot emerged during the pilot.

What we did

We analyzed product data to identify caseworkers who frequently used the assistive chatbot. We then conducted in-depth interviews with these caseworkers to better understand their motivation and experience interacting with the chatbot.

What we learned

1

Consistent exposure to chatbot-related content

In the early weeks, caseworkers received frequent communications from our team, including training materials, product updates, and requests for feedback about the chatbot. These regular touchpoints helped to maintain interest and awareness of the chatbot as a valuable resource for caseworkers.

2

Seasonal variation in demand for benefit navigation support

Site C, a higher education institution, follows an academic calendar (i.e., school in session from September to May). Caseworkers at this site began to use the chatbot in week 5 (early April), followed by gradually increasing usage over time that peaked in week 13 (early June). Interviews with caseworkers revealed that their usage is driven in part by students' needs — some students have to transition from their usual living arrangement (e.g., on-campus housing) to short-term summer arrangements, including housing, transportation, and potentially insurance coverage. Navigating these transitions led to increased use of the assistive chatbot.

3

Peer and organizational reinforcement

Some sites listed internal goals for chatbot usage, which further motivated caseworkers to increase adoption. Caseworkers at site E reported sharing tips and tricks on prompting with one another, which also likely contributed to increased adoption. Endorsement from supervisors also lent an extra layer of buy-in to promote use among caseworkers.

4

Domain 4
Acceptability

Why it matters

The emergence of these "super users" highlight how strategically fostering acceptability — or the extent to which stakeholders view a tool as agreeable, satisfactory, or useful — shaped caseworkers' chatbot usage during the pilot. Factors including hands-on training, frequent reminders, and peer-to-peer support from colleagues and supervisors increased caseworkers' willingness to integrate the chatbot into their workflows. In contrast, without these enabling factors, the chatbot may have been viewed as less acceptable, which may explain the lower adoption rates at some sites. Overall, these findings suggest that building acceptability requires not only a well-designed product but also ongoing support to build and maintain relevance and value over time.

4

Domain 4
Acceptability

Outcome 4.3:
The Net Promoter Score suggests a fairly successful pilot with opportunities for improvement.

What we did

After the chatbot pilot ended, we surveyed caseworkers and asked how likely they were to recommend the assistive chatbot to colleagues using a scale from 0 to 10, with 0 being "extremely unlikely" and 10 being "extremely likely." This measure is known as the Net Promoter Score (NPS), and it identifies three types of users:[17, 18]

"Promoters" consist of users who give a rating of 9–10. They are loyal advocates who are likely to recommend the product to others.

"Passives" consist of users who give a rating of 7–8. While they may be satisfied, they are unlikely to actively promote the product.

"Detractors" include users who give a rating of 0–6. They are generally dissatisfied and may deter others from using the product.

What we learned

Out of the 31 caseworkers with access to the assistive chatbot, 18 (58%) responded to this NPS question. Figure 8 displays the distribution of user types, with 40% of caseworkers categorized as promoters, 33% as passives, and the remaining 28% as detractors. Based on their responses, we derived an NPS of 11.

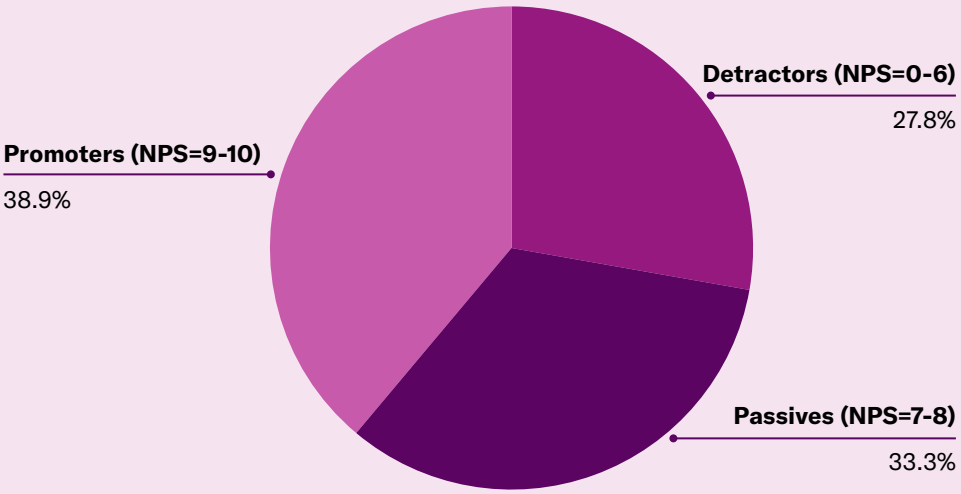
Caseworkers were moderately inclined to recommend the chatbot. While nearly 40% of caseworkers were "promoters," the presence of passives and detractors reflects mixed adoption and room for improvement.

Domain 4
Acceptability

Why it matters

With 40% of caseworkers as promoters, there's promising early signal of the chatbot's acceptability in practice. These findings also underscore the importance of continuing to improve the chatbot and enhance features that provide positive user experience for caseworkers.

Figure 8.
Distribution of user types based on the Net Promoter Score



Outcomes

5

Domain 5
Acceptability

Outcome 5.1:
There is promising evidence that the assistive chatbot may reduce learning and psychological costs, but results were inconclusive due to small sample size and wide variation in experiences.

What we did

We analyzed data from the baseline and endpoint surveys from caseworkers in both intervention and control groups. We utilized the Administrative Burden Scale in both surveys, which consisted of questions on learning, compliance, and psychological costs. Our hypotheses included the following:

1
Learning cost
Given the chatbot's ability to retrieve and summarize complex information about benefit programs, we hypothesized the chatbot would reduce learning cost among caseworkers.

2
Compliance cost
As the chatbot shows detailed guidance on how to access benefit programs (e.g., step-by-step instructions, acceptable documents to verify eligibility), this can increase the amount of effort needed to fulfill these requirements. As such, we hypothesized that the chatbot would increase compliance cost among caseworkers.

3
Psychological cost
The chatbot streamlines caseworkers' information-finding process, which in turn enables them to have more time to interact directly with clients. Given the expected increase in client-centered interactions, we hypothesized that psychological cost would decrease among caseworkers.

In addition to analyzing survey responses from caseworkers, we conducted seven semi-structured interviews with caseworkers who used the chatbot. These interviews aimed to deepen our understanding of how caseworkers experience administrative burden and the extent to which the chatbot may help reduce it. We analyzed interview transcripts using qualitative methods to identify common themes and insights related to chatbot use and its potential to streamline tasks or workflows that may be particularly burdensome.

Domain 5
Acceptability

What we learned

Baseline and endpoint surveys

In the subsample of 39 caseworkers who completed both baseline and endpoint surveys, we observed modest shifts in administrative burden across both the intervention and control groups. Learning costs declined in both groups and psychological costs also decreased, indicating some potential improvements in caseworkers' experience over time. Compliance costs increased slightly for both groups, though the change remained quite small. While none of these pre-post differences were statistically significant, the overall pattern of findings offer promising insight into how administrative burden evolved during the pilot and can help guide future improvements to the chatbot.

Interviews with caseworkers

Qualitative analysis of caseworker interviews uncovered more nuanced insights into how the assistive chatbot helped to reduce administrative burden during benefit navigation. The chatbot altered caseworkers' experience of administrative in three ways:

1
Reducing learning costs by improving information retrieval and summary

2
Reducing psychological costs by enhancing case management activities

3
Bridging knowledge gaps in complex and evolving policy areas

Caseworkers also expressed concerns about data currency, particularly given the broader context of evolving federal priorities and policies that occurred during our pilot period. Below is a summary of these qualitative findings.

1
Reducing learning costs by improving information retrieval and summary

A major source of administrative burden for caseworkers is the cognitive effort required to sort through large volumes of complex and often fragmented policy information. Caseworkers reported that the chatbot reduced this learning cost by streamlining how they find and synthesize relevant policies. Unlike search engines — which require users to review results and filter out irrelevant information — the chatbot delivered curated responses that distilled key details from reliable sources.

When comparing the chatbot to search engines, one caseworker noted the difference in time and effort spent to answer clients' benefit questions:

"[A search engine] will come up with a hundred different answers that you don't want and to try and click on every single one to find out what you actually want, the chatbot narrows it down to where you don't have such a large variety to try and search through."

Similarly, another caseworker highlighted the accessibility of the chatbot's output, which converted dense information into more comprehensible language:

"I think [using the chatbot is] easier because it would just give you the little bullet points of what it is instead of going to [a search engine] and then a whole bunch of documents come up."

For another caseworker, the chatbot's ability to quickly summarize dense information contributed to efficient and clear communication with clients:

"I feel the chatbot is a lot easier because if you plug in your question, it'll give you a brief rundown. It's quicker. Instead of Google — where you're reading and searching for relevant information to relay to a [client] — the [chatbot's] answer is just right there. Especially for questions like, 'What documents do you need for [a particular benefit program]?' It breaks [information] down. It's so much easier than reading through an entire article."

Domain 5
Acceptability

Learning costs can also show up as delays — i.e., the time between a caseworker receiving a question and obtaining the information needed to answer it reliably. Caseworkers noted that the chatbot reduced efforts spent on researching information, contacting agencies to seek clarification, and following up with clients to share information. For clients who may be difficult to reach after the initial appointment, the chatbot's ability to provide real-time information at the point-of-care may be particularly valuable.

A caseworker explained how the chatbot decreased a backlog of research questions accumulated from navigating less familiar client circumstances:

"Being able to have that [chatbot] response on hand within seconds was amazing, instead of 'let me write down all the [client] questions, search for them, and then I'll read them back to [the client] later,' [the chatbot] minimizes the amount [of] back-and-forth with the [client]."

For another caseworker, the chatbot quickly became a helpful resource during busy workdays, when limited time and competing demands made it challenging to conduct in-depth research to find reliable answers for complex client cases.

"Being able to ask specific questions [about] public benefits, I found really helpful. [...] It's a tool for us to just be more knowledgeable and not have to dig for answers. [...] My day-to-day looks different every day and a lot of times it can be chaotic, so having something to kind of rely on and just ask a quick question versus having to Google it [...] is really helpful in the work that we do and the intakes that we do with [clients]."

2
Reducing psychological costs by enhancing case management activities

Caseworkers often work in high-stakes situations where giving incorrect information can have serious consequences for clients. These moments can create significant psychological pressure, including uncertainty, anxiety, and hesitation. The assistive chatbot served as a risk-mitigation tool by allowing caseworkers to double-check their knowledge and verify complex eligibility rules before advising clients.

Two caseworkers described the chatbot as a trustworthy tool to verify accuracy:

"Let me triple check for [the client] because I don't want to give [the client] the wrong answer. [...] I trust that [the chatbot] gives me the right information [because] it gives you the source."

"I kind of already knew [about a particular benefit program's eligibility criteria]. [But using the chatbot] was just confirming [...] I just wanted to double-check."

One of these caseworkers, who used the assistive chatbot to confirm her existing knowledge about medical coverage for therapy, also shared this feedback. This quick validation provided her with a peace of mind, knowing that she gave the client accurate information to help inform the client's decision-making.

"The [chatbot prompt] I asked was more of just double confirming because I do know that medical [insurance] does cover therapy sessions. [...] I just mainly wanted to triple check my own knowledge."

3
Bridging knowledge gaps in complex and evolving policy areas

Caseworkers often serve as generalists who are required to navigate complex client scenarios amidst evolving policy landscapes (e.g., foster youth tax credits, changes in immigration policies). The learning cost and cognitive effort needed to become an expert across multiple benefit programs can be prohibitively high and difficult to maintain. The assistive chatbot enabled caseworkers to make informed recommendations to clients without requiring extensive training or knowledge.

One caseworker used the chatbot to quickly verify details about tax credits for foster youth clients:

"I wasn't super knowledgeable about tax credits and the details before utilizing [the chatbot]. [I was] getting new information for the first time [and] this was the information I was looking for — how much [tax credits] would be and eligibility requirements."

Similarly, another caseworker used the chatbot to address a client's misunderstanding about child support enforcement in the CalWORKs program, helping alleviate a significant barrier for the client to gain access to CalWORKs benefits.

"A [client] had said [...] in order to get [CalWORKs] benefits, she needed to get child support. I asked the chatbot if child support [was needed], the chatbot said that you don't need it. [...] What's the balance of understanding the technical aspects of getting benefits, but also trying to understand where the [client] is coming from, because they're asking for help. They want help."

Concerns about data currency

While the chatbot helped reduce administrative burden, caseworkers identified areas where burden may still persist. In particular, they questioned how current the information in the chatbot's responses was. Although citations increased the perceived trustworthiness of the output, caseworkers noted that they often lacked the time or capacity to manually verify each source. As a result, they expressed general trust in the chatbot's responses but also lingering concerns about whether the information was fully up to date.

One caseworker voiced concerns about the rapidly changing policy landscape and suggested that including a "last updated" timestamp may help to build greater trust.

"Even though the website says it's up-to-date, knowing how case-by-case some of the [clients] I meet with, I get nervous [about] wanting to give [accurate] information. [...] I think especially right now in this political climate with things changing all the time, there's a part of me that [is] looking out for reassurance that information has been updated."

Overall, our interview findings suggest that the assistive chatbot meaningfully reduced dimensions of administrative burden by helping caseworkers access information more efficiently, feel more confident in their case management work, and bridge knowledge gaps in complex policy areas. These benefits were especially valuable in time-sensitive and high-pressure situations, where access to accurate information supported both caseworker workflows and client needs. At the same time, caseworker concerns about data currency point to a critical opportunity: while the chatbot can streamline information retrieval, its usefulness ultimately depends on maintaining up-to-date policy content in a fast-changing environment. Taken together, these insights suggest that the chatbot has strong potential to enhance benefit navigation. However, its ability to achieve meaningful impact will depend partly on making ongoing improvements to ensure sustained trust, relevance, and long-term adoption.

Outcomes

6

Domain 6
Accessibility

Outcome 6.1:
Chatbot responses averaged a 10th to 12th grade reading level.

What we did

We conducted a readability analysis to better understand the accessibility of the chatbot's responses, an important factor in its usefulness to caseworkers. We analyzed 271 chatbot responses using four grade-level readability measures (i.e., Flesch-Kincaid, Gunning Fog Index, SMOG Index, and Automated Readability Index), as well as the Flesch Reading Ease score and estimated reading time. Table 2 summarizes the readability tests used in the analysis.

What we learned

On average, the chatbot generated responses corresponding to a 10th- to 12th-grade reading level, which means caseworkers would require at least a high school-level reading proficiency to effectively engage with the chatbot's content. We also observed wide variation in the readability of chatbot responses, with some written at an elementary school level and others requiring college and even post-graduate levels.

Table 2.
Readability metrics

Readability metrics	Description
Flesch Reading Ease	How easy the text is to read; scores range from 0 to 100, with higher score indicating easier to read
Flesch Kincaid Grade Level	School grade (U.S.) needed to comprehend text
Gunning Fog Index	Years of education needed to grasp text on first read
Simple Measure of Gobbledygook (SMOG) Index	Grade level based on usage of complex words
Automated Readability Index	Grade level based on sentence and word length
Estimated reading time	Time needed to read through text (minutes)

Domain 6
Accessibility

Table 4 presents summary statistics from each readability assessment. The Flesch Reading Ease test reported an average score of 49, placing most chatbot responses in the "difficult [to read]" range (see Figure 9 for the distribution of Flesch Reading Ease scores). A large standard deviation hints at a wide distribution of reading difficulty levels. Similarly, the Gunning Fog, SMOG, and Automated Readability indices reported an average of 11th- to 12th-grade reading level, with some responses reaching into post-secondary levels of difficulty.

On average, a chatbot response took caseworkers about 31 seconds to read. However, this measure doesn't fully account for the cognitive effort needed to interpret technical or complex information.

Table 3.
Readability scores of chatbot outputs based on common readability metrics

Readability metrics (n=271 responses)	**Mean**	**Std dev.**	**Min**	**Median**	**Max**
Flesch Reading Ease (score 1–100; higher = easier to read)	49.3	16.9	1.0	50.3	86.7
Flesch-Kincaid Grade Level	10.3	3.1	3.7	10.2	27.9
Gunning Fog Index (grade level)	11.5	2.9	5.3	11.1	27.7
SMOG Index (grade level)	11.6	3.9	0.0	12.2	19.7
Automated Readability Index (grade level)	11.8	3.6	3.9	11.6	34.6
Estimated reading time (minutes)	0.5	0.3	0.1	0.5	1.4
Estimated reading time (seconds)	31.4	18.7	4.0	28.0	86.0

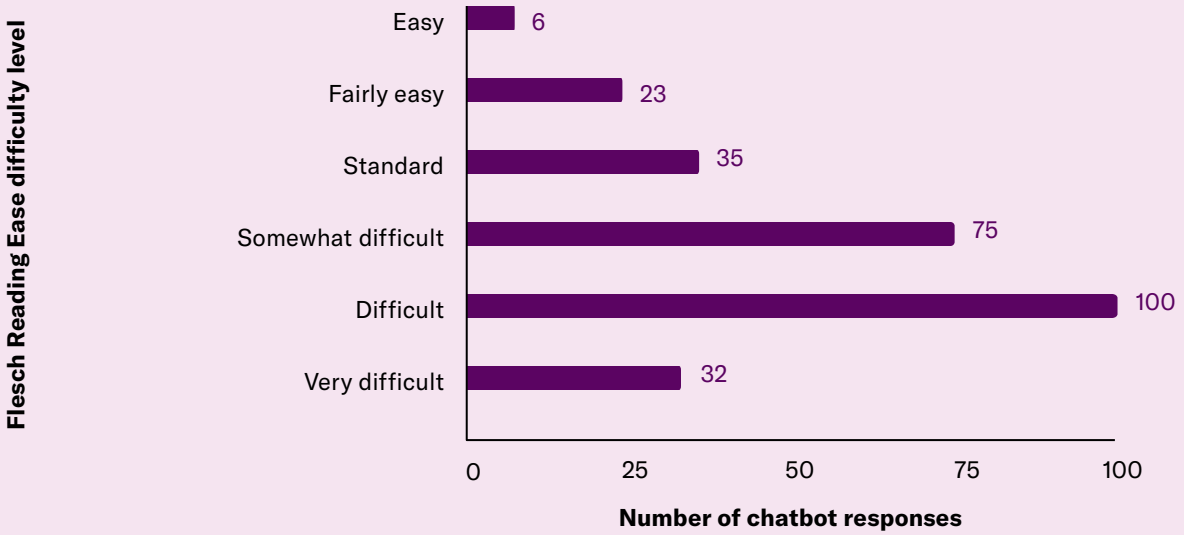
Domain 6
Accessibility

Why it matters

These findings have important implications for accessibility, particularly in relation to the recommendations outlined in the Plain Language Act of 2010.[19, 20] While the chatbot's responses may meet the needs of more experienced caseworkers, their relatively high reading level can introduce confusion and create barriers to adoption. Caseworkers with fewer years of experience, limited familiarity with written English, or those working under significant time pressure may find dense or highly technical responses harder to interpret and use effectively.

There are several opportunities to improve accessibility. Ensuring that chatbot outputs consistently align with recommended readability levels would make the tool more usable for a wider range of caseworkers. Additional strategies — such as reducing jargon, shortening sentences, and presenting information in layered or scaffolded formats (e.g., brief takeaways followed by more detailed explanations) — would further enhance clarity and ease of use. Improving readability would not only support caseworkers in their daily work but also increase the value of the information they relay to clients.

Figure 9.
Chatbot responses by difficulty category on the Flesch Reading Ease readability metric



6

Domain 6
Accessibility

Outcome 6.2:
Chatbot responses are more accessible than source policy manual text.

What we did

While 76% of chatbot responses were categorized as at least "somewhat difficult" to read, the source materials the chatbot draws from (e.g., policy manuals) can be even more difficult to read. To test this hypothesis, we applied the same set of readability metrics to a randomly selected section from the CalFresh policy manual published by the Los Angeles County Department of Public Social Services (DPSS).[21]

What we learned

Across all five readability measures, the policy manual consistently required a higher reading level than the chatbot responses. The comparative results are shown in Table 4. Although both the DPSS manual and the chatbot responses fell into the "difficult" category based on the Flesch Reading Ease score, the chatbot responses scored higher (i.e., easier to read) by an average of 17 points. Based on grade-level readability metrics, the chatbot responses required fewer years of education to understand than the original policy manual. In fact, across the five readability measures, the chatbot reduced the required reading level by an average of two to five grade levels — representing a 12-36% improvement in readability.

Table 4.
Average readability scores for chatbot outputs and DPSS CalFresh manual across common readability metrics

Readability metrics (average)	**DPSS CalFresh manual**	**Chatbot responses**	**Difference**	**% change**
Flesch Reading Ease (score 1–100; higher = easier to read)	32.1	49.3	+17.2	+54%
Flesch-Kincaid Grade Level	15.1	10.3	-4.8	-32%
Gunning Fog Index (grade level)	17.9	11.5	-6.5	-36%
SMOG Index (grade level)	13.1	11.6	-1.6	-12%
Automated Readability Index (grade level)	14.4	11.8	-2.6	-18%

6

Domain 6
Accessibility

Why it matters

These results suggest that while chatbot responses could be further simplified, they represent meaningful improvements in making complex benefit program information more readable. At the same time, our findings point to a broader opportunity for government agencies to improve the readability of their policy documents, which would make crucial program-related information more accessible for caseworkers and benefit-seeking clients.

6

Domain 6
Accessibility

Outcome 6.3:
Multilingual capabilities are enhanced by caseworkers' language and cultural translation.

What we did

While the chatbot's underlying LLM is capable of supporting multiple languages, caseworkers rarely used this feature. During the pilot, less than 1% of prompts submitted by caseworkers used another language other than English. We conducted semi-structured interviews with several bilingual caseworkers to better understand why they primarily used the chatbot in English rather than in other languages (e.g., Spanish).

What we learned

These bilingual caseworkers shared that translating English responses from the chatbot in real-time enabled them to verbally simplify complex details, adapt explanations to cultural contexts, and build trust with clients. This "human-in-the-loop" approach was particularly useful when chatbot responses included technical, lengthy, or abstract information. Caseworkers emphasized serving as ad hoc interpreters helped maintain rapport with clients and minimized risks of miscommunication or loss of nuance that could occur with automated translation.

In addition, because the Benefit Navigator platform uses English, some caseworkers reported being unaware of or forgetting about the chatbot's multilingual features. This finding reveals an insightful implementation gap between the chatbot's technical capabilities and its actual use in day-to-day workflows.

"Usually I would just search in English and [either] translate or give them a summary [in Spanish]. The majority of the time the [client] will know what I am referencing."

Caseworker

6

Domain 6
Accessibility

Why it matters

From an implementation science perspective, accessibility includes not just whether a tool can serve diverse linguistic needs, but also whether users find it practical, trustworthy, and culturally responsive in real-world contexts. Making multilingual features more visible and user-friendly could promote wider adoption. At the same time, keeping a "human-in-the-loop" for language translation may be an essential feature to maintain, particularly given how caseworkers' judgment and cultural competence can contribute to social and relational benefits that cannot be replicated by or substituted with technology tools.

References

1
Giannarelli L, Minton S, Wheaton L, Knowles S.
A Safety Net with 100 Percent Participation: How Much Would Benefits Increase and Poverty Decline?
Urban Institute; 2023.
https://www.urban.org/research/publication/safety-net-100-percent-participation

2
Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM.
An introduction to implementation science for the non-specialist.
BMC Psychol. 2015;3(1):32.

3
Proctor E, Silmere H, Raghavan R, et al.
Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda.
Adm Policy Ment Health Ment Health Serv Res. 2011;38(2):65-76. doi:10.1007/s10488-010-0319-7

4
Haffield L.
Tackling turnover: How agencies are supporting and sustaining their workforce.
APHSA. August 30, 2022.
Accessed November 26, 2025.
https://aphsa.org/resources/tackling-turnover-how-agencies-are-supporting-and-sustaining/

5
Seefeldt KS.
Waiting it out: Time, action, and the process of securing benefits.
Qual Soc Work. 2017;16(3):300-316.

6
Holt SB, Vinopal K.
Examining inequality in the time cost of waiting.
Nat Hum Behav. 2023;7(4):545-555.

7
Lieb D.
Federal officials raise concerns about long call center wait times as millions are dropped from Medicaid.
PBS News. August 17, 2023.
Accessed November 26, 2025.
https://www.pbs.org/newshour/politics/federal-officials-raise-concerns-about-long-call-center-wait-times-as-millions-are-dropped-from-medicaid

8
Amplifi.
Benefit Navigator. Benefit Navigator.
Accessed October 28, 2025.
https://www.benefitnavigator.us

9
Lindbom E.
Tech for good: Families using new Benefit Navigator tool average $10,000 in additional public benefits.
EIN Presswire. October 24, 2025.
Accessed October 28, 2025.
https://www.einpresswire.com/article/859161752/tech-for-good-families-using-new-benefit-navigator-tool-average-10-000-in-additional-public-benefits

10
Schweitzer J.
How to address the administrative burdens of accessing the safety net.
Center for American Progress. May 5, 2022.
Accessed June 27, 2025.
https://www.americanprogress.org/article/how-to-address-the-administrative-burdens-of-accessing-the-safety-net/

11
Herd P, Moynihan D.
Administrative Burden: Policymaking by Other Means.
Russell Sage Foundation; 2019.

12
Herd P, Moynihan D.
Administrative burdens in the social safety net.
J Econ Perspect. 2025;39(1):129-150.

13
Jilke S, Bækgaard M, Herd P, Moynihan D.
Short and sweet: Measuring experiences of administrative burden.
J Behav Public Adm. 2024;7.

14
The "Other" category includes programs that received less than 5 prompts each, which included state disability insurance (SDI), workers' compensation, Cash Assistance Program for Immigrants (CAPI), Lifeline (utility assistance), Earned Income Tax Credit (EITC), CalWORKs Emergency Assistance to Prevent Eviction (EAPE), Medicare, utility assistance (EZ-SAVE), and Women, Infacts, and Children (WIC).

15
The "Other" category includes inquiry types that appeared less than 6 times each, which included appeals, utility assistance, and housing assistance.

16
The "Other" category includes client needs that appeared less than 3 times each, which included tenant rights, childcare, household items, and mental health.

17
Reichheld FF.
The one number you need to grow.
Harv Bus Rev.
Published online 2003.
Accessed November 26, 2025.
https://hbr.org/2003/12/the-one-number-you-need-to-grow

18
Baehre S, O'Dwyer M, O'Malley L, Story VM.
Customer mindset metrics: A systematic evaluation of the net promoter score (NPS) vs. alternative calculation methods.
J Bus Res. 2022;149:353-362.

19
Plain Language Guide Series – Digital.gov.
September 23, 2025.
Accessed October 2, 2025.
https://digital.gov/guides/plain-language

20
An Act to Enhance Citizen Access to Government Information and Services by Establishing That Government Documents Issued to the Public Must Be Written Clearly, and for Other Purposes.;
2010. Accessed October 2, 2025.
https://www.govinfo.gov/app/details/PLAW-111publ274

21
Los Angeles County Department of Public Social Services.
DPSS ePolicy.
Accessed October 2, 2025.
https://my.dpss.lacounty.gov/public/en/home/epolicy/home.html

Nava Labs

Nava Public Benefit Corporation

601 13th St NW, Floor 12
Washington, DC 20005

Email:
labs@navapbc.com

Web:
navapbc.com

For more information

Please get in touch if you're interested in learning more about the evaluation design, data collection tools, and data analysis. We'd be happy to discuss and share!