A Computational Analysis of Millions of Scam and Spam Robocalls Targeting Older Adults and the General Public

Ryan C. Moore
Assistant Professor
School of Information
The University of Texas at Austin
ryanmoore@utexas.edu

Doug Shadel
Managing Director
Fraud Prevention Strategies, LLC
doug@fraudpreventionstrategies.com

Abstract

Robocalls that use autodialers or prerecorded messages remain a dominant channel for spam (commercial solicitations) and scams (attempts to steal money or information), yet their psychological characteristics at scale remain poorly understood. We leveraged a honeypot (decoy phone lines to receive and transcribe unsolicited calls) of 295,749 lines to analyze 4.49 million transcripts over five years (2018–2022). A large language model, validated against fraud expert judgments, classified calls as scam, spam, or neither, and as targeting older adults or the general public. Transcripts were profiled using LIWC-22, a fraud tactics dictionary (promise of gain, threat of loss, time pressure, authority), and emotional arousal indices. Scam robocalls showed higher arousal, time pressure, threats of loss, and authority appeals than spam. Scams targeting older adults showed even higher arousal, time pressure, loss-framing, and health content – patterns linked to scam susceptibility in older adults – with implications for theory and intervention design.

Keywords: scams, spam, robocalls, older adults, emotional arousal, loss aversion, interventions, honeypot, large language model

Americans receive approximately four billion robocalls per month[1]. Robocalls are defined as calls made with an autodialer or that contain a prerecorded or artificial voice message[2]. According to the National Consumer Law Center, 68 million Americans lost $29 billion to robocall scams in 2022[3]. In fact, even with advances in other technologies used to disseminate spam and perpetrate scams (e.g., social media accounts, text messaging), robocalls continue to make up the majority of complaints received by the Federal Communications Commission (2025). Despite regulators' efforts to thwart robocalls, robocallers remain difficult to stop for several reasons. First, in the case of the United States, many, if not the majority of, robocallers operate outside U.S. borders[4], limiting the U.S.'s ability to enforce domestic regulations. Second, robocallers frequently change the phone numbers they use to avoid detection and blocking. They often employ a tactic known as neighbor spoofing, in which the robocaller "spoofs" their outgoing number to mimic the target's area code, increasing the likelihood that the call will be answered[5].

Older adults, who have greater financial resources on average than younger adults, are often deliberately targeted by fraudsters[6,7], and this dynamic is also observed among robocall scams. Older adults are more likely to receive robocalls[8] and report higher median losses to robocall scams than other age groups[9]. In fact, the scams that older adults report the highest median losses to are scams that begin over the phone[10]. The consequences of fraud victimization can be financially and psychologically devastating for older adults[6,11,12], who often have more to lose in the form of retirement and other assets.

While there has been research on robocalls, the existing literature faces two critical limitations. First, much of the research relies on self-reports of exposure to (e.g.,[13,14]) and victimization by (e.g.,[10,15]) robocalls. The reliance on self-reports is problematic due to (a) limitations in people's ability to recall and accurately report what media content they are exposed to and the content of that media[16] and (b) underreporting due to refusal to acknowledge victimization or lack of awareness that fraud occurred[15,17]. This limitation is important because it restrains our understanding of the breadth of robocalls targeting consumers, potentially distorting the recommendations or resources that are generated to protect them. Second, while older adults are more likely to be targeted by robocall scams and report greater losses[8,9], little existing work has detailed, objectively, how robocall scams that target older adults differ in their content from those targeting the general public. Objectively analyzing the landscape of robocall scams targeting older adults is particularly important because older adults are especially prone to underreporting fraud[12,18]. Given this, understanding what robocall scams older adults face without relying on self-reporting is critical for the development of resources to protect older consumers' financial and psychological well-being.

While robocall scams can be damaging and need to be better understood, robocalls are also frequently used to disseminate spam. For the purposes of this work, we define spam robocalls as calls attempting to sell a product or service where there is a reasonable chance the caller will actually deliver something of value. Scam robocalls, on the other hand, are defined as calls attempting to get money or personal information where there is little to no chance that a product or service will be delivered. Examples of spam robocalls include unsolicited marketing for business listing services, healthcare insurance products, and home improvement installations, while examples of scam robocalls include impersonated government agencies, fake service suspension threats, and fraudulent debt forgiveness programs.

In this paper, we overcome limitations of existing work on robocalls, namely not knowing the content of robocalls at scale and in the wild, by leveraging a large honeypot of 295,749 phone lines. This honeypot records incoming robocalls and transcribes them to text. The honeypot does not consist of phone lines of real users and there is no response from the honeypot phone lines aside from a verbal "hello" when answering the call. Thus, the honeypot call data represent robocall "pitches" rather than recorded interactions between robocallers and consumers. We analyzed approximately 4.5 million robocalls received in the honeypot from 2018-2022. We leveraged large language models (LLMs), validated against fraud expert judgments, to distinguish scam robocalls from spam robocalls, as well as to identify robocalls which targeted older adults compared to robocalls which target the general public. Our approach, novel in the study of robocalls, allows for analysis of key robocall types of interest (e.g., scams targeting older adults) at a scale not achieved in prior work. We then used a variety of tools for computational text analysis (including LIWC-22[19], a dictionary representing several fraud tactics [promise of gain, threat of loss, time pressure, authority], and an index of the emotional arousal of text[20]) to analyze the psychological dynamics and persuasion tactics of the robocall pitches.

Our results indicate that robocall scams targeting the public contain significantly greater levels of emotional arousal and use the tactics of time pressure and authority (i.e., naming of government entities) more often than robocall spam. These findings are significant because these features have been demonstrated to drive susceptibility to scams in past research[21–24]. Moreover, scams targeting older adults were even higher in arousal, more likely to frame scams as threats of loss (and less likely to frame scams as promises of gain), and instill time pressure, tactics that older adults in particular may be especially vulnerable to due to cognitive and psychological changes that tend to occur with age[22,25,26].

Our findings have implications for both researchers and practitioners. For researchers, we demonstrate that LLMs can be used to characterize dynamics in robocalls and robocall scams. Importantly, an LLM could reliably reproduce the judgments of a fraud expert using one of the most inexpensive models commercially available at the time (OpenAI's GPT-4o-mini) and via prompt engineering alone, without any custom models or fine-tuning. Our approach could be adapted to analyze existing databases of robocalls and may be able to be scaled to analyze scam/spam in other modalities (e.g., robotexts[27]). Our empirical results provide evidence that tactics prevalent in scams in other modalities (e.g., email) extend to robocalls. Rather than relying on anecdote, assumption, or extrapolation about the contents of robocalls, our work characterizes five years of naturalistic robocall pitches targeting consumers at scale.

Furthermore, our findings can focus future research on particular dynamics of robocalls that may induce susceptibility to victimization, especially among older adults.

For practitioners, our findings can inform the design of interventions to build individuals' resilience to robocall scams. We show the tactics interventions and consumer education can focus on in building the public's resilience to robocall scams. In particular, we highlight aspects of robocall scams specifically targeting older adults that can be used to tailor interventions for that population. Our focus on general psychological characteristics and tactics such as emotional arousal and time pressure allow educational materials to benefit users regardless of the specific types of scams targeting them or circulating at a given point in time (as has been the focus of consumer education efforts to date; e.g.,[13]).

Background and Research Questions

Some robocalls are automated calls for specific, legitimate purposes (e.g., pharmacy refill reminders), while others are unsolicited telemarketing spam calls trying to sell products or services. Still others are scam campaigns which attempt to rob consumers of money or personal information. While all these types of robocalls work to make up the approximately four billion robocalls Americans alone receive in a given month[1], a great deal of scholarly and practitioner interest in robocalls concentrates on scams.

As was discussed previously, most research on robocall scams relies on self-reports of exposure and victimization. Much of this work tends to focus on documenting the types of scams targeting consumers (e.g., sweepstakes, IRS impersonation, charity scams)[13,28]. One goal of this work is making consumers aware of the types of scams that may target them so that they're prepared and able to hang-up/disengage if these scams target them via phone. However, this work may not paint a complete portrait of the range of scam types targeting consumers and may instead largely reflect those which people are willing or likely to report to authorities or in surveys.

Another thread of research, related to but broader than robocall scams, examines the psychological dynamics and persuasion tactics that scammers use to influence victims across different types of scams. This work has identified characteristics of scams that make people more susceptible to them. For instance, emotional arousal increases susceptibility to fraud in older and younger adults, regardless of positive or negative valence[22]. A Better Business Bureau survey revealed that scams offering the promise of gain were more likely to elicit consumer responses and financial losses than scams threatening loss[29], although a laboratory study manipulating promise of gain and threat of loss message type and assessing responses found the opposite[23]. In online fraud (e.g., phishing emails), messages pressuring recipients to respond quickly were more likely to be responded to than messages which were less urgent[21,23,30]. Scammers also often attempt to appeal to authority, for example by impersonating government agencies or other official authorities and such signals credibility can make individuals more likely to respond[24,31].

While these studies about characteristics of scams that increase susceptibility are valuable in illuminating the psychology of victimization, they typically rely on artificial settings to assess susceptibility (e.g., participants respond to mock scam messages in a controlled laboratory setting[22]) or rely on existing consumer complaints about scams[29]. These studies cannot reveal the extent to which these characteristics which render consumers more susceptible actually appear in the messages targeting them and thus which factors are most important for consumers to watch out for. Furthermore, the studies include scams targeting consumers across a mix of modalities (e.g., mail, email, text message, phone calls) and thus cannot isolate the prevalence or impact of tactics in a specific modality such as robocalls.

A small body of work has used honeypots to study robocalls beyond self-reports, but this work has largely focused on metadata or call types rather than psychological content. Prasad et al. used honeypot data to characterize the phone numbers robocallers dial from[32] and to evaluate the effect of the FCC's STIR/SHAKEN call authentication policy on robocall volume[33].

Marzuoli et al.[34] used cluster analysis on 40,000 unsolicited calls to show that a small number of robocallers were responsible for the majority received. Closest to the present work, Prasad et al.[35] analyzed ~233,000 robocall transcripts to quantify the prevalence of social security, tech support, political, business listing, and financial robocalls. To our knowledge, however, no honeypot-based study has examined the psychological characteristics or tactics of robocalls that drive fraud susceptibility in the broader literature.

In the current study, we wanted to leverage a honeypot data collection approach to explore the psychological and tactical characteristics of real robocall scams that target consumer phone numbers. What characteristics and tactics differentiate robocall scams from robocall spam, which make up most of the robocalls Americans receive[28]? Do the characteristics of scam robocalls Americans receive reflect some of the previously studied features of scams in general (not necessarily robocalls specifically) which render people more susceptible to them (e.g., higher arousal, more time pressure)? Specifically, we ask RQ1:

> **RQ1**: How do the psychological dynamics, tactics utilized, and emotional arousal differ between spam and scam robocall pitches?

Fraudsters often deliberately target older adults in part because they tend to have greater levels of wealth than younger adults[6,7]. Indeed, in the context of robocalls, older adults are more likely to receive them[8] and report higher median losses to robocall scams than other age groups[9].

Aging is also marked by changes in the psychology of judgment and decision-making that may render older adults vulnerable to certain scam tactics. For instance, older adults are more likely than younger adults to behave in a way that minimizes losses when making decisions[25,36,37]. Thus, older adults may be especially vulnerable to scams framed as threats of loss compared to promises of gain. Time pressure has also been shown to inhibit decision-making in older adults[26,38], and older adults may be more vulnerable to scams which invoke time pressure.

Because older adults are more likely to receive robocalls *and* may be more vulnerable to specific scammer tactics, we also wanted to examine whether there were differences in the robocall scams that targeted older adults versus those that targeted the general public. Specifically, we ask RQ2:

> **RQ2**: How do psychological dynamics, tactics utilized, and emotional arousal differ between scam robocall pitches targeting older adults and scam robocall pitches targeting the general public?

Results

To illustrate the kinds of pitches our LLM-based approach distinguished, we describe representative examples drawn from our LLM prompts (see Methods). Spam pitches offered products or services with some commercial substance (e.g., home security installations, solar enrollments, extended auto warranties). Scam pitches, by contrast, often presented pretexts in which the recipient's funds, identity, or legal standing were ostensibly at risk, with no genuine product or service to be delivered. Examples of scam pretexts included Social Security Administration impersonations ("Your Social Security number has been suspended"), IRS or FBI warrant threats ("You are being listed as the primary suspect in a case being filed by I.R.S."), and fraudulent account-charge notifications ("You have been charged $1,279.99 on your default card saved on Amazon.com"). Within scam calls, examples treated as targeting older adults invoked Medicare, Social Security, or age-associated health conditions, often in combination — for example: "We have tried numerous times to contact you…regarding your eligibility for top of the line braces to alleviate your pain…this is your final notice if you do not act soon Medicare will label you unavailable for coverage." Examples of scams targeting the general public invoked consumer financial transactions or workforce-relevant programs (e.g., credit card interest rate reduction, student loan forgiveness, utility refunds, fabricated vehicle accident compensation).

RQ1: How do psychological dynamics, tactics utilized, and emotional arousal differ between spam and scam robocall pitches?

First, we used the computational text analysis program LIWC-22 to broadly profile the psychological dynamics of robocall scams (see Methods section for more details about LIWC measures). Specifically, we sought to identify which psychological dynamics were most diagnostic of a robocall scam relative to robocall spam. For this initial, broad examination of robocall scams, we took a data-driven approach. We compared robocall scams and spam across all 100+ LIWC linguistic variables. After computing the average score across all LIWC variables, we filtered down to variables where the average difference between scam and spam robocalls was statistically significant (i.e., $p < .05$) and then sorted the variables by the largest effect size of the average difference in terms of Cohen's d. The resulting list of 10 LIWC variables is presented in Table 1. Some of the most diagnostic linguistic indicators that a call was scam (as opposed to spam) according to LIWC were words about social behavior (e.g., love, say, care) ($d = 0.71$), including prosocial words (e.g., help, thank, please) ($d = 0.59$) and words about communication processes (e.g., said, tell) ($d = 0.57$) (all p's < .001).

[TABLE 1]

Next, we analyzed specific tactics used in robocall scams relative to spam to see whether tactics of scams in general are also likely to appear in robocall scams and could be used to distinguish robocalls scams from spam. We compared the average percentage of words and phrases in each transcript appearing in three tactics dictionaries: promise of gain, threat of loss, and time pressure, as well as the percentage of transcripts containing at least one of the words or phrases in the authority dictionary between scam and spam robocalls. Compared to spam robocalls, scam robocalls featured significantly greater use of promise of gain ($t(1,684,807) = 53.25$, $p < .001$, $d = 0.065$), threat of loss ($t(1,194,211) = 298.9$, $p < .001$, $d = 0.45$), and time pressure ($t(1,105,901) = 512.42$, $p < .001$, $d = 0.83$) language. Although scam robocalls featured both more promise of gain and threat of loss language than spam robocalls, the difference between scam and spam was significantly larger for threat of loss than promise of gain (interaction term $b = -0.00363$, $SE = 0.000022$, 95% CI [-0.00367, -0.00358], $p < .001$), suggesting that threat of loss language is especially diagnostic that a robocall is a scam. The percentage of transcripts containing authority words or phrases (i.e., naming government entities) was also significantly higher for scam robocalls than spam ($z = 234.06$, $p < .001$, Cohen's $h = 0.27$). Results for these four tactics between scam and spam robocalls are visualized in Figure 1.

[FIGURE 1]

Finally, we assessed the emotional arousal of robocall scams and whether arousal could distinguish robocall scams from spam. We found that the average arousal score of scam robocall transcripts ($M = 3.76$) was significantly higher than the average arousal score of spam robocalls ($M = 3.65$; difference = 0.110, 95% CI [0.1099, 0.1106]; $t(1,896,855) = 547.41$, $p < .001$, $d = 0.63$). This comparison is visualized in Figure 3A.

RQ2: How do psychological dynamics, tactics utilized, and emotional arousal differ between scam robocall pitches targeting older adults and scam robocall pitches targeting the general public?

To answer RQ2, we repeat the analyses answering RQ1, but instead of comparing scam robocalls to spam robocalls, we compare scam robocalls targeting older adults to scam robocalls targeting the general public. The 10 LIWC variables with the largest effect sizes for the average difference between scams targeting older adults and scams targeting the general public are presented in Table 2. Robocall scams targeting older adults were significantly more likely than scams targeting the general public to feature words representing health (e.g., medic, patients, physician, health) ($d = 1.58$), risk (e.g., secure, protect, pain, risk) ($d = 1.58$), and illness (e.g., hospital, cancer, sick, pain) ($d = 1.19$) (all p's < .001).

[TABLE 2]

Compared to scam robocalls targeting the general public, scam robocalls targeting older adults featured significantly less use of promise of gain language ($t(628,464) = 164.43$, $p < .001$, $d = 0.32$) and significantly greater threat of loss ($t(496,998) = 49.44$, $p < .001$, $d = 0.11$) language. Additionally, scams targeting older adults featured more time pressure ($t(332,711) =$

159.08, $p < .001$, $d = 0.43$) language. The percentage of transcripts containing words or phrases appealing to authority (i.e., representing government entities) was also significantly higher for scam robocalls targeting older adults than scams targeting the general public ($z = 657.02$, $p < .001$, $h = 1.53$). Results for these four tactics between scam robocalls targeting older adults and the general public are visualized in Figure 2.

[FIGURE 2]

The average emotional arousal score of robocall scams targeting older adults ($M = 3.78$) was significantly higher than the average arousal score of robocall scams targeting the general public ($M = 3.75$; difference = 0.030, 95% CI [0.0295, 0.0308]; $t(518{,}858) = 89.81$, $p < .001$, $d = 0.19$). This comparison is visualized in Figure 3B.

[FIGURE 3]

Discussion

In this paper, we present one of the first large-scale analyses of the psychological characteristics and tactics of robocalls. Our analysis of 4.49 million robocall pitches collected via honeypot from 2018-2022 (to our knowledge, the largest study of robocalls to-date in terms of number of calls, even among honeypot-based studies) revealed a profile of the characteristics of robocalls targeting Americans and particularly older adults. While previous work has drawn on small samples of scam pitches (e.g.,[39]), anecdotes from fraudsters (e.g.,[40]), and complaints from scam victims (e.g.,[29]), our data paint a picture of naturalistic robocall scams at scale. Our data present synergies with past work done at smaller scales suggesting that hallmarks of scams are the use of emotional arousal[22], time pressure[21,30], and threats of loss[23,29]. While these features have been studied at smaller scales as drivers of people's susceptibility to responding or losing money to scams, we show that the features appear in real robocall scams circulating in the telephone ecosystem.

We also examine unique challenges older adults face when it comes to robocall scams. Concerningly, we found that characteristics of messages and situations that researchers have shown can impair older adults' decision making in the domain of fraud (and otherwise), namely time pressure, threats of loss, and emotional arousal, are *more likely* to appear in the content of the robocall scams targeting older adults compared to those targeting the general public. Scams targeting older adults were also much more likely to focus on health topics compared to scams targeting the general public, which is worrisome given health burdens tend to increase with age[41] and that older adults may be more likely to attend to health information than younger adults[42]. Our findings can focus future research into the mechanisms underlying pathways of older adults' (as well as the general public's) susceptibility to robocall fraud.

The fact that scams targeting older adults were more likely to include threat of loss language is intriguing given theoretical perspectives, namely socioemotional selectivity theory[43], and a large body of empirical evidence[44,45] suggesting that older adults exhibit a preference (e.g., more likely to attend to and remember) for positive information (deemed the "positivity effect"). This pattern would seem more consistent with promise-of-gain language rather than threat-of-loss language. Given that scammers likely learn through trial and error which tactics are most effective, why would scams targeting older adults, then, focus more on threats of loss and less on promises of gain than scams targeting the general public?

We posit two complementary explanations. First, as noted earlier, older adults exhibit greater loss aversion than younger adults[25,36,37]. Second, threats of loss may also pose a threat to one's emotional equilibrium. If older adults are motivated to maintain positive emotional states, as work on socioemotional selectivity theory suggests, they may respond to threats of loss in an effort to restore their emotional equilibrium towards a more positive state. Our findings suggest that the combined motivation to prevent losses and restore emotional equilibrium, particularly under a state of high emotional arousal such as that induced by a scam pitch, could be stronger than the motivation to pursue the promise of positive gains, although we acknowledge this interpretation is speculative given that we lack data from actual victims. This suggests a potential boundary condition for the positivity effect: when faced with emotionally arousing threats of loss, older adults' typical preference for positive information may be overridden by the combined forces of heightened loss aversion and the motivation to restore emotional equilibrium. While boundary conditions for the positivity effect have been identified (e.g., when negative information is relevant to situation-specific goals[45]), including attenuation of the positivity effect under threat[46], to our knowledge, how loss aversion interacts with emotional arousal to override the positivity effect has not been examined in older adults. Yet, our findings suggest this interaction may be common in real-world fraud situations where older adults face loss-framed, emotionally arousing robocall scam pitches.

Our findings suggest directions for the development of educational interventions and campaigns to build people's resilience to robocall scams. In particular, interventions for older adults should base intervention content and recommendations on things that existing research has shown make older adults more vulnerable to victimization by scams *and* are more likely to show up in robocalls targeting older adults (i.e., time pressure, threat of loss, emotional arousal). While there are still many questions regarding how best to convey specific recommendations to consumers in such interventions, and how these recommendations can best be tailored for older adults[47], our work provides a concrete set of tactics interventions can focus on to combat the threats posed by robocall scams. Given difficulties in scaling interventions to build people's resilience to deceptive content[48], it is critical that interventions are designed to be efficient, focusing on the recommendations and educational content most relevant to consumers. Furthermore, the GPT prompt that we developed for sorting the robocalls from the honeypot data into call types at a level of performance consistent with that of a fraud expert could be used by a consumer with ChatGPT when asking it about an unfamiliar call or voicemail they are trying to determine whether to trust or act on. Of course, it remains an open question as to whether consumers would engage in this behavior and how judgments from AI after receiving our engineered prompt will be perceived or used by consumers, but it is an approach that could be included in an educational intervention as a practical tool to protect oneself from robocall scams.

Like any study, ours should be interpreted in light of its limitations, and a number of questions remain unanswered and are ripe for future research. First, because our honeypot data represent only robocall “pitches” (i.e., there is no prospective victim or consumer picking up the phone), our findings are generally restricted to the robocall pitches themselves and cannot speak to the detailed nature of the process of engagement with a robocall or scammer. We know that just as characteristics of scams may influence susceptibility, so do characteristics of the targets[17,29], and in the “real world,” the characteristics of the scam likely interact with the characteristics of the target to shape vulnerability and intervention. Second, we distinguish scams targeting older adults from scams targeting the general public based upon the contents of the scam messages rather than the actual intended target of the scam communication (again, in the case of the honeypot data, there is no target). As such, our approach and the limitations of the honeypot data likely lead to undercovering the robocalls that actually reach older adults because robocalls targeting the general public may also reach older adults (and vice-versa). Relatedly, because our classification of scams as targeting older adults relies on substantive content cues (e.g., references to Medicare, age-associated health conditions), several of which overlap conceptually with LIWC categories that emerged as diagnostic (e.g., health, illness, risk), the LIWC-level differences in those categories may reflect this operationalization rather than serve as fully independent evidence of differential targeting. The fraud-tactics measures and the emotional arousal index, however, are not part of the operational definition of either target category.

Overall, we provide comprehensive evidence about the contents, psychological dynamics, and tactics of contemporary robocalls targeting consumers. Our findings that tactics especially likely to target older adults are also tactics likely to lead to scam victimization among older adults reiterate the critical need to research and develop resources for older adults to build their resilience to scams[47], in this case robocall scams specifically. We hope that our substantive findings and the research infrastructure we built to improve the ability for researchers to study robocalls at large scales can further both science and practice in this area, ultimately contributing to the goal of preserving the psychological and financial well-being of consumers of all ages.

Methods

Data

We analyze 4.49 million robocalls collected from 295,749 phone lines in a honeypot from January 2018 through December 2022. This honeypot is maintained by the company Nomorobo, who use the honeypot to collect robocalls to inform its consumer call blocking software. Nomorobo provided data from their honeypot for analysis in this project. The honeypot receives and picks up calls, recording the ensuing pitch from the caller. Each call recording is transcribed to text in real time, resulting in the data that we analyze in this paper. The honeypot received, recorded, and transcribed over 10 million calls between January 2018 - December 2022, but the corpus of transcripts was filtered down to those ≥ 25 words following best practices for analysis of psychological dimensions of text[49–51], resulting in 4.49 million robocall transcripts for analysis. This study was deemed not to qualify as human subjects research by the Institutional Review Board at the first author's institution.

Measures

Sorting Robocalls Into Scams / Spam / Neither Scam nor Spam

To sort the millions of robocall transcripts into calls representing scams, spam, and neither spam nor scam, we used an LLM-based approach. Specifically, we followed Rathje et al.'s[52] protocol and materials for using OpenAI's GPT models for psychological text analysis. The workflow consisted of five steps. First, the second author, a fraud expert with over three decades of experience working in fraud investigation and prevention, coded a random sample of 1,499 transcripts from the honeypot (a "training set") into scam, spam, and neither scam nor spam. Second, we generated a prompt for GPT that contained the definitions the expert used to label the transcripts into the three categories and gave that prompt with just definitions to GPT-4o-mini to label the same transcripts (a "zero-shot learning" approach). The zero-shot learning approach did not align with the expert judgments at a conventionally high rate of reliability of F1 score ≥ .7[53,54]. Therefore, the third step in the process was expanding our prompting approach to "few-shot learning," adding exemplar call transcripts representing scam, spam, and neither scam nor spam to our prompt in addition to the definitions. We iterated on the few-shot learning prompt, testing how including different examples improved the performance of the model until we achieved a prompt with which the outputs from GPT-4o-mini aligned with the expert judgments for the training set of transcripts at a macro-averaged F1 score ≥ .7 (final average F1 score = .77, precision = .76, recall = .78; see Supplemental Materials A1 for the final few-shot learning prompt). Fourth, to ensure that the prompt we engineered to reliably align with expert judgments in the training set was not overfit to the specific transcripts included in the training set, we drew another random sample of 650 robocall transcripts from the honeypot (a "test set") for both the expert to label and GPT to label using our engineered prompt. This process achieved an average F1 score of .7 (average precision = .68, recall = .75) for the test set. This achieved F1 score is noteworthy given past work demonstrating that it is notoriously difficult to use automated approaches to identify deceptive content at scale[55]. Fifth, and finally, we gave the engineered prompt to GPT-4o-mini to label the remaining millions of transcripts using the OpenAI API platform. Given the scale of our data that needed to be processed by the API, to keep our prompts as economical as possible, we did not provide any special instructions in our prompts (e.g., "act as if you are a fraud investigator"). That said, following Rathje et al.[52], we achieved strong performance with the out-of-the-box GPT-4o-mini model.

Sorting Robocalls Into Calls Targeting Older Adults versus Calls Targeting the General Public

We used the same LLM-based approach used to sort robocall transcripts into scam/spam/neither scan nor spam to sort transcripts into calls targeting older adults and calls targeting the general public. Specifically, for our prompts, we defined calls targeting the two categories as (1) calls targeting older adults (generally defined as people 55 years of age or older) and (2) not targeting older adults. A few-shot learning prompt (see Supplemental Materials A2 for the final prompt) achieved an average F1 score of .83 (average precision = .82, recall = .85) on a training set of 1,559 transcripts and that same prompt achieved an average F1 score of .82 (average precision = .86, recall = .79) on a test set of 149 transcripts. The engineered prompt was then given to GPT-4o-mini to label the remaining millions of call transcripts. See Table 3 for the confusion matrix of resulting call labels across these two criteria (scam/spam/neither spam nor scam; targeting older adults/targeting general public), along with full definitions for each category as were contained in the prompts.

Although our prompt defined the target category in terms of age (people 55 years of age or older), the examples provided in our few-shot learning prompt grounded that definition in the substantive content cues that distinguish calls targeting older adults from those targeting the general public. Specifically, calls were treated as targeting older adults when they invoked content that is unambiguously age-relevant. These cues fell into a small number of recurring types: explicit references to age-restricted federal programs (e.g., Medicare benefits and consultations, Social Security "suspension" notices), products and services predominantly marketed to older consumers (e.g., final expense or burial insurance, pain-relief braces), and sweepstakes appeals (e.g., Publishers Clearing House–style prize notifications). Calls were treated as targeting the general public when their content was either age-neutral (e.g., business listing services, telecom bill reduction promotions, residential energy and solar upgrade pitches) or pitched at non-age-specific financial concerns (e.g., IRS back-tax assistance and debt relief). Examples of each call type, along with full definitions, are provided in the few-shot learning prompt in Supplemental Materials A2. We note that this operationalization of "targeting older adults" is largely grounded in topical cues, several of which may overlap conceptually with LIWC categories analyzed in our results (e.g., health, illness). The fraud-tactics measures (promise of gain, threat of loss, time pressure, authority) and the emotional arousal index, however, are conceptually independent of these topical cues, and the comparative findings on those measures are therefore not driven by the classification scheme itself. We address this point further in the Discussion section.

[TABLE 3]

Linguistic Inquiry and Word Count (LIWC)

To broadly index the psychological characteristics of the honeypot robocall transcripts, we used the computational text analysis tool LIWC-22[19]. LIWC counts words as a percentage of total word count per text against its internal dictionary of social, psychological, and part of speech dimensions[56] and has undergone extensive psychometric validation[19]. LIWC has been used widely across the social sciences to characterize psychological dynamics in a variety of texts, including open-ended survey responses[57], online journals[58], advertisements[59], social media posts[60], and, most relevant to the present study, investment scam websites[39] and phishing emails[61] (although at far smaller scales than the size of the honeypot data we draw from). Using the LIWC software, LIWC-22 was applied to each transcript in the honeypot data, documenting over 100 attributes (e.g., positive and negative emotion, health, prosociality) per transcript. The LIWC scores along these attributes were then analyzed in R (version 4.6.0).

Tactics Utilized

To assess the use of specific tactics associated with scams, we developed a dictionary of 327 words and phrases associated with four tactics: promise of gain, threat of loss, time pressure, authority (i.e., naming of legitimate government entities)[21,23,24,29–31]. Dictionary development was informed by findings from Pratkanis & Shadel[40] and a US Department of Justice study[62] that involved analysis of hundreds of tapes of scammers’ interactions with undercover investigators posing as older adult targets, revealing how these scam tactics work in action. Those findings were then adapted to apply to the current study by Shadel and a team of telemarketing fraud experts including Aaron Foss, founder of Nomorobo, and Karla Pak, an AARP fraud researcher who has co-authored numerous fraud studies[62,63]. Examples of words and phrases in each category are as follows: promise of gain (claim your prize, at no cost to you, eliminate debt, live your dream, rebate, winner), threat of loss (billed on your card, charged, auto renewed, incoming charge, unusual activity), time pressure (last chance, limited time, expires today, act soon, final attempt, urgent), authority (IRS, FBI, Social Security, Medicare, tax department). Promise of gain, threat of loss, and time pressure were calculated for each transcript as a percentage of total words in the transcript that appeared in that category in the dictionary, following LIWC's calculation approach[19]. Authority was measured at the transcript level as a binary (1 = a word in the authority dictionary appeared in the transcript, 0 = no words in the authority dictionary appeared in the transcript).

Emotional Arousal

Given the importance of emotional arousal in driving susceptibility to scams[22], we measured the emotional arousal of each robocall transcript using arousal metrics developed by Warriner et al.[20]. Warriner et al.[20] provide arousal scores for nearly 14,000 English lemmas (base forms of words) on a 9-point scale. To compute arousal scores for each of our transcripts, we first lemmatized both the Warriner dictionary terms and the words in each transcript using the textstem package in R. We then calculated the mean arousal score across all words in the transcript that matched lemmas in the Warriner et al.[20] database. For example, the scam robocall transcript "Please listen if you want to make $100.00 every 24 hours press 1 now press 1 right now or press 2 if there's no interest" contains 11 words that corresponded to lemmas in the Warriner et al.[20] database (e.g., "please" = 4.73, "listen" = 3.76, "want" = 5.29), yielding a mean arousal score of 3.849. The resulting average arousal score across honeypot transcripts was 3.69 (SD = 0.18, range = 2.49 - 6.79).

Statistical Analysis

Each transcript analyzed represents an independent robocall recording. All statistical analyses were conducted in R (version 4.6.0). All Welch two-sample t-tests and z-tests for proportions were two-sided. The normality of distributions was not formally tested. No corrections for multiple comparisons were applied. Cohen's d was computed using the pooled standard deviation of both groups as the denominator. Cohen's h was computed as the difference between arcsine-transformed proportions. The interaction between call type (scam vs. spam) and tactic type (promise of gain vs. threat of loss) was tested using a linear mixed effects model with a random intercept by unique transcript ID. Group means and 95% confidence intervals are reported in the Results section text, figure legends, and tables.

Claude large language models were used to assist with the analysis code and replication files. The authors verified all outputs and take responsibility for the final content.

Acknowledgements

The authors would like to thank Marti DeLiema, Brad Reaves, and Martha Deevy for their feedback on an earlier version of this manuscript. Claude large language models were used for light editing of the manuscript text.

Conflicts of Interest

Ryan C. Moore has no conflicts of interest to declare. Doug Shadel has provided consulting services for Nomorobo, the company that maintains the honeypot which collected the robocall transcripts analyzed in this study. Nomorobo imposed no conditions or restrictions on the analysis or publication of findings.

References

1. Federal Communications Commission. Robocall Response Team: Combating Scam Robocalls & Robotexts | Federal Communications Commission. https://www.fcc.gov/spoofed-robocalls (2024).
2. Federal Communications Commission. Stop Unwanted Robocalls and Texts. https://www.fcc.gov/consumers/guides/stop-unwanted-robocalls-and-texts (2025).
3. Semuels, A. Robocalls Finally Have the U.S. Government's Attention. *TIME* https://time.com/6513036/robocalls-government-action/ (2023).
4. Federal Trade Commission. FTC Ramps Up Fight to Close the Door on Illegal Robocalls Originating from Overseas Scammers and Imposters. *Federal Trade Commission* https://www.ftc.gov/news-events/news/press-releases/2023/04/ftc-ramps-fight-close-door-illegal-robocalls-originating-overseas-scammers-imposters (2023).
5. Federal Communications Commission. Caller ID Spoofing | Federal Communications Commission. https://www.fcc.gov/consumers/guides/spoofing (2024).
6. DeLiema, M., Deevy, M., Lusardi, A. & Mitchell, O. S. Financial Fraud Among Older Americans: Evidence and Implications. *J. Gerontol. Ser. B* **75**, 861–868 (2020).
7. Robinson, J. & Edwards, M. Fraudsters target the elderly: Behavioural evidence from randomised controlled scam-baiting experiments. *Secur. J.* **37**, 1173–1196 (2024).
8. Truecaller. *America Under Attack: The Shifting Landscape of Spam and Scam Calls in America*. https://publicinterestnetwork.org/wp-content/uploads/2024/10/US-SpamScam-Report_2024_0307.pdf (2024).
9. Federal Trade Commission. Age and Fraud | Tableau Public. https://public.tableau.com/app/profile/federal.trade.commission/viz/AgeandFraud/Infographic (2025).

10. Federal Trade Commission. Protecting Older Consumers 2024-2025 (A Report of the Federal Trade Commission). 56 (2025).

11. Kemp, S. & Erades Pérez, N. Consumer Fraud against Older Adults in Digital Society: Examining Victimization and Its Impact. *Int. J. Environ. Res. Public. Health* **20**, 5404 (2023).

12. Shao, J., Zhang, Q., Ren, Y., Li, X. & Lin, T. Why are older adults victims of fraud? Current knowledge and prospects regarding older adults' vulnerability to fraud. *J. Elder Abuse Negl.* **31**, 225–243 (2019).

13. Federal Communications Commission. 'Grandparent' Scams Get More Sophisticated | Federal Communications Commission. https://www.fcc.gov/consumers/scam-alert/grandparent-scams-get-more-sophisticated (2025).

14. Gottfried, J., Park, E. & Anderson, M. Online Scams and Attacks in America Today. *Pew Research Center* https://www.pewresearch.org/internet/2025/07/31/online-scams-and-attacks-in-america-today/ (2025).

15. DeLiema, M., Shadel, D. & Pak, K. Profiling Victims of Investment Fraud: Mindsets and Risky Behaviors. *J. Consum. Res.* **46**, 904–914 (2020).

16. Parry, D. A. *et al.* A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nat. Hum. Behav.* **5**, 1535–1547 (2021).

17. Houtti, M., Roy, A., Gangula, V. N. R. & Walker, A. M. A Survey of Scam Exposure, Victimization, Types, Vectors, and Reporting in 12 Countries. Preprint at https://doi.org/10.48550/arXiv.2407.12896 (2024).

18. Pak, K. & Shadel, D. *AARP Foundation National Fraud Victim Study*. 93 https://www.aarp.org/content/dam/aarp/research/surveys_statistics/econ/2011/2011-aarp-national-fraud-victim-study.pdf (2011).

19. Boyd, R. L., Ashokkumar, A., Seraj, S. & Pennebaker, J. W. *The Development and Psychometric Properties of LIWC-22*. 50 https://www.liwc.app/static/documents/LIWC-22%20Manual%20-%20Development%20and%20Psychometrics.pdf (2022).
20. Warriner, A. B., Kuperman, V. & Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res. Methods* **45**, 1191–1207 (2013).
21. Jones, H. S., Towse, J. N., Race, N. & Harrison, T. Email fraud: The search for psychological predictors of susceptibility. *PLOS ONE* **14**, e0209684 (2019).
22. Kircanski, K. *et al.* Emotional Arousal May Increase Susceptibility to Fraud in Older and Younger Adults. *Psychol. Aging* **33**, 325–337 (2018).
23. Lyu, C., Gao, S. & Zhang, Q. The impact of time pressure and type of fraud on susceptibility to online fraud. *Front. Psychol.* **16**, (2025).
24. Yu, L. *et al.* Vulnerability of Older Adults to Government Impersonation Scams. *JAMA Netw. Open* **6**, e2335319 (2023).
25. Horn, S. Adult age differences in value-based decision making. *Curr. Opin. Psychol.* **55**, 101765 (2024).
26. Salthouse, T. A. Selective review of cognitive aging. *J. Int. Neuropsychol. Soc. JINS* **16**, 754–760 (2010).
27. AARP. Watchdog Alert: Robotexts Are On The Rise. https://states.aarp.org/wisconsin/watchdog-alert-robotexts-are-on-the-rise (2023).
28. Murray, T., Tracy, L. & Mela, J. *Ringing in Our Fears 2025*. 56 https://publicinterestnetwork.org/wp-content/uploads/2025/10/RIOF-2025-REPORT-10151035.pdf (2025).
29. DeLiema, M., Li, Y. & Mottola, G. Correlates of responding to and becoming victimized by fraud: Examining risk factors by scam type. *Int. J. Consum. Stud.* **47**, 1042–1059 (2023).

30. Wang, J., Herath, T., Chen, R., Vishwanath, A. & Rao, H. R. Research Article Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email. *IEEE Trans. Prof. Commun.* **55**, 345–362 (2012).

31. Fischer, P., Lea, S. E. G. & Evans, K. M. Why do individuals respond to fraudulent scam communications and lose money? The psychological determinants of scam compliance. *J. Appl. Soc. Psychol.* **43**, 2060–2072 (2013).

32. Prasad, S., Bouma-Sims, E., Mylappan, A. K. & Reaves, B. Who’s Calling? Characterizing Robocalls through Audio and Metadata Analysis. *Proc. 29th USENIX Secur. Symp.* 19 (2020).

33. Prasad, S., Nahapetyan, A. & Reaves, B. Characterizing Robocalls with Multiple Vantage Points. Preprint at https://doi.org/10.48550/arXiv.2410.17361 (2024).

34. Marzuoli, A., Kingravi, H. A., Dewey, D. & Pienta, R. Uncovering the Landscape of Fraud and Spam in the Telephony Channel. in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)* 853–858 (2016). doi:10.1109/ICMLA.2016.0153.

35. Prasad, S., Dunlap, T., Ross, A. & Reaves, B. Diving into robocall content with SnorCall. in *Proceedings of the 32nd USENIX Conference on Security Symposium* 427–444 (USENIX Association, USA, 2023).

36. Ebner, N. C., Freund, A. M. & Baltes, P. B. Developmental changes in personal goal orientation from young to late adulthood: From striving for gains to maintenance and prevention of losses. *Psychol. Aging* **21**, 664–678 (2006).

37. Mrkva, K., Johnson, E. J., Gächter, S. & Herrmann, A. Moderating Loss Aversion: Loss Aversion Has Moderators, But Reports of its Death are Greatly Exaggerated. *J. Consum. Psychol.* **30**, 407–428 (2020).

38. Sarno, D. M., Lewis, J. E., Bohil, C. J. & Neider, M. B. Which Phish Is on the Hook? Phishing Vulnerability for Older Versus Younger Adults. *Hum. Factors* **62**, 704–717 (2020).

39. Aziz, A. A. A. *et al.* Linguistic Cues of Deception in Malaysian Online Investment Scams' Promotional Materials. *GEMA Online® J. Lang. Stud.* **23**, (2023).

40. Pratkanis, A. R. & Shadel, D. *Weapons of Fraud: A Source Book for Fraud Fighters*. (AARP Washington, 2005).

41. Prince, M. J. *et al.* The burden of disease in older people and implications for health policy and practice. *The Lancet* **385**, 549–562 (2015).

42. Liu, K., Su, P., Wang, H. & Tao, D. Contextualizing Visualizations of Digital Health Information among Young and Older Adults Based on Eye-Tracking. *Sustainability* **14**, 16506 (2022).

43. Carstensen, L. L., Isaacowitz, D. M. & Charles, S. T. Taking time seriously. A theory of socioemotional selectivity. *Am. Psychol.* **54**, 165–181 (1999).

44. Carstensen, L. L. & DeLiema, M. The positivity effect: a negativity bias in youth fades with age. *Curr. Opin. Behav. Sci.* **19**, 7–12 (2018).

45. Reed, A. E. & Carstensen, L. L. The Theory Behind the Age-Related Positivity Effect. *Front. Psychol.* **3**, 339 (2012).

46. Mather, M. & Knight, M. R. Angry faces get noticed quickly: threat detection is not impaired among older adults. *J. Gerontol. B. Psychol. Sci. Soc. Sci.* **61**, P54-57 (2006).

47. Scam Prevention Research Committee. *A Review of Scam Prevention Messaging Research: Takeaways and Recommendations*. 22 https://consumer.ftc.gov/system/files/consumer_ftc_gov/pdf/A%20Review%20of%20Scam%20Prevention%20Messaging%20Research.pdf (2024).

48. Moore, R. C. Short educational videos improve older adults' digital literacy and resilience to online deception at scale. (Stanford University, Stanford, California, 2025).

49. Aghazadeh, S., Hoang, K. & Pomeroy, B. Using LIWC to Analyze Participants' Psychological Processing in Accounting JDM Research. *Audit. J. Pract. Theory* **41**, 1–20 (2022).

50. Boyd, R. L. Psychological Text Analysis in the Digital Humanities. in *Data Analytics in Digital Humanities* (ed. Hai-Jew, S.) 161–189 (Springer International Publishing, Cham, 2017). doi:10.1007/978-3-319-54499-1_7.

51. Markowitz, D. M. Psychological trauma and emotional upheaval as revealed in academic writing: The case of COVID-19. *Cogn. Emot.* **36**, 9–22 (2022).

52. Rathje, S. *et al.* GPT is an effective tool for multilingual psychological text analysis. *Proc. Natl. Acad. Sci.* **121**, e2308950121 (2024).

53. Brown, J. B. Classifiers and their Metrics Quantified. *Mol. Inform.* **37**, 1700127 (2018).

54. Shar, L. K. *et al.* AutoConf: Automated Configuration of Unsupervised Learning Systems Using Metamorphic Testing and Bayesian Optimization. in *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)* 1326–1338 (2023). doi:10.1109/ASE56229.2023.00094.

55. Conti, M., Lain, D., Lazzeretti, R., Lovisotto, G. & Quattrociocchi, W. It's always April fools' day!: On the difficulty of social network misinformation classification via propagation features. in *2017 IEEE Workshop on Information Forensics and Security (WIFS)* 1–6 (IEEE, Rennes, France, 2017). doi:10.1109/WIFS.2017.8267653.

56. Markowitz, D. M., Hancock, J. T. & Bailenson, J. N. Linguistic Markers of Inherently False AI Communication and Intentionally False Human Communication: Evidence From Hotel Reviews. *J. Lang. Soc. Psychol.* **43**, 63–82 (2024).

57. Moore, R. C., Lee, A. Y., Hancock, J. T., Halley, M. C. & Linos, E. Age-Related Differences in Experiences With Social Distancing at the Onset of the COVID-19 Pandemic: A Computational and Content Analytic Investigation of Natural Language From a Social Media Survey. *JMIR Hum. Factors* **8**, e26043 (2021).

58. Cohn, M. A., Mehl, M. R. & Pennebaker, J. W. Linguistic markers of psychological change surrounding September 11, 2001. *Psychol. Sci.* **15**, 687–693 (2004).

59. Markowitz, D. M. Words for Sale: Linguistic Complexity Associates with Higher Housing Prices in Online Realty Advertisements. *J. Lang. Soc. Psychol.* **43**, 717–729 (2024).

60. Kelley, S. W. & Gillan, C. M. Using language in social media posts to study the network dynamics of depression longitudinally. *Nat. Commun.* **13**, 870 (2022).

61. Xu, T. & Rajivan, P. Determining psycholinguistic features of deception in phishing messages. *Inf. Comput. Secur.* **31**, 199–220 (2023).

62. AARP Foundation. *Off the Hook: Reducing Participation in Telemarketing Fraud*. 166 https://assets.aarp.org/rgcenter/consume/d17812_fraud.pdf (2003).

63. Williams, A., Choi-Allum, L., Shadel, D., Pak, K. & Raineville, C. A MOMENT'S NOTICE: RECOGNIZING THE STRESSFUL LIFE EVENTS, EMOTIONS, AND ACTIONS THAT MAKE US SUSCEPTIBLE TO SCAMS. *Innov. Aging* **6**, 615 (2022).

Table 1: Top ten LIWC categories differing between scam and spam robocalls

LIWC category	Description / exemplars	Average % of words in scam robocalls	Average % of words in spam robocalls	Difference, 95% CI	|Cohen's d|	p-value
Negations	not, no, never, nothing	0.80	2.35	−1.549, [−1.552, −1.545]	0.78	< .001
Technology	car, phone, email	0.75	2.37	−1.624, [−1.628, −1.620]	0.71	< .001
Apostrophes		0.80	2.03	−1.224, [−1.227, −1.220]	0.71	< .001
Social behavior	love, say, care	7.75	4.94	2.814, [2.805, 2.824]	0.71	< .001
Visual	see, look, saw	0.27	0.92	−0.645, [−0.647, −0.643]	0.66	< .001
Other punctuation	punctuation other than periods, commas, question marks, exclamation points, apostrophes	1.70	0.58	1.127, [1.121, 1.133]	0.64	< .001
Culture	car, united states, govern	1.13	2.60	−1.462, [−1.467, −1.457]	0.61	< .001
Attention	look, watch, check	0.52	1.34	−0.817, [−0.820, −0.814]	0.60	< .001

Prosocial behavior	help, thank, please	2.38	1.29	1.091, [1.086, 1.096]	0.59	< .001
Communication	said, say, tell	4.23	2.79	1.442, [1.436, 1.449]	0.57	< .001

Note: Top ten LIWC variables differing between scam and spam robocalls, sorted by largest absolute value of effect size (Cohen's d). Before LIWC variables were ranked, variables where the difference between scam and spam robocalls was not significant at $p < .05$ were discarded. Words and phrases in "Description / exemplars" column are some of the most frequently used words and phrases from that LIWC category according to the LIWC-22 development manual [19]. Difference and 95% CI are for scam minus spam (average % of words).

Table 2: Top ten LIWC categories differing between scam robocalls targeting older adults and scam robocalls targeting the general public

LIWC category	Description / exemplars	Average % of words in scam robocalls targeting older adults	Average % of words in scam robocalls targeting general public	Difference, 95% CI	|Cohen's d|	p-value
Health	medic, patients, physician, health	2.76	0.17	2.585, [2.573, 2.597]	1.58	< .001
Risk	protect, pain, risk	1.99	0.28	1.713, [1.706, 1.719]	1.58	< .001
Physical	food, patients, eye	2.97	0.40	2.571, [2.558, 2.583]	1.46	< .001
Illness	hospital, cancer, sick, pain	0.84	0.06	0.787, [0.782, 0.792]	1.19	< .001
Money	business, pay, price, market	1.41	3.87	−2.461, [−2.475, −2.447]	0.78	< .001
Negative emotion	bad, hate, hurt, tired	0.63	0.14	0.495, [0.491, 0.499]	0.75	< .001
Lifestyle	work, home, school, working	4.19	7.04	−2.848, [−2.865, −2.832]	0.71	< .001
Numbers	one, two, first, once	3.37	5.17	−1.801, [−1.812, −1.790]	0.68	< .001
Feeling	feel, hard, cool, felt	0.51	0.12	0.388, [0.384, 0.392]	0.66	< .001

Prosocial behavior	help, thank, please	1.43	2.70	−1.272, [−1.281, −1.262]	0.61	< .001

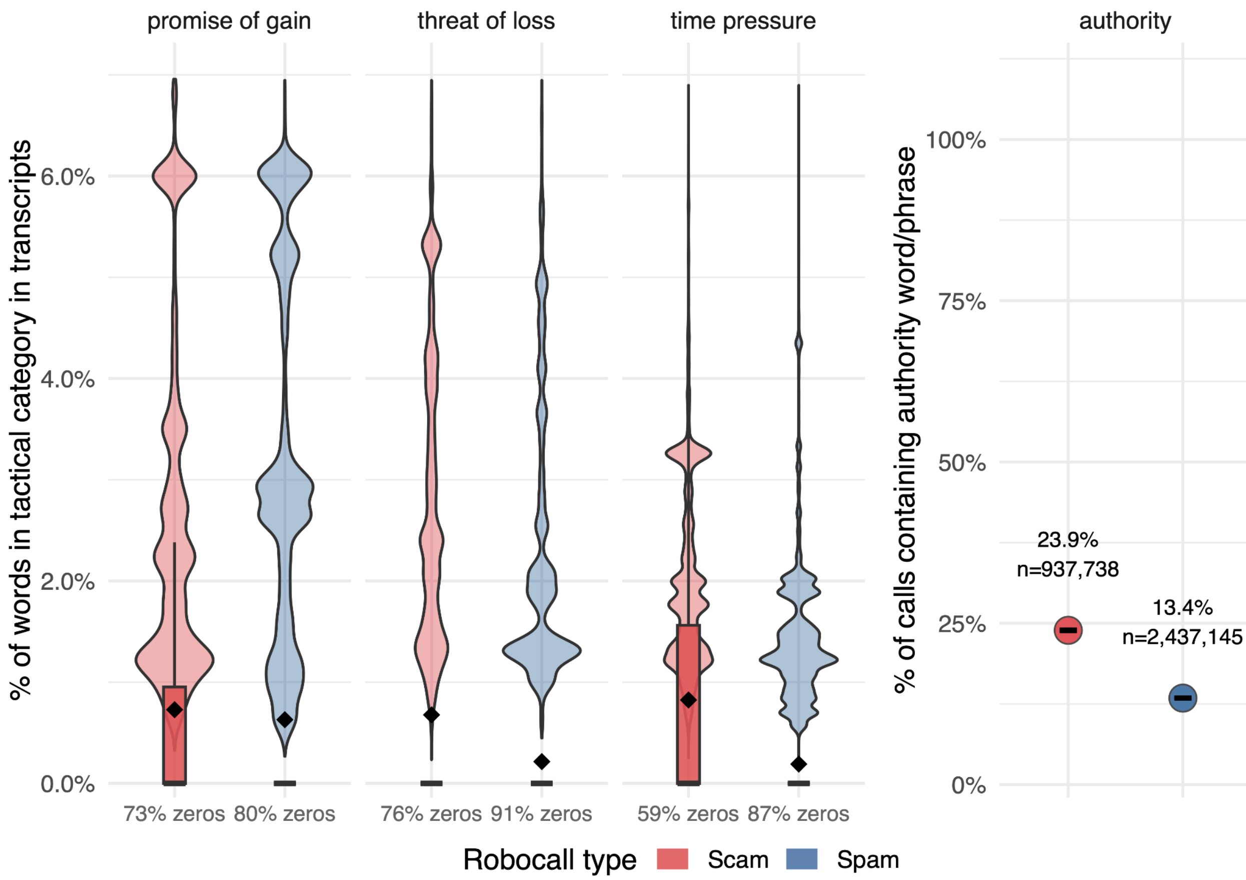
Note: Top ten LIWC variables differing between scam robocalls targeting older adults and scam robocalls targeting the general public, sorted by largest absolute value of effect size (Cohen's d). Before LIWC variables were ranked, variables where the difference between scam robocalls targeting older adults and scam robocalls targeting the general public was not significant at $p < .05$ were discarded. Words and phrases in "Description / exemplars" column are some of the most frequently used words and phrases from that LIWC category according to the LIWC-22 development manual [19]. Difference and 95% CI are for scams targeting older adults minus scams targeting the general public (average % of words).

Table 3: Confusion matrix of honeypot calls by call type as labeled by GPT

	Call targets general public	**Call targets older adults**
Scam	701,280	236,458
Spam	1,991,478	445,667
Neither scam nor spam	1,005,905	110,045

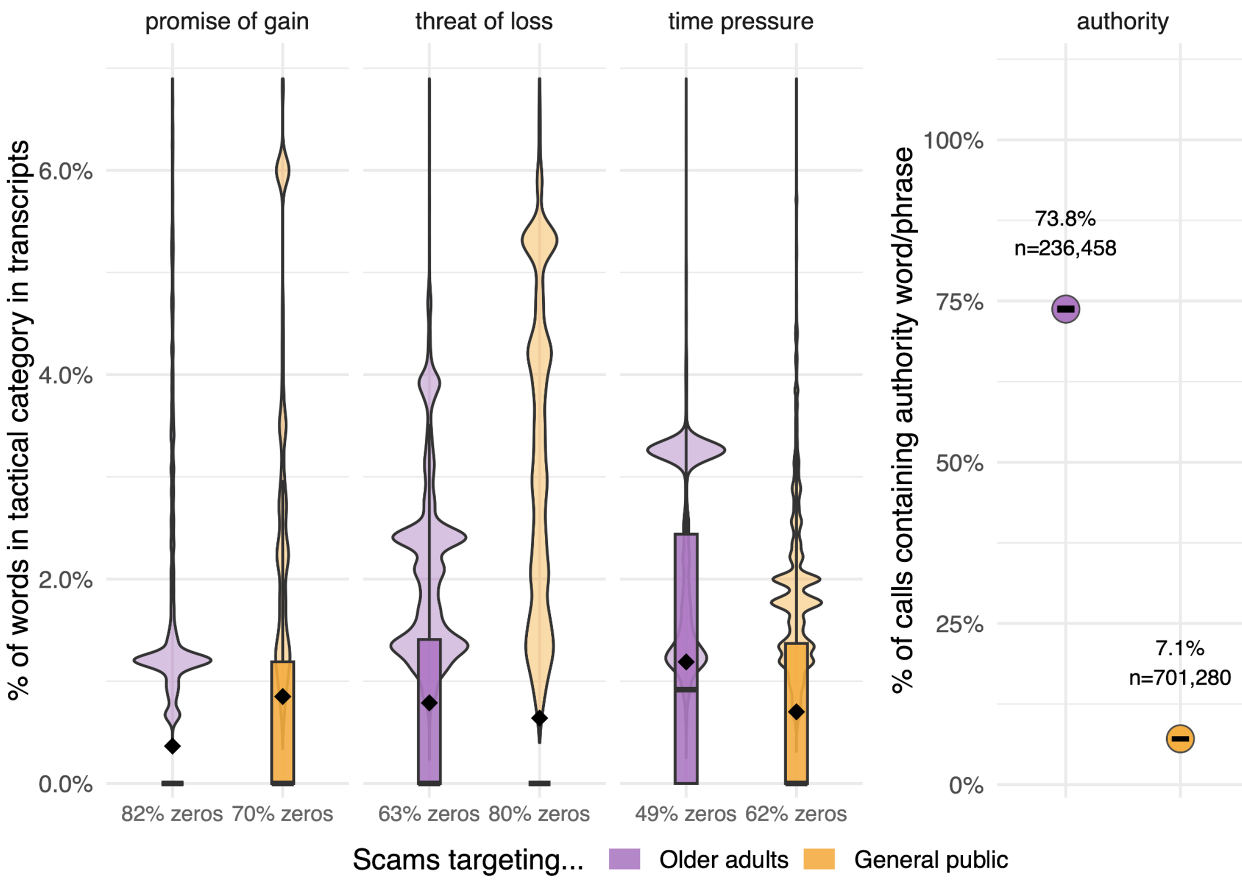
Note: Breakdown of 4,490,833 honeypot call transcripts from 2018-2022 into categories labeled by GPT (see Methods section for prompt engineering details and metrics for validation against fraud expert judgments). Few-shot learning prompts are available in the Supplemental Materials.

Figure 1: Tactics utilized between robocall scam and spam



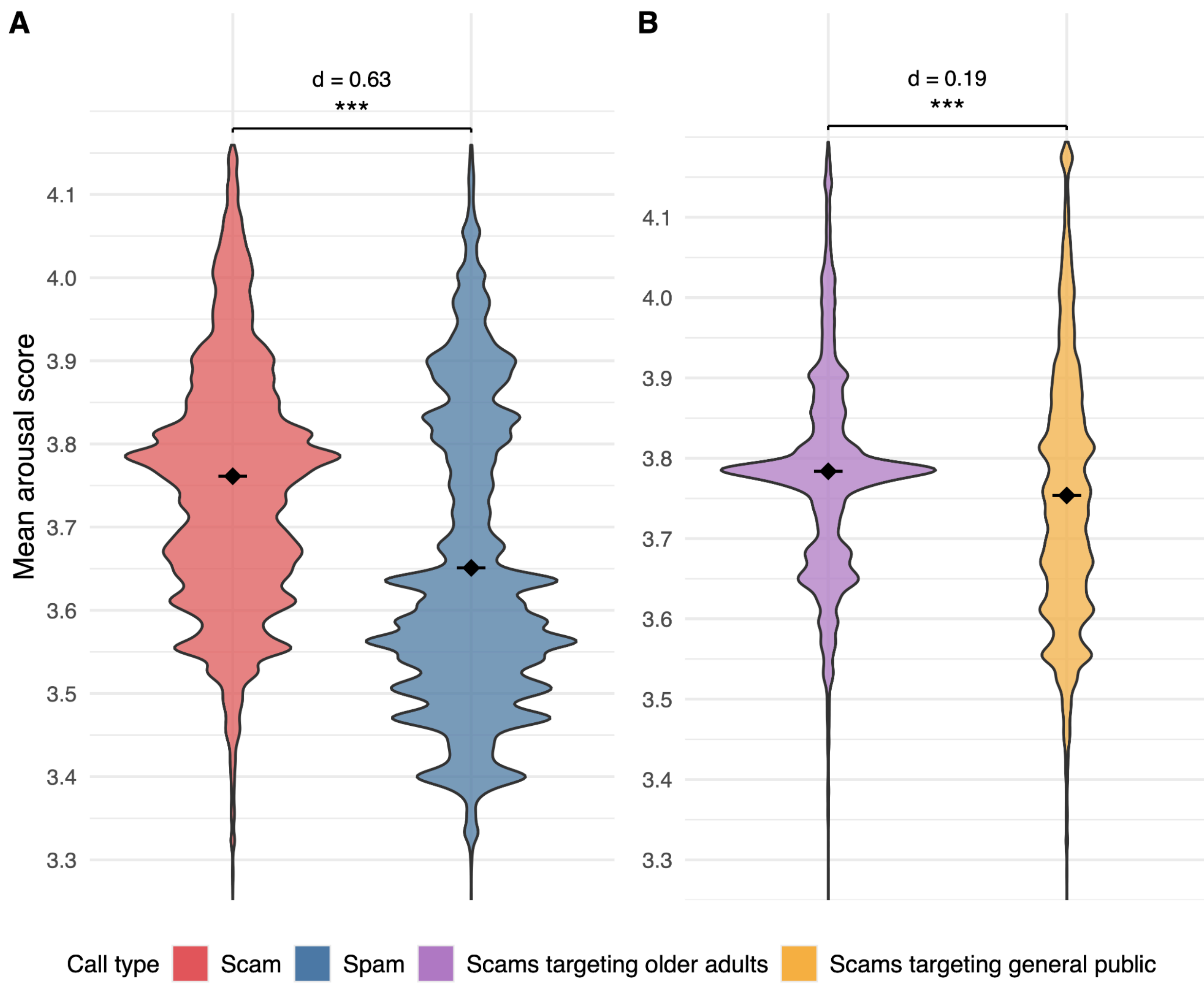
Note: For promise of gain, threat of loss, and time pressure, violin plots show the distribution of per-transcript word proportions among calls with non-zero usage; the embedded box indicates the interquartile range and median across all calls (including those with zero usage), and the diamond marker denotes the group mean. The percentage of calls with zero usage of each tactical category is shown below each violin. For the tactic of invoking authority, points show the proportion of calls containing at least one word or phrase from the authority (i.e., government entity) dictionary, with error bars representing 95% confidence intervals. Sample sizes are shown adjacent to each point. Group means (diamonds) and mean differences (scam minus spam) by tactic: promise of gain — scam M = 0.73%, spam M = 0.63%, difference = 0.10 pp, 95% CI [0.0954, 0.1027]; threat of loss — scam M = 0.68%, spam M = 0.22%, difference = 0.46 pp, 95% CI [0.4587, 0.4648]; time pressure — scam M = 0.82%, spam M = 0.19%, difference = 0.63 pp, 95% CI [0.6311, 0.6359].

Figure 2: Tactics utilized between robocall scams targeting older adults and the general public



Note: For promise of gain, threat of loss, and time pressure, violin plots show the distribution of per-transcript word proportions among scam robocalls targeting older adults and scam robocalls targeting the general public with non-zero usage; the embedded box indicates the interquartile range and median across all calls (including those with zero usage), and the diamond marker denotes the group mean. The percentage of calls with zero usage of each tactical category is shown below each violin. For the tactic of invoking authority, points show the proportion of calls containing at least one word or phrase from the authority (i.e., government entity) dictionary, with error bars representing 95% confidence intervals. Sample sizes are shown adjacent to each point. Group means (diamonds) and mean differences (older adult targeting minus general public targeting) by tactic: promise of gain — older adult scams M = 0.37%, general public scams M = 0.85%, difference = −0.49 pp, 95% CI [−0.4908, −0.4792]; threat of loss — older adult M = 0.79%, general public M = 0.64%, difference = 0.15 pp, 95% CI [0.1431, 0.1549]; time pressure — older adult M = 1.19%, general public M = 0.70%, difference = 0.49 pp, 95% CI [0.4823, 0.4944].

Figure 3: Emotional arousal by call type



Note: Violin plots show the distribution of mean arousal scores per transcript within each call type, truncated at the 99th percentile. The diamond marker denotes the group mean (Panel A: scam M = 3.76, spam M = 3.65; Panel B: scams targeting older adults M = 3.78, scams targeting the general public M = 3.75); error bars represent ±1 SEM. Significance brackets indicate Welch two-sample t-test results; d = Cohen's d; *** = p < .001.

Supplemental Materials for

A Computational Analysis of Millions of Scam and Spam Robocalls Targeting Older Adults and the General Public

A. Few-shot learning prompts for GPT-4o-mini

A1. Prompt to classify robocalls into scam/spam/other

"The following text is a transcript of a phone call. Please code it as either scam, spam, or neither scam nor spam. Spam is defined as any message that is attempting to sell a product or service where there is a reasonable chance the caller will actually deliver something of value. Scam is defined as any message that is attempting to get money or personal information where there is little to no chance that a product or service will be delivered. Answer only with a number: 1 if scam, 2 if spam, and 3 if neither scam nor spam. Here are some examples to be aware of:

Text: This is an important message do not hang up the phone your Google business listing is it risk of displaying incorrectly We've tried numerous times to contact you by telephone so we can check the status of your listing this will only take a few minutes of your time press one so we can quickly check the status of your Google listing to ensure it's active if your Google listing is suspended or deleted your business will not be shown causing customers to think you're no longer open for business press one now to check the status of your local listing if you're the business owner press one now Press 2 to be removed from this list.
Rating: 2

Text: Our records show that you have not updated your free google business listing. Press one now to verify and update your Google listing. Press 9 to be removed from the list. Again, press one to verify and update your Google listing. Our records show that you have not updated your free google business listing. Press one now to verify and update your Google listing. Press 9 to be removed from the list. Again, press one to verify and update your Google listing.
Rating: 2

Text: Hi, this is Lynette with the Medicare department at healthcare benefits. How are you today? Great. I'm calling because the updated plan for Medicare has been released, and it may give you some better access to things like dental, vision, hearing, and over the counter benefits. Are you there?
Rating: 2

Text: Hello. Hi, my name is Tom. I'm calling on behalf of childhood leukemia foundation. The foundation is delivering wish baskets to sick children and cancer treatment centers. These contain toys, games, and even a generous gift card. I'm with innovative, it's a paid fundraiser and tomorrow were mailing out pledge cards to generous people like

yourself. And when you received a pledge card, can we count on you to make a small tax deductible donation.
Rating: 2

Text: Hello. This is John with back blue lives pack. As you know, our police officers are being defunded demoralized and disrespected, and they put their lives on the line every day without the support of many politicians. It's time to say enough is enough. The organization is having its fundraising drive to support lawmakers who will stand with brave men and women in blue. So can you help us protect our police officer's lives and race with a one time donation for the Dr.
Rating: 2

Text: Hello, this is Amy with Medical Services. You were recently recommended by a medical professional to receive an emergency medical alert system at no cost to you, and your system is ready to be shipped. Press one now and hold the line. To receive your medical alert system at no cost. One in three Americans over the age of 65 fall every year in the US. And it could be the last. Don't be a statistic. Press one now and hold the line to receive your or no cost medical alert system, press one now, press nine to opt out. Rating: 2

Text: Hello, this is Ashley calling you from road traffic accident. How have you been today? This call is in relation to your vehicle accident which happened in past two years and that was not your fault. Is that the correct information we have got? All right, let me transfer the call to the senior supervisor. He will guide you further about the compensation amount. Thank you so much.
Rating: 2

Text: Hi this is Allison I'm calling because we said she to several notifications about extending coverage on your car warranty but didn't hear back and this is one last courtesy call to let you know that you can still keep your vehicle under coverage would you like to hear more about your options.
Rating: 2

Text: Hey how are you today my name is Jessica with Home Security promotions The reason I'm calling you today as we're giving away up to 20 cellular home security systems in your area absolutely free and we're also updating all of the old outdated landline systems to wireless video systems again absolutely free because I don't know if you know that the of us have learned that by cutting the phone lines outside of your home will disable your outdated alarm system studies show that one in 17 homes are

being broken into every single day so does this sound like something you're interested in.
Rating: 2

Text: Hello. Hi, my name is Dalia calling from american solar. How are you doing today? Well, I am calling you to tell you little about solar program. The benefits, tax rebates and grants from government. That typically means going solar is absolutely free for most of people. We are helping out the homeowner who are paying more than $100 bill. Are you still paying over dollar 100 on your electric bill? I understand you are the homeowner of single family regular house, right? Great. You are qualified to get rebate option.
Rating: 2

Text: Hello? This is Lisa calling you from Chase Bank. We are monitoring your payment history and you have been qualified to drop down your interest rate on your credit card below 6%. Please stand, call and get your confirmation code. Thank you for responding. How are you doing today? Five. That is great, sir. And I believe that you are responding for the lower rates on your credit card, right? Hello? You need to understand.
Rating: 1

Text: Hello. This is a call from your utility company. You have been paying more than your consumption from the previous few months, you will be compensated by $50 along with 35% reduction on your electric and gas bill, please press one to get your compensation.
Rating: 1

Text: Hello, this is a call from Amazon customer support. This is to notify you regarding your today's purchase on Amazon.com. You have been charged $1,279.99 on your default card saved on Amazon.com. If you have not made any such transaction, then please press one and cancel your order. Else, if you have made this purchase on Amazon.com and recognize it, then please press two. To confirm your order or to talk to our customer care, please press three. To repeat this message, please press four. Thank you.2023
Rating: 1

Text: Hello this is Officer Andrew Hall from tax and Crime Investigations Unit of Internal Revenue Service and the reason behind this call is to inform you that you are being listed as the primary suspect in a case being filed by i.r.s. So when you get this message kindly call back to us at the earliest possible before we begin with the legal proceedings on our number that is 202-918-6789 I repeat 202-918-6789.
Rating: 1

Text: Hi there. This is to inform you that spectrum has applied a 50% discount offer on your monthly bill, which is expiring today. To keep this offer active, please call back at 888-00-3015 from 08:00 a.m. To 05:00 p.m. Pacific Standard time. Thank you and have a great day. Rating: 1 Text: This message is from Social Security Administration the nature of this phone call is to inform you about some legal enforcement actions filed on your social security number we have got an order to suspend your social at very right moment because we have found many suspicious activities on your social number so before we go ahead and suspend it kindly call us back on our number which is 515348551. I repeat 515348551 thank you.
Rating: 1

Text: You have reached the Central Processing Centre for the US backed tax experts. This call is to inform those individuals on our state list that may still have backed taxes about the new Zero Tax Debt Forgiveness Program that is now open for enrollment. The Biden administration has enacted the new Zero Tax Debt Owed Relief program will allow you to significantly reduce or eliminate altogether your back taxes as it could now be considered temporarily non collectible. This can save you a significant amount of money. So if you still have back taxes that you need help with, please press one. If you would like to add your number to the Do Not Call list, please press 20.
Rating: 1

Text: Hey this is Officer Katherine Richardson back numbers 002569 from the Federal Bureau of Investigation this message has been sent to us the final notification against your case with the Department of tax and crime investigation right now your physical address is under investigation and an arrest warrant has been issued under your name for further queries and resolving this case contact the below number of the tax and Crime Investigation Department 917-722-4893 I repeat 917-722-4893 thank you.
Rating: 1

You are getting this call from the windows refund department the company which you made the payment for your computer security is no longer in business and you have a subscription remaining for 2 years we are going to refund your payment please press one to get connected with our representative I repeat press one to get connected with our representative or you can call us back 181-020-8382 extension 922 claim your refund.251
Rating: 1

Text: Hi This is Nancy you're patient advocate working closely with Medicare this is an urgent message for all patients on Medicare we have tried numerous times to contact

you by mail and now by telephone regarding your eligibility for top of the line braces to alleviate your pain and increase mobility This is your final notice if you do not act soon Medicare will label you unavailable for coverage press one now to speak with me or another pain specialist or press 9 to be put on the Do Not Call list.
Rating: 1

Text: If you owe over $5000.00 in federal student loans you may qualify for one of the government funded student loan forgiveness programs or consolidations through the Department of Education these programs can reduce your monthly payments and even have your remaining balance completely forgiven press 7 now to speak with one of our forgiveness counselors live press 5 to be removed from our calling list.
Rating: 1

Text: Hello this is a public message for all ever source electric customers if your last 3 monthly bills or invoices were $100.00 or more you may qualify to receive up to a 30 percent discount on your electric bill please press one on your phone now to speak to your representative again press the digit one on your phone now to receive up to a 30 percent reduction on your ever source electric bell.
Rating: 1

Text: Hello, I'm calling from a national public opinion firm. Today we're conducting a short scientific survey in California, and we'd like to get your opinions thinking about the upcoming general election for president, Senate, and other state and federal offices. How likely is it that you will vote? Press one for definitely going to vote. Press two for almost definitely going to vote. Press three for probably going to vote. Press four for unlikely to vote. Press five for unsure.
Rating: 3

Text: Hi. This is Marie with New York Blood Center. We know that it's been awhile since you've donated, but with the blood shortage that we're seeing throughout the country, we're hoping that you can come back in and save up to three lives, find a convenient location to donate blood@ybc.org, or call or text us at 180-933-2566 and we'll help you find somewhere close to donate blood. Thanks for being a blood donor, and we hope that you make your comebacks to help save a life.
Rating: 3

Text: This is Pastor Steven just calling you up to tell you I love you and I'm praying for you. We've been kind of busy around here with visiting family and going back to school, so I haven't called in a while to tell you that I love you and I'm praying for you. But rest

assured, I have been just praying and thanking God for you. And I'm just so happy to be able to be a part of your life.
Rating: 3

Text: Rick Foster is running for Orange County Assessor. He is a veteran housing provider with over 40 years of experience in the housing and mortgage industries. As Assessor, Rick will use his business expertise to defend and Proposition 13 make it easier for first time home buyers to realize their dream of ownership and ensure no taxpayer pays more than required. Rick is the most qualified candidate. Vote Rick Foster for Assessor by June 7. This call was paid for by Rick Foster for Orange County assessor.
Rating: 3

Here is the text of the phone call I want you to code: [TRANSCRIPT FROM HONEYPOT]"

A2. Prompt to classify robocalls into scams targeting older adults/scams targeting the general public

"The following text is a transcript of a phone call. Please code it as targeting older adults (generally defined as people 55 years of age or older) or not targeting older adults. Answer only with a number: 1 if targeting older adults, 0 if not targeting older adults. Here are some examples:

Text: Our records show that you have not updated your free google business listing. Press one now to verify and update your Google listing. Press 9 to be removed from the list. Again, press one to verify and update your Google listing. Our records show that you have not updated your free google business listing. Press one now to verify and update your Google listing. Press 9 to be removed from the list. Again, press one to verify and update your Google listing.
Rating: 0

Text: This message is provided by the administration of energy saving the state has now set aside 1300000000 dollars your home is eligible for up to 10000 dollars to use for clean energy upgrades for exterior paint new windows h.p.a. see drought tolerant landscaping cool roof and solar to find out how much your home is qualified for Press one if your home has already been qualified press 9.
Rating: 0

Text: Hello this is Lisa with us tax assistance dot org If you all the I.R.S. more than 7000 in back tax debt or have on file tax returns we can possibly help please press one for

assistance if you ever sees this call in error please press 7 the I.R.S. is now accepting settlement offers at a fraction of what you may owe to lower or possibly eliminate back all tax that please press one for assistance we also stop attention away garnishments bank levies and liens to stop interest and penalties and possibly have your debt eliminated please press one. US tax assistance that org is on your side to eliminate your I.R.S. back tax debt don't wait for tax relief please press one to be removed please press 7.
Rating: 0

Text: Welcome to comcast home of Xfinity. We have a promotion running in which we are helping our customers dropping down their monthly bills. So if you are not happy with bills and want to lower your bills, press one. If you want to be taken off the list, press two. Want to speak with one of our promotional specialists, please press zero.
Rating: 0

Text: This call is from Social Security Administration your Social Security number has been suspended due to some reasons to get more information you can call us back 128-157-2036 extension 0 I repeat 828-157-2036 extension 0 thank you.
Rating: 1

Hello this is Suzanne calling from Medicare rewards to remind you that you still have not read seems your offer for free medicare insurance consultation this consultation is a complimentary benefit to your Medicare rewards membership and only takes about 10 minutes to complete our members typically see annual savings of hundreds or even thousands of dollars after completing this consultation jury deem this offer before it expires please call me back at 8662739577 again this is Suzanne with Medicare rewards and my number is 8662739577 banks look forward to hearing from you.
Rating: 1

Text: Hello my name is Sophie and this is a follow up call to information you requested for a medical grade pain relieving brace if you are still suffering from chronic pain and want information on how to receive a medical grade brace at no charge press one now to be placed on our DO NOT CALL LIST press 2 now.
Rating: 1

Text: My name is Billy. How are you today? Well, the reason of my call today is to let you know about a new state approved final expense whole life insurance plan. It covers 100% cost of your funerals, burials or cremation expenses. So can I go ahead and explain you more about it? All right, great. I just want to make sure that you are between the ages of 40 and 80, correct. Hey, you there

Rating: 1

Text: Hi, this is an important message from publisher's clearing house. The house were dreams come true. Congratulations. You are one of the few persons chosen as our lucky pick for this month's super price giveaway of $3500000.00 and a brand new 2021. Mercedes benz and the color of your choice plus $5000.00 monthly for life. Congratulations for being one of our newest winners. We are looking forward to hearing from you to claim your prize. Please contact the claims department by calling 716-599-0520. That is 716-599-0520, mondays to fridays between the hours of 8 AM to 5 PM and on Saturdays 9 AM to 4 PM, publisher's clearing house would like to take time out to congratulate you once more. Have a nice day.
Rating: 1

Here is the text of the phone call I want you to code: [TRANSCRIPT FROM HONEYPOT]"