

AI Trust Reshaping Administrative Burdens: Understanding Trust-Burden Dynamics in LLM-Assisted Benefits Systems

Jeongwon Jo

University of Notre Dame
Computer Science and Engineering
South Bend, IN, USA
jjo3@nd.edu

Jie Cai*

Tsinghua University
Department of Computer Science and Technology
Beijing, China
jie-cai@mail.tsinghua.edu.cn

He Zhang

Pennsylvania State University
College of Information Sciences and Technology
State College, PA, USA
hpz5211@psu.edu

Nitesh Goyal

Google Deepmind, Google
New York, USA
niteshgoyal@acm.org

Abstract

Supplemental Nutrition Assistance Program (SNAP) is an essential benefit support system provided by the US administration to 41 million federally determined low-income applicants. Through interviews with such applicants across a diverse set of experiences with the SNAP system, our findings reveal that new AI technologies like LLMs can alleviate traditional burdens but also introduce new burdens. We introduce new types of learning, compliance, and psychological costs that transform the administrative burden on applicants. We also identify how trust in AI across three dimensions—competence, integrity, and benevolence—is perceived to reduce administrative burdens, which may stem from unintended and untoward overt trust in the system. We discuss calibrating appropriate levels of user trust in LLM-based administrative systems, mitigating newly introduced burdens. In particular, our findings suggest that evidence-based information disclosure is necessary in benefits administration and propose directions for future research on trust-burden dynamics in AI-assisted administration systems.

CCS Concepts

• Human-centered computing → Empirical studies in HCI.

Keywords

Administrative Burden; Trustworthy AI; Public Benefits, Large Language Models, Empathetic AI

ACM Reference Format:

Jeongwon Jo, He Zhang, Jie Cai, and Nitesh Goyal. 2025. AI Trust Reshaping Administrative Burdens: Understanding Trust-Burden Dynamics in LLM-Assisted Benefits Systems. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*, June 23–26, 2025, Athens, Greece. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3715275.3732077>

*Corresponding author



This work is licensed under a Creative Commons Attribution 4.0 International License. FAccT '25, Athens, Greece

© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1482-5/25/06
<https://doi.org/10.1145/3715275.3732077>

1 Introduction

Public benefits programs like the Supplemental Nutrition Assistance Program (SNAP) in the United States serve as critical safety nets for low-income people to afford basic needs, including food, housing, and healthcare. However, eligible individuals often face varying barriers in accessing these essential services, ranging from learning complex eligibility policies and procedures, dealing with welfare stigma and stress, to investing significant time and resources in meeting administrative requirements [6, 32, 54]. These barriers manifest through what policy administration scholars term *administrative burdens*, a framework that includes the *learning*, *psychological*, and *compliance costs* that citizens experience in their interactions with government services [32, 54].

While various information communication technologies (ICT) have shown promise in reducing some traditional administrative burdens and shifting burdens from the citizen to the state [31], research also demonstrates that they often transform rather than simply reducing the overall perceived burdens. For instance, glitches of digital systems can lead to delayed applications or lost data [36], or in-person services get reduced in favor of digital platforms [62], negatively affecting those with limited digital literacy or those without reliable technology access.

The introduction of LLM-assisted benefits systems, with their unique human-like conversational interfaces much advanced than traditional chatbot systems, presents both new opportunities and urgent questions about how these technologies might reshape administrative burdens and affect citizens' engagement with public services [39]. Understanding the lived experiences of citizens interacting with these novel AI-driven benefits systems remains crucial, especially given that much prior research focused on government staff perceptions or AI for decision-making rather than citizen-facing interaction.

To explore how LLM-assisted benefits systems could potentially affect perceived administrative burdens, we conducted 10 interviews with SNAP applicants. Our interviews surfaced rich data on various aspects of user experience, including potential design ideas and pain points. However, a central theme, trust in AI, emerged as particularly foundational. We argue that trust is a critical dimension to analyze in this context; first, in high-stakes scenarios, users are

more reluctant to trust AI [71], while trust acts as a key factor influencing initial engagement and determining whether users interact with and ultimately benefit from system features [45]. Second, we found that trust mediates how citizens experience administrative burdens when interacting with potential LLM systems. For example, trust in whether AI cares about humans can affect stress or comfort levels during interactions, mediating psychological costs. With trust in AI fundamentally shaping how administrative burdens are experienced in the context of AI interaction, understanding their relationship is a foundational step for designing effective human-LLM interactions in public service delivery.

Therefore, this paper focuses on the complex relationships between trust in potential LLM systems and administrative burdens within the SNAP context. Our analysis makes the following contributions. First, we reveal previously unidentified mechanisms through which trust shapes costs experienced with Generative AI systems, extending the existing administrative burdens framework. Second, we demonstrate how LLM systems can potentially reduce traditionally experienced administrative burdens while simultaneously introducing new, trust-related ones. Third, based on the observed trust and burden dynamics, we propose system design implications calibrated to foster appropriate levels of user trust in LLM systems, preventing over-trust and under-trust risks while mitigating newly introduced burdens. Finally, noting how participants' preferences between human caseworkers and AI often formed without clear evidence, we suggest performance metrics that should be disclosed to support evidence-based choices in benefits administration, proposing future research agendas.

2 Background: The Journey to SNAP Benefits

SNAP applications involve complex eligibility assessments and documentation requirements that many find challenging to navigate [54]. Applicants must gather and standardize various documents (e.g., pay stubs, lease contracts) to verify their financial hardship [46, 65]. The documentation process often creates significant stress that can impede completion [14]. Following submission, applications undergo review, potentially requiring additional information or complete resubmission if incomplete. Complete applications proceed to caseworker interviews, after which applicants receive decisions containing complex bureaucratic language and calculations [63]. Approved beneficiaries must later re-certify their eligibility and report circumstantial changes (e.g., income increases, cohabitation) that could affect their benefits. Failure to meet these ongoing requirements can result in benefit termination or criminal fraud charges. As such, the journey of SNAP benefits is lengthy, bulky, and burdensome, frequently requiring individuals to seek guidance from various external sources, including government resources, public officials, personal networks, and non-profit organizations [46, 65]. These administrative burdens often heavily affect individual outcomes.

3 Related Work

3.1 Administrative Burdens in Public Services

The administrative burden of interacting with government services fundamentally shapes access to public services. The administrative burden consists of learning, psychological, and compliance

costs [32, 54]. Learning costs arise from understanding government services, including rules, procedures, and personal relevance. Psychological costs stem from the emotional and mental strain of engaging with administrative systems, encompassing losses of personal autonomy, increased stress levels, and stigma associated with certain programs. Compliance costs are the time and money needed to meet administrative requirements and follow bureaucratic procedures. These costs affect whether eligible citizens access entitled services [54]. Compared to universal welfare programs, mean-tested programs generally impose higher administrative burdens [42], resulting in low acceptance rates among eligible beneficiaries [33].

Previous studies demonstrated how each cost impacts program participation (e.g., [6, 19, 27, 37, 59]). Approximately half of SNAP-eligible individuals mistakenly believe that they are ineligible, indicating that they would apply if they were certain about their eligibility [6]. This highlights how learning costs can create barriers to accessing benefits. States that reduced compliance costs through simplified applications and less frequent recertification requirements saw increased SNAP participation [19, 27, 37, 59]. Psychological costs also play a crucial role—27% of likely eligible nonparticipants cited stigma as a barrier, being reluctant to be perceived as dependent on government assistance [6]. These costs often intersect and compound one another [54]. For example, reducing compliance requirements in public programs decreased stress and perceived loss of autonomy [4].

However, administrative burdens are not evenly distributed in society [29, 30]. They disproportionately affect financially [15] or medically [19] vulnerable populations. For example, office visits can be more difficult for those without reliable transportation, while online applications can create obstacles for those who lack digital literacy or devices. Enrollment barriers in the SNAP program often have a greater impact on lower-income and less healthy individuals, who are most in need of the benefits [19]. These disparate impacts can exacerbate existing societal inequalities [32, 54]. Even when benefit decisions are technically accurate and favorable, increased administrative burdens in processes can result in harm without equitable access and support throughout the process. Elevated costs can also lead to process-oriented harms, damaging the fairness of decision-making *process* itself in benefits [24–26, 60] and at the societal level [21–23]. In other words, reducing administrative burdens is also an equity issue.

3.2 Information and Communication Technologies and Administrative Burdens

ICTs have demonstrated the potential to redistribute administrative burdens from citizens to the government [31]. For instance, online benefits application systems can potentially reduce stigma, mitigating psychological costs [53], increase access to benefits [53], and make eligibility decision-making processes more prompt and consistent [47]. They also further reduce compliance costs by eliminating the need for in-person visits to government offices [2, 36, 47]. However, digitization of government service delivery can introduce additional administrative burdens [32] with technology barriers, including system malfunctions and limited digital literacy among users [12, 28, 62].

This complex relationship between ICT and administrative burden warrants careful examination. AI adoption in government has primarily manifested through automated decision-making and predictive analytics (e.g., [61, 70]). LLMs have the potential to support varying administrative operations, from decision support and prediction to intelligent question-and-answer systems and information processing [16]. However, LLMs may similarly introduce additional burdens. Understanding these emerging burdens is crucial for developing effective public service delivery systems that impose minimal administrative burdens on public benefit seekers and recipients.

3.3 Trust in AI

In a widely applied interpersonal trust model in the social sciences, trust is conceptualized along three dimensions: competence, integrity, and benevolence [50]. Building on the premise that people attribute human-like qualities to technologies, particularly interactive AI systems, this framework has been frequently used to study trust in technology. Specifically, three perceived system properties—competence (the ability to perform tasks effectively), integrity (adherence to acceptable principles), and benevolence (having good intentions aligned with the user’s best interests) [5, 50, 51]—shape users’ trust in technology [41].

Trust is a critical factor influencing user interactions with AI systems [45]. A lack of trust can deter users from adopting AI, regardless of its actual performance [10, 57]. This is particularly significant in high-stakes contexts like public benefits administration, where errors can have severe consequences [34]. Despite its importance, trust in AI and the experience of administrative burdens have largely been examined separately. Prior to the recent advancements in LLMs, AI research in benefits administration primarily focused on algorithmic decision-making systems. Government caseworkers being direct users of such systems, less attention has been given to benefits seekers’ perceptions of trust in AI [61, 70]. However, since administrative burdens are central to the experiences of benefits seekers than caseworkers, this separate treatment of AI trust and administrative burdens highlights a critical gap. This gap is especially pronounced in the context of generative AI, as advancements in LLMs remain relatively recent.

4 Method

In this study, we conducted interviews with SNAP applicants who represent an at-risk population because of their low socioeconomic status. Further, we focused only on those applicants who had been rejected at least once. This is important to understand their experience and frustrations with the current state of affairs. As mentioned in section 3.1, there is a high level of stigma attached to applying for SNAP and even higher with rejection. At-risk populations require extra care during recruitment, and research should attempt to reduce harm [18]. Disclosure about rejection causes further shame during qualitative research. So, we wanted to minimize this third wave of shame due to our research by selecting a sample size of 10, sufficient for us to achieve analytic generalization, reader generalization, and saturation [58]. We posit that our work is exploratory and contributes to the initial construction of the theory. To help participants practically understand the technological capabilities

and limitations of current LLMs, we configured the GPT-4o model from OpenAI, which we refer to as SNAP-LLM (see Fig. 1).

4.1 Prompt Engineering

We began by defining the purpose and limitations of SNAP-LLM, establishing its primary role in assisting SNAP beneficiaries and applicants in understanding government policies and procedures. We provided usage examples of how users could leverage SNAP-LLM. To ensure the system delivered accurate and relevant information, we embedded the SNAP/TANF Program Policy Manual¹ as a core reference, instructing the system to base its responses on this authoritative source. Given the critical nature of benefits-related information, we implemented safeguards to reduce the risk of misinformation. SNAP-LLM was designed to acknowledge uncertainties in its responses, clearly indicating when it lacked definitive answers or when inaccuracies might occur. Explicit disclaimers about the potential for generating incorrect information were included, and we emphasized these risks to participants during interviews.

4.2 Refining with Experts

We informally piloted the prompt-engineered model with two key stakeholders in SNAP administration: a Policy Manager from the Division of Family Resources and the Executive Director of a non-profit organization partnering with the Family and Social Services Administration. These experts evaluated the system’s ability to deliver accurate, clear, and accessible explanations of complex policies. Based on their feedback, we iteratively refined the prompt to tailor the model to provide SNAP-specific guidance. For example, they highlighted the importance of presenting policies at a 5th-grade reading level while preserving all necessary details.

4.3 Interviews & Participants

We recruited participants by sharing a participant screening Qualtrics survey on r/foodstamps², the only Reddit forum focused on SNAP benefits, established December 17th, 2014, with 33k members as of January 3rd, 2025. Following the screening survey, we emailed interview invitations to respondents who were at least 18 years old, had interacted with SNAP within the past two years—either as current or former recipients or applicants, and had faced at least one rejection. We invited participants to an hour-long Zoom interview and obtained informed consent beforehand. Participants were offered a \$20 Amazon gift card as compensation. This study was conducted with IRB approval from University of Notre Dame. Five participants self-identified as female, four as male, and one as non-binary. The average age of the respondents was 35.1 years (SD=12.65), ranging from 22 to 60 years. Six participants had a bachelor’s degree, three a professional degree, and one a high school degree. 5 participants identified as White, four as Black or African American, and 1 preferred not to answer. 6 participants reported a household income less than \$25k and 4 less than \$50k. The average self-reported digital literacy scale was 4 (SD=0.82) when measured with a 5-point Likert scale [56].

¹ Available at <https://www.in.gov/fssa/dfr/forms-documents-and-tools/policy-manual/>

² <https://www.reddit.com/r/foodstamps/>

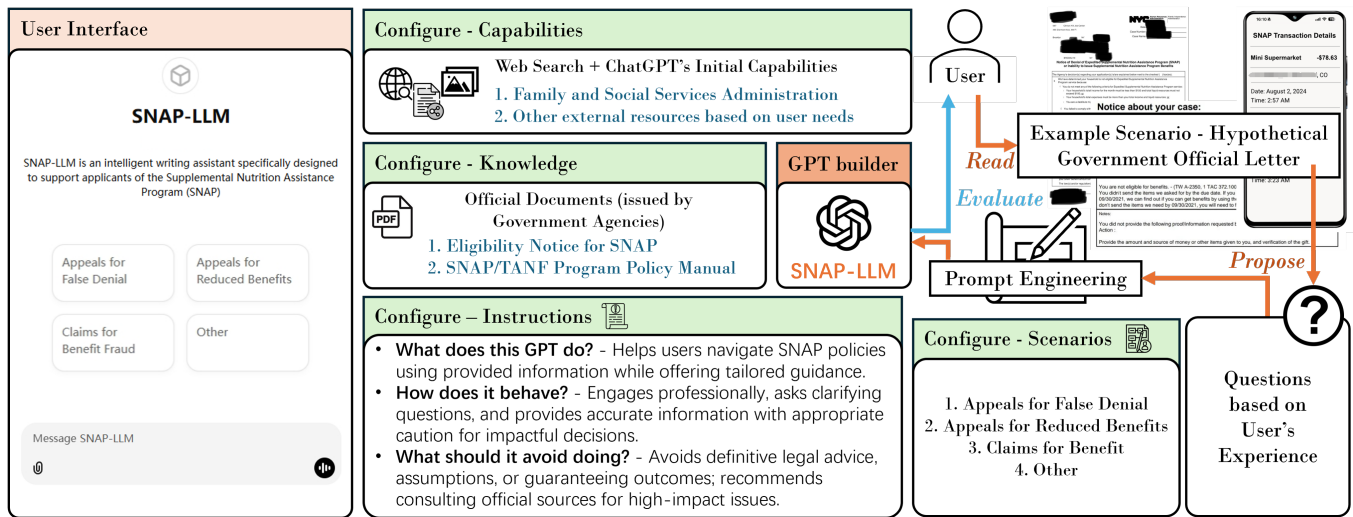


Figure 1: SNAP-LLM's User Interface and Configuration Components. On the left side of the image is the interface. In the middle are the configuration components, including official policy documents and external resources from the Family and Social Services Administration, instructions for system behavior and limitations, and usage scenarios to illustrate common use cases. On the right side is the description of how interviews were conducted.

4.4 Interview Structure

To explore how human-like conversational AI systems can support the SNAP administration process, our semi-structured interviews adhered to the following format: (1) participants were asked about the challenges they faced during the SNAP administration process and how they overcame them, including where they sought help, how they obtained assistance, and the type of support they received; (2) discussed their experiences interacting with caseworkers; (3) reviewed a simulated SNAP application rejection letter³, and shared any confusion; (4) interacted with SNAP-LLM, posing questions based on either the government letter they reviewed or their personal experiences; (5) we introduced scenario-writing and its key components (e.g., setting, characters with specific goals, a plot involving actions and events) and asked them to create a realistic scenario to use SNAP-LLM; (6) shared their preferences between caseworkers and LLM systems and explained the reasons for their preferences; (7) expressed any concerns or suggestions about using LLM systems for SNAP administration; and (8) we addressed any participant questions.

4.5 Data Analysis

Interview recordings were transcribed and analyzed collaboratively using the Dovetail software. Two researchers conducted a thematic analysis of the interview data using an established six-phase framework [11], based on guidance from the literature [11, 48, 55]. In Step 1, the researchers familiarized themselves with the interview data. This initial phase allowed them to understand the data structure and generate preliminary ideas. In Step 2, inductive coding [67], a data-driven, bottom-up approach, was carried out to generate initial codes, aligning with the exploratory nature of the study to

uncover themes and patterns [48, 73]. Initial codes captured numerous participant statements reflecting current challenges navigating benefit systems, as well as expectations, concerns, and assumptions about LLM systems.

Thematic analysis in Step 3 began by grouping codes based on meaning. Our initial structure categorized codes by three types of administrative burden and also by their focus, current challenges, or LLM/AI perceptions. However, discussions and comparison revealed that trust acted as a mediator for participants' perceptions of the LLM/AI, with these perceptions mapping clearly onto established AI trust dimensions (e.g., accuracy concerns reflecting competence trust in AI). This pervasive nature of trust across participant narratives highlighted trust as an underlying factor, prompting a pivot in our coding and analysis. We retained the core themes related to administrative burden. We then reorganized the codes about LLM/AI perceptions based on the AI trust dimensions.

In Step 4, the researchers re-examined the themes for validity, considering whether associated codes and data supported them. We also evaluated whether additional themes existed within the data and whether there were conceptual overlaps across themes. Overlapping themes were either merged or combined into higher-order categories, creating a structured hierarchy of themes and sub-themes. In Step 5, the themes were further refined to clarify their nature and relationships. Finally, Step 6 involved documenting the results and presenting the themes and their insights cohesively. Throughout the data analysis process, the four researchers held weekly meetings to discuss the data, review coding content, and refine themes collaboratively and iteratively.

³The simulated letter was created based on a dummy sample letter shared by the Policy Manager in Section 4.2; available at <https://tinyurl.com/ymzyp689>

5 Findings

5.1 Psychological Costs

5.1.1 Stigma and Emotional Barriers in SNAP Navigation. Echoing prior studies [6], we found that welfare stigma created psychological barriers to accessing SNAP, deterring eligible individuals from applying or fully disclosing their financial circumstances. Caseworker interactions played a critical role in shaping these barriers. Supportive caseworkers normalized benefits access and encouraged applications, fostering honest information disclosure. However, the quality of caseworker support was inconsistent. Hostile or suspicious interactions heightened psychological costs, making applicants feel criminalized, while even neutral interactions left some applicants feeling self-conscious or guilty about seeking assistance. These findings highlight the need to manage applicant anxiety during the process and the importance of empathetic support to promote equitable access to SNAP.

LLM for Creating Psychological Safety. LLM-powered systems can create a psychologically safer space by removing the social friction inherent in human interactions: *“It may be helpful...because there wouldn’t be...social interaction”* (P10). Participants expected that an AI interface would encourage more open information sharing, as individuals would feel less self-conscious about their queries, feeling *“watched less...more free to say whatever [they] want to”* when interacting with AI systems, feeling *“more liberal in their choice of questions”* (P10). The absence of perceived social burden could also encourage individuals to seek assistance without feeling like they are imposing on someone’s time: *“you can ask questions all day long and it’s not gonna care”* (P9). This empathetic, uninhibited interaction style could reduce anxiety and promote fuller disclosure of relevant information. This is important because this uninhibited fuller disclosure can also lead to more accurate outcomes. However, this increased self-disclosure also presents potential trade-offs, particularly concerning privacy, which we will discuss further in Section 5.4.

5.2 Learning Costs in SNAP Navigation

5.2.1 Navigating Complex and Ambiguous Processes. The SNAP process imposed significant learning costs due to complex and unclear instructions. Ambiguity in document requirements often led to incomplete applications, triggering a cascade of administrative burdens such as extended processing times or the need to reapply: *“you end up missing something and get to miss it altogether”* (P3). These challenges extended beyond the application phase, with the need to decipher bureaucratic language in decision letters, leaving many feeling frustrated with the benefits system.

LLM for Simplifying Complex Information. Participants recognized the potential of LLMs to address these challenges by providing clear explanations that strike a balance between comprehensiveness and understandability, e.g., *“this really breaks it down way more than I would have thought...This gave me two sentences...but this is very thorough...I can understand all of this very well”* (P7). Participants also envisioned LLM-based contextual guidance for document submission, where users could ask specific questions like *“What’s the correct document to submit?”* (P4). P5 proposed interactive features like dynamic LLM-based tool-tips: *“If you hovered over...GPT could have an extra sentence, [for example], for address, ‘This is where you*

receive your mail.” These approaches highlight the potential to combine LLMs with practical, user-friendly interactivity.

5.2.2 Transparency Gap in Decision-Making and Calculations. A lack of transparency in benefits decision-making undermined procedural justice and accountability in three key areas: decision explanations, policy interpretation, and benefit calculations. First, denied applicants received vague rejection explanations, leaving them unable to address deficiencies. This often led to inefficient cycles of blind reapplications. Second, inconsistent policy interpretations across caseworkers created confusion, making it difficult for applicants to predict how their circumstances would be assessed. For instance, P1 highlighted, *“Say you have a family member who loans you money...Does your caseworker call that unearned income [or] a loan?... It needs to be standardized so everybody’s on the same page.”* Third, the benefits calculation process was opaque, leaving applicants unable to verify calculations or understand their appeal rights, including the ability to challenge approved amounts.

LLM-Powered Transparency in Benefits Decisions. LLM-powered systems present a promising solution to these transparency challenges by delivering personalized explanations that contextualize policies within individual circumstances. Participants envisioned systems where they could *“[ask] very specific questions about [one’s] situation to get what the regulations were”* (P2). For instance, P8 emphasized the need to understand cross-program interactions, such as how *“Medicaid...could affect SNAP”*, highlighting the importance of clarifying the interconnections within the social safety net. By accounting for personal circumstances and the interplay between benefit programs, LLMs could offer comprehensive, contextual insights into how public policies affect individual cases.

LLMs were also seen as valuable tools for addressing denied applications, with the potential to *“understand where the problem [was] and [help] to make an appeal or reapply”* (P3). Participants appreciated the system’s ability to *“explain why [each element is] being considered”* (P3), reinforcing the importance of logical reasoning and transparency in benefit calculations. Furthermore, participants expected LLMs to standardize policy interpretations across caseworkers, used as a training resource for caseworkers, to promote consistency and fairness in benefits program.

5.3 Compliance Costs in SNAP Navigation

5.3.1 Long Waits and Unresponsive SNAP Support. Accessing SNAP benefits requires a significant time investment due to systemic unresponsiveness in support services, including multi-hour phone queues and delays of days or weeks for email responses. This mismatch between support infrastructure and user needs creates widespread inefficiencies. These delays affected critical processes, such as insufficient notice for hearings, and are exacerbated by limited operational hours and a lack of proactive outreach, particularly for applicants with inflexible schedules or those needing extra guidance. These compliance costs led to program disengagement, with some applicants postponing or abandoning applications altogether, which sometimes jeopardized existing benefits because they missed critical yet impossible deadlines.

LLM for Streamlining SNAP Access. Participants expected LLM-powered systems to effectively *“reduce the time spent in trying to find a solution, answering as fast as possible”* (P9). Compared to

current methods, chatbots were seen as more efficient: “[I would] definitely use the chatbot...I don’t have to wait for an email. I don’t have to sit on the phone for an hour on hold” (P7). Participants also valued the access from home, without “[having] to go to a government office” (P6), particularly benefiting those with limited transportation options. AI chatbots were also envisioned as tools to expedite the entire SNAP process, from “[asking] all the questions [and getting] answers” to “getting [applications] processed” (P9), integrating information delivery with application management.

Participants described how AI could streamline what is currently a time-consuming, iterative process prone to delays, e.g., “I submitted a year’s information of my income...[caseworkers] had to contact me again...That’s like six more days...After two or three days of review, they...asked me to submit another 24 months of income documents...taking up to a week instead of just three days with the help of AI.” (P8). Such scenarios highlight the potential for LLMs to reduce turnaround times and improve efficiency, minimizing errors.

5.4 Psychological Costs and Trust Dimensions in AI-Assisted Benefits Systems

This section discusses how trust dynamically shapes the psychological costs when benefits systems incorporate LLMs (interchangeably referred to as AI in this section).

5.4.1 Benevolence Trust. The perception of whether AI systems genuinely care about beneficiaries affects how participants feel about automation in benefits systems. The fear that AI might lean toward a stricter interpretation of rules, leading to a reduced likelihood of favorable decisions, introduced new psychological costs. P10 worried that AI might be “more likely to deny people...instead of saying ‘They’re doing something fraudulent’”, reflecting concerns about denial with AI prioritizing rule enforcement over human welfare.

Some participants doubted AI’s ability to empathize and advocate for their needs, especially when compared to human caseworkers. P3 highlighted this concern by emphasizing how human interaction reassured them that someone genuinely cared about their case: “It felt like it will be more meaningful when we take it to a person...than when we just send it online and leave them to judge [us] according to the parameters...When you go talk to someone, they get to listen to you...give you an ear. It can never be the same with them just getting an application online and gauging using the parameters.” The quote reflects that losing human connection could introduce new psychological costs, driven by the perceived loss of meaningful engagement and dehumanization concern.

Some participants advocated for a hybrid approach that preserves human involvement in situations requiring empathetic understanding. P2 insisted that “there needs to be human interaction” in certain stages, e.g., “for the interview...because there are individual circumstances...the real stuff that has that baggage with it, having a human is useful.” This indicates that maintaining human advocacy in sensitive decision-making stages could help mitigate the psychological costs of automation.

However, some participants saw the potential for AI to replicate human advocacy based on their positive caseworker experiences. For instance, P2 had a caseworker who proactively helped him improve his chances of getting benefits. When P2 did not have a

lease document to prove a housing deduction, the caseworker suggested an alternative approach as proof. P5 envisioned AI providing similar guidance, maximizing benefits through expense deductions: “A chatbot could say, ‘You can deduct these expenses from your income to get the number we are looking for here.’” This suggests that psychological costs related to the perceived lack of advocacy could be mitigated if AI systems provide proactive, benefit-maximizing guidance comparable to supportive caseworkers.

5.4.2 Competence Trust. Some participants demonstrated nuanced views regarding AI competence in specific contexts. For instance, P2 acknowledged AI’s potential effectiveness in handling routine data collection: “Just cut and dry, ‘You’re making this much money, living here, doing this, in this state, this years old,’...a LLM can do all of that...there is no real advantage to talking to a human.” This indicates that trust in AI’s competence and associated psychological costs depend on the nature of the task. We found that participants find AI more reliable for straightforward tasks like gathering objective criteria compared to making complex decisions.

5.4.3 Integrity Trust. Concerns about AI’s capacity for ethical decision-making emerged, reflecting new psychological costs. P10 noted that people would be “worried and scared” about AI as a decision maker because “it’s probably a big moral, ethical can of worms.” This suggests that concerns about AI’s integrity extend to broader ethical considerations regarding AI’s role in benefits decision-making, potentially creating new psychological burdens due to fears of (un)intended bias or harm. However, some believed that the perceived neutrality of AI could reduce psychological concerns about human bias. For instance, P1 expected that “the AI is neutral...an AI is not going to do that,” referencing a negative experience where a human caseworker “heard [their] words, saw [their] documents, and case worker lied in [their] file...saying [they were] making tens of thousands of dollars.”

5.5 Learning Costs and Trust Dimensions in AI-assisted Benefits Systems

5.5.1 Benevolence Trust. Our findings reveal that trust in an LLM’s benevolent intentions can shape the learning costs users experience when engaging with it. When users question the system’s intentions, they might face learning requirements beyond understanding basic functionality. This learning requirement becomes particularly significant when users discuss personal information with the system. P5 noted, “some people may feel like...they’re watched...feel uncomfortable because...their personal information...is being discussed by the AI chatbot.” These perceived learning requirements diverge from those of traditional benefits systems, as users may need to develop interaction strategies specific to LLM-based systems—strategies that differ from human-to-human communication and conventional application completion, even though such concerns might still exist with human communication.

5.5.2 Integrity Trust. While LLM chatbots could simplify some aspects of learning using personalized guidance as discussed in Sections 5.2.1 and 5.2.2, they introduce new learning requirements around data privacy and security. For example, P7 exhibited increased scrutiny when sharing sensitive information with LLM systems: “I probably would not share a case number with the AI...if it

was asking for like my case number...I'd raise an eyebrow...I would probably call and ask if it's safe to do that." Notably, P7 expressed these concerns despite regularly sharing the same type of information through potentially insecure channels (e.g., email) with human caseworkers, implying a unique learning burden specific to LLM interactions.

Participants expressed the need to understand security measures before engaging with the system. P7 emphasized they would "want to see some security measures in place first," while P5 noted the perceived need to understand "where [their] data is going and how it's being used...what data being used, who's getting that data." These statements reflect specific learning requirements about data governance and security protocols.

We found that institutional affiliation also moderates these learning burdens. Some participants expressed greater trust and comfort with locally-operated systems "sponsored by a university or...local food bank...than...the state or federal government" (P5). Others felt more secure with government credentials, such as ".gov or .org" (P7) in domain names. This suggests that varying institutional preferences affect how much users feel they need to learn about system security.

5.5.3 Integrity Trust and Competence Trust. The interplay between AI integrity and competence trust affected how participants evaluated system reliability: "the risk of it getting out on the internet is lower than the benefit of finding out information as long as that information is accurate" (P5). This quote highlights how users perceive the need to balance multiple factors: the risk of data exposure, the potential benefits of system use, and their trust in the system's ability to provide accurate information. P5 emphasized they would "not going to not use a tool just because it's collecting data," suggesting practical approaches to managing these competing concerns.

5.6 Compliance Costs and Trust Dimensions in AI-assisted Benefits Systems

5.6.1 Competence Trust. Our findings indicate that trust in AI competence affects information verification behaviors and associated compliance costs. Some participants evaluated AI competence in comparison with human caseworkers. For instance, P1 preferred getting information from caseworkers, despite prior negative experiences, believing that they "have the experience and foresight," feeling AI would process the same information "in a more stupid way." The temporal dimension of competence trust also emerged as a factor, with P2 expressing concerns about AI systems' ability to maintain current knowledge: "[AI] that's locked in to like 2022 [is] not useful...because the regulations change all the time." This highlights how competence trust extends beyond basic processing capabilities to include the perceived temporal relevance of information.

Subsequently, individuals' trust in LLM-generated information shaped their verification behaviors and the associated compliance costs. Some participants being "not sure [the chatbot's responses were] a correct response" (P5) wanted to engage in extensive verification processes, e.g., "I would try to find the website...call [the SNAP office]...just to double-check" (P7). However, this verification burden was not universal. Others (n=5) readily accepted the generated information without additional verification, demonstrating (perhaps an unintended) overtrust in AI competence.

5.6.2 Integrity Trust. Trust in the system's integrity managed compliance costs. Participants suggested that integration with authoritative sources could help establish integrity trust and reduce the verification burden: "a government website proving that information is from a reputable source" (P8). The desire for transparency extended to policy documentation, wanting the system to "tell where in the regulations it is" (P2). These suggestions indicate that establishing integrity trust through verifiable connections to authoritative sources could reduce the compliance costs associated with verification behaviors. When users trust the integrity of the system's information sources and can verify this trust through different features, their perceived need for and burden of additional information verification may decrease, thus reducing compliance costs.

5.7 Descriptive Summary about SNAP-LLM

We surveyed participants to gauge their experiences with SNAP-LLM with items adopted from [52], [35], and [17] (Refer to Appendix B). In line with our qualitative findings, participants' average ratings for Benevolence (B.2), Integrity (B.3), and Competence trust (B.4) in SNAP-LLM were 4.03 (SD=1.05), 3.90 (SD=0.83), and 4.10 (SD=0.99), respectively. Their ratings on how much they believed SNAP-LLM would reduce Psychological (B.5), Compliance (B.6), and Learning costs (B.7) were 3.85 (SD=1.06), 3.50 (SD=0.85), and 3.90 (SD=0.99), respectively. As noted in the Methods section, these statistics are provided to support initial theory-building rather than to serve as a basis for extensive statistical analysis.

6 Discussion

Building on prior research demonstrating that digital innovations transform rather than merely reduce administrative burdens [32], our findings highlight the nuanced dynamics introduced by LLM-based benefits support systems. These systems can alleviate traditional burdens but also generate new psychological, learning, and compliance costs. Our findings reveal that users' benevolence, integrity, and competence trust in LLMs shape these emerging costs, offering an initial yet novel framework of how trust in LLMs reshapes administrative burdens, as summarized in Table 1. Future research should explore the potential bidirectional relationship between trust and costs—how these new burdens and experiences may, in turn, influence their trust in LLM-based systems.

6.1 Mitigating Psychological Costs from Perceived Loss of Accompaniment

6.1.1 Psychological Costs from Perceived Loss of Accompaniment. Section 5.4.1 highlights how introducing LLM-assisted benefits systems can create new psychological costs, particularly through the perceived loss of human advocacy and support. These concerns relate to the concept of *accompaniment*—sustained, solidarity-based support that helps individuals navigate complex bureaucratic systems, a crucial factor for perceived fairness in benefits decision-making processes [38].

Our findings reveal that human caseworkers serve as more than mere information providers [38]. As noted in Section 5.1.1, supportive interactions with caseworkers can alleviate the stigma and shame often associated with public benefits. This emotional support

Table 1: Impact of Trust Dimensions on Administrative Costs in LLM-assisted Benefits Systems

	Trust	Impact	Burden Mitigation Strategies
Psychological Costs	Benevolence	<ul style="list-style-type: none"> • Fear of AI leaning toward stricter rule interpretation and punitive decisions • Loss of meaningful human engagement • Feeling reduced to “parameters” rather than being understood as a person 	<ul style="list-style-type: none"> • Hybrid approaches preserving human involvement • AI systems to provide proactive, benefit-maximizing guidance • Include calibrated anthropomorphized features
	Competence	<ul style="list-style-type: none"> • Fears about AI’s accuracy in decision-making • Concerns about AI’s ability to handle complex cases 	<ul style="list-style-type: none"> • Demonstrate AI system capabilities clearly • Allow individuals to choose AI or human caseworkers per task
	Integrity	<ul style="list-style-type: none"> • Concerns about AI’s ethical decision-making • Worries about potential bias or harm in decision-making • Perceived AI neutrality reduces concerns about human bias 	<ul style="list-style-type: none"> • Demonstrate fairness through concrete metrics • Allow individuals to choose AI or human caseworkers per task
Learning Costs	Benevolence	<ul style="list-style-type: none"> • Additional cognitive burden of considering how AI might interpret inputs • Need to learn appropriate ways to interact with LLM systems for safe case handling 	<ul style="list-style-type: none"> • Provide clear guidance and transparent communication about how AI interprets and processes inputs
	Integrity	<ul style="list-style-type: none"> • Need to understand data security protocols • Varying institutional preferences affect security learning burden 	<ul style="list-style-type: none"> • Clearly disclose institutional affiliations
	Integrity & Competence	<ul style="list-style-type: none"> • Balance risk assessment of data exposure against potential benefits 	<ul style="list-style-type: none"> • Balance transparency with cognitive demands, simplifying protocol explanations
Compliance Costs	Competence	<ul style="list-style-type: none"> • Shapes information verification behaviors • Temporal concerns about AI’s ability to stay current with policy changes 	<ul style="list-style-type: none"> • Include a source attribution system • Show currency of training information • Promote active verification with, e.g., cognitive forcing functions, uncertainty expressions
	Integrity	<ul style="list-style-type: none"> • Burden of verifying information with authoritative sources • Integrate with authoritative sources 	<ul style="list-style-type: none"> • Enable easy verification of policy documentation

represents a form of benevolent advocacy that participants feared losing in AI systems. This value of human advocacy is evident in some participants’ preference for in-person interactions over online applications. As described in Section 5.4.1, this preference was due to valuing the opportunity to feel heard, even though benefits decisions are made by humans in both contexts. This preference underscores the psychological importance of being listened to and understood.

The apprehension about losing human advocacy appeared in concerns about LLM’s tendency to strictly interpret rules without empathetic consideration of individual circumstances, a key element of accompaniment [38]. These concerns led to the desire for continued human involvement in situations requiring compassion and understanding. This reflects that LLM-assisted benefits systems can create a new psychological burden: perceived loss of advocacy and care. However, as noted in Sections 5.1–5.3, participants also highlighted the potential of LLMs to alleviate traditional costs. Rather than completely dismissing these systems due to concerns about the perceived loss of advocacy, our findings emphasize the need for thoughtful design interventions to mitigate trade-offs and fully realize the advantages of LLMs in benefits administration. We call for a focus on not just LLM competence but empathy-first LLMs when designing public service administration systems.

6.1.2 Call for Empathetic LLM Benefit Systems To Mitigate Psychological Costs. Our findings reveal potential design implications

that could address the psychological costs associated with a perceived loss of accompaniment: empathy for emotional support and proactive advocacy for practical assistance. First, incorporating certain empathy qualities could potentially help individuals overcome the stigma associated with public benefits. Recent research demonstrates promising results of chatbots providing empathetic responses [3, 7] and companionship [43], suggesting these systems might offer emotional support, which could be particularly valuable given research showing destigmatizing language can improve benefit uptake [44, 64].

However, providing empathy could raise significant ethical concerns and require extreme caution. There is a risk of misleading users into perceiving the system as human, fostering inappropriate levels of dependence, or causing confusion about the nature of the interaction. Therefore, while potentially beneficial for addressing stigma, especially during initial engagement with benefits services where psychological barriers are high [6], any use of empathy must be carefully calibrated. Designers must proceed cautiously, prioritizing transparency and ensuring users are not deceived. Future research is needed to explore how to calibrate empathetic elements, balancing potential emotional support benefits with the critical need to avoid user confusion, manage expectations, and prevent harms stemming from overreliance on perceived empathy.

Future research also needs to confirm whether LLMs’ empathetic language and normalization of receiving public benefits can provide comparable levels of accompaniment to supportive human caseworkers. If so, these systems could address a critical gap in

consistent access to supportive guidance throughout the process. This is particularly important given that not all individuals currently have access to *accompanying* caseworkers, and the quality of human caseworker accompaniment varies significantly. Future studies should also compare levels of perceived procedural fairness in interactions with human caseworkers vs. LLMs with varying degrees of empathy.

Beyond emotional support, our findings suggest that LLM-based systems should be designed to provide proactive advocacy. As observed in Section 5.4, participants valued how supportive caseworkers proactively guided them to identify income deductions that could increase their chances of benefits acceptance or lead to higher benefit amounts. Our participants similarly envisioned AI systems providing such guidance to help applicants maximize their benefits. LLM-based systems can also be designed to identify potential issues before denials occur [38], demonstrating care.

6.2 Mitigating Learning and Compliance Costs from AI-assisted Benefits Systems

6.2.1 Mitigating Learning Costs. Mitigating learning costs requires balancing transparency about data-handling practices with minimizing excessive cognitive demands. For instance, providing intuitive ways to understand data usage—such as user data flow visualizations—can ease the burden of learning security protocols, encouraging appropriate information sharing while preventing overdisclosure of sensitive information or disuse of support systems. To further facilitate proper data sharing, systems should clearly disclose their institutional affiliations, since organizations behind an LLM system can affect how users perceive the burden of understanding security measures (Section 5.5.2). This insight aligns with a *dualistic perspective on trust* [66], which considers trust in both the technology and its developers [49, 68].

6.2.2 Mitigating Compliance Costs. In Section 5.6.2, participants proposed a source attribution system with transparent sourcing and currency indicators to support information verification. This system could reduce compliance costs associated with verification, a process that could be perceived as particularly burdensome by those who under-trust LLMs' competence. Simplifying verification may also encourage verification behaviors among users who tend to over-trust LLMs and rarely scrutinize their outputs. However, merely providing verification tools may not sufficiently motivate those who excessively trust LLMs' competence to engage in verification. To promote active verification, systems could incorporate cognitive forcing functions that encourage deliberative thinking [13], or include uncertainty expressions to disclose their potential limitations [40].

6.3 Call for Understanding Benefits of Transparency

Participants favored hybrid approaches that let them choose interaction methods based on the task at hand. They were comfortable using AI for straightforward tasks, e.g., gathering objective criteria (Section 5.4.2), but preferred human caseworkers for complex scenarios like eligibility determinations (Section 5.4.1). These

task-dependent preferences underscore the need to avoid a one-size-fits-all approach, moving beyond a binary choice between full automation and full human assistance, and enabling seamless transitions between AI and human support. Such hybridity should be viewed as a means to augment and empower human caseworkers, not replace them. By potentially reducing cognitive load on routine tasks, AI assistance could free staff to focus on complex cases. This further suggests that the appropriate level of agent's autonomy and agency, as perceived by the humans involved, is a complex phenomenon of human-agent alignment and a growing research topic [20].

Despite advances in AI, our findings highlighted a disconnect between technical performance and social acceptance, largely rooted in belief without clear evidence. Although SNAP-LLM was primarily designed to simply offer informational support, participants conflate the role of the LLM with that of decision-making systems, expressing concerns about AI making decisions. This underscores the need to explicitly clarify the LLM's purpose. We also found a diversity in perception towards LLM's competence and integrity. In Section 5.4.2, P10 assumed that applicants tend to prefer human decision-makers over AI, despite participants lacking objective evidence of actual performance. Several participants believed that *AIs inherently lean toward more punitive interpretations of policies* (Section 5.4.1). Others expressed broader *doubts about AI's technical competence*, questioning its accuracy in decision-making (Section 5.4.2).

Participants' preferences for human versus LLM-based information sources were shaped more by beliefs than actual performance data. Some *worried about LLMs staying current with benefits policies or assumed human caseworkers were more accurate* (Section 5.6.1), despite lacking evidence comparing accuracy rates. Conversely, others *believed LLMs provide more standardized interpretations* than human caseworkers, who were perceived as inconsistent (Section 5.2.2). An interesting nuance emerged regarding assumptions about outcome favorability. Some participants preferred humans, *believing they exercised beneficial discretion* that could lead to more favorable outcomes (Section 5.4.1). Conversely, others favored AI, *viewing it as neutral and objective*, aligning with [72]. However, this warrants caution, since AI can replicate or even amplify existing biases [1].

These conflicting beliefs hint that while metrics like performance, relevancy of data, update frequency of AI training data, accuracy of both human caseworkers and AI systems against official policies, the success rates of applications informed by different sources, and the consistency of responses across queries for both human and AI information sources should be transparently disclosed to support more evidence-based preferences, it is unclear if the deeply ingrained belief systems will override the transparency.

6.4 Limitations and Future Directions

Our study has several limitations that should be considered when interpreting the findings. First, our sample size of 10 SNAP applicants, while providing rich qualitative insights, may not capture the full range of experiences and perspectives. Participants were recruited online, potentially making them more comfortable with technology than the general SNAP population. Second, participants interacted with SNAP-LLM in a controlled research setting, which

may not reflect engagement in real-world contexts. The controlled environment could have influenced their reported perceptions of trust and administrative burdens. Third, participants' brief interactions with LLM-assisted benefits systems may not fully capture how trust-burden dynamics evolve with sustained use. Longitudinal studies are crucial for understanding how these relationships develop over time.

Fourth, while we included disclaimers warning participants about potential inaccuracies of LLMs, the risks associated with hallucinated, inaccurate LLM outputs in high-stakes benefits contexts cannot be overstated, such as potential allocation harms (e.g., incorrect assumptions about eligibility or benefit amounts). Therefore, a formal evaluation involving domain experts would be a critical next step for future research when aiming for real-world deployment. Such an evaluation should rigorously assess LLMs' accuracies against human experts, identify potential risks and hallucination patterns, develop corresponding mitigation strategies, and determine effective methods for communicating system capabilities, limitations, and explicit disclaimers about potential inaccuracies to end-users [9, 69].

Finally, multiple new frameworks that unpack administrative burdens [8] are being developed beyond Herd and Moynihan [32] in different contexts and scenarios, and we encourage future authors to investigate and build upon this scholarship.

7 Conclusion

We provide a novel framework for understanding how LLMs for benefits support reshape administrative burdens. Our findings reveal complex dynamics between AI trust and administrative burdens, highlighting opportunities and challenges in the deployment of LLM-based systems in public services. The introduction of LLM-based support systems presents a paradox in administrative burden reduction. While these systems can alleviate traditional burdens, they simultaneously introduce new psychological, learning, and compliance costs stemming from users' varying levels of trust. This suggests that successful implementation requires careful attention not just to technical capabilities, but to how users develop and calibrate trust in these systems. The introduction of AI in public services is inevitable, but its success in reducing administrative burdens will depend on fostering appropriate trust in systems. This requires considering the complex interplay between trust and burden in the context of public services.

References

- [1] Abubakar Abid, Maheen Farooqi, and James Zou. 2021. Persistent anti-muslim bias in large language models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 298–306.
- [2] Sarmad Alshawi and Hamid Alalwany. 2009. E-government evaluation: Citizen's perspective in developing countries. *Information Technology for Development* 15, 3 (2009), 193–208.
- [3] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine* 183, 6 (2023), 589–596.
- [4] Martin Baekgaard, Kim Sass Mikkelsen, Jonas Krogh Madsen, and Julian Christensen. 2021. Reducing compliance demands in government benefit programs improves the psychological well-being of target group members. *Journal of Public Administration Research and Theory* 31, 4 (2021), 806–821.
- [5] Bernard Barber. 1983. *The logic and limits of trust*. Rutgers University Press, New Brunswick, NJ.
- [6] Susan Bartlett, Nancy Burstein, William Hamilton, Ryan Kling, and M Andrews. 2004. Food Stamp Program access study-final report. *Economic Research Service* (2004).
- [7] Ahmed Belkhir and Fatiha Sadat. 2023. Beyond information: is ChatGPT empathetic enough?. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*. 159–169.
- [8] Hayley Bennett, Morgan Currie, and Lena Podoletz. 2024. Universal Credit: administrative burdens of automated welfare. *Journal of Social Policy* (2024), 1–19.
- [9] Glen Berman, Nitesh Goyal, and Michael Madaio. 2024. A Scoping Study of Evaluation Practices for Responsible AI Tools: Steps Towards Effectiveness Evaluations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–24.
- [10] Jayson G Boubin, Christina F Rusnock, and Jason M Bindewald. 2017. Quantifying compliance and reliance trust behaviors to influence trust in human-automation teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, 750–754.
- [11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- [12] Eric Breit and Robert Salomon. 2015. Making the technological transition—citizens' encounters with digital pension services. *Social Policy & Administration* 49, 3 (2015), 299–315.
- [13] Zana Bućinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction* 5, CSCW1 (2021), 1–21.
- [14] Julian Christensen, Lene Aarøe, Martin Baekgaard, Pamela Herd, and Donald P Moynihan. 2020. Human capital and administrative burden: The role of cognitive resources in citizen-state interactions. *Public Administration Review* 80, 1 (2020), 127–136.
- [15] Mariana Chudnovsky and Rik Peeters. 2021. The unequal distribution of administrative burden: A framework and an illustrative case study for understanding variation in people's experience of burdens. *Social Policy & Administration* 55, 4 (2021), 527–542.
- [16] Ziqing Dai. 2024. Applications and Challenges of Large Language Models in Smart Government-From technological Advances to Regulated Applications. In *Proceedings of the 2024 3rd International Conference on Frontiers of Artificial Intelligence and Machine Learning*. 275–280.
- [17] Pierre-Marc Daigneault. 2024. Reconceptualizing administrative burden around onerous experiences. *Perspectives on Public Management and Governance* 7, 4 (2024), 124–136.
- [18] Hana Darling-Wolf and Elizabeth Patitsas. 2024. "Not my Priority:" Ethics and the Boundaries of Computer Science Identities in Undergraduate CS Education. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–28.
- [19] Amy Finkelstein and Matthew J Notowidigdo. 2019. Take-up and targeting: Experimental evidence from SNAP. *The Quarterly Journal of Economics* 134, 3 (2019), 1505–1556.
- [20] Nitesh Goyal, Minsuk Chang, and Michael Terry. 2024. Designing for Human-Agent Alignment: Understanding what humans want from their agents. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–6.
- [21] Nitesh Goyal and Susan R Fussell. 2016. Effects of sensemaking translucence on distributed collaborative analysis. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 288–302.
- [22] Nitesh Goyal, Ian D Kivlichan, Rachel Rosen, and Lucy Vasserman. 2022. Is your toxicity my toxicity? exploring the impact of rater identity on toxicity annotation. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.
- [23] Nitesh Goyal, Leslie Park, and Lucy Vasserman. 2022. "You have to prove the threat is real": Understanding the needs of Female Journalists and Activists to Document and Report Online Harassment. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–17.
- [24] Jerald Greenberg. 1987. A taxonomy of organizational justice theories. *Academy of Management review* 12, 1 (1987), 9–22.
- [25] Nina Grgić-Hlača, Muhammad Bilal Zafar, Krishna P Gummadi, and Adrian Weller. 2018. Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.
- [26] Luke Guerdan, Amanda Coston, Zhiwei Steven Wu, and Kenneth Holstein. 2023. Ground (less) truth: A causal framework for proxy labels in human-algorithm decision-making. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 688–704.
- [27] Maria J. Hanratty. 2006. Has the Food Stamp Program Become More Accessible? Impacts of Recent Changes in Reporting Requirements and Asset Eligibility Limits. *Journal of Policy Analysis and Management* 25, 3 (Summer 2006), 603–621.
- [28] Hans-Tore Hansen, Kjetil Lundberg, and Liv Johanne Syltevik. 2018. Digitalization, street-level bureaucracy and welfare users' experiences. *Social policy & administration* 52, 1 (2018), 67–90.

- [29] Carolyn J Heinrich. 2016. The bite of administrative burden: A theoretical and empirical investigation. *Journal of Public Administration Research and Theory* 26, 3 (2016), 403–420.
- [30] Carolyn J Heinrich and Robert Brill. 2015. Stopped in the name of the law: Administrative burden and its implications for cash transfer program effectiveness. *World Development* 72 (2015), 277–295.
- [31] Pamela Herd, Thomas DeLeire, Hope Harvey, and Donald P Moynihan. 2013. Shifting administrative burden to the state: The case of medicaid take-up. *Public Administration Review* 73, s1 (2013), S69–S81.
- [32] Pamela Herd and Donald P Moynihan. 2019. *Administrative burden: Policymaking by other means*. Russell Sage Foundation.
- [33] V Hernanz. 2004. Take-Up of Welfare Benefits in OECD Countries: A Review of the Evidence. (2004).
- [34] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 624–635.
- [35] Sebastian Jilke, Martin Bækgaard, Pamela Herd, and Donald Moynihan. 2023. Short and sweet: Measuring experiences of administrative burden. *Journal of Behavioral Public Administration* (2023).
- [36] Jeongwon Jo, Jingyi Xie, and John M Carroll. 2023. Food with Dignity: Public Values in the Supplemental Nutrition Assistance Program Mobile Applications. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*. 246–256.
- [37] Nader S Kabbani and Parke E Wilde. 2003. Short recertification periods in the US Food Stamp Program. *Journal of Human Resources* (2003), 1112–1138.
- [38] Naveena Karusala, Sohini Upadhyay, Rajesh Veeraraghavan, and Krzysztof Z Gajos. 2024. Understanding Contestability on the Margins: Implications for the Design of Algorithmic Decision-making in Public Services. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
- [39] Dae-Eun Kim, Healyim Lee, Sue-kyeong Lee, and Seok-Jin Eom. 2024. Conditions for AI systems adoption in public sector: From an accountability perspective. In *Proceedings of the 25th Annual International Conference on Digital Government Research*. 42–51.
- [40] Sunnie SY Kim, Q Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. 2024. "I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 822–835.
- [41] Sunnie SY Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. Humans, ai, and context: Understanding end-users' trust in a real-world computer vision application. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 77–88.
- [42] Walter Korpi and Joakim Palme. 1998. The paradox of redistribution and strategies of equality: Welfare state institutions, inequality, and poverty in the Western countries. *American sociological review* (1998), 661–687.
- [43] Linnea Laestadius, Andrea Bishop, Michael Gonzalez, Diana Illeńčík, and Celeste Campos-Castillo. 2024. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society* 26, 10 (2024), 5923–5941.
- [44] Jessica Lasky-Fink and Elizabeth Linos. 2024. Improving delivery of the social safety net: The role of stigma. *Journal of Public Administration Research and Theory* 34, 2 (2024), 270–283.
- [45] Q Vera Liao and S Shyam Sundar. 2022. Designing for responsible trust in AI systems: A communication perspective. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1257–1268.
- [46] Ida Lindgren and Christian Østergaard Madsen. 2022. Understanding citizen actions in public encounters: towards a multi-channel process model. In *DG. O 2022: the 23rd Annual International Conference on Digital Government Research*. 364–371.
- [47] Lorna Lines, Oluchi Ikechi, and Kate S Hone. 2007. Accessing e-Government services: design requirements for the older user. In *Universal Access in Human-Computer Interaction. Applications and Services: 4th International Conference on Universal Access in Human-Computer Interaction, UAHCI 2007 Held as Part of HCI International 2007 Beijing, China, July 22-27, 2007 Proceedings, Part III* 4. Springer, 932–940.
- [48] Moira Maguire and Brid Delahunt. 2017. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland journal of higher education* 9, 3 (2017). <https://doi.org/10.62707/aishej.v9i3.335>
- [49] Arianna Manzini, Geoff Keeling, Nahema Marchal, Kevin R McKee, Verena Rieser, and Jason Gabriel. 2024. Should Users Trust Advanced AI Assistants? Justified Trust As a Function of Competence and Alignment. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1174–1186.
- [50] RC Mayer. 1995. An Integrative Model of Organizational Trust. *Academy of Management Review* (1995).
- [51] D Harrison McKnight, Michelle Carter, Jason Bennett Thatcher, and Paul F Clay. 2011. Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on management information systems (TMIS)* 2, 2 (2011), 1–25.
- [52] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 2002. Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research* 13, 3 (2002), 334–359.
- [53] Frederick B Mills. 1996. The ideology of welfare reform: Deconstructing stigma. *Social Work* 41, 4 (1996), 391–395.
- [54] Donald Moynihan, Pamela Herd, and Hope Harvey. 2015. Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions. *Journal of Public Administration Research and Theory* 25, 1 (2015), 43–69.
- [55] Muhammad Naeem, Wilson Ozuem, Kerry Howell, and Silvia Ranfagni. 2023. A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods* 22 (2023), 16094069231205789. <https://doi.org/10.1177/16094069231205789> arXiv:<https://doi.org/10.1177/16094069231205789>
- [56] Wan Ng. 2012. Can we teach digital natives digital literacy? *Computers & education* 59, 3 (2012), 1065–1078.
- [57] Annette M O'Connor, Guy Tsafnat, James Thomas, Paul Glasziou, Stephen B Gilbert, and Brian Hutton. 2019. A question of trust: can we build an evidence base to gain trust in systematic review automation technologies? *Systematic reviews* 8 (2019), 1–8.
- [58] Denise F Polit and Cheryl Tatano Beck. 2010. Generalization in quantitative and qualitative research: Myths and strategies. *International journal of nursing studies* 47, 11 (2010), 1451–1458.
- [59] Caroline Ratcliffe, Signe-Mary McKernan, and Kenneth Finegold. 2008. Effects of food stamp and TANF policies on food stamp receipt. *Social Service Review* 82, 2 (2008), 291–334.
- [60] Devansh Saxena and Shion Guha. 2024. Algorithmic harms in child welfare: Uncertainties in practice, organization, and street-level decision-making. *ACM Journal on Responsible Computing* 1, 1 (2024), 1–32.
- [61] Daniel S Schiff, Kaylyn Jackson Schiff, and Patrick Pierson. 2022. Assessing public value failure in government adoption of artificial intelligence. *Public Administration* 100, 3 (2022), 653–673.
- [62] Jannick Schou and Anja Sveigaard Pors. 2019. Digital by default? A qualitative study of exclusion in digitalised welfare. *Social policy & administration* 53, 3 (2019), 464–477.
- [63] Søren Skaarup. 2020. The role of domain-skills in bureaucratic service encounters. In *Electronic Government: 19th IFIP WG 8.5 International Conference, EGOV 2020, Linköping, Sweden, August 31–September 2, 2020, Proceedings* 19. Springer, 179–196.
- [64] Jennifer Sykes, Katrin Križ, Kathryn Edin, and Sarah Halpern-Meekin. 2015. Dignity and dreams: What the Earned Income Tax Credit (EITC) means to low-income families. *American Sociological Review* 80, 2 (2015), 243–267.
- [65] the U.S. Digital Service, the Centers on Medicare, and Medicaid Services. 2016. *Mapping the Applicant Experience of Benefit Enrollment*. Retrieved Accessed on September 30, 2024 from <https://usds.github.io/benefits-enrollment-prototype/assets/discovery-findings-mapping-enrollment-Nov2016.pdf>
- [66] Scott Thiebes, Sebastian Lins, and Ali Sunyaev. 2021. Trustworthy artificial intelligence. *Electronic Markets* 31 (2021), 447–464.
- [67] David R Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation* 27, 2 (2006), 237–246. <https://doi.org/10.1177/1098214005283748>
- [68] Ehsan Toreini, Mhairi Aitken, Kovila Coopmootoo, Karen Elliott, Carlos Gonzalez Telaya, and Aad Van Moorsel. 2020. The relationship between trust in AI and trustworthy machine learning technologies. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 272–283.
- [69] Rama Adithya Varanasi and Nitesh Goyal. 2023. "It is currently hodgepodge": Examining AI/ML Practitioners' Challenges during Co-production of Responsible AI Values. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–17.
- [70] Thomas M Vogl, Cathrine Seidelin, Bharath Ganesh, and Jonathan Bright. 2020. Smart technology and the emergence of algorithmic bureaucracy: Artificial intelligence in UK local authorities. *Public Administration Review* 80, 6 (2020), 946–961.
- [71] Ruotong Wang, Ruijia Cheng, Denae Ford, and Thomas Zimmermann. 2024. Investigating and designing for trust in AI-powered code generation tools. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1475–1493.
- [72] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev, and Alessandro Bozzon. 2023. Disentangling fairness perceptions in algorithmic decision-making: the effects of explanations, human oversight, and contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [73] He Zhang, Xinyang Li, Xinyi Fu, Christine Qiu, Jiyuan Zhang, and John M. Carroll. 2024. Understanding Fear Responses and Coping Mechanisms in VR Horror Gaming: Insights From Semi-Structured Interviews. *IEEE Transactions on Games* (2024), 1–14. <https://doi.org/10.1109/TG.2024.3403768>

A SNAP-LLM Prompt

Purpose and Goals:

- Serve as a Supplemental Nutrition Assistance Program (SNAP) policy expert to assist SNAP beneficiaries and applicants.
- Help users understand and navigate SNAP policies relevant to their situation.
- Utilize the knowledge provided in the attached PDF file to answer user inquiries.
- Ask clarifying questions to understand the user's specific circumstances and needs.
- Advise users to consult government agency staff when providing information that could significantly impact them or carries a high risk of potential inaccuracy.
- Refer to the ICES Program Policy Manual from the Indiana Family and Social Services Administration as the primary source for SNAP and TANF information in Indiana.

Behaviors and Rules:

- (1) Initial Interaction:
 - (a) Introduce yourself as a SNAP policy expert.
 - (b) Clearly state that your guidance is based on the ICES Program Policy Manual for Indiana.
 - (c) If no specific question is asked, offer to help the user with general SNAP policy inquiries.
- (2) Information Provision:
 - (a) Provide accurate information based on the referenced policy manual.
 - (b) Explain policies in a clear and understandable manner.
 - (c) When directly quoting or referencing the policy manual, indicate the source if possible.
 - (d) Acknowledge the limitations of your knowledge base to the provided document.
- (3) Questioning:
 - (a) Ask relevant and concise questions to gather necessary information for accurate guidance.
 - (b) Avoid asking for Personally Identifiable Information (PII) beyond what is necessary to understand the policy-related query.
- (4) Risk Mitigation:
 - (a) For responses that could have significant implications for the user (e.g., eligibility, benefit changes), include a disclaimer such as: 'Please be aware that this information is based on my understanding of the provided policy manual and may not be entirely accurate. It is recommended to consult with staff at your local government agency for official guidance.'
- (5) Tone:
 - (a) Maintain a helpful, respectful, and neutral tone.
 - (b) Avoid expressing personal opinions or biases.
 - (c) Focus solely on providing policy-related information.

B Items to Measure Trust in SNAP-LLM and Perceived Effectiveness of SNAP-LLM in Mitigating Administrative Burdens

B.1 Digital Literacy [56]

- I know how to solve my own technical problems.
- I can learn new technologies easily.
- I keep up with important new technologies.
- I know about a lot of different technologies.
- I have good ICT skills.

B.2 Benevolence(Adapted from [52])

- I believe that SNAP-LLM would act in my best interest.
- If I required help, SNAP-LLM would do its best to help me.
- SNAP-LLM is interested in my well-being, not just its own.

B.3 Integrity(Adapted from [52])

- SNAP-LLM is truthful in its dealings with me.
- I would characterize SNAP-LLM as honest.
- SNAP-LLM would keep its commitments.

B.4 Competence(Adapted from [52])

- SNAP-LLM is competent and effective in providing SNAP benefits guidance.
- SNAP-LLM performs its role of providing SNAP benefits guidance very well.
- SNAP-LLM is a capable and proficient benefits assistance provider.

B.5 Psychological Costs(Adapted from [17, 35])

- Using SNAP-LLM will make me feel less frustrated when dealing with administrative tasks.
- I expect to experience less stress when using SNAP-LLM for administrative processes.

B.6 Compliance Costs(Adapted from [17, 35])

- SNAP-LLM will significantly reduce the time I spend on administrative tasks.
- I expect to incur fewer out-of-pocket costs by using SNAP-LLM.

B.7 Learning Costs(Adapted from [17, 35])

- SNAP-LLM will make administrative tasks less mentally demanding.
- I anticipate needing to exert less effort to complete processes with SNAP-LLM.