# When Scammers Talk Back: Understanding Potentially Unwanted Calls via LLM-Based Interaction


Zhuoer Lyu
Arizona State University
Tempe, AZ, USA
zlyu15@asu.edu

Marzieh Bitaab*
Amazon
Tempe, AZ, USA
mbitaab@amazon.com

Shuyi Huang
Arizona State University
Tempe, AZ, USA
shuan213@asu.edu

Alireza Karimi
Arizona State University
Tempe, AZ, USA
akarimi6@asu.edu

William Robertson
Arizona State University
Tempe, AZ, USA
wjrob@asu.edu

Zeming Yu
Arizona State University
Tempe, AZ, USA
zemingyu@asu.edu

Moritz Schloegel
CISPA Helmholtz Center for Information Security
Saarbrücken, Germany
moritz.schloegel@cispa.de

Yan Shoshitaishvili
Arizona State University
Tempe, AZ, USA
yans@asu.edu

Gail-Joon Ahn
Arizona State University
Tempe, AZ, USA
gahn@asu.edu

Ananta Soneji
University of Nebraska at Omaha
Omaha, NE, USA
asoneji@unomaha.edu

Ruoyu Wang
Arizona State University
Tempe, AZ, USA
fishw@asu.edu

Adam Doupé
Arizona State University
Tempe, AZ, USA
doupe@asu.edu

Tiffany Bao
Arizona State University
Tempe, AZ, USA
tbao@asu.edu



## Abstract

Despite stricter regulations being gradually enforced in telecommunications, potentially unwanted calls (PUCs), including telemarketing and vishing, remain a persistent and evolving threat. However, understanding how PUCs operate in practice has been limited by the lack of real-world datasets with rich conversational content, hindering the development of effective defenses. To fill this gap, we developed an LLM-based phone call honeypot for collecting real-world PUCs, to our knowledge among the first systems to collect such data in the wild. Based on the development, we conducted thematic analysis on 115 hours of conversations collected over 20 months, where we understand how PUCs operate in practice by studying the conversational structure, content, and campaign operations.

Our analysis reveals that PUC conversations are highly structured across diverse call topics. We identify a widespread early-call pattern, which we term the *hook phase*, that callers use to engage and prescreen targets before eliciting sensitive information. Within this phase, we identify an emerging entity, *call-to-interaction robocalls*. Unlike *call-to-action robocalls*, which rely on prerecorded audio, *call-to-interaction robocalls* sustain engagement through interactive questioning, more closely resembling *human agents*. Beyond the *hook phase*, we find that malicious callers progressively escalate the sensitivity of personal information over the course of conversations while using persuasion techniques to establish authority and rapport. Our longitudinal analysis further identifies eight major topics that remain present throughout the nearly two-year study period (e.g., Medicare), along with major campaigns that stay active over time, suggesting that PUC operations are durable and underscoring the need for stronger interventions against them.



## CCS Concepts

• **Security and privacy → Social engineering attacks**; **Usability in security and privacy**.

## Keywords

Unwanted Calls; Scam Calls; Honeypot System; Social Engineering

**ACM Reference Format:**
Zhuoer Lyu, Marzieh Bitaab, Shuyi Huang, Alireza Karimi, William Robertson, Zeming Yu, Moritz Schloegel, Yan Shoshitaishvili, Gail-Joon Ahn, Ananta Soneji, Ruoyu Wang, Adam Doupé, and Tiffany Bao. 2026. When


*This work does not relate to the author's position at Amazon.




This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.


*CCS '26, The Hague, Netherlands*


© 2026 Copyright held by the owner/author(s).


ACM ISBN 979-8-4007-2871-6/2026/11
https://doi.org/10.1145/3830454.3846770

Scammers Talk Back: Understanding Potentially Unwanted Calls via LLM-Based Interaction. In *Proceedings of the 2026 ACM SIGSAC Conference on Computer and Communications Security (CCS '26), November 15–19, 2026, The Hague, Netherlands.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3830454.3846770

## 1 Introduction

Phone calls remain a cornerstone of modern communication, but the same low cost and ubiquitous reach that make telephony useful also make it attractive to malicious actors. With virtually everyone carrying a phone and little ability to verify a caller's identity, users routinely face *potentially unwanted calls (PUCs)*, ranging from unsolicited telemarketing to outright scams such as vishing, where attackers impersonate trusted entities to exfiltrate sensitive information or convince victims to take actions that cause financial loss or beyond. According to the Federal Trade Commission, scam calls are the second most commonly reported fraud method and cause the highest median per-person loss of $1,480, accounting for up to $850 million in losses in 2023 alone [20]. Despite years of public-awareness campaigns [19], regulatory action [15, 18], and stricter carrier obligations [16], PUCs continue to threaten users and to evolve at a pace that current defenses struggle to match [29].

A central obstacle to studying PUCs is the lack of real-world datasets that capture *what actually happens during calls*. Phone conversations are ephemeral, and victims rarely report them [38]. Unlike phishing emails or malicious websites, which leave textual records, documenting the content of a phone call requires significant effort from victims. This high reporting cost discourages disclosure and contributes to a self-reinforcing cycle where limited data availability further hinders systematic study. As a result, there is no widely shared corpus of PUC content [40, 45].

The honeypot proposed by Prasad et al. [39, 40] took an important first step by passively answering incoming calls, but its lack of interactive response limits collection to *call-to-action robocalls*—pre-recorded prompts asking the callee to press a button or call back. As a result, prior conversation-level analyses have relied on cherry-picked scambaiter videos [45, 54], leaving the broader landscape of how PUCs unfold in practice largely unexamined.

We address this gap by designing and developing CallTrap, an LLM-based telephony honeypot that *engages* PUC callers and records full conversations. The system transcribes caller speech, uses a large language model (LLM) to generate responses, and converts them back into speech for real-time interaction. Built on a Voice over IP (VoIP) platform, CallTrap autonomously answers calls and sustains interaction with callers.

We deploy CallTrap across 29 phone numbers (12 VoIP, 16 carrier, and 1 donated) over a 20-month period from late July 2024 through mid-February 2026. During this period, the system answered 22,697 calls, of which 8,962 contained substantive conversations totaling 281.8 hours of audio. To our knowledge, this is among the first systems to collect interactive, real-world PUC conversations at scale.

Building on this dataset, we perform large-scale analysis of a stratified 115-hour sample (3,630 calls) using a codebook developed inductively from the data and deductively from prior work on robocalls, vishing, and social engineering [23, 33, 39, 40, 45, 54]. Through this analysis, we set out to answer three questions about PUCs that have remained open in the absence of conversational data: **(i)** How are PUC conversations structured, and how does that structure vary across topics and caller types? **(ii)** How do callers persuade targets and elicit sensitive information once a conversation is underway? **(iii)** What do these conversations reveal about the operational dynamics of PUC campaigns over time?

Our analysis surfaces several findings that we believe are new to the literature. First, despite enormous variation in topic and identity, PUC conversations follow a remarkably consistent two-stage structure: an early *hook phase* that establishes presence, claims an identity, and prescreens the target, followed by an *operational phase* in which the caller pursues the underlying objective. Within the hook phase we identify an emerging entity that we term *call-to-interaction robocalls*: synthesized callers that, unlike the *call-to-action robocalls* studied by prior work, sustain engagement through realistic voices and interactive questioning, and that account for at least 56.40% (1,370/2,429) of observed hook phases. Second, we show that malicious callers *progressively escalate* the sensitivity of requested information over the course of a call, and that *authority* (AUTH), *commitment, reciprocation, consistency* (CRC), and *distraction* (DIS) are heavily front-loaded as persuasion tools to lower the callee's guard before sensitive requests appear. Third, our longitudinal analysis identifies eight call topics (e.g., Medicare, final expense, subsidy) and dozens of distinct campaigns that remain active across nearly the full 20-month window, indicating that PUC operations are durable and resistant to existing interventions.

**Contributions.** This paper makes the following contributions:

- We develop CallTrap, an LLM-based telephony honeypot that enables automated interaction with callers. Based on CallTrap, we collect the first large-scale dataset of real-world PUC conversations.
- We show that PUCs follow a consistent two-stage structure consisting of a hook phase and an operational phase, identify the emergence of *call-to-interaction robocalls*, and uncover key behavioral patterns such as progressive escalation of information requests.
- We reveal that PUCs are driven by persistent campaigns and that existing protections such as the National Do Not Call Registry are insufficient on their own. We further discuss the mitigation implications of our study, including how campaign analysis can support ecosystem-level disruption and how conversational findings can improve emerging end-user protections such as call screening and in-call detection.

## 2 Terminology & Related Work

**Terminology.** We use *potentially unwanted calls (PUCs)* as an umbrella term for any call generally undesired by the called party. Three sub-types are particularly relevant to this work:
*Robocalls* are automated calls that play pre-recorded audio [39], typically asking the callee to press a key or call back. They are often the entry point of scam or telemarketing campaigns, though they may also be legitimate (e.g., public-service announcements).
*Scam calls* are a form of social engineering in which malicious actors elicit sensitive or financial information by impersonating trusted entities and manufacturing urgency or isolation [4, 49].

*Telemarketing calls* promote products or services over the phone [27, 44]. They are often legal but may employ misleading or aggressive tactics, or violate regulations such as The National Do Not Call Registry [21, 25, 55]. The line between telemarketing and scam blurs when scammers pose as telemarketers to push fictitious goods [13].

**Related Work.** Prior work studies PUCs through honeypot data collection, content analysis, and AI-based countermeasures. Due to the nature of the approach, the call data analyzed in prior work either lacks interaction, is limited to metadata or scambaiter footage, or is evaluated in controlled settings. How PUCs unfold as real-time, interactive conversations at scale remains largely unexamined.

Honeypots lure malicious actors into a controlled environment so researchers can observe their behavior [41]. Phoneypot [27] and Mobipot [5] pioneered telephony honeypots but focused on metadata rather than call content. Prasad et al. [39, 40] shifted attention to content by greeting each incoming call with "Hello" and recording the response for 60 seconds. This approach captures *call-to-action robocalls* but cannot sustain a conversation, which leaves how scams actually operate in interactive settings unobserved. Other studies obtain interactive data by having human pose as victims and manually engage with scammers [2, 33], but these approaches require substantial effort, limiting their scalability.

Lacking real conversational data, prior content studies rely on publicly posted scambaiter videos. Sahin et al. [45] examined 200 interactions with the Lenny chatbot and identified scammer strategies such as urgency and scripted dialogue, though Lenny's fixed responses constrained how far scammers progressed. Wood et al. [54] applied a Hidden Markov Model to 90+ hours of scambaiter audio to predict scam stages. Both analyses are shaped by scambaiter behavior, which alters the natural progression of conversations and limits scalability.

AI systems have been explored for detection, automated handling, and simulation of unwanted calls. Pandit et al. [37] proposed *RoboHalt*, an NLP-based assistant that distinguishes humans from robocallers using questions easy for humans but hard for automated systems—an approach that does not generalize to human-driven scams. More recently, LLM-based systems have appeared on both sides of the interaction: Figueiredo et al. [24] use ChatGPT to play the scammer role (*Viking*), while Charnsethikul et al. [11] use it to play the victim (*Puppeteer*). Both rely on lab simulations with recruited participants without real-world callers.

## 3 Method

This paper aims to systematically characterize potentially unwanted calls (PUCs) by analyzing their conversational structures, attacker behaviors, and operational dynamics, with the goal of understanding their risks and informing effective defenses. We design a hybrid approach for the study, composed of three steps: data collection (§ 3.1) using our proposed system CallTrap, filtering (§ 3.2), and analysis (§ 3.3), as shown in Figure 1.

### 3.1 Data Collection

To capture real-world PUCs in the wild on a large scale and enable systematic analysis, we design CallTrap, an automatic system to interact with incoming calls. CallTrap contains three modules: *speech-to-text*, *response generation*, and *text-to-speech*, hosted on a *telephony system*. This design enables CallTrap to autonomously answer incoming calls, transcribe caller audio, generate contextually appropriate responses, synthesize natural-sounding speech, and interact with callers in real time. Each module uses existing machine learning services to handle caller–system interaction, prioritizing both real-time efficiency and human-like quality.

**Speech-To-Text (STT) Module.** This module transcribes incoming caller audio into text. Our primary goal is to enable low-latency, real-time transcription while maintaining sufficient accuracy for downstream text processing.

We evaluated several STT services based on latency and word error rate (WER) (see Table 1). We measured WER on the public CallHome English corpus [28]. Telephone disfluencies such as repetitions and filler words can increase WER without affecting semantic meaning. During manual review, we observed no STT errors that caused incorrect or off-topic downstream responses, and coders cross-checked transcripts against call audio to mitigate potential STT errors in the coding results.

Given the latency requirements of real-time interaction, we prioritize low latency for downstream response generation. We therefore selected Deepgram [12] for the initial deployment from July 2024 to August 2025. After August 2025, we switched to OpenAI's Realtime API [35] as part of the system update described below.

**Response Generation Module.** Provided with the transcribed caller input, LLMs generate a textual response fitting the conversational context. During initial system development in May 2024, ChatGPT was the leading provider among limited options, so we selected it for its fast response and its streaming option where output is provided while generation is ongoing. This enables earlier conversion of text to speech, thus reducing latency. When upgrading the system in August 2025, significantly more options were available. We evaluated LLaMA, Grok, Claude, and Gemini before settling on OpenAI's new Realtime API that combines speech-to-text and text-generating LLM into a single model. This proved significantly faster than using two dedicated models for STT and text generation, providing excellent latency. In addition, we reviewed CallTrap's responses during manual coding. We found that the responses remained relevant to the callers' utterances and consistent with the provided persona, synthetic personal information and conversation context. We did not observe off-topic or hallucinated responses in our manual analysis.

**Text-To-Speech (TTS) Module.** Finally, TTS converts the generated response back into speech. Similar to the STT module, we explored multiple TTS solutions (see Table 1) to find one with low latency and natural voices. Based on our evaluation, ElevenLabs [14] and Deepgram [12] offered the best trade-offs. Deepgram generated speech faster but relied on a limited set of prebuilt voices, whereas ElevenLabs provided more natural and customizable voices across perceived gender and age. From July 2024 to August 2025, we selected Deepgram TTS to prioritize low latency. After the system upgrade in 2025 reduced latency via OpenAI's Realtime API, we switched to ElevenLabs to improve voice naturalness and diversify voices across feminine-sounding and masculine-sounding voices as well as three age ranges (young, middle-aged, and senior). We discuss the effects of different voices on calls in Section 6.2.

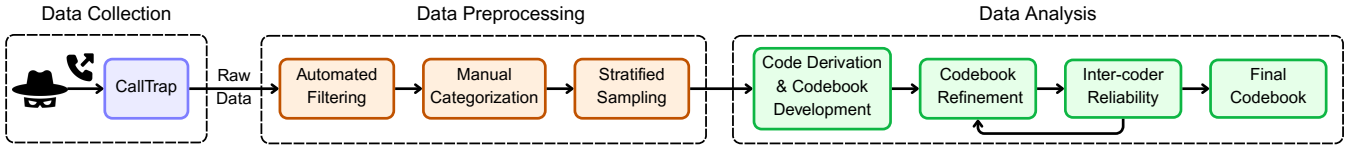



**Figure 1: Study Procedure for Understanding Potentially Unwanted Calls (PUCs)**

**Table 1: Evaluation of candidate components for each module. Experiments were conducted in August 2025. (* indicates models used from July 2024 to August 2025. ** indicates models since August 2025.)**

| | Model | Latency (s) | WER (%) |
|---|---|---|---|
| **STT** | **Deepgram Nova-3*** | 1.16 | 54.34 |
| | AssemblyAI Streaming | 1.20 | 58.55 |
| | Google Telephony | 1.28 | 69.64 |
| | SpeechMatics Realtime | 4.50 | 53.73 |
| | **Model** | **Latency (s)** | **Context Window** |
| **LLM** | Grok-3 | 0.40 | 131k |
| | Llama-3.3-70b | 0.42 | 128k |
| | **ChatGPT-4o*** | 0.64 | 128k |
| | Claude-Sonnet-4 | 1.56 | 200k |
| | Gemini-2.5-Flash | 1.97 | 1M |
| **STT & LLM** | **Model** | **Latency (s)** | **Context Window** |
| | **Realtime API**** | 0.71 | 128K |
| | **Model** | **Latency (s)** | **Customization Level** |
| **TTS** | **Deepgram-Aura*** | 0.27 | Prebuilt |
| | **ElevenLabs-Flash-v2**** | 0.31 | Customized |
| | Google-Chirp3 | 0.48 | Prebuilt |
| | Fishaudio-Speech-1.6 | 0.69 | Customized |
| | Minimax-Speech-02-Turbo | 1.19 | Customized |
| | PlayHT | 1.19 | Prebuilt |

**Telephony System.** To receive calls, we implemented a telephony system using Twilio [50]. This Voice-over-IP (VoIP) telephony platform allows us to integrate call routing, audio streaming, recording, and real-time interaction powered by our three modules. Initially in July 2024, we provisioned 12 phone numbers through this system. We later incorporated one donated phone number whose owner had been heavily affected by unwanted calls and voluntarily provided the number for research purposes. In September 2025, we added 16 phone numbers from four major carriers, AT&T, Verizon, Mint Mobile, and T-Mobile, to study whether the National Do Not Call (DNC) Registry is respected (discussed in Section 6.2).

Overall, CallTrap operated with 29 phone numbers: 12 Twilio numbers, one donated number, and 16 carrier numbers. As ordinary end users, we cannot determine the prior exposure of these numbers to PUCs because their prior contact histories are maintained by service providers and are not available to customers. We therefore diversified the number sources across multiple service providers and retained all numbers regardless of their acquisition channel. This approach better approximates the experience of a typical end user obtaining a phone number from a common service provider.

Data collection has been ongoing since late July 2024, and for the purpose of this paper, we analyze data collected through mid-February 2026 as a single continuous dataset. Throughout the data collection period, we did not seed or publicize any of the numbers through websites, advertisements, online forms, or other channels to attract calls.

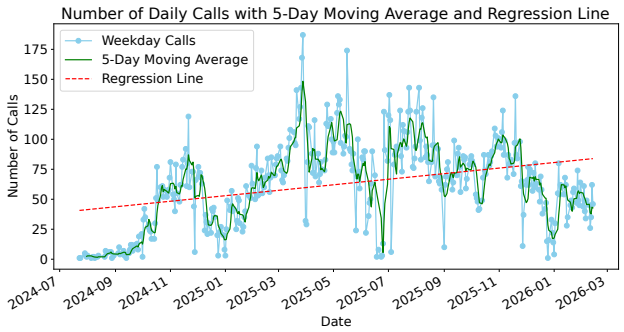



**Figure 2: Daily incoming calls (blue), 5-day moving average (green), and linear regression trend line (red). The regression slope (0.08) indicates a gradual increase over time.**

**Conversation Management.** Sustaining caller engagement requires CallTrap to maintain natural and coherent interactions, particularly because it relies on AI for real-time speech understanding and response generation. We address this challenge through the following design components.

*Cooperative Persona.* We intentionally designed CallTrap with a cooperative persona to sustain engagement with unwanted callers. We did not evaluate alternative personas, such as resistant or uncooperative ones. Because these alternative personas would likely produce shorter interactions and fewer analyzable conversations, collecting a dataset of comparable scale would require substantially more time and resources. Accordingly, the persuasion strategies (Section 5) and escalation patterns (Section 4.2.3) observed in our study were elicited under this cooperative persona and may not generalize to interactions with other personas.

*Synthetic PII.* CallTrap never uses or discloses real personally identifiable information (PII), instead, it randomly generates fictitious PII, which is provided to ChatGPT as part of the prompt. The complete prompt consisting of the *cooperative persona* and *synthetic PII* is provided in Supplementary Material. We also discuss more details about safeguards in Ethics.

## 3.2 Data Preprocessing

In the following, we describe the pipeline used to filter, categorize, and sample calls prior to data analysis.

**Raw Dataset Overview.** We briefly describe the raw dataset to provide additional context on our data collected beyond Figure 2.

Overall, the data shows clear temporal variation in call volume. In both 2024 and 2025, call volume increases from mid-year to just before the holiday season, drops sharply during the holiday period, and rises again afterward. These recurring patterns suggest seasonal effects. To quantify the overall trend, we fit a linear regression model to the daily call volume. The model yields a slope of 0.08, indicating that the system received, on average, 0.08 additional calls per day over the data collection period. The regression also has a $p$-value below 0.001, suggesting that this increasing trend is unlikely to be explained by random variation alone.

**Automated Filtering.** We now discuss how to filter out calls that do not contain meaningful conversational content. Among the 22,697 answered calls, we identify two major types of non-conversational calls: *silent calls* and *hello calls*. *Silent calls* contain no caller audio, even after attempting to trigger a response through greetings. *Hello calls* contain only greeting exchanges and do not progress into a substantive conversation. We filter these calls using conservative, deterministic criteria, including missing caller transcripts and fixed greeting patterns, such as "hello", to avoid excluding study relevant calls.

In total, we identify 11,839 silent calls and 1,896 hello calls. After removing these non-conversational calls, we obtain 8,962 calls with conversational content, totaling 281.8 hours.

**Manual Categorization.** While automated filtering removes non-conversational calls, the remaining conversational calls still vary. To understand caller intent and distinguish different types of PUC activity, we perform manual categorization based on the caller's apparent objective and the evidence present in the conversation. We define four high-level categories: *telemarketing*, *vishing*, *nuisance*, and *others*. Supplementary Material details labeling criteria.

*Telemarketing.* Calls that promote goods, services, or enrollment opportunities. Our definition is grounded in regulatory frameworks such as the Telemarketing Sales Rule (TSR).[1] TSR distinguishes *deceptive* and *abusive* telemarketing based on rule violations. For our study, we operationalize these categories using observable features (e.g., misrepresentation of identity or violations of calling rules) derived from TSR and FTC compliance guidance [21] to further categorize telemarketing calls into *deceptive* and *abusive*. Calls that do not exhibit these features are labeled as *legitimate telemarketing*.

*Vishing.* Calls that clearly impersonate trusted entities (e.g., government agencies or well-known companies) to obtain information, induce action, or create a false sense of urgency. Following prior work [7, 8], we distinguish impersonation-based calls (vishing) from promotion-based calls (telemarketing).

*Nuisance.* Calls do not exhibit clear malicious intent or sufficient evidence for classification (e.g., short or incomplete calls).

*Others.* Calls that are unrelated to PUC activity (e.g., misdials or irrelevant conversations).

**Sampling for Data Analysis.** Given the scale of 281.8 hours of captured conversation, we sample a subset for detailed thematic analysis. More specifically, we randomly sample 40% of the conversational calls, resulting in 3,630 calls and 115 hours of audio for manual analysis. To ensure that the sampled data remains representative of the original dataset, we divide all calls into six duration buckets, ranging from less than one minute to more than 20 minutes. We then apply *proportionate stratified random sampling* within these buckets, so that the sampled dataset maintains the duration distribution of the original conversational dataset. As shown in Supplementary Material, the sampled dataset closely follows the overall duration distribution of the original conversational dataset, while substantially reducing the amount of data that requires manual coding.

*Overview of Call Types.* Overall, we identify 1,977 *telemarketing* calls throughout the study period. Among them, only 33 were classified as *legitimate*, 1,772 as *deceptive*, and 172 as *abusive*. We also identify 485 *vishing* calls. For *nuisance* calls, we identified 1,024 calls, including 520 *short* calls. Most of the 144 calls in *others* category were misdials or conversations unrelated to PUC activity.

*Analysis Scope.* For the rest of the paper, we focus on calls posing security and privacy threats, that is *deceptive telemarketing*, *abusive telemarketing*, and all *vishing* calls, resulting in 2,429 calls.

## 3.3 Data Analysis

We analyze the sampled dataset using qualitative methods. Specifically, we apply thematic analysis that combines both inductive and deductive approaches. Our analysis follows a three-stage process.

**Initial Code Derivation and Codebook Development.** The lead researcher first derived an initial set of codes by reviewing prior work on unwanted calls, including scam calls [33, 45], robocalls [39, 40], social engineering [54], and user-reported experiences on Reddit [9], as well as FTC [21, 22] and FCC regulations [17]. These sources informed the initial deductive codes such as persuasion techniques. Codes describing conversational structures and requested information had to be developed inductively, as these aspects have been less explored due to the lack of real-world datasets. To do so, the lead researcher reviewed 1,000 transcripts (approx. 23 hours) randomly selected from the sampled dataset to understand conversational structures, flows, and tactical patterns. During this process, the researcher created the initial codes, identified code definitions, and refined them accordingly, introducing new open codes as needed. This process revealed dataset-specific features, including accent cues, background noise characteristics, and the progression of sensitive information requests, which were incorporated into the preliminary codebook.

**Codebook Refinement.** After developing the preliminary codebook, two additional coders applied it to 60 transcripts (10 hours) sampled across different duration buckets. Coders met regularly to discuss disagreements, resolve discrepancies, and refine code definitions. Following this process, we organized the final codebook to align with our study goals and capture aspects of PUCs that have received limited attention in prior work. These include how PUCs unfold in real-world settings, including high-level structures, role transitions, and identity disclosure. Beyond call structure, we also focused on caller behaviors, including requested information and persuasion techniques. Full codebook details are provided in Supplementary Material.

**Inter-coder Reliability.** After finalizing the codebook, we assessed inter-coder reliability on a subset of the sampled dataset.

[1] "A plan, program, or campaign which is conducted to induce the purchase of goods or services or a charitable contribution, by use of one or more telephones and which involves more than one interstate telephone call" [22].

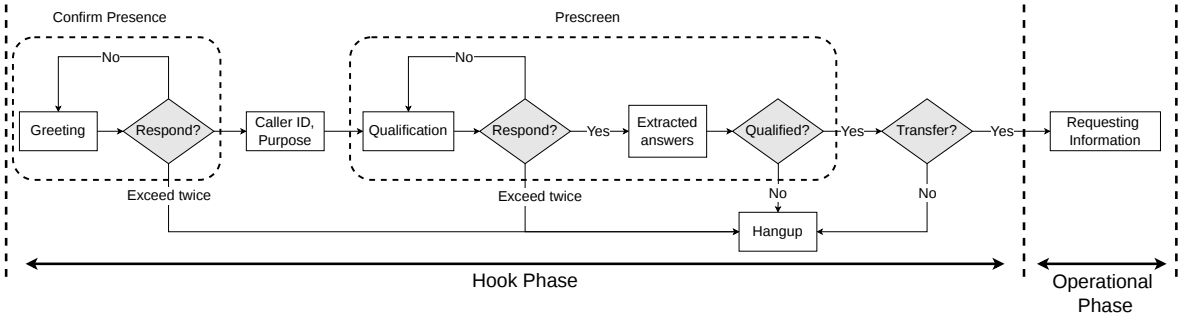



**Figure 3: The synthesized structure of collected PUCs.**

We randomly selected 337 new transcripts (10 hours), following the temporal distribution of the dataset, and each transcript was independently coded by two coders.

We evaluated agreement using Cohen's Kappa ($\kappa$) [32]. The average $\kappa$ scores across coder pairs are 0.96, 0.96, and 0.97, indicating substantial agreement [34].

After establishing reliability, we applied the finalized codebook to the full sampled dataset. Given the size of the dataset, transcripts were distributed across coders for independent annotation. Coders continued to meet regularly to resolve uncertainties and ensure consistency in code application. Once coding was complete, the lead author analyzed the annotated dataset to identify recurring patterns, behaviors, and characteristics of PUC interactions, which are presented in the results section.

## 4 PUC Conversation Structure

In this section, we analyze the conversational structure of PUCs to understand how these interactions unfold in practice. Figure 3 shows the synthesized structure derived from our data. Despite variation in topics, caller identities, and execution strategies, PUCs exhibit a consistent high-level pattern, which is a two-phase process. In the first phase (*Hook Phase*), callers establish contact, engage the callee, and assess whether the callee is a viable target. If successful, the interaction proceeds to a second phase (*Operational Phase*), where callers pursue operational objectives such as eliciting information, inducing actions, or transferring the callee.

### 4.1 Hook Phase

We focus on three aspects regarding the hook phase:

- *What* the hook phase consists of: The canonical conversational structure.
- *Who* executes the hook phase: The entities responsible for initiating and conducting this stage.
- *How* different entities realize the hook phase: The strategies used to implement each step of the structure.

*4.1.1 Canonical Hook Phase Structure.* In our study, we observe that 2,378/2,429 (97.90%) follow a consistent sequence across calls in the hook phase, as shown in Figure 3. This structure appears across different call topics and execution entities. It is a stable workflow for engaging and prescreening callees. Each step serves a distinct role in the hook phase as we state below:

**Step Greeting** establishes the presence of the callee and confirms that the call has reached an active and responsive target. Callers often use repeated prompts such as *"Hello? Can you hear me?"* to elicit a response. Repeated greetings are commonly used before the caller proceeds to disclose identity or purpose.

**Step Identification** introduces the caller's name, organizational affiliation, or both. In some cases, callers explicitly provide both pieces of information at the beginning of the interaction, e.g., *"Hi, this is Susan. I'm a Medicare advisor on a recorded line."* On the other hand, depending on the type of callers, some provide incomplete information. Extent of identity disclosure for different caller types is discussed in Section 4.1.3.

**Step Purpose** frames the interaction by stating the reason for the call, which typically links the conversation to a specific topic such as Medicare, insurance, or financial services. For example, *"The reason for the call is that there are some new state-approved 2026 Medicare benefits with a zero dollar premium that might be available to you."* This step provides the context that guides subsequent interaction. In our dataset, we identified 58 distinct topics, eight of which appeared in more than 50 calls. Details of these topics are shown in Supplementary Material.

**Step Qualification** assesses whether the callee matches a desired target profile through a series of screening questions. These questions are often topic-dependent and may include demographic or eligibility-related inquiries. For example, following the above *medicare* example, a common qualification question is *"So do you have Medicare part A and B?"*

In practice, qualification does not always function as a strict filter. Callers may continue the interaction even when responses do not match the expected profile, which may suggest that qualification also serves to maintain engagement. For example, in an *accident claim* call, the caller claimed that the callee had been involved in a car accident within the last two years and asked for confirmation. When the callee responded, *"I don't really remember any accident. Can you give me more details?"*, the caller did not terminate the call. Instead, the caller immediately attempted to initiate transfer by saying *"Let me bring a claim investigator on the line. He will give you more information on how you can get this compensation money."* This example illustrates that qualification questions may be used less to verify eligibility than to create a plausible transition to the next stage of the call.

**Table 2: Statistics of calls across conversation stages.**

| Entity | Calls (Total #) | Greeting (Mean #) | Identification | | | | Purpose (Top3 Topics) | Qualification (Mean #) | Transfer (Total #) |
|---|---|---|---|---|---|---|---|---|---|
| | | | Name-only | Org-only | Both | None | | | |
| Action | 105 | - | 9 (8.57%) | 43 (40.95%) | 41(39.05%) | 12(11.43%) | Online shopping<br>Tax relief<br>Auto finance | 0.18 | 14 (13.33%) |
| Interaction | 1,370 | 1.18 | 584 (42.63%) | 26 (1.90%) | 730 (53.28%) | 30 (2.19%) | Medicare<br>Final expense<br>Subsidy | 1.27 | 802 (58.54%) |
| Human | 954 | 0.94 | 307 (32.18%) | 38 (3.98%) | 546 (57.23%) | 63 (6.60%) | Final expense<br>Health insurance<br>Medicare | 2.05 | 314 (32.91%) |

**Step Transfer** advances the call to a subsequent stage or agent after initial engagement or qualification. *call-to-action robocalls* and *call-to-interaction robocalls* are expected to transfer most substantive calls because they cannot handle later-stage interaction themselves. Sometimes due to caller-side backend failures, unavailable agents, or abrupt dropped connections, transfer may fail, which often generates a robotic message, e.g., *"The number you have dialed is not in service. Please check your number and dial again."* Compared with robocalls, human agents transfer calls much less, as they can continue handling complex interactions directly.

*4.1.2 Execution Entities.* We identify three types of callers that execute the hook phase. Besides traditionally known human agents and *call-to-action robocalls* identified in prior work [40], we discover a new type of caller, which we name as *call-to-interaction robocalls*. Compared to *call-to-action robocalls*, which rely on prerecorded messages and primarily instruct the callee to perform an action (e.g., pressing a button or calling back) without conversational interaction, *call-to-interaction robocalls* support limited back-and-forth interaction by asking simple questions and processing responses.

We discovered that *call-to-interaction robocalls* and human agents occurred more often than *call-to-action robocalls* (Table 2). This indicates that interactive engagement has become more prevalent than static, instruction-based approaches in contemporary PUCs.

We further observe that the use of execution entities varies across call topics. As shown in Table 2, topics such as Medicare and final expense insurance are frequently associated with *call-to-interaction robocalls* and human agents, where callers often prescreen callees before advancing the call. In contrast, *call-to-action robocalls* appear in a narrower set of topics and are typically used in calls that require no conversational prescreening, such as *tax relief* calls claiming unresolved tax debt or *online shopping* calls claiming that a high-value order has been placed. This suggests that execution entities are tied to the underlying objectives of the call, with more complex or prescreening-required topics favoring interactive engagement.

*4.1.3 Entity Tactics.* While the hook phase follows a consistent structure, different entities employ distinct tactics to realize each operation step.

**Greeting.** *Call-to-action robocalls* typically deliver a fixed, one-way opening without waiting for a response. In contrast, *call-to-interaction robocalls* actively confirm the callee's presence through repeated prompts such as *"Hello, can you hear me?"* or *"Hello, are you there?"* Human agents adopt more natural conversational openings and often jump into later stages with fewer rounds of interactions than *call-to-interaction robocalls* (Table 2, column *Greeting*).

**Identification.** We observe that callers differ substantially in the way of disclosing their identity. Also, *call-to-action robocalls* disclose organization names more frequently than the other two entities: combined with *Org-only* and *Both*, 80.00% of *call-to-action robocalls* disclosed at least an organization name, compared with 55.18% of *call-to-interaction robocalls* and 61.22% of human agents (Table 2, column *Identification*).

However, disclosure of business name alone is not a reliable indicator of legitimacy, as callers may provide imprecise, generic, or unverifiable business names. We further examined the legitimacy of disclosed business names by searching for external evidence, including official websites and impersonation disclaimers, BBB profiles [6], and customer complaints on platforms such as Google Reviews. Among 443 disclosed names, we found 21 impersonated government agencies, such as *ACA*, *Marketplace*, and *US Medicare Department*. In the remaining 422 names, we found 364 (86.26%) exhibited malicious characteristics. These included direct impersonation of well-known companies, such as shopping websites and mobile carriers; generic and official sounding names, such as *US Final Expenses* and *Senior Center*; and resembled existing businesses by typosquatting, such as *Medical Alert Systems* impersonating *medicAlert*, *Senior Financial Services* impersonating *Senior Financial solutions*. Overall, 385/443 (86.91%) suspicious business names appeared fabricated or difficult to verify.

We infer that different entities rely on different strategies to establish legitimacy. While *call-to-action robocalls* establish legitimacy through the precise and well-known business (e.g., Amazon, AT&T), *call-to-interaction robocalls* and human agents may build trust based on conversational interaction. We discuss more details about authority and legitimacy in § 5.

**Qualification.** *Call-to-action robocalls* rarely perform meaningful screening, while *call-to-interaction robocalls* and human agents ask qualification questions to assess the callee. Considering human agents can handle much more complex situations, the average number of qualification questions asked by human is also more than both robocalls (Table 2, column *Qualification*). In practice, qualification does not always function as a strict filter. Interactions often continue even when responses do not match the expected profile, suggesting that qualification also serves to maintain engagement. For example, in an *accident claim* call, the *call-to-interaction robo* asked whether the callee had been involved in a car accident

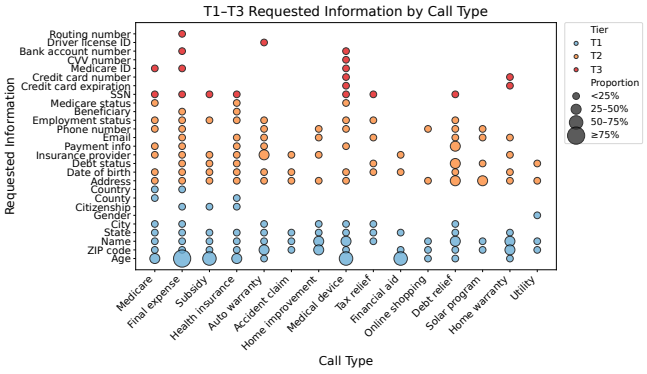



**Figure 4: Requested information across topics (within-tier ordering does not imply relative sensitivity).**

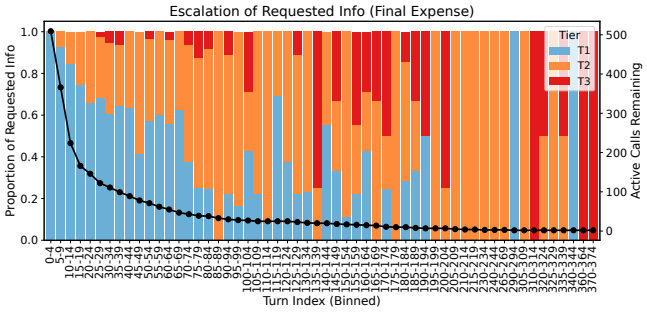



**Figure 5: Progression of requested information sensitivity over the course of conversations (*Final expense* calls as an example).**

and had received compensation. When the callee responded, *"Oh… well, I don't really remember any accident. Are you sure it's me you're talkin' about?"* The caller did not hang up but instead initialized a transfer by *"Let me bring a claim investigator on the line. He will give you more information on how you can get this compensation money."*

Overall, these findings show that while the hook phase is structurally consistent, its execution varies substantially across entities. Different entities implement the same workflow using distinct strategies, reflecting different approaches to engaging and prescreening callees before advancing to later stages of the call.

## 4.2 Operational Phase

After the hook phase establishes engagement and filters potential targets, calls proceed to the operational phase, where callers pursue their underlying objectives. In this phase, the interaction moves beyond initial contact toward concrete outcomes, such as eliciting sensitive information or requesting payment. We analyze the operational phase along three aspects: the objectives pursued, the types of information requested, and how these requests evolve over the course of the interaction.

*4.2.1 Operational Objectives.* The primary objective of the operational phase is to advance the caller's goal after a callee has been successfully engaged. Across our dataset, we observe that these

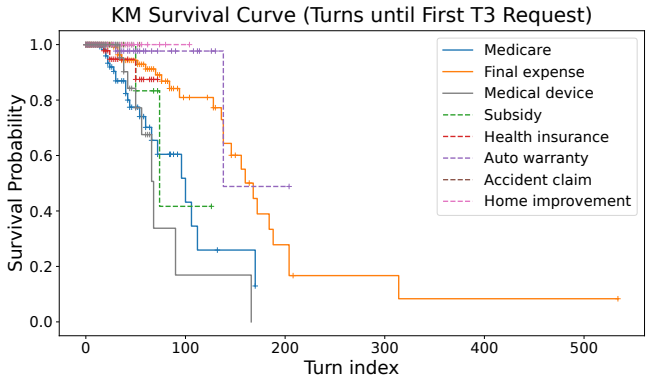



**Figure 6: Survival analysis of eight major call topics.**

objectives generally fall into two categories: information elicitation and action induction. Information elicitation involves collecting personal or contextual information from the callee, whereas action induction includes prompting the callee to take specific actions, such as enrolling in a service or interacting with another agent.

Importantly, the stated purpose introduced during the hook phase does not always align with the information requested in this phase. We observe cases where callers request information beyond what their stated purpose would reasonably require, suggesting that the initial framing can serve as a pretext for collecting more sensitive or exploitable information. For example, calls framed as offering free medical devices may request highly sensitive information such as *Medicare ID*, *Social Security numbers*, and *payment related details*, which are not strictly necessary for the stated service regarding the free medical device.

*4.2.2 Requested Information.* To understand what callers seek and their tactics during the operational phase, we analyze the types of information requested and categorize them by sensitivity. We use prior privacy regulations and guidance as reference, such as the CCPA [48] and Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) [31]. A key principle from this guidance is that the confidentiality impact of personally identifiable information depends on the potential harm caused by its disclosure. In particular, the guidance defines low, moderate, and high impact levels based on the consequences of disclosure of such information.

Following the privacy guidance, we group requested information per sensitivity into three tiers: low sensitivity (T1), moderate sensitivity (T2), and high sensitivity (T3). Specifically, T1 tier includes basic attributes such as name, age, and ZIP code, T2 tier includes more identifying details such as date of birth and full address, and T3 tier includes highly sensitive data such as Social Security numbers, payment credentials, and financial account information. The complete mapping is provided in Supplementary Material.

Figure 4 shows the distribution of requested information across different call topics. We observe that requested information varies substantially by topic, with some attributes (e.g., name and age) appearing broadly across calls, while others are concentrated in specific topics. In addition, although T1 information is less sensitive than T3 information, it should not be considered harmless as individuals are more likely to disclose their ZIP code than their

bank account. In practice, T1 information can be combined with other data to enable more severe attacks. For example, details such as pharmacy address, doctor name, or recent medical visits can support fraudulent billing, impersonation, or future targeting in Medicare-related scams. Therefore, protecting T1 information also becomes important.

*4.2.3 Information Escalation Tactics.* We next examine how information requests evolve over the course of a call. As discussed in Section 3.1, these escalation patterns reflect interactions with CallTrap's cooperative persona.

Across call types, we observe a consistent pattern of gradual escalation from lower-sensitivity to higher-sensitivity information, which reflects a deliberate tactic in which callers incrementally build engagement and trust before eliciting more sensitive data, as shown in Figure 5. In particular, lower-sensitivity information dominates early turns, while higher-sensitivity requests emerge later, indicating a staged approach to information elicitation.

To quantify differences across call types, we conduct survival analysis on the time until the first T3 request (Figure 6). The survival curves differ significantly across topics ($p < 0.001$), indicating that escalation behavior is not uniform. For example, final expense calls tend to reach T3 requests later than Medicare and medical device calls ($p < 0.001$), while Medicare and medical device calls do not differ significantly ($p = 0.56$). These results suggest that callers adapt escalation timing based on the context and objectives of calls.

Overall, under CallTrap's cooperative persona, the operational phase reveals a structured process in which callers progressively increase the sensitivity of requested information, rather than requesting highly sensitive data immediately. This gradual escalation allows callers to build engagement before attempting to obtain more valuable information.

## 5 Persuasion Techniques

Persuasion techniques are a core component of social engineering. Callers use them to establish authority and rapport. Prior work has examined persuasion techniques in social engineering and phone scams [54], but real-world phone-call datasets remain limited. As a result, less is known about how frequently these techniques appear in unwanted calls, where they occur in the conversation, and what conversational purposes they serve.

Our persuasion analysis mainly builds on Ferreira et al. [23], which combines three prior frameworks into five broad principles: *Authority* (AUTH), *Commitment, Reciprocation, and Consistency* (CRC), *Distraction* (DIS), *Liking, Similarity, and Deception* (LSD), and *Social Proof* (SP). To capture finer-grained persuasion particularly relevant to scam calls, such as immediacy and urgency, we complement this framework with three principles from Stajano and Wilson [47]: *Need and Greed* (NG), *Time* (TIME), and *Dishonesty* (DSH). Detailed definitions are provided in Supplementary Material.

As discussed in Section 3.1, these persuasion findings should be interpreted in the context of CallTrap's cooperative persona.

### 5.1 Overall Usage

Across all PT assignments, AUTH is the most common technique, accounting for 3,928 assignments (40.33%), followed by DIS with 2,713 (27.85%) and CRC with 2,562 (26.30%). LSD and SP occur much less frequently, appearing in 471 (4.84%) and 66 (0.68%) assignments, respectively. We incorporate three complementary persuasion techniques as fine-grained refinements under the scam call context. Specifically, 2,616 NG and 97 TIME are treated as refinements of DIS, and 43 DSH are treated as refinements of CRC.

AUTH appears primarily in the early stages of the call, especially during caller identification, purpose disclosure, and transfer. In these contexts, callers invoke official-sounding organizations, roles, or programs to establish legitimacy and reduce suspicion. For example, one caller stated: *"Hi, this is [Caller Name] with **American [Org Name]**. This call is about a new **state-regulated** final expense insurance plan, which covers 100% of your burial, funeral, or cremation expenses."* During transfer, callers may similarly invoke authority by saying that they will connect the callee to a *"**senior supervisor**," "**licensed agent**," "**product specialist**," or "**verification department**."*

CRC appears mainly during purpose disclosure and qualification. Its reciprocation component is reflected when callers present themselves as offering help, benefits, or compensation before asking simple qualification questions. For example, a caller stated: *"For a limited time, we are **offering a free device** with just a **small monitoring fee**. May I ask how old you are?"* Its commitment-and-consistency component appears when callers begin with low-friction questions, such as name or age, and later escalate toward more sensitive information, such as Medicare identifiers or SSNs, as shown in Figure 5. DSH captures instances in which scammers guide us to provide false answers during qualification. For example, after we disclosed that we had Medicare, the scammer instructed us to conceal this fact and promised the benefit: *"You have to simply **say no**. You have to be very confident in front of them … And once you are done with this process, you will be receiving the subsidy card, okay?"* Similarly, after we disclosed that we were retired, the caller instructed us to lie about our employment status and income: *"You can say that you are **self-employed** and your **annual income** should be $22,000 in a year."*

DIS captures how callers distract callees' attention from potential risks by emphasizing either what they can get or what they might lose soon unless they act immediately, corresponding to NG and TIME, respectively. NG is reflected in promises of unrealistic benefits, such as *"**free benefits**," "**extra support**," "**compensation**."* For example, one vishing call impersonating government claimed: *"Your number has been selected from government to get **some free money** which you don't have to pay back again, alright?"* TIME captures attempts to pressure callees to act quickly by threatening the loss of an offered benefit. For example, in a final expense call, after we expressed hesitation, the caller urged: *"That is actually with a discounted rate I'm not sure that I can go ahead and give you the same prices tomorrow or the day after tomorrow. So you need to decide in just **two minutes**."*

LSD is most often observed around the transfer stage. When a delay is expected, callers use light social conversation, such as asking about the weather or where the callee grew up, to avoid awkward silence and maintain engagement.

SP is the least frequent technique in our dataset. When it appears, it usually takes the form of broad validation claims, such as *"**most people**," "**in your area**," "**a lot of seniors**,"* suggesting social acceptance rather than concrete evidence. Examples include claims

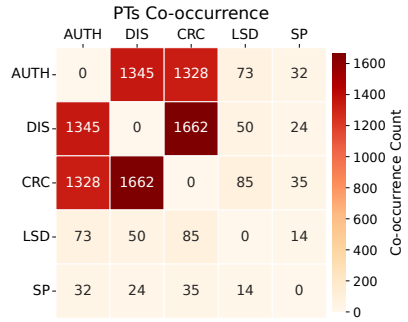



**Figure 7: Co-occurrence of PTs over caller turns.**

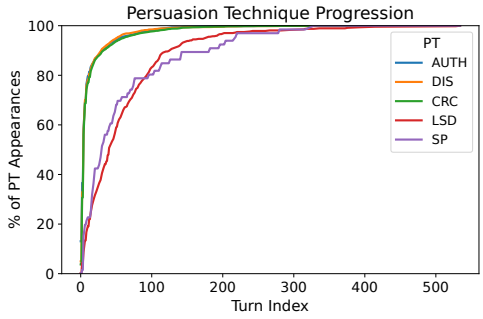



**Figure 8: Progression of persuasion techniques over calls.**

such as *"**a lot of seniors** are qualified," "helping more than **5,000 families** a year,"* or *"**most people** go with $20,000."*

## 5.2 Combination of Persuasion Techniques

Persuasion techniques often appear together within the same utterance. We show their co-occurrence in Figure 7. A common pattern combines AUTH, DIS, and CRC, especially during the *hook phase*. In such cases, callers typically invoke an official-sounding program or organization, frame the call around a potential benefit, and then ask a simple eligibility question.

The following example illustrates how AUTH, CRC, and DIS are combined in practice. In a vishing based government subsidy call, the caller stated: *"The call is about the **ACA program**, in which the **government** provides you **additional benefits**, like a **welcome package and rewards spending card**, to help pay for your daily expenses like rent, groceries, utilities, gas, and even dental or vision. Are you **receiving these benefits** right now?"* This utterance uses AUTH through references to the ACA program and the government, uses DIS by emphasizing attractive benefits, and uses CRC by ending with a low-friction eligibility question that keeps the conversation moving.

## 5.3 Persuasion Techniques Progression

We further examine how persuasion techniques are used during the course of calls. As shown in Figure 8, AUTH, DIS, and CRC are intensively used at the beginning: all three accumulate 75% of their occurrences within the first 10 turns and 90% by turn 34. This suggests that callers rely on these techniques early to establish legitimacy, present the call purpose, and keep the callee engaged.

By contrast, SP and LSD accumulate more gradually. SP reaches 75% around turn 74 and 90% around turn 182, while LSD reaches 75% around turn 80 and 90% around turn 126. This pattern is consistent with their supporting roles in reinforcing persuasion and maintaining rapport during longer interactions.

Overall, persuasion in PUCs is front-loaded, but different techniques serve different temporal roles: AUTH, CRC, and DIS are primarily opening-stage techniques to keep callees engaged, whereas SP and LSD tend to emerge later for supporting roles.

# 6 Mitigation Implications

Our analysis of PUC conversations reveals not only how individual calls unfold, but also how these calls are embedded within larger operational systems. The structured nature of the hook and operational phases and the comprehensive use of persuasion techniques suggest that PUCs are not isolated incidents, but organized and evolving operations.

These observations motivate the need to consider mitigation from *two complementary perspectives*. At the ecosystem level, defenses must account for how campaigns are organized, distributed, and sustained over time. At the end-user level, defenses must consider how individual users experience these calls, including how attackers engage, persuade, and target specific groups. In this section, we explore implications for both perspectives based on our empirical findings. We first analyze ecosystem-level characteristics of PUC operations, including campaign structure, longevity, and infrastructure. We then examine end-user-level factors, including targeting behavior and potential defenses on detecting PUCs in real time and mechanisms that prevent users from receiving PUCs.

## 6.1 Ecosystem-Level Mitigation

Ecosystem-level mitigations have the potential to protect users at scale. A first step towards implementing a mitigation on this level is the identification of *campaigns*. We follow prior work [5, 30, 36, 39] to define a campaign as a set of calls that share the same topic and similar scripts. This definition characterizes recurring call content rather than attributing calls to a common operator, as multiple operators may share the same script, while one operator may run multiple campaigns [39]. In the following, we outline how such campaigns can be identified in our dataset before we then discuss our observations and how they can inform mitigation design.

**Campaign Identification.** Via our rich conversational content analysis, we identify several signals for campaign clustering. We start by scoping the campaign analysis to topics with at least 50 calls. Within each topic, we first group calls that use the same normalized claimed organization name. For calls without an identifiable claimed organization, we rely on features extracted from the hook phase (discussed in Section 4.1), including *purpose utterances, opening entities, and qualification questions*, to identify content-based campaign clusters. Specifically, we encode the purpose utterances using a sentence-embedding model [42, 43] and calculate their cosine similarity. We represent each hook phase structure based on two features: which entity delivers the opening and which qualification questions are asked. We then calculate the Jaccard similarity

between these representations. We group calls whose purpose similarity is at least 0.85 and whose hook-structure similarity is at least 0.75. When the hook-structure similarity is unavailable or below 0.75, we require a stricter purpose-similarity threshold of 0.90. We show the pseudocode in Supplementary Material.

**Campaign Validation.** To ensure a robust basis for discussion of our results, we conduct a dedicated validation of identified campaign clusters [39]. To this end, we randomly sampled 20 calls[2] from each of the top three campaigns in each topic (shown in Table 3), yielding 376 calls across 24 campaigns. The primary coder then manually verifies whether sampled calls within a campaign indeed shared similar scripts. The decision is based on the core pitch, underlying offer, conversational strategies, and qualifications. Overall, 345 of 376 calls were consistent within the same campaign. The remaining 31 calls were classified as misclustered: nine calls followed different recurring scripts, 14 calls contained only generic topic-level language rather than a campaign-specific pattern, and eight calls contained insufficient conversational evidence as they were grouped by the claimed organization names. This corresponds to an accuracy of 91.8% (345/376), indicating that calls grouped into the evaluated clusters generally shared similar scripts and conversational patterns.

**Observations.** We now discuss the campaigns observed within our dataset. As Table 3 summarizes, we find that PUC activity is organized into recurring campaigns across multiple topics, including Medicare, final expense, and subsidy. The recurring structure and reuse of similar scripts indicate that PUCs are structured around persistent and organized activity rather than isolated calls [33, 39].

When taking a closer look at the involved calls, we discovered that 1,492 out of 1,751 (85.21%) of human-agent calls contain human-noisy backgrounds that imply call-center environments (shown in Table 4) [33]. We find that 91.39% of human agents exhibit non-American accents. Table 5 further shows that the most common caller profile corresponds to human agents with non-American accents and human-noisy backgrounds, while automated callers dominate early-stage interactions.

Looking at the bigger picture, we observe from Table 3 that only 3%–18% of campaigns account for 50% of calls across topics. For example, 5 out of 39 accident-claim campaigns and 14 out of 94 auto-warranty campaigns generate half of the calls. This suggests that targeting a small number of high-volume campaigns could significantly disrupt overall exposure to PUCs.

When considering the underlying motivations, we observe persistent economic incentives and structural support that can sustain PUC activity. In particular, we identify 24 calls where overseas call centers forward targets to U.S.-based agents or companies through three patterns: a three-layer structure where call centers transfer targets to U.S.-based quote companies and then legitimate insurance providers (Figure 9); direct transfers to U.S.-based agents; and callback scheduling for follow-up calls. These patterns suggest that some campaigns connect to legitimate U.S.-based businesses through multi-stage transfer chains.

[2]Or the maximum amount of calls if less than 20.

[3]`org` denotes campaigns grouped using a recurring claimed organization name, whereas `no_org` denotes content-based clusters without an identifiable claimed organization. We pseudonymize claimed organization names for de-identification purposes.

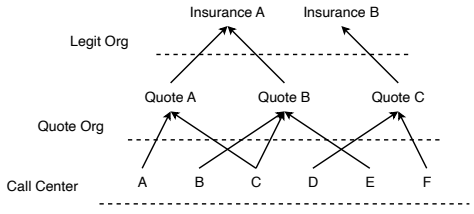



**Figure 9: Hierarchical business model and call-transfer structure.**

Our observations of such business models are also consistent with cases documented in public court and enforcement records, involving medicare, health insurance, medical device and final expense [51–53]. For example, federal court records describe a U.S.-based lead generation company [52, 53] that established and operated call centers in Pakistan, developed their scripts and trained their telemarketers, and contracted with a U.S. final expense insurance company to make illegal telemarketing calls to U.S. customers. By delivering final expense through live transfers (transfers when callees remained on the line), the lead generation company priced each live transfer at $55–$75 [53].

Therefore, we believe that parts of the PUC ecosystem are intertwined with business operations in the U.S., often through opaque, cross-border and multi-entity lead-generation pipelines.

**Mitigation of PUC Campaigns.** We now discuss how such recurring campaign activity could be disrupted at the ecosystem level. A key technique is telephone traceback to identify the source and routing path of an abusive call. Prior work found such tracebacks to be a time-consuming process requiring coordination across multiple carriers [40]. Even though recent research demonstrated the feasibility of conducting tracebacks within seconds [1], investigators must still determine which calls should be traced and prioritized. CallTrap can help prioritize such campaigns: As shown in Table 3, campaigns persist for extended periods. All major topics spanned over one year, 16 campaigns remained active for over a year, 12 of which continued into 2026. Such persistence provides ample opportunity to correlate recurring conversational activity and prioritize highly active campaigns for tracebacks and downstream investigation.

We note that effective campaign-level disruption requires collaboration among multiple ecosystem actors. CallTrap can only help to identify recurring conversational patterns, carriers and traceback organizations then need to combine this evidence with routing and provider records to identify call sources. Regulators can in turn use the combined evidence to investigate telemarketing and lead-generation practices.

## 6.2 End-User-Level Mitigation

As ecosystem-level mitigation remains insufficient, and eliminating individual campaigns does not fundamentally prevent PUC activity, it is critical to develop end-user-level mitigation to complement ecosystem-level defenses. In this subsection, we present

**Table 3: Campaign concentration and longevity across major topics. Representative examples are provided in Supplementary Material. The 50% column reports the minimum number of campaigns accounting for half of the calls in each topic.**

| Topic | Calls | Campaigns | 50% | First | Last | Span | Top Campaigns[3] | Calls | First | Last | Span |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Medicare | 716 | 177 | 6 | 08/12/24 | 02/12/26 | 550 | medicare_no_org_A | 188 | 09/30/24 | 11/20/25 | 417 |
| | | | | | | | medicare_org_A | 60 | 08/12/24 | 02/11/26 | 549 |
| | | | | | | | medicare_no_org_B | 53 | 02/25/25 | 01/29/26 | 339 |
| Final expense | 512 | 182 | 10 | 09/12/24 | 02/14/26 | 521 | final_expense_org_A | 80 | 10/02/24 | 02/02/26 | 489 |
| | | | | | | | final_expense_org_B | 52 | 09/12/24 | 12/17/25 | 462 |
| | | | | | | | final_expense_org_C | 27 | 11/25/24 | 01/30/26 | 432 |
| Subsidy | 263 | 110 | 7 | 07/29/24 | 02/07/26 | 559 | subsidy_no_org_A | 44 | 10/02/24 | 01/23/25 | 114 |
| | | | | | | | subsidy_no_org_B | 34 | 11/11/24 | 06/12/25 | 214 |
| | | | | | | | subsidy_no_org_C | 21 | 11/25/24 | 11/04/25 | 345 |
| Health insurance | 185 | 97 | 9 | 10/02/24 | 01/28/26 | 484 | health_insurance_org_A | 46 | 10/23/24 | 01/28/26 | 463 |
| | | | | | | | health_insurance_org_B | 23 | 10/28/24 | 01/07/26 | 437 |
| | | | | | | | health_insurance_org_C | 5 | 04/04/25 | 09/29/25 | 179 |
| Auto warranty | 162 | 94 | 14 | 08/22/24 | 02/02/26 | 530 | auto_warranty_org_A | 22 | 08/23/24 | 01/13/26 | 509 |
| | | | | | | | auto_warranty_org_B | 12 | 02/07/25 | 11/04/25 | 271 |
| | | | | | | | auto_warranty_org_C | 6 | 03/06/25 | 01/26/26 | 327 |
| Accident claim | 158 | 39 | 5 | 10/01/24 | 02/12/26 | 500 | accident_claim_org_A | 26 | 10/09/24 | 02/11/26 | 491 |
| | | | | | | | accident_claim_org_B | 19 | 11/25/24 | 02/12/26 | 445 |
| | | | | | | | accident_claim_org_C | 14 | 11/19/24 | 02/11/26 | 450 |
| Home improvement | 93 | 38 | 4 | 10/23/24 | 02/09/26 | 475 | home_improvement_org_A | 30 | 11/15/24 | 01/08/26 | 420 |
| | | | | | | | home_improvement_org_B | 8 | 03/08/25 | 01/14/26 | 313 |
| | | | | | | | home_improvement_org_C | 7 | 10/23/24 | 11/17/25 | 391 |
| Medical device | 65 | 33 | 6 | 09/10/24 | 02/13/26 | 522 | medical_device_org_A | 11 | 09/10/24 | 12/09/25 | 456 |
| | | | | | | | medical_device_org_B | 9 | 11/15/24 | 02/13/26 | 456 |
| | | | | | | | medical_device_org_C | 5 | 02/06/25 | 02/12/26 | 372 |

**Table 4: Call distribution over accent, noise, and entity.**

| Entity | Background Noise | | | Accent | |
|---|---|---|---|---|---|
| | Human-noisy | Noisy | Silent | American | Non-American |
| Human | 1,492 | 39 | 220 | 150 | 1,593 |
| Bot | 540 | 62 | 870 | 1,347 | 122 |

**Table 5: Top five caller profile ranked by frequency.**

| Caller Profile (Entity, Accent, Background Noise) | Count |
|---|---|
| Human, non-American, human-noisy | 1,417 |
| Bot, American, silent | 826 |
| Bot, American, human-noisy | 452 |
| Human, non-American, silent | 137 |
| Bot, non-American, human-noisy | 80 |

**Table 6: Telemarketing calls and top topics for DNC and non-DNC numbers.**

| Observation Period | DNC | | Non-DNC | |
|---|---|---|---|---|
| | # Calls | Top3 Topics (#) | # Calls | Top3 Topics (#) |
| 10/15/2025 | | Medicare (36) | | Auto finance (56) |
| - | 121 | Final expense (30) | 135 | Accident claim (9) |
| 02/15/2026 | | Medical device (13) | | Auto warranty (8) |

our findings on end-user-level mitigation from both prevention and real-time detection perspectives.

**Limitations of Current Protections.** The National Do Not Call (DNC) registry is an important protection intended to help users avoid unwanted telemarketing calls. To examine its effectiveness,

**Table 7: Call statistics for vocal gender and age profiles.**

| Attribute | Category | Mean↓ ($m's''$) | # Calls |
|---|---|---|---|
| Vocal Gender | Masculine | 2′29″ | 1,821 |
| | Feminine | 2′06″ | 1,795 |
| Age | Senior | 2′36″ | 1,031 |
| | Middle | 2′23″ | 1,255 |
| | Young | 1′58″ | 1,330 |
| Voice Profile | Masculine_Senior | 2′44″ | 590 |
| | Masculine_Middle | 2′38″ | 574 |
| | Feminine_Senior | 2′24″ | 441 |
| | Feminine_Middle | 2′10″ | 681 |
| | Masculine_Young | 2′06″ | 657 |
| | Feminine_Young | 1′50″ | 673 |

we added 16 phone numbers, four each from AT&T, Mint Mobile, Verizon, and T-Mobile, to CallTrap, randomly assigning half to the DNC group within each provider.

The DNC numbers were registered on September 14, 2025. After the required 31-day waiting period [18], we compared telemarketing calls from October 15, 2025 to February 15, 2026. As shown in Table 6, the DNC group received 121 calls versus 135 for the non-DNC group, showing no clear reduction in our deployment.

Despite the limited scale of our experiment, our results demonstrate a practical limitation of relying on DNC alone: users continued to receive telemarketing calls after registering with DNC, thus motivating complementary defenses during interaction.

**Targeted Prevention for Vulnerable Users.** Our analysis shows that vulnerability to PUCs is not uniform across users, which hints that targeted prevention for more vulnerable users can be effective

for PUC mitigation. To examine whether vocal gender and age were associated with call duration, we fitted a linear regression model with log-transformed call duration as the dependent variable and vocal gender and age as predictors, using feminine and young voices as reference categories. Masculine ($\beta = 0.076$, 95% CI [0.006, 0.146], $p = 0.032$), middle-aged ($\beta = 0.228$, 95% CI [0.149, 0.308], $p < 0.001$), and senior voices ($\beta = 0.178$, 95% CI [0.090, 0.265], $p < 0.001$) were associated with longer call durations. We then added gender-by-age interaction terms, but found no significant interactions. Descriptive statistics are shown in Table 7. Longer call duration suggests that certain user groups are more likely to remain engaged with callers, increasing their exposure to subsequent information requests.

These findings suggest that prevention mechanisms can be improved by accounting for user-specific risk. For example, systems could apply stronger filtering or warnings for users who are more likely to engage with PUCs. More broadly, these results indicate that PUC mitigation should move beyond uniform defenses and consider adaptive strategies that reflect differences in user vulnerability.

**Emerging Content-Based Protections.** Beyond call filtering, content-based protections are emerging on users' phones, where CallTrap can support these protections in several ways.

*Call Screening Before Answering.* Call screening asks unknown callers to identify themselves and state the purpose before users engage [3]. This process aligns with our analysis of the hook phase, where caller identification and purpose provide important information. By collecting and analyzing hook phase content at scale, CallTrap can provide real-world data for developing a risk scoring system that helps users better decide whether to answer a call.

*In-Call Detection During Engagement.* In-call detection monitors an ongoing interaction and warns users when suspicious behavior is detected. For example, Google's Scam Detection uses on-device LLMs to analyze calls in real time and warn users of potential scam patterns [26]. The complete conversations collected by CallTrap can support evaluation and improvement of such systems, including which early-stage signals should trigger a warning and how the detected risk and recommended actions should be communicated to users [10, 46].

*CallTrap's role.* CallTrap captures rich conversational information across interaction, including different entities conducting hook phase, information request escalation, persuasion techniques, and other audio information such as accent and background noise. Feature selection should build on concrete conversational signals, while audio features such as accent or background noise should only be used as supporting signals in a multimodal detection system. These data can support training and evaluation of new techniques.

In conclusion, these findings suggest that effective end-user mitigation requires a shift from static protections to adaptive, content-focused defenses. Prevention should incorporate user-specific risk factors, while detection should leverage early-stage conversational signals. Most importantly, defenses must operate early in the interaction to prevent escalation toward high-risk outcomes. By combining these approaches, we believe that end-user-level mitigation has the potential to complement ecosystem-level efforts and reduce the overall impact of PUCs.

## 7 Discussion

Our study shows that potentially unwanted calls (PUCs) are not merely discrete call incidents, but part of a big abuse ecosystem that combines automation, human agents, conversational manipulation, and supporting infrastructure. Unlike prior studies focusing on metadata, our findings highlight the importance of observing the conversation itself. The structure and progression of these calls reveal how callers engage targets, check eligibility, and gradually move toward harvesting sensitive information.

**Implication.** CallTrap has implications for both measurement and defense. First, conversational data provides visibility into behaviors that are difficult to infer from metadata alone, such as impersonation, qualification, persuasion, and escalation. Future measurement of PUCs should therefore move beyond call metadata and incorporate even longer conversational data. Second, the boundary between robocalls and human-operated calls is becoming less clear, as interactive systems can sustain realistic exchanges and perform early stage screening. This challenges the conventional view of robocalls as static prerecorded messages and suggests that defenses should account for increasingly interactive automated callers. Third, as LLM agents become more widely available, future PUC operations may become more adaptive and conversationally capable, including the possible use of LLMs in real-world vishing operations. This possibility highlights the need to proactively study and defend against more advanced attacks. Fourth, our conversational dataset can support the development of more conversation-focused defense mechanisms, such as detecting suspicious progression patterns, identifying staged information requests, and intervening before victims provide sensitive information. Fifth, our study focuses on systematically collecting real-world PUC conversations and identifying their common conversational structures. Building on these data and findings, future work can examine how callers conduct deceptive narratives and how these deceptions progress toward financial fraud or identity theft. Such analysis would support more targeted characterization and detection of deception-driven risks. In summary, these insights can inform future systems for detection, intervention, and policy evaluation.

**Limitations.** Our study aims to provide a large-scale, real-world view of PUC interactions through an automated conversational system, and our findings are shaped by how those calls are collected and interpreted.

On the collection side, CallTrap operates on a small set of phone numbers and includes only calls that were answered and produced some interaction. Because call targeting depends on a number's history and exposure, both the mix of incoming calls and how conversations unfold may differ from what a larger and more diverse pool would yield. In addition, all interactions are conducted by an automated system rather than a human: although designed to mimic natural conversation, synthesized responses and audio may still influence caller behavior, and our LLM is not fine-tuned on phone-scam conversations. Moreover, CallTrap uses a cooperative persona to sustain engagement and does not evaluate alternative personas, such as resistant ones. Therefore, the observed persuasion strategies and requested information progression may differ when callers interact with other personas. While the approach is scalable,

future work should expand to more numbers, fine-tune on real conversational data, and diversify the personas.

On the analysis side, our qualitative analysis is conducted on a stratified sample of the dataset to enable in-depth manual coding; while this preserves key characteristics such as call duration, the sample may miss rare or emerging behaviors present in the full corpus. Coding itself relies on human interpretation: despite high inter-coder reliability and an iteratively refined codebook, some ambiguity remains where intent is hard to infer, and our categories (e.g., telemarketing, vishing) are grounded in prior work and regulatory frameworks but real-world calls may combine tactics that do not fit neatly into these definitions. Our labels should therefore be read as structured approximations of caller behavior rather than exhaustive descriptions of all PUC tactics.

Finally, our dataset is primarily English-language and reflects a specific (U.S.) regulatory context, while PUCs and regulations vary across countries; our findings thus offer a partial view of the global PUC ecosystem. Future work can extend this approach by deploying multiple coordinated systems and incorporating multi-channel and multilingual data to better capture how these activities operate in practice.

**Replicability and Future Effectiveness of CallTrap.** By open-sourcing our framework, other researchers can reproduce or replicate our study using comparable systems and methodologies, although the collection outcomes may vary depending on the available phone numbers, country of operation, and time-based fluctuations in services and PUC ecosystems. We further discuss ethical considerations of open-sourcing our system in Ethics. At the same time, releasing CallTrap may make malicious actors aware of its presence, and they may attempt to circumvent or detect it. While this may reduce CallTrap's effectiveness, we believe the impact can be limited: personas and conversational strategies can be continuously modified and improvements in LLMs and TTS technology will further blur identifiable differences between human and automated conversation partners. While they may attempt prompt injection or similar techniques, exaggerated attempts will alert any human callees to unusual activity.

## 8 Conclusion

In this paper, we conduct a comprehensive study of PUCs to understand how these operations function in practice and to identify potential mitigation strategies. We develop an LLM-based telephony honeypot to collect real-world conversational data at scale. Our findings show that PUCs exhibit structured interaction patterns and persistent campaign activity. We further demonstrate that effective mitigation requires a combination of ecosystem-level and end-user-level defenses, and discuss how CallTrap can support the development and improvement of these defenses.

## Acknowledgments

We thank the anonymous reviewers for their feedback and suggestions. This work was partly supported by the U.S. National Science Foundation (NSF) under grants NSF-CICI-2613459, 2232915, and 2146568; Department of Navy award N00014-24-1-2193 issued by the Office of Naval Research; the Advanced Research Projects Agency for Health (ARPA-H) under contract SP4701-23-C-0074; and the Institute of Information & Communications Technology Planning & Evaluation (IITP) through grants RS-2024-004398199 and RS-2024-00442085; the Google Research Award: AI-Mediated Collaborative Software Understanding under grant G12404-300; and the generous support of the US Department of Defense. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research or ARPA-H.

## Open Science

To support reproducibility, we will release the CallTrap implementation (excluding deployment credentials), the annotation codebook, and the redacted coded transcripts used in our analysis. We will not release raw audio, unreviewed transcripts, or telephony logs because they may contain sensitive or identifying information, including voice characteristics that cannot be reliably anonymized. Detailed artifact documentation and access information are available at https://github.com/sefcom/calltrap-artifact.

## Ethics

Our institution's IRB approved this study after reviewing its operation, recording, data-handling, and risk-mitigation procedures. CallTrap operated passively: it placed no outbound calls, did not seed its numbers to attract callers, and interacted only with incoming calls. Because disclosure could alter callers' behavior and undermine the rigor of the study, we did not attempt to obtain callers' consent or debrief before or after calls. Since we used a fictional persona and fabricated personal information, callers who misdialed could recognize the mismatch and disconnect. In addition, we did not seek callers' real identities during conversations, and we also reviewed applicable recording requirements with institutional oversight. We also protected collected data, opted out of provider model training where applicable, and considered the dual-use risks of releasing CallTrap. A detailed ethical discussion is available at https://github.com/sefcom/calltrap-artifact.

## Supplementary Material

The full supplementary material includes the PUC classification criteria, ChatGPT prompt, annotation codebook, persuasion technique definitions, information sensitivity tiers, call topics, sampling distribution, campaign identification algorithm, AI disclosure, and representative conversation examples. These materials are available at https://github.com/sefcom/calltrap-artifact.

## References

[1] David Adei, Varun Madathil, Sathvik Prasad, Bradley Reaves, and Alessandra Scafuro. 2024. Jäger: Automated telephone call traceback. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*. 2042–2056.

[2] Sharad Agarwal, Emma Harvey, Enrico Mariconti, Guillermo Suarez-Tangil, and Marie Vasek. 2025. 'Hey mum, I dropped my phone down the toilet': Investigating Hi Mum and Dad {SMS} Scams in the United Kingdom. In *34th USENIX Security Symposium (USENIX Security 25)*. 4879–4896.

[3] Apple. 2026. Screen and Block Calls on iPhone. https://support.apple.com/guide/iphone/screen-and-block-calls-iphe4b3f7823/ios

[4] Farhanim Mohamad Asri and Tengku Elena Tengku Mahamad. 2023. Anatomy of phone scams: Victims' recall on the communication phrases used by Phone Scammers. In *International Conference on Communication and Media 2022 (i-COME 2022)*. Atlantis Press, 498–509.

[5] Marco Balduzzi, Payas Gupta, Lion Gu, Debin Gao, and Mustaque Ahamad. 2016. Mobipot: Understanding mobile telephony threats with honeycards. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*. 723–734.

[6] Better Business Bureau. 2025. BBB. https://www.bbb.org/

[7] Marzieh Bitaab, Haehyun Cho, Adam Oest, Zhuoer Lyu, Wei Wang, Jorij Abraham, Ruoyu Wang, Tiffany Bao, Yan Shoshitaishvili, and Adam Doupé. 2023. Beyond phish: Toward detecting fraudulent e-commerce websites at scale. In *2023 ieee symposium on security and privacy (sp)*. IEEE, 2566–2583.

[8] Marzieh Bitaab, Alireza Karimi, Zhuoer Lyu, Adam Oest, Dhruv Kuchhal, Muhammad Saad, Gail-Joon Ahn, Ruoyu Wang, Tiffany Bao, Yan Shoshitaishvili, and Adam Doupé. 2025. SCAMMAGNIFIER: Piercing the Veil of Fraudulent Shopping Website Campaigns. In *Proceedings of the 32nd Network and Distributed System Security Symposium (NDSS)*.

[9] Elijah Bouma-Sims, Mandy Lanyon, and Lorrie Faith Cranor. 2025. 'Is this a scam?': The Nature and Quality of Reddit Discussion about Scams. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security*. 2444–2458.

[10] Elijah Bouma-Sims, Enze Liu, Alexandra Xinran Li, and Lorrie Faith Cranor. 2026. From "Be Careful" to "Here's Why": Investigating User Reasoning with Context-Specific SMS Scam Warnings. In *2026 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2589–2608.

[11] Pithayuth Charnsethikul, Jelena Mirkovic, Rishit Saiya, Jeffrey Liu, Benjamin Crotty, and Genevieve Bartlett. 2025. Puppeteer: Leveraging a Large Language Model for Scambaiting.. In *HICSS*. 1–10.

[12] Deepgram. 2025. Deepgram. https://deepgram.com

[13] Jeffrey H Doocy, David Shichor, Dale K Sechrest, and Gilbert Geis. 2001. Telemarketing fraud: Who are the tricksters and what makes them trick? *Security Journal* 14 (2001), 7–26.

[14] ElevenLabs. 2025. ElevenLabs. https://elevenlabs.io

[15] Federal Communications Commission. 2020. FCC Mandates That Phone Companies Implement Caller Id Authentication To Combat Spoofed Robocalls. (2020). https://docs.fcc.gov/public/attachments/DOC-363399A1.pdf.

[16] Federal Communications Commission. 2021. STIR/SHAKEN Small Provider Report and Order. (2021). https://docs.fcc.gov/public/attachments/FCC-21-122A1.pdf.

[17] Federal Communications Commission. 2025. Stop Unwanted Robocalls and Texts. https://www.fcc.gov/consumers/guides/stop-unwanted-robocalls-and-texts.

[18] Federal Trade Commission. 2003. National Do Not Call Registry. (2003). https://www.donotcall.gov/

[19] Federal Trade Commission. 2023. Phone Scams. (2023). https://consumer.ftc.gov/articles/phone-scams.

[20] Federal Trade Commission. 2024. Consumer Sentinel Network. (2024). https://www.ftc.gov/system/files/ftc_gov/pdf/CSN-Annual-Data-Book-2023.pdf.

[21] Federal Trade Commission. 2025. Complying with the telemarketing sales rule. https://www.ftc.gov/business-guidance/resources/complying-telemarketing-sales-rule

[22] Federal Trade Commission. 2026. Telemarketing Sales Rule. https://www.ecfr.gov/current/title-16/chapter-I/subchapter-C/part-310.

[23] Ana Ferreira, Lynne Coventry, and Gabriele Lenzini. 2015. Principles of persuasion in social engineering and their use in phishing. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 36–47.

[24] João Figueiredo, Afonso Carvalho, Daniel Castro, Daniel Gonçalves, and Nuno Santos. 2024. On the Feasibility of Fully AI-automated Vishing Attacks. *arXiv preprint arXiv:2409.13793* (2024).

[25] Chester S Galloway and Steven P Brown. 2004. Annoying, intrusive,… and constitutional: telemarketing and the national Do-Not-Call Registry. *Journal of Consumer Marketing* 21, 1 (2004), 27–38.

[26] Google. 2026. Use Scam Detection. https://support.google.com/phoneapp/answer/15654065?hl=en

[27] Payas Gupta, Bharat Srinivasan, Vijay Balasubramaniyan, and Mustaque Ahamad. 2015. Phoneypot: Data-driven understanding of telephony threats.. In *NDSS*, Vol. 107. 108.

[28] Linguistic Data Consortium. 2008. CABank English CallHome Corpus. TalkBank. doi:10.21415/T5KP54

[29] Jienan Liu, Pooja Pun, Phani Vadrevu, and Roberto Perdisci. 2023. Understanding, measuring, and detecting modern technical support scams. In *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*. IEEE, 18–38.

[30] Aude Marzuoli, Hassan A Kingravi, David Dewey, and Robert Pienta. 2016. Uncovering the landscape of fraud and spam in the telephony channel. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 853–858.

[31] Erika McCallister. 2010. *Guide to protecting the confidentiality of personally identifiable information*. Diane Publishing.

[32] Mary L McHugh. 2012. Interrater Reliability: The kappa Statistic. *Biochemia medica* 22, 3 (2012), 276–282.

[33] Najmeh Miramirkhani, Oleksii Starov, and Nick Nikiforakis. 2016. Dial one for scam: A large-scale analysis of technical support scams. *arXiv preprint arXiv:1607.06891* (2016).

[34] Cliodhna O'Connor and Helene Joffe. 2020. Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods* 19 (2020).

[35] OpenAI. 2025. Introducing the Realtime API. https://openai.com/index/introducing-gpt-realtime/.

[36] Sharbani Pandit, Roberto Perdisci, Mustaque Ahamad, and Payas Gupta. 2018. Towards Measuring the Effectiveness of Telephony Blacklists.. In *NDSS*.

[37] Sharbani Pandit, Krishanu Sarker, Roberto Perdisci, Mustaque Ahamad, and Diyi Yang. 2023. Combating robocalls with phone virtual assistant mediated interaction. In *32nd USENIX Security Symposium (USENIX Security 23)*. 463–479.

[38] Katalin Parti and Faika Tahir. 2023. "If We Don't Listen to Them, We Make Them Lose More than Money:" Exploring Reasons for Underreporting and the Needs of Older Scam Victims. *Social Sciences* 12, 5 (2023), 264.

[39] Sathvik Prasad, Elijah Bouma-Sims, Athishay Kiran Mylappan, and Bradley Reaves. 2020. Who's calling? characterizing robocalls through audio and metadata analysis. In *29th USENIX Security Symposium (USENIX Security 20)*. 397–414.

[40] Sathvik Prasad, Trevor Dunlap, Alexander Ross, and Bradley Reaves. 2023. Diving into Robocall Content with {SnorCall}. In *32nd USENIX Security Symposium (USENIX Security 23)*. 427–444.

[41] Niels Provos et al. 2004. A Virtual Honeypot Framework.. In *USENIX Security Symposium*, Vol. 173. 1–14.

[42] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*. 3982–3992.

[43] Nils Reimers and Iryna Gurevych. 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*. 4512–4525.

[44] Merve Sahin, Aurélien Francillon, Payas Gupta, and Mustaque Ahamad. 2017. Sok: Fraud in telephony networks. In *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 235–250.

[45] Merve Sahin, Marc Relieu, and Aurélien Francillon. 2017. Using chatbots against voice spam: Analyzing {Lenny's} effectiveness. In *Thirteenth symposium on usable privacy and security (SOUPS 2017)*. 319–337.

[46] Filipo Sharevski, Jennifer Vander Loop, Bill Evans, and Alexander Ponticello. 2025. (Blind) Users Really Do Heed Aural Telephone Scam Warnings. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE, 74–92.

[47] Frank Stajano and Paul Wilson. 2011. Understanding scam victims: seven principles for systems security. *Commun. ACM* 54, 3 (2011), 70–75.

[48] State of California Department of Justice. [n. d.]. California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa

[49] Judith L Tabron. 2016. Linguistic features of phone scams: A qualitative survey. In *11th Annual Symposium on Information Assurance (ASIA'16)*. 52–58.

[50] Twilio. 2025. Twilio. https://www.twilio.com

[51] U.S. Department of Justice. 2026. Former NFL Player Sentenced to Over 16 Years in Prison for $197M Medicare Fraud. https://www.justice.gov/opa/pr/former-nfl-player-sentenced-over-16-years-prison-197m-medicare-fraud

[52] U.S. District Court for the District of Massachusetts. 2025. Robert A. Doane v. Python Leads, LLC et al. https://www.govinfo.gov/content/pkg/USCOURTS-mad-1_24-cv-12947/pdf/USCOURTS-mad-1_24-cv-12947-0.pdf

[53] U.S. District Court for the Middle District of Florida. 2026. Final Expense Direct v. Python Leads, LLC et al. https://cases.justia.com/federal/district-courts/florida/flmdce/8%3A2023cv02093/418482/211/0.pdf

[54] Ian Wood, Michal Kepkowski, Leron Zinatullin, Travis Darnley, and Mohamed Ali Kaafar. 2023. An analysis of scam baiting calls: Identifying and extracting scam stages and scripts. *arXiv preprint arXiv:2307.01965* (2023).

[55] M Bayer Ya'akov. 2019. Older adults, aggressive marketing, and unethical behavior: A sure road to financial fraud? *Ethical Branding and Marketing* (2019), 1–18.