RescueLens: LLM-Powered Triage and Action on Volunteer Feedback for Food Rescue

Naveen Janaki Raman,[1] Jingwu Tang,[1] Zhiyu Chen,[2] Zheyuan Ryan Shi,[3] Sean Hudson,[4] Ameesh Kapoor,[4] Fei Fang[1]

[1] Carnegie Mellon University
[2] University of Texas at Dallas
[3] University of Pittsburgh
[4] 412 Food Rescue

Abstract

Food rescue organizations simultaneously tackle food insecurity and waste by working with volunteers to redistribute food from donors who have excess to recipients who need it. Volunteer feedback allows food rescue organizations to identify issues early and ensure volunteer satisfaction. However, food rescue organizations monitor feedback manually, which can be cumbersome and labor-intensive, making it difficult to prioritize which issues are most important. In this work, we investigate how large language models (LLMs) assist food rescue organizers in understanding and taking action based on volunteer experiences. We work with 412 Food Rescue, a large food rescue organization based in Pittsburgh, Pennsylvania, to design **RESCUELENS**, an LLM-powered tool that automatically categorizes volunteer feedback, suggests donors and recipients to follow up with, and updates volunteer directions based on feedback. We evaluate the performance of RESCUELENS on an annotated dataset, and show that it can recover 96% of volunteer issues at 71% precision. Moreover, by ranking donors and recipients according to their rates of volunteer issues, RESCUELENS allows organizers to focus on 0.5% of donors responsible for more than 30% of volunteer issues. RESCUELENS is now deployed at 412 Food Rescue and through semi-structured interviews with organizers, we find that RESCUELENS streamlines the feedback process so organizers better allocate their time.

1 Introduction

Despite advances in food production, 800 million people remain chronically undernourished worldwide (UNICEF et al. 2021). At the same time, 40% of the food produced worldwide is wasted, demonstrating that food insecurity is a problem of distribution rather than production (WWF 2021; FAO 2013). Food rescue organizations simultaneously tackle both food waste and insecurity by redistributing food from those who have excess to those in need. Food rescue organizations work with volunteers to organize rescue trips, where volunteers transport food from donors, such as grocery stores, to recipients, including shelters. Food rescue organizations have been massively successful, saving over 150 million pounds of food in the United States since 2016 (Kelly 2021).

Food rescue organizations rely on volunteer feedback to understand the issues faced by volunteers. Volunteer feedback is critical because it serves as the primary mechanism for volunteers to learn about volunteer behaviors. Without it, these organizations have little visibility into relational issues between volunteers, donors, and recipients. For example, volunteer feedback can alert food rescue organizations to situations where a grocery store repeatedly misses their food pickup or if volunteers have repeated issues dropping off food for a particular recipient. Organizers at food rescue organizations can then intervene by contacting donors, changing pickup schedules, or updating directions.

While volunteer feedback is important, going from feedback to action is difficult. Although only a small fraction of volunteers leave feedback, this quickly adds up. For example, 412 Food Rescue, a large food rescue organization in Pittsburgh, received more than 1800 pieces of feedback per year, which costs the organization time. Moreover, manual feedback tracking makes it difficult to determine how best to allocate organizer time because it is unclear which issues are the most pressing or occur most frequently.

In this work, we investigate how large language models (LLMs) can help food rescue organizations understand volunteer feedback. LLM-based tools are typically faster and more efficient than human analysts, which can free up organizer time (Song, Agarwal, and Wen 2024). For example, LLM-based tools can quickly alert organizers to situations when intervention is needed, allowing organizers to focus on the most pressing pieces of feedback. At the same time, food rescue organizations are typically resource-limited, limiting their ability to develop large datasets for LLM applications. Moreover, volunteer feedback can often be ambiguous and involve domain-specific context.

In response to the challenges faced by food rescue organizations, we develop **RESCUELENS**, an LLM-powered tool that automatically analyzes feedback and enables organizers to take actions based on these insights (see Figure 1 for a summary). We built RESCUELENS in coordination with 412 Food Rescue, a large food rescue organization based in Pittsburgh, Pennsylvania, that has rescued over 30 million pounds of food since 2015. We first conducted a need-finding study with organizers at 412 Food Rescue, where we found that there is currently little formal documentation of volunteer feedback analyses, making it difficult to track which issues are most pressing. Based on this user study, we designed RESCUELENS to consist of two components: 1) an

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

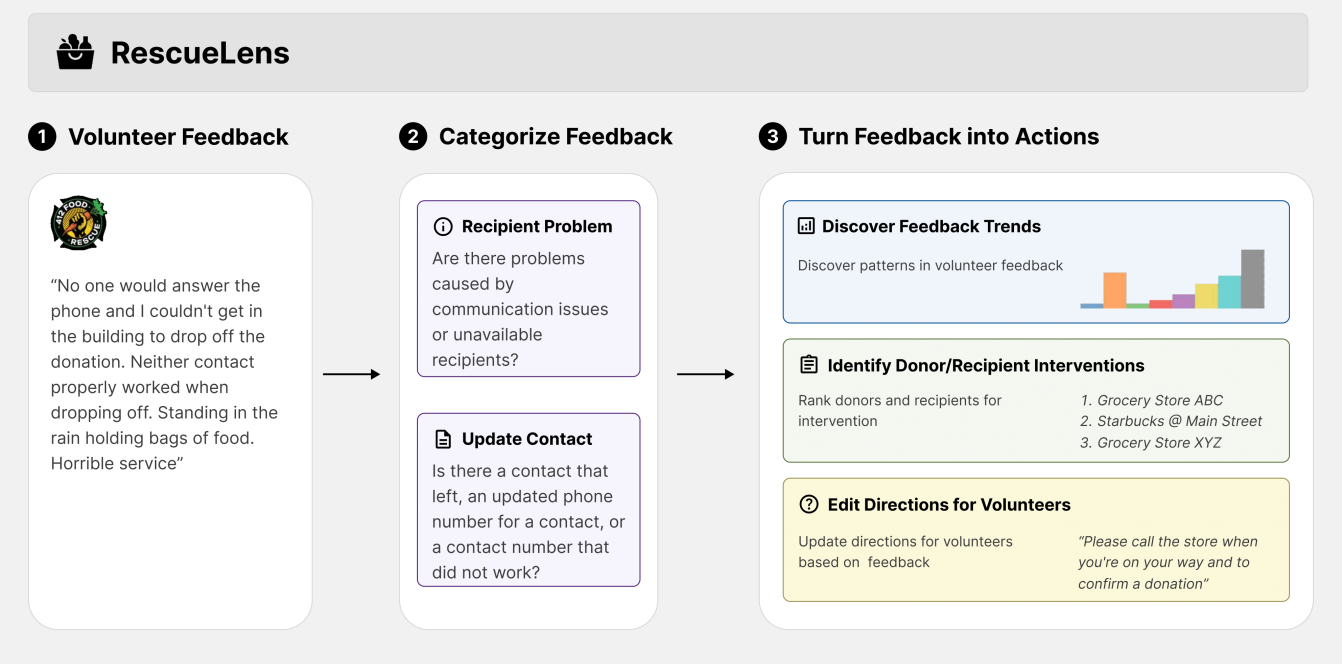

Figure 1: We introduce RESCUELENS, an LLM-powered tool that automatically analyzes volunteer feedback in food rescue. Our tool first categorizes volunteer issues into different categories, such as Recipient Problem and Update Contact. RESCUELENS then uses these predictions to discover trends in volunteer feedback, identify which donors and recipients require interventions, and suggest updates to the directions based on feedback.

LLM-based system which uses LLM with few-shot learning to efficiently categorize feedback and 2) a set of action modules that leverage feedback categorizations to a) identify which donors and recipients require intervention and b) update volunteer directions based on feedback. Through a mixed methods study, we demonstrate that RESCUELENS achieves a 96% recall and 71% precision, while allowing organizers to focus on the 0.5% of donors responsible for more than 30% of volunteer issues. Through interviews with organizers, we show that RESCUELENS helps quantify which problems are most pressing and determine how best to allocate their time. Our tool has been deployed at 412 Food Rescue since May 2025, and has analyzed more than 1,200 pieces of feedback from volunteers so far. Beyond food rescue, RESCUELENS can be broadly applied across non-profits to better understand their text-based feedback.

2 Related Work

Food Rescue Computational work in food rescue can be categorized into algorithmic research, which analyzes matching algorithms between volunteers, donors, and recipients, and system-level research, which investigates platform design (Shi, Wang, and Fang 2020). Food rescue organizations leverage algorithms both for matching donors with recipients (Mertzanidis, Psomas, and Verma 2024) and notifying volunteers about rescue trips (Raman, Shi, and Fang 2024; Shi, Lizarondo, and Fang 2021; Tang et al. 2025). Matching is difficult because organizers balance volunteer engagement, allocation efficiency, and fairness (Aleksandrov et al. 2015; Raman, Shi, and Fang 2024; Shi et al. 2020). On the other hand, system-level research investigates how to improve system design by analyzing volunteer interactions. Examples include one work that deploys a notification system at a university to reduce food waste (Silvis, Sicilia, and Labrinidis 2018) and another that informs users about task difficulty (Shi et al. 2024). Our work can be viewed as an extension of Shi et al. (2024), where we investigate how to design tools that mitigate task difficulties.

LLMs and Non-Profits Beyond food rescue, our work is broadly situated in the field of using LLMs to assist non-profits. Developing LLMs in non-profit scenarios is more challenging due to limitations in computational power and dataset size (Pu et al. 2025). In our situation, this required us to use in-context learning rather than fine-tuning. Additionally, developing LLMs for non-profits requires a balance between an LLM's abilities and its risks for hallucinations and bias (Goldkind, Ming, and Fink 2025). Moreover, non-profits themselves might have varied perspectives or desiderata for using LLMs. For example, some non-profits are concerned with the governance of LLMs, while others are concerned with the environmental impact of LLMs (Johnson 2025). These diverse perspectives inspire us to conduct interviews with organizers to understand *how* organizers planned to use RESCUELENS and tailor modules.

Automatic Analysis of User Feedback Non-profit stakeholders often express subjective feedback that can help improve systems, but automatic approaches are necessary to

gain insights from large-scale data (Chen et al. 2025; Behari et al. 2024). Subjective stakeholder feedback is prevalent across domains, including education (Parker et al. 2024), e-commerce (Kushwaha et al. 2024), and mobile applications (Abedini and Heydarnoori 2025). While stakeholder feedback is prevalent, automatic analysis is hard due to domain-specific language and ambiguity (Shaik et al. 2022). To circumvent this, prior work categorized feedback into different topics through supervised methods such as BERT and LLMs (Assi, Hassan, and Zou 2024) and unsupervised methods such as topic modeling (Perez-Encinas and Rodriguez-Pomeda 2018). Our work extends these feedback analysis techniques to scenarios with little labeled training data, then combines them with tools that transform user feedback into actionable insights for organizers.

3 Motivating RescueLens: A Needfinding Study at 412 Food Rescue

Study Procedure To better understand current practices for processing volunteer feedback, we conducted a series of needfinding studies with three organizers at 412 Food Rescue. Volunteer attrition is a large issue in food rescue organizations (Shi et al. 2024). A better understanding of volunteer feedback allows organizers to combat volunteer attrition by understanding volunteer problems. To understand current practices around volunteer feedback, we recruited three organizers (P1, P2, and P3) from 412 Food Rescue through social media and email advertisements. We asked each organizer a series of questions related to their role at 412 Food Rescue, current practices for processing volunteer feedback, and any issues that organizers face. We then discuss a potential tool that automatically analyzes volunteer feedback and ask organizers for feedback. We received approval from our institution's IRB, and we compensated $30 for the 30-minute study.

User Study Results Our user study revealed that the current feedback analysis procedure is largely manual with little formal documentation. Because the process lacks formal documentation, organizers mention difficulties in understanding which issues are most pressing and which only occur rarely. For example, P2 notes "We can't keep track of [feedback] as much as we need to." Such issues are amplified due to the scale of food rescue operations, with P2 noting that "800 recipients and 500 donors are too much to keep up with."

When we ask organizers to envision the benefits an automated system could potentially have, we find that organizers value tools that identify problems and fix volunteer directions based on feedback. P1 notes that "It's helpful to have a clear-cut list of issues you may experience with the app." P3 similarly notes that "coding these issues is helpful because so many different people make touch points", revealing the utility in categorizing feedback. Organizers also detail the utility of having tools for identifying which donors and recipients require intervention, with P3 stating "It would be useful to have a database that points out where they need to work and what problems we need to focus on." Organizers also note that having a system automatically edit directions would be useful, with P2 noting that "It would be useful if we could get pinged for the [updated] directions."

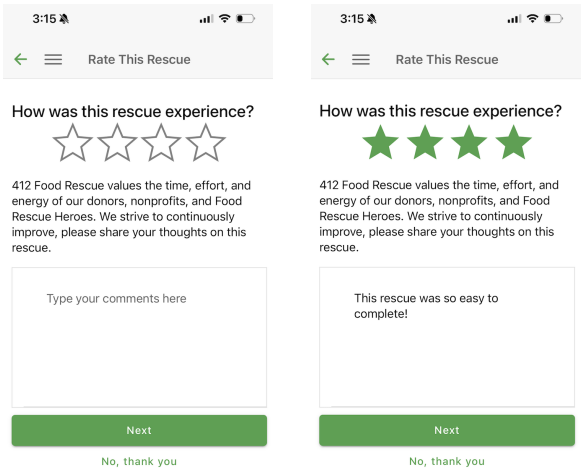

Figure 2: After each rescue trip, 412 Food Rescue collects a rating out of four along with text-based feedback.

4 RescueLens System Design

RescueLens is an LLM-based system that first categorizes volunteer feedback, then uses these classifications to recommend actions for food rescue organizers.

4.1 Categorizing User Feedback

Volunteer Dataset To construct RescueLens, we begin with a dataset of volunteer feedback from rescue trips. 412 Food Rescue elicits user feedback after each rescue trip along with a rating out of four (see Figure 2). The dataset starts from 2018 and includes over 200000 rescue trips. In total, this results in 14,439 pieces of text feedback in the database.

Defining User Feedback Categories As a first step towards automatic feedback analysis, we perform open coding to understand the types of feedback present. We start with a random sample of 250 volunteer feedback from rescue trips in 2024, then employed open coding, iteratively refining the codebook as new themes emerged. We stopped coding after observing exhaustion, where no new categories arose. After discarding categories that did not influence organizer action (e.g., comments that gave positive reviews of their trip), we arrived at seven categories, which we define below:

1. Inadequate Food: Assess whether the reported challenges or failures in the food rescue process were caused by inadequate food quantities provided by the donor. For example, "*Nothing to donate. Everything they had put aside was burned.*"
2. Earlier Pickup: Assess whether the reported challenges or failures in the food rescue process were caused by someone else (e.g., another volunteer) picking up the food earlier, leading to little or no available food. For ex-

ample, "*Per the store manager, someone else was already there today and picked up everything.*"

3. Donor Problem : Assess whether the reported challenges or failures in the food rescue process were caused by communication issues with donors. For example, "*Terrible pickup! Nobody knew who 412 was. After 1/2 hour, I was given 3 boxes of apples. As I left, I was flagged down and given a cart full of leftover Easter candy.*"
4. Recipient Problem : Assess whether communication issues or unavailable recipients caused the reported challenges. For example, "Food pantry was closed."
5. Update Contact : Assess whether the feedback discussed the need to update contact information for a donor or recipient. For example, "*The manager contact at Walmart has a new job and won't be there starting next week.*"
6. System Problem : Assess whether bugs or glitches on the food rescue app caused the reported challenges. For example, "*System not responsive.*"
7. Direction Problem : Assess whether unclear or inaccurate directions or information caused the reported challenges. For example, "*The map directions took me to Alexander Street. Please adjust pick up location to Powell.*"

Employing LLMs for Classification We construct prompts for each category. Prompts consist of background data, general task description, specific task details, and a few-shot examples with manually written explanations. We use in-context learning, which improves LLM performance without the need for large annotated corpora (Brown et al. 2020).

4.2 Turning Feedback into Actions

RESCUELENS converts feedback classifications into actionable insights through a pair of tools: the first ranks donors and recipients for intervention, and the second rewrites volunteer directions.

Determining Where to Intervene We built a module in RESCUELENS to help food rescue organizers identify which donors and recipients require attention. The module produces a ranked list of donors and recipients, each scored using two types of feedback: volunteer ratings and reported issues. The score represents the rate at which volunteers have issues when completing rescue trips for a donor or recipient. Higher scores represent higher priorities for intervention. We use these scores to rank donors and recipients, which can help organizers decide where to focus their efforts.

We compute the score based on the volunteer rating score, on a 1-4 scale, and a set of predictions from RESCUELENS. From the predictions made by RESCUELENS, we produce a comment score that represents whether any issue is present. For donors, we compute whether Update Contact , Inadequate Food , Earlier Pickup , Direction Problem , or Donor Problem is present, while for recipients, we check whether Update Contact or Direction Problem or Recipient Problem is present. The final score is then the fraction of rescue trips where either the comment score is non-zero or the rating score is below 4.

Suggesting Direction Rewrites We constructed a tool to automatically rewrite volunteer directions for feedback labeled as Direction Problem . Each donor or recipient has a set of directions that includes information on driving directions, points of contact, and delivery details. While directions are critical to volunteer success, they can become outdated over time. To assist with this, we periodically process the latest batch of new feedback and, for each feedback item, use an LLM to rewrite the relevant volunteer directions. For each feedback item, we first prompt the LLM to determine whether it contains new information that warrants updating the directions; only if the LLM identifies relevant updates do we generate a revised direction. We prompt LLMs with the original directions, the new feedback, and explicit constraints to incorporate all new information (e.g., updated contact, entrance, or address), preserve correct existing details, and remove information only if directly contradicted. We use seven manually curated few-shot examples—covering both donor and recipient cases—to guide the rewrite style and scope.

5 RescueLens Evaluation

We evaluate the performance of RESCUELENS through comparisons with baselines on historical data.

5.1 Evaluation Setup

We evaluate the classification module of RESCUELENS by comparing with baselines on expert-annotated data. We randomly sample 125 data points consisting of volunteer feedback from 2024, and then we annotate feedback for each category from Section 4.1. We use two annotators, having them each independently rate metrics, before coming to consensus. We assess performance according to four metrics: accuracy, precision, recall, and F1 score. We measure the inter-annotator agreement through Cohen's κ, and find that it is $\kappa = 0.73$, indicating significant agreement across independent annotators.

We compare RESCUELENS against LLM and non-LLM baselines. For LLM baselines, we maintain the prompts used by RESCUELENS while varying the underlying LLM. By default, RESCUELENS uses GPT-4o mini (Hurst et al. 2024), and we compare this against GPT-4o (Hurst et al. 2024), Llama 3 (Dubey et al. 2024), and DeepSeekR1 (Guo et al. 2025). We additionally compare RESCUELENS against a term frequency-inverse document frequency (TF-IDF) baseline with logistic regression, which computes the relative frequency of different words, and a DistilBERT module (Sanh et al. 2019), which finetunes BERT using a small dataset of labeled volunteer feedback. All evaluations are across three random seeds.

Model	Acc.	Prec.	Recall	F1	Cost
RescueLens					
GPT-4o mini	93.1%	83.3%	97.4%	89.8%	$15
GPT-4o	94.4%	88.1%	94.9%	91.4%	$250
Llama 3.1	90.5%	76.5%	100.0%	86.7%	$10
DeepSeek R1	79.4%	68.6%	61.5%	64.9%	$80
TF-IDF	31.0%	31.0%	100.0%	47.3%	-
DistilBERT	31.0%	31.0%	100.0%	47.3%	-

Table 1: We evaluate RESCUELENS on an annotated dataset. The first four rows are LLM-based approaches, while the latter two are non-LLM approaches. We find that RESCUELENS, which relies on a GPT-4o mini backbone, outperforms all alternatives on accuracy and F1 score except GPT-4o. Moreover, RESCUELENS is significantly cheaper than both GPT-4o and DeepSeek R1, demonstrating that RESCUELENS is both cost-efficient and well-performing.

5.2 Classification Evaluation

Comparison against Baselines In Table 1, we compare the performance and cost per year when varying the model. We compare the rates of finding any issue; that is, the rate of detecting whether any of the seven categories are true, given a comment. We focus on this metric because it represents the success of RESCUELENS on finding which comments require further organizer intervention. We show that GPT-4o maximizes accuracy and F1 score, while GPT-4o mini sacrifices a bit of performance for cost reduction. RESCUELENS. Moreover, LLM-based algorithms outperform non-LLM baselines; all LLM-based variants of RESCUELENS achieve at least 75% accuracy, while TF-IDF and DistillBERT achieve 31% accuracy, as they always predict true. Comparing between LLMs, we find that RESCUELENS, which is built on top of GPT-4o mini, is 2% worse in F1 score compared to GPT-4o, while reducing costs from $250 to $15 per year. While Llama 3.1 costs $5 less than GPT-4o mini, it performs worse on the F1 score, demonstrating that GPT-4o mini is Pareto efficient among these models.

Ablating RescueLens Components To understand which components are most responsible for the performance of RESCUELENS, we ablate different components. The prompt for RESCUELENS consists of a description of the task and scenario, a set of detailed guidelines stating what to label or not label, and a set of few-shot examples with explanations.For example, one part of the guidelines for Donor Problem states to "mark comments where the interaction with the donor was delayed as donor problems." We ablate the guidelines and few-shot examples, then re-evaluate the performance of RESCUELENS after this. In Figure 3, we compare the model performance across these three variants, and find that removing the guidelines reduces F1 score by 4%, while removing few-shot learning reduces the F1 score by 2%. Most of this reduction is concentrated in a lower precision, as precision reduces by 6% and 3% when removing guidelines and few-shot examples, respectively.

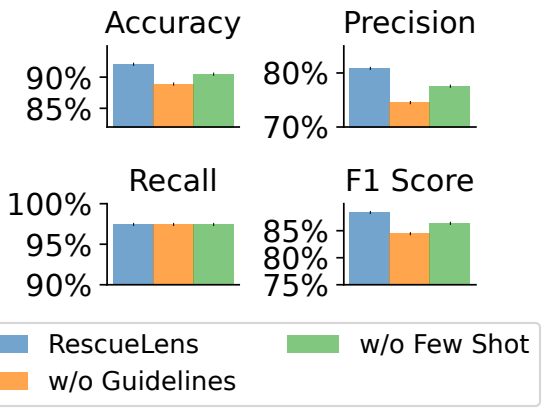

Figure 3: We conduct an ablation study to understand the importance of different aspects of RESCUELENS. We find that RESCUELENS performs well because of a combination of few-shot learning and task-specific guidelines.

Category	Acc.	Prec.	Recall	F1
Any Issue	93.1%	83.3%	97.4%	89.8%
Donor Problems	95.2%	68.9%	100.0%	81.5%
Inadequate Food	94.2%	73.3%	100.0%	84.6%
Earlier Pickup	100.0%	100.0%	100.0%	100.0%
Donor Problem	93.7%	42.5%	83.3%	56.0%
Recipient Problem	96.6%	48.1%	100.0%	65.0%
Update Contact	97.9%	27.8%	100.0%	43.3%
System Problem	98.4%	75.0%	75.0%	75.0%
Direction Problem	98.7%	75.4%	100.0%	85.9%

Table 2: We evaluate the performance of RESCUELENS across different categories. We find that RESCUELENS achieves at least a 75% recall across categories, and achieves over 93% accuracy across all categories.

Performance by Category We stratify the performance of RESCUELENS by each of the seven feedback categories from Section 4.1. In Table 2, we find that RESCUELENS achieves at least a 75% recall rate across categories, and at least a 92% accuracy score as well. We intentionally design RESCUELENS to optimize for recall rather than precision after internal discussions because organizers can filter down false positives. High recall ensures that organizers are able to find all volunteer comments that require further action. Categories like Earlier Pickup and Inadequate Food have high rates for recall and precision because they are less ambiguous, and these categories also have the highest inter-annotator agreements. While RESCUELENS has a low precision for Donor Problem, we note that this occurs because of mix-ups within other donor-related issues, as shown through the higher score across all donor problems. Across all categories, RESCUELENS achieves a higher recall than precision; we intentionally construct RESCUELENS in such a way that organizers can pare down false positives when intervening. Moreover, for some categories, such as the low F1 score, it is due to mislabeling between different types

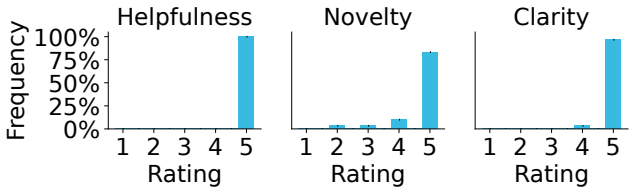

Figure 4: We assess the performance of our direction rewrite module according to three criteria: helpfulness, novelty, and clarity. Across these criteria, we find that RESCUELENS performs well, averaging over a 4.7/5 across all three categories.

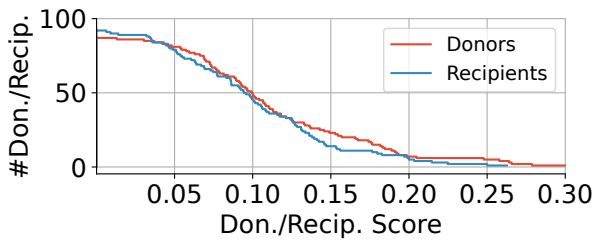

Figure 5: We compute scores for donors and recipients using the formula from Section 4.2, then plot the distribution of scores. We show that only a few donors and recipients require intervention, and by focusing on these few, we can reduce organizer efforts.

of donor issues, as shown by the higher F1 score across all donor problems.

5.3 Evaluating Direction Rewrites

We evaluate the performance of the direction rewrite module against a set of three criteria:

1. **Helpfulness** - Does the new direction contain information that is important for completing the trip? For example, does it contain information on directions, contact information, or important logistics?
2. **Novelty** - How well does the new direction incorporate information from both the original direction and the feedback? Does it properly incorporate the new information, without removing any important previous information?
3. **Clarity** - How clear are the directions written; can they be easily understood without much thought?

We have two authors annotate 30 rewritten directions. Each direction is assessed on a 1-5 scale for each criterion.Our inter-annotator agreement is $\kappa = 0.39$, indicating fair agreement.

In Figure 4, we plot the distribution of scores across each of the three criteria after averaging scores between annotators.. We find that RESCUELENS performs well across categories; for all three categories, the average score is at least 4.7/5. Moreover, over 70% of the rewritten sentences achieved a perfect score of 5/5 across all three categories, demonstrating that RESCUELENS can construct rewritten directions that are helpful, novel, and clear. Rewritten directions are almost always clear and can incorporate new volunteer feedback most of the time.

6 Deploying RescueLens

We provide details on our deployment of RESCUELENS to 412 Food Rescue, then perform a mixed methods study to evaluate the impact of RESCUELENS.

6.1 Deployment to 412 Food Rescue

We deployed RESCUELENS at 412 Food Rescue through a series of stages where different versions of RESCUELENS were deployed. We began development of RESCUELENS in Spring 2024 and produced the first working prototype early in Fall 2024. We spent the next several months integrating RESCUELENS into 412 Food Rescue, and we produced our first report on February 11th, 2025. We worked on further updates to better integrate RESCUELENS, and officially launched our RESCUELENS on May 15th, 2025, where it analyzed over 600 pieces of feedback so far.

We integrate RESCUELENS into the existing database for 412 Food Rescue through a new table titled "Rescue Feedback." This table includes information on each piece of feedback, its categorization into seven different categories, and notes left by organizers during analysis. By integrating directly into the database, we ensure easy access and usage by organizers at 412 Food Rescue. To automatically populate this table, we run a daily Ruby script that calls RESCUELENS to populate the database with the previous day's feedback. Each feedback is labeled daily with the seven categories from Section 4.1. To incorporate our action modules into 412 Food Rescue, we send the direction rewrites and the ranked list of donors and recipients on a monthly basis. We update the action modules monthly because the donor and recipient rankings are aggregations of historic comments that operate over longer timescales, while direction rewrites occur only a few times per week.

During deployment, we made changes to RESCUELENS to improve usability. We introduce a user interface that allows organizers to search historical data. Organizers can query the user interface across date ranges and combinations of volunteer feedback classifications. The user interface also allows organizers to track which pieces of feedback require intervention and to leave notes. We also present information from RESCUELENS to organizers via an automatic bot in Slack that fetches a daily list of analyzed rescues.

6.2 Impact on Practice

We assess the impact of RESCUELENS through a mixed methods study. We first demonstrate how RESCUELENS reduces organizer workload by guiding their actions towards intervening on the most impactful donors and recipients. We complement this with a semi-structured interview with an organizer at 412 Food Rescue, who discusses the impact of RESCUELENS in reducing the organizer's workload and streamlining the feedback analysis process.

RESCUELENS ranks donors and recipients, which enables organizers to focus on the donors and recipients who lead to

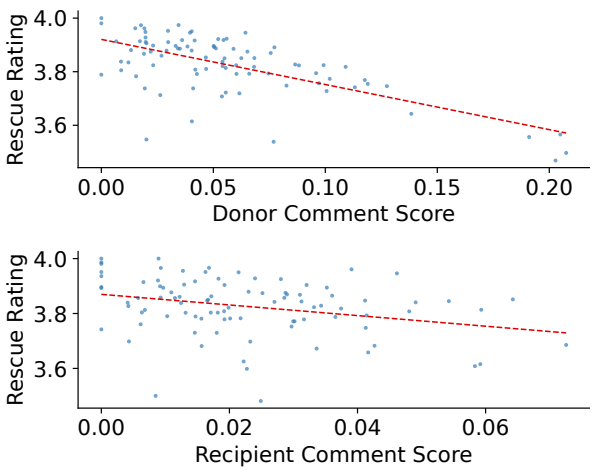

Figure 6: We plot the correlation between donor and recipient comment scores and the average rating for rescue trips associated with that donor or recipient. We find a large negative correlation ($r^2 = 0.45$) for donor comment scores and a moderate correlation for recipient comment scores ($r^2 = 0.09$). Fixing donor-related issues can potentially improve overall volunteer experiences.

the most volunteer issues. In Figure 5, we compute the distribution of scores for donors and recipients with at least 100 rescue trips. Here, we use the same score from Section 4.2, where a higher score indicates a higher rate of volunteer issues with the donor or recipient. We find that the vast majority of donors and recipients have a low score, indicating positive volunteer experiences, and less than 20% of donors and recipients have scores larger than 0.20. Addressing these issues can improve volunteer experiences with 412 Food Rescue; in Figure 6, we show that rescue trip ratings correlate with a donor's comment score ($r^2 = 0.45$) and somewhat correlate with a recipient's comment score ($r^2 = 0.09$). Moreover, addressing these issues only requires intervention on a few donors. In Figure 7, we show that 5 donors, representing 0.5% of all donors, are responsible for at least 25% of all comment issues (across each category) for each of the 4 categories. Moreover, these donors have an issue rate disproportionate to their size; they only cover 2.5% of rescue trips but are responsible for 25% of issues. We stratify this trend by donor issue in Figure 7 and show that five donors, representing 0.5% of all donors, are responsible for more than 30% of volunteer issues across four categories. By providing organizers with a ranked list of donors and recipients for intervention, we guide organizers towards donors most responsible for issues.

We additionally find that they can include valuable information based on volunteer feedback. We show three examples of rewritten directions in Figure 8. Each direction includes vital information, such as phone numbers, drop-off details, and driving directions. Including these rewritten directions helps improve volunteer experiences by making it clearer how to complete each rescue. Such instructions can also help volunteers avoid common pitfalls.

To understand the impact of RESCUELENS in practice, we conducted a semi-structured interview with organizers at 412 Food Rescue. We conducted a 30-minute interview with the donor relations coordinator at 412 Food Rescue. The donor relations coordinator handles reach-outs to donors based on volunteer feedback, making them a natural fit to assess the impact of RESCUELENS. We asked a series of questions related to the impact and performance of RescueLens in practice.

Through the interview, we find that RESCUELENS reduces the manual effort needed to understand user feedback while streamlining the outreach process. The organizer reported on the daily impact of RESCUELENS:

> We look at RESCUELENS every day so we can track everyday trends with food donors, such as shifts in store leads or changes in ordering.

RESCUELENS is helpful because organizers are able to understand volunteer patterns and turn these into actions:

> Because of RESCUELENS, we are able to see where the trends are happening seasonally or quarterly, and it helps us quantify and qualify what needs to be fixed.

RESCUELENS enables organizers to take actions based on volunteer feedback, as the organizer noted:

> Around 50% of the reported comments from RESCUELENS are actionable, and we try to get to every actionable thing.

Our interview demonstrates the power of RESCUELENS in enabling organizers at 412 Food Rescue to better understand volunteer feedback and take action based on this.

7 Lessons Learned

Our experience deploying RESCUELENS has provided a set of valuable insights, which we detail below:

1. **Importance of Integration and Presentation** - Much of the work in developing RESCUELENS focused on integrating it into food rescue systems. Seamless integration within the food rescue platform is essential because it reduces the barriers for organizers to access and leverage RESCUELENS for their decision-making; the easier it is to use RESCUELENS, the more of an impact it will have. As a simple example, we found that visually presenting these results is critical so organizers can better understand trends in an accessible manner.
2. **Metrics Beyond Accuracy** - We experimented with different configurations and underlying models during the development of RESCUELENS, where these settings varied along dimensions including precision, recall, and cost. While larger models (e.g., GPT-4o) tend to perform better, we found that these models were significantly more expensive. We engaged in conversations around tradeoffs between models, where we found that accuracy is not the sole arbiter. We encourage future researchers to understand which metrics are most important to stakeholders rather than relying on a metric of convenience.

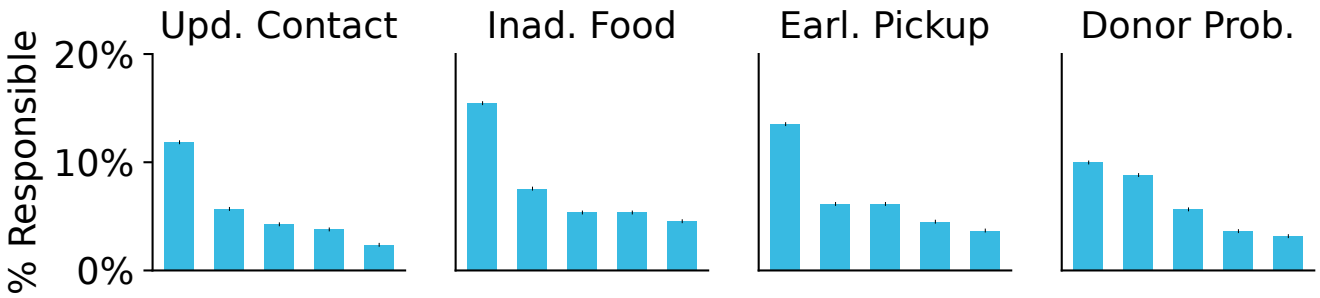

Figure 7: Here, we plot the five donors with the largest number of comments for each of four categories. Although there are hundreds of donors, we show that a small subset of donors is responsible for many of the issues. By directing organizers towards these donors, we can reduce the organizer workload by focusing on the donors who require intervention.

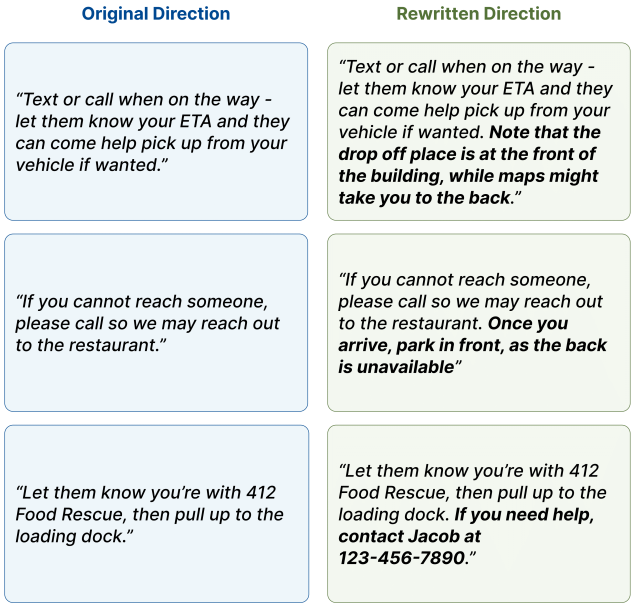

Figure 8: We plot examples of rewritten directions. These directions incorporate valuable information on contact information and drop-off details.

3. **Heterogeneity of Volunteer Feedback** - We built RESCUELENS around volunteer feedback, and through the development of RESCUELENS, we find that volunteer feedback is a rich source of information for a nonprofit. Volunteer feedback in RESCUELENS not only informs which donors and recipients require intervention but also details direction modifications and contact updates. Such an idea holds across non-profits, as text-based feedback can reveal information about a non-profit's underlying health. At the same time, human feedback is hard to parse due to inherent ambiguity. We overcame this issue by developing clear standards when defining categories and using them to retrieve predictions.

8 Discussion and Conclusion

Food rescue organizations tackle food insecurity by redistributing excess food from donors to recipients via volunteers. Volunteer feedback is critical for a healthy relationship between non-profits and volunteers, yet manually processing feedback becomes cumbersome due to scale. To tackle this problem, we develop RESCUELENS, an LLM-based tool that automatically categorizes feedback at food rescue organizations and helps organizers take actions based on feedback. We deploy RESCUELENS with our partners at 412 Food Rescue, and demonstrate that RESCUELENS maintains high accuracy, precision, and recall. Through a mixed methods study, we find that RESCUELENS can help guide organizers at 412 Food Rescue identify which donors and recipients require intervention, thereby streamlining the feedback process. Throughout the process, we learned valuable lessons on how AI-based tools can be best integrated into non-profits. Future improvements to RESCUELENS include a tracking system to measure which donors and recipients have been intervened upon and a system for answering volunteer questions based on prior feedback.

Our results demonstrate the efficacy of LLMs in helping organizers better understand volunteer feedback. Beyond food rescue, RESCUELENS can be extended to other non-profits, and we hope our work outlines the steps needed for deployment. At the same time, non-profits vary in their access to data, types of feedback, and intervention priorities (e.g., how they intervene based on volunteer feedback). RESCUELENS is flexible enough to adapt based on non-profit specifics. For example, the donor and recipient scores can be modified based on non-profit priorities. To help with RESCUELENS adoption, we include the code and prompts necessary for replication, which can help organizations with integration. Through the use of such a tool, non-profits can better quantify volunteer opinions and perspectives and improve non-profit health.

Acknowledgements

We thank Rex Chen for comments on an earlier draft of this paper. Co-author Raman is also supported by an NSF Fellowship. This work was supported in part by NSF grant IIS-2046640 (CAREER).

References

Abedini, Y.; and Heydarnoori, A. 2025. Leveraging Large Language Models for Classifying App Users' Feedback. *arXiv preprint arXiv:2507.08250*.

Aleksandrov, M.; Aziz, H.; Gaspers, S.; and Walsh, T. 2015. Online fair division: Analysing a food bank problem. *arXiv preprint arXiv:1502.07571*.

Assi, M.; Hassan, S.; and Zou, Y. 2024. LLM-Cure: LLM-based Competitor User Review Analysis for Feature Enhancement. *ACM Transactions on Software Engineering and Methodology*.

Behari, N.; Zhang, E.; Zhao, Y.; Taneja, A.; Nagaraj, D.; and Tambe, M. 2024. A decision-language model (dlm) for dynamic restless multi-armed bandit tasks in public health. *Advances in Neural Information Processing Systems*, 37: 3964–4002.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901.

Chen, R.; Wu, K.; McCartney, J.; Sadeh, N.; and Fang, F. 2025. Out of the Past: An AI-Enabled Pipeline for Traffic Simulation from Noisy, Multimodal Detector Data and Stakeholder Feedback. *arXiv preprint arXiv:2505.21349*.

Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Yang, A.; Fan, A.; et al. 2024. The llama 3 herd of models. *arXiv e-prints*, arXiv–2407.

FAO. 2013. Food Wastage Footprint: Impacts on Natural Resources: Summary Report.

Goldkind, L.; Ming, J.; and Fink, A. 2025. AI in the nonprofit human services: Distinguishing between hype, harm, and hope.

Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Hurst, A.; Lerer, A.; Goucher, A. P.; Perelman, A.; Ramesh, A.; Clark, A.; Ostrow, A.; Welihinda, A.; Hayes, A.; Radford, A.; et al. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Johnson, J. 2025. How Nonprofits Can Help Shape AI Governance. Accessed: 2025-08-02.

Kelly, K. 2021. How One App Saved Over 40 Million Pounds of Food from the Landfill. https://blog.techsoup.org/posts/how-one-app-saved-over-40-million-pounds-of-food-from-the-landfill. TechSoup Blog — Impact Stories.

Kushwaha, A. K.; Jadon, S.; Kamal, P.; Saini, M. L.; and Shrimal, V. M. 2024. Comments and feedback verification system using large language model. In *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, 1–6. IEEE.

Mertzanidis, M.; Psomas, A.; and Verma, P. 2024. Automating food drop: The power of two choices for dynamic and fair food allocation. *arXiv preprint arXiv:2406.06363*.

Parker, M. J.; Anderson, C.; Stone, C.; and Oh, Y. 2024. A large language model approach to educational survey feedback analysis. *International journal of artificial intelligence in education*, 1–38.

Perez-Encinas, A.; and Rodriguez-Pomeda, J. 2018. International students' perceptions of their needs when going abroad: Services on demand. *Journal of Studies in International Education*, 22(1): 20–36.

Pu, C.; Kim, M.; Derrick-Mills, T.; and Faulk, L. 2025. Challenges and Experiences in Data Integration to Support Research on Nonprofit Organizations. *ACM Transactions on Internet Technology*.

Raman, N.; Shi, Z. R.; and Fang, F. 2024. Global rewards in restless multi-armed bandits. *Advances in Neural Information Processing Systems*, 37: 24625–24658.

Sanh, V.; Debut, L.; Chaumond, J.; and Wolf, T. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Shaik, T.; Tao, X.; Li, Y.; Dann, C.; McDonald, J.; Redmond, P.; and Galligan, L. 2022. A review of the trends and challenges in adopting natural language processing methods for education feedback analysis. *Ieee Access*, 10: 56720–56739.

Shi, Z. R.; Lizarondo, L.; and Fang, F. 2021. A recommender system for crowdsourcing food rescue platforms. In *Proceedings of the web conference 2021*, 857–865.

Shi, Z. R.; Wang, C.; and Fang, F. 2020. Artificial intelligence for social good: A survey. *arXiv preprint arXiv:2001.01818*.

Shi, Z. R.; Yuan, Y.; Lo, K.; Lizarondo, L.; and Fang, F. 2020. Improving efficiency of volunteer-based food rescue operations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 13369–13375.

Shi, Z. R.; Zhi, J.; Zeng, S.; Zhang, Z.; Kapoor, A.; Hudson, S.; Shen, H.; and Fang, F. 2024. Predicting and presenting task difficulty for crowdsourcing food rescue platforms. In *Proceedings of the ACM Web Conference 2024*, 4686–4696.

Silvis, M.; Sicilia, A.; and Labrinidis, A. 2018. PittGrub: a frustration-free system to reduce food waste by notifying hungry college students. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 754–763.

Song, F.; Agarwal, A.; and Wen, W. 2024. The impact of generative AI on collaborative open-source software development: Evidence from GitHub Copilot. *arXiv preprint arXiv:2410.02091*.

Tang, A.; Raman, N.; Fang, F.; and Shi, Z. R. 2025. Contextual Budget Bandit for Food Rescue Volunteer Engagement. *arXiv preprint arXiv:2509.10777*.

UNICEF; et al. 2021. *The state of food security and nutrition in the world 2021*. FAO;.

WWF, U. 2021. Driven to waste: The global impact of food loss and waste on farms. *Retrieved from WWF*.