Original Paper

Alzheimer Disease and Related Dementias Caregiver Burden on Reddit: Large Language Model–Assisted Measurement Study

Yan Zhang, PhD; Mingfei Zhang, PhD

Texas State University, San Marcos, TX, United States

Corresponding Author:

Yan Zhang, PhD
Texas State University
461 Trauth-Huffman Hall, 604 N Guadalupe St
San Marcos, TX 78666
United States
Phone: 1 7164286780
Email: zhangy@txstate.edu

Abstract

Background: Informal caregivers of older adults with Alzheimer disease and related dementias (ADRD) often face high burden and chronic stress. Reddit provides naturally occurring caregiver narratives to study caregiving burden, but prior studies have often relied on small manual coding or aggregate topic models that may not easily connect caregiving context, burden, and stressors at scale.

Objective: This study aimed to explore what a large language model (LLM)–assisted measurement workflow can recover from ADRD caregiving-related Reddit discourse, such as caregiver burden, caregiving contexts, stressors, temporal patterns, and high-burden subgroups.

Methods: We processed monthly Reddit submission archives from January 2020 through June 2025. Posts were selected from 5 ADRD-specific subreddits and 4 broader caregiving or aging subreddits, yielding a balanced 13,200-post corpus. A locally hosted open-weight LLM inferred structured post-level variables, including caregiver sociodemographic characteristics, caregiver role, ADRD stage, care setting, support need, dominant emotion, sentiment, burden, and primary and secondary stressors. The workflow included schema validation, uncertainty ratings, field-level correction prompts, source matching, a 150-post double-coded human validation benchmark, and pairwise subgroup analyses.

Results: The source-matched analytic sample included 13,179 LLM-coded posts. Human agreement was high for caregiver role, ADRD stage, care setting, dominant emotion, support need type, burden group, and stressor categories (Cohen κ=0.77-1.00), but poor for the initial in-scope screen (κ=0.00). Adult children were the largest inferred caregiver group (8,266/13,179, 62.7%), followed by grandchildren (2,012/13,179, 15.3%). Moderate ADRD was the modal inferred stage (6,554/13,179, 49.7%), and home was the modal care setting (7,145/13,179, 54.2%). Mean burden increased from early-stage ADRD (5.38, SD 2.47) to moderate-stage ADRD (6.96, SD 2.17) and late-stage ADRD (7.77, SD 2.16). Behavioral strain was the dominant primary stressor, while family conflict, financial strain, emotional strain, and social isolation were leading secondary stressors. Pairwise subgroup analysis showed high burden in late-stage ADRD with emotional support needs (mean 8.42, SD 1.46), care transitions with emotional support needs (mean 8.45, SD 1.12), and late-stage care transitions (mean 8.40, SD 1.32).

Conclusions: LLM-assisted measurement can help convert unstructured ADRD caregiving-related Reddit discourse into post-level, theory-guided quantitative measures when paired with validation and cautious interpretation. This workflow extends prior online caregiver research by linking caregiving context, burden, and stressor pathways at scale. The findings highlight care transitions, emotional support needs, financial navigation, and family role negotiation as potential targets for caregiver-support interventions.

JMIR Aging 2026;9:e103305; doi: 10.2196/103305

Keywords: Alzheimer disease; dementia; caregivers; social media; natural language processing; large language models; psychological distress; digital health

Introduction

In the United States, an estimated 7.4 million adults aged 65 years and older were living with clinical Alzheimer dementia in 2026, and 12.7 million family members and other unpaid caregivers provided care for people living with Alzheimer disease or related dementias (ADRD) in 2025 [1]. Caring for people with ADRD is usually long-term and demanding because caregivers must manage progressive cognitive decline, behavioral symptoms, functional dependence, and repeated care transitions [2]. Informal ADRD caregivers help support aging in place and reduce demand for formal long-term care [3,4], but caregiving can also expose them to social isolation, role overload, financial strain, family conflict, and health decline [5,6]. The stress process model developed by Pearlin et al [7] describes stress as a process in which life strains, social roles, coping resources, and social support shape psychological distress. Its caregiving-specific extension distinguishes primary stressors arising directly from care work, secondary stressors that spill over into the caregiver's broader life, and resources that may buffer burden [8]. This framework is useful for ADRD caregiving because dementia-related symptoms such as wandering, sundowning, hallucinations, aggression, and care refusal can create both immediate care demands and downstream conflicts over money, work, and family.

Despite the importance of ADRD caregivers' experiences, conventional surveys and interviews may only partially capture the complexity of informal caregiver burden. Population surveys are essential for inference but often rely on relatively thin, pre-specified measures [9,10]. Qualitative interviews and focus groups provide depth but are usually small, difficult to replicate, and less suited to tracking change over time [11,12]. ADRD caregiving also involves stigmatized emotions and difficult moral concerns, including guilt, ambivalence, and resentment, that may be under-disclosed in structured interviews because of social desirability pressures [13]. Online communities therefore provide a complementary source of naturally occurring caregiver narratives. Recent studies using online platform data show that caregivers disclose health concerns, relationship strain, emotional problems, financial burden, support-seeking needs, and positive aspects of caregiving in these spaces [14-16]. These studies show the value of online narratives, but traditional manual coding methods remain limited in scale because they require intensive human review. For example, 2 studies by Huang et al [14,15] analyzed >700 Reddit posts and >3000 comments to summarize caregivers' stressors and experiences. These sample sizes are appropriate for close content analysis, but they make it harder to infer caregiving contexts and study subgroup comparisons.

Machine learning, traditional natural language processing (NLP), and broader text-as-data approaches have helped researchers analyze larger collections of online narratives. This literature shows that computational methods can convert large volumes of language into analyzable evidence, which can make large-scale analysis faster, less costly, and more reproducible when paired with careful validation [17]. In dementia caregiving research, sentiment analysis, supervised classification, and topic modeling have been used to study online posts. These studies have identified core themes in large online corpora, including family issues, care facilities, emotional distress, advice seeking, financial struggles, abuse concerns, and sleep problems [18-24]. For example, Cheng and Yin [18] used structural topic modeling to analyze Reddit and ALZConnected posts and found recurring needs related to family issues, disease stages, medical concerns, and care facilities. Ni et al [19] used a 2-stage machine learning framework to identify ADRD caregiving posts on Reddit and showed that caregivers discussed learning challenges, care experiences, and emotional support across ADRD and non-ADRD communities. This work is important because it moves beyond small samples and manual coding. However, many of these approaches are better suited to summarizing broad themes than to preserving more complex post-level information, such as how stressors cluster, how caregiver characteristics relate to stressors, or how burden varies across care contexts.

Large language models (LLMs) and generative AI tools are now being applied in caregiving and dementia research. Some studies have used LLMs to organize caregiving tasks, map digital caregiving functions, or build caregiver-facing support tools. For example, Kim et al [25] used ChatGPT 3.5 (OpenAI) to analyze caregiving education materials and mobile app descriptions across ADRD stages. Hasan et al [26] developed a conversational AI system for Alzheimer disease caregivers and found stronger usability and more relevant support than a baseline ChatGPT model in a user study. He et al [27] used GPT-4o in a 2-stage topic-modeling workflow for more than 230,000 dementia-caregiver tweets and found topics related to emotional burden, cognitive decline, family relationships, caregiving activities, and adaptation or support resources. These studies show the growing role of LLMs in dementia-caregiver research, but much of this work has focused on support tools, educational content, or aggregate topic discovery.

LLMs may also help address a long-standing scale-depth tradeoff in caregiver research. Manual qualitative coding offers depth but is slow and costly, while surveys and fixed-label classifiers offer scale but often reduce narratives to a small set of predefined measures [11,17]. Compared with traditional BERT (Bidirectional Encoder Representations from Transformers)-style classifiers, keyword filters, or bag-of-words topic models, LLMs can use broader context to identify caregiving contexts, burden, and distress even when caregivers use indirect language, euphemisms, or situational descriptions [28-30]. This may reduce reliance on a short list of predefined keywords and allow richer constructs than broad topics alone. LLMs can also provide a relatively fast and inexpensive way to apply structured interpretation

to large text corpora while still preserving narrative detail [27]. They can function as measurement tools that convert unstructured text into structured quantitative data for subgroup comparison, such as caregiver role, demographic attributes, burden level, and stressor types [31]. In ADRD caregiving research, these advantages point to a remaining gap: prior studies have identified WHAT questions (main caregiving themes or patterns), but less is known about HOW questions with respect to how caregiver contexts are associated with their needs, burden, and stressors, and how large-scale narrative data can be used to identify the most vulnerable subgroups. LLM-assisted measurement may help describe these patterns more comprehensively and inform policy design, social implications, and caregiver support interventions.

To address these gaps, this study examines what an LLM-assisted measurement workflow can recover from ADRD caregiving-related Reddit discourse. Guided by the stress process model, we sought to move beyond aggregate topic discovery by measuring caregiving contexts, burden, sentiment, and primary and secondary stressors at the post level. Specifically, we asked the following research questions (RQs):

- RQ1: What caregiver contexts, such as roles, ADRD stages, care settings, support needs, sentiment levels, and burden levels, can be identified in ADRD caregiving posts on Reddit?
- RQ2: How do sentiment and burden differ across inferred caregiving contexts?
- RQ3: How do primary stressors and secondary stressors appear and pair together within posts?
- RQ4: How do temporal patterns in caregiving burden and sentiment change over time?
- RQ5: Which subgroups of caregiver contexts and stressor types are associated with especially high burden?

To answer these questions, we first assembled a long-term Reddit corpus from ADRD-specific subreddits and ADRD-keyword-filtered posts in broader caregiving communities. We then used a locally hosted instruction-tuned LLM to infer structured post-level variables, including demographic attributes, caregiver role, ADRD stage, care setting, sentiment, burden, and stressors. We next validated key fields with a human-coded benchmark, summarized confidence coverage, examined post-level patterns, assessed temporal trends, and identified high-burden subgroup intersections relevant to caregiver support. This study contributes a scalable workflow for measuring ADRD caregiving-related Reddit posts and shows how LLM-assisted measurement can support clinical aging research, digital health design, and future caregiver-support interventions.

Methods

Study Design

The workflow had 2 steps. Step 1 built the Reddit post corpus. Step 2 applied LLM coding to the retained posts.

Step 1: Data Collection and Post Extraction

In step 1, we constructed the source corpus from monthly Reddit submission archives provided through Pushshift, a large-scale Reddit data collection infrastructure widely used in social media research [32]. In the completed scrape, we processed 66 monthly archives spanning January 2020 through June 2025. For each monthly archive, we restricted the initial search space to 9 subreddits selected a priori for likely ADRD caregiving content. Five were ADRD-specific communities: r/Alzheimers, r/dementia, r/lewyBody, r/FrontotemporalDementia, and r/vascular_dementia. Posts from these communities were treated as potentially relevant. Four were broader aging or caregiving communities: r/AgingParents, r/CaregiverSupport, r/eldercare, and r/elderlycare. Posts from these broader communities were retained only if the title or body matched an ADRD keyword list covering disease names, cognitive or behavioral symptoms, and dementia-specific care terms. The subreddit selection was informed by prior Reddit caregiver and ADRD caregiver studies that sampled dementia-specific forums as well as broader aging and caregiver forums where dementia caregiving is discussed [14,15,23,33,34].

Across all 66 monthly archives, the pipeline scanned 2,318,164,645 submission records. Subreddit and keyword filtering yielded 41,612 candidate posts, including 33,264 posts from ADRD-specific subreddits and 8348 posts from general caregiving subreddits that satisfied the ADRD keyword filter. Before final ranking, the extraction script excluded posts that were deleted, removed, or empty (n=14,328) and posts with fewer than 50 characters (n=471), leaving 26,813 eligible posts. Candidate posts substantially increased over time, from 207 eligible posts in April 2020 to 1409 in March 2025. Table S1 in the Multimedia Appendix 1 displays the corpus funnel for post extraction.

To prevent later, higher-volume months from dominating the corpus, we retained the top 200 eligible posts by Reddit score within each month after filtering and post-ID deduplication. Because author metadata was not retained in the monthly extracted files, we could not estimate unique contributors or deduplicate by author. A conservative text-level audit identified 112 exact or near-duplicate posts (0.85%); because their removal had minimal impact on the results, they were retained. Accordingly, all reported percentages and means describe posts rather than unique caregivers and may include multiple posts from the same author. This produced a final step-1 corpus of 13,200 posts (200 posts×66 months), including 10,278 (77.9%) from ADRD-specific subreddits and 2922 (22.1%) from broader caregiving communities. The retained source corpus was concentrated in r/dementia (61.1%), followed by r/Alzheimers (16.8%), r/AgingParents (12.0%), and r/CaregiverSupport (8.5%); r/eldercare contributed 1.5%, and r/elderlycare contributed fewer than 0.1%. After LLM output was joined back to the final monthly source corpus by post-ID, 13,179 source-matched LLM-coded posts remained for analysis (99.8% of the retained source corpus). This ID-matching step excluded inference rows that did not correspond to the final retained monthly

source files and prevented nonsource records from entering the analytic sample.

Step 2: LLM Inference Procedure

In step 2, we converted each retained Reddit post into a structured measurement record using a locally hosted open-weight instruction-tuned large language model trained by Google DeepMind (Gemma 4 26B A4B) [35,36]. The model was served locally through Ollama on researcher-controlled hardware using the quantized model identifier gemma4:26b-a4b-it-q4_K_M. Local deployment kept post text within the researchers' computing environment, enabled use of a fixed model version, and supported large-scale inference without per-request commercial API charges. The prompting framework treated the LLM as a narrative measurement instrument rather than a generator of free text, following recent work on LLMs as scalable text annotation and computational social science tools [29,30]. This differs from recent LLM dementia-caregiver studies that used LLMs for digital-caregiving task mapping, retrieval-augmented caregiver support, or aggregate topic modeling of caregiver social-media text [25-27]. Rather, each post was supplied with explicit instructions to infer attributes of the caregiver and to return a constrained structured output.

Informed by prior Reddit caregiver studies [14,15,18,20,23], we inferred 15 variables spanning 4 domains. First, caregiver sociodemographic variables included gender, age, race, education, marital status, and income. Second, caregiving-context variables included caregiver role, ADRD stage, and care setting. Third, stress-process variables included primary stressor and secondary stressor. Fourth, socioemotional variables included dominant emotion, support need type, sentiment, and burden level.

We defined primary stressor and secondary stressor according to the Stress Process Model before running the LLM [8]. The prompt instructed the LLM to read each post and select a short phrase from the narrative that best matched the predefined concept. Primary stressor was defined as the main direct caregiving burden described in the post, such as symptom management, care refusal, supervision demands, or behavioral strain. Secondary stressor was defined as the main spillover burden affecting the caregiver's broader life, such as sibling conflict, financial strain, work disruption, or social isolation. To reduce variation across posts, both stressor fields were constrained to a single phrase of no more than 2 words, or (not available) N/A when no such stressor was clearly present. Dominant emotion was constrained to 1 of 7 categories: anger, sadness, fear, joy, bittersweet, hope, or neutral, following the emotion categories used in prior analysis of supportive communication among Reddit caregivers [15]. Support need type was constrained to 1 of 7 support categories plus N/A: emotional, informational, instrumental, financial, respite, medical, or legal.

Most inferred variables were assigned a 3-level confidence rating (low, mid, or high) intended to capture the degree to which the variable was explicitly stated versus contextually inferred. Confidence ratings were assigned for all sociodemographic variables, caregiving-context variables, support need type, sentiment, and burden level. However, dominant emotion and the 2 stressors were returned without a confidence rating because they were framed as direct summarizations of the narratives. Sentiment and burden level were constrained to a 0-10 scale, with higher values indicating more positive tone and greater caregiver burden, respectively. The workflow recorded the model's structured label and a low/mid/high confidence category where applicable, but it did not collect explanatory rationales or feature attributions. Confidence therefore reflects the model's requested assessment of textual support; it should not be interpreted as a calibrated probability or an explanation of the internal reasoning that produced a label.

The full prompt template used for LLM inference is provided in Multimedia Appendix 2. An abbreviated prompt logic is summarized below:

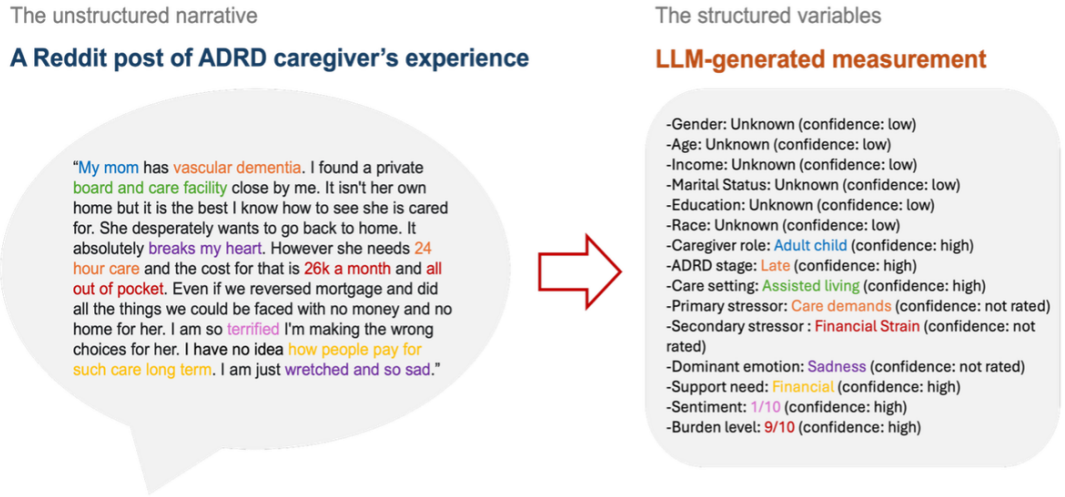
> *Analyze each post as a caregiver narrative about ADRD. Infer caregiver attributes only, not care-recipient attributes. Return one structured record per post. Primary_Stressor=main direct caregiving stressor (1-2 words or N/A) Secondary_Stressor=main spillover stressor (1-2 words or N/A) Dominant_Emotion=anger/sadness/fear/joy/bittersweet/hope/neutral Support_Need_Type=Emotional/Informational/Instrumental/Financial/ Respite/Medical/Legal/N/A Sentiment=0-10 Burden_Level=0-10*

After the model returned a record, we checked whether the output followed the study rules. These checks included whether category labels came from the allowed list, confidence ratings were present when required, sentiment and burden scores stayed within the 0-10 range, and stressor phrases were no longer than two words. If the whole record was unusable, we asked the model to process the post again. If only one field had a problem, we asked the model to correct only that field. We repeated this field-level correction up to 3 times. This process helped keep correct parts of a record while fixing specific errors.

This design had 2 purposes. First, it made the model output easier to analyze because each post had the same set of variables and the same allowed response options. Second, the confidence ratings helped separate information that was clearly stated in a post from information that required more interpretation. Records with lower confidence or repeated correction needs could then be treated more cautiously in later analyses.

Figure 1 illustrates how the LLM converted a Reddit caregiver narrative into structured inferred variables with confidence ratings.

Figure 1. Example of LLM (large language model)-assisted measurement converting an unstructured narrative into structured inferred variables. The narrative shown in the figure was rephrased to reduce traceability. ADRD: Alzheimer disease and related dementias.



Human Validation Benchmark

To assess the reliability of the LLM-derived measurement fields, we created a 150-post human validation benchmark from the source-matched analytic corpus. The benchmark was selected to include variations in support need, ADRD stage, caregiver role, and burden level. Two coders (YZ and MFZ) independently reviewed the post title and narrative and assigned labels for whether the post was in scope, caregiver role, ADRD stage, care setting, dominant emotion, burden level, support need type, and primary and secondary stressor categories. For each field, we calculated percent agreement and Cohen kappa between the 2 human coders. We then compared LLM labels with human consensus labels for fixed-label fields when both coders agreed, and the LLM output used the same label space. Open-vocabulary primary and secondary stressor phrases were not directly compared with consensus labels because the human benchmark coded broader stressor categories rather than exact LLM phrases.

Analytic Strategy

We conducted post-level descriptive and exploratory analyses after merging the LLM output with monthly metadata. We first examined inference coverage and confidence ratings to show which variables were usually stated clearly in posts. We then evaluated the human validation benchmark to assess whether the LLM-derived fields aligned with independent human coding. After these measurement checks, we described inferred caregiving contexts and burden variables captured in the Reddit corpus. These summaries provided the baseline portrait of who was posting in related subreddits, what care situations they described, what kinds of support they appeared to seek, and how much emotional strain was present.

We then examined whether burden and sentiment varied across nondemographic caregiving groups. This step linked the descriptive fields to caregiver strain by comparing mean burden and sentiment across caregiver contexts. Next, we analyzed primary and secondary stressors. Rather than assigning every post to a fixed stressor checklist, the LLM first summarized the main direct caregiving stressor and the main spillover stressor in each post as short phrases. We then grouped similar phrases after inference using token overlap, shared leading terms, and string similarity. Primary and secondary placements were retained separately so that we could distinguish stressors tied directly to care tasks from stressors that reflected broader effects on the caregiver's family, finances, work, or social life. We also examined quarterly temporal patterns in burden, sentiment, and caregiving contexts, using the post as the unit of analysis. This step was used to assess whether the emotional and practical content of posts changed over the 2020-2025 study period.

Finally, we used the structured variables to identify high-burden subgroup intersections that may be most relevant for support intervention. We examined all pairwise combinations of caregiver role, ADRD stage, care setting, and support need type. For each pair, we calculated the number of posts and mean burden for each subgroup and reported the highest-burden cell among those with more than 200 posts. We also examined the 3 most frequent primary stressor topics and the 3 most frequent secondary stressor topics, reporting subgroup patterns only when they had more than 200 posts and mean burden above 8.

Ethical Considerations

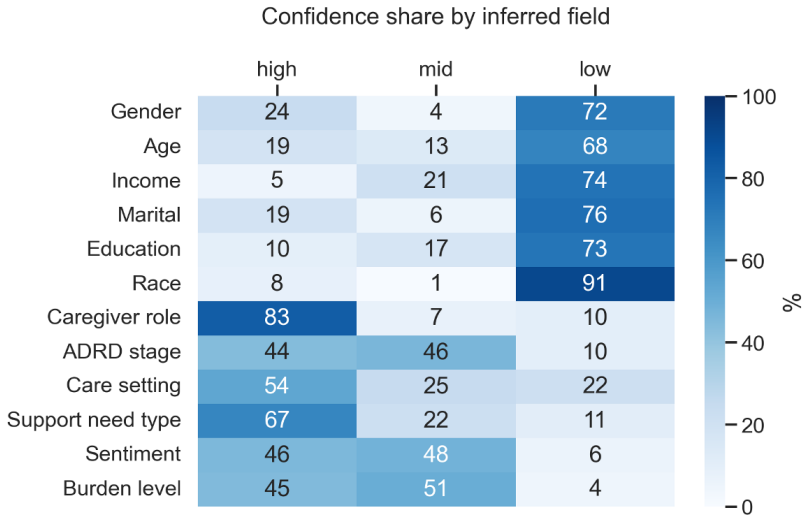
This study was reviewed by Texas State University Research Integrity and Compliance and was determined not to meet the criteria for human subjects research Institutional Review Board (IRB-26-161). The project involved a secondary analysis of publicly accessible Reddit data and did not involve the recruitment of human participants; therefore, informed consent and participant compensation were not required. To protect confidentiality and anonymity, the paper does not report usernames, URLs, post identifiers, or individual user-level information. Results are reported mainly in aggregate form at the post level, and post excerpts were rephrased and shortened to reduce traceability. These procedures are consistent with ethical approaches used in published Reddit caregiver studies involving public data and IRB-exempt or secondary-data designs [14,20].

Results

Inference Coverage

A high proportion of sociodemographic fields were coded as low confidence, including gender, age, income, marital status, education, and race (8962/13,179, 68% to 11,993/13,179, 91%) (Figure 2). By contrast, several variables were more consistently identified from narratives: caregiver role, ADRD stage, care setting, and support need type had combined mid- and high-confidence shares of 90% (11,861/13,179), 90% (11,861/13,179), 78% (10,280/13,179), and 89% (11,729/13,179), respectively. Sentiment and burden level also had relatively high confidence coverage, with low-confidence ratings in only 6% (791/13,179) and 4% (527/13,179) of posts, respectively. In the following analysis, we skipped reporting analysis by demographic characteristics as they show low confidence ratings.

Figure 2. Confidence coverage across inferred variables. ADRD: Alzheimer disease and related dementias.



Human Validation of LLM-Derived Measures

Human coders had perfect agreement for caregiver role, ADRD stage, care setting, and dominant emotion (all 100.0% agreement; Cohen κ=1.00). Agreement was also high for support need type (99.3%; κ=0.99) and burden level group (97.3%; κ=0.94). Agreement for stressor categories was lower but still substantial: 80.7% agreement for primary stressor category (κ=0.77) and 84.0% for secondary stressor category (κ=0.80). These results support the reliability of the structured constructs used in the main analyses, while also showing that stressor category assignment remains more interpretive than role, stage, setting, emotion, support need, or burden coding.

The main validation weakness was the initial in-scope screen. One human coder (MFZ) applied the broad "from or about an informal caregiver" criterion to all 150 benchmark posts, whereas the second coder applied a stricter authorship threshold and labeled 87 posts "yes," 49 "no," and 14 "unclear." Thus, human coders agreed on only 58.0% of in-scope labels (κ=0.00), with disagreement concentrated in posts that were clinically or dementia-relevant but less clearly first-person informal caregiver narratives. This finding is consistent with the extraction design: posts from ADRD-specific subreddits were retained without a caregiver-perspective keyword requirement, which increased sensitivity but admitted some general discussion, patient-perspective, news, or ambiguous community posts. Accordingly, corpus-level proportions should be interpreted as describing a high-engagement ADRD caregiving corpus rather than a sample of confirmed caregiver-authored posts. The benchmark was diagnostic; no adjudication or post hoc in-scope filtering was used to alter the 13,179-post analytic corpus.

For fixed-label fields with human consensus labels, the LLM labels matched the human consensus labels in all compared cases: caregiver role (150/150), ADRD stage (150/150), care setting (150/150), dominant emotion (150/150), support need type (149/149), and burden level group (146/146). Because primary and secondary stressors were extracted as open-vocabulary phrases and humans coded broader categories, LLM–human agreement for stressors was evaluated indirectly through the human category reliability results rather than exact phrase matching. Table A2 in the Multimedia Appendix 3 presents agreement statistics for the 150-post double-coded validation benchmark.

Inferred Caregiving Contexts, Emotion, Sentiment, and Burden

Table 1 presents the inferred caregiving contexts, emotional tone, sentiment, and burden level in the source-matched LLM-coded corpus. Among 13,179 posts, caregiver role was dominated by adult children (8266/13,179, 62.7%), followed by grandchildren (2012/13,179, 15.3%) and spouses (523/13,179, 4.0%). Smaller role groups included formal caregivers (202/13,179, 1.5%), friends (75/13,179, 0.6%), and siblings (57/13,179, 0.4%), while caregiver role was unknown in 15.5% (2044/13,179) of posts. ADRD stage was most often coded as moderate (6554/13,179, 49.7%), followed by late

stage (2935/13,179, 22.3%) and early stage (1675/13,179, 12.7%); stage was unknown in 15.3% (2015/13,179) of posts. Care recipients were most often described as living at home (7145/13,179, 54.2%). Care transitions were identified in 6.8% (891/13,179) of posts. Institutional settings included memory care (728/13,179, 5.5%), nursing homes (531/13,179, 4.0%), and assisted living (452/13,179, 3.4%), while care setting was unknown in 26.0% (3432/13,179) of posts. The modal combination of adult-child caregivers and moderate-stage ADRD likely suggests that much of the Reddit disclosure captured by this workflow occurs during the period when cognitive and behavioral symptoms are prominent, but care has not yet fully shifted to institutional settings.

Table 1. Inferred characteristics in the source-matched LLM-coded[a] corpus (N=13,179).

Variable and categories/statistics	Values
Caregiver role, n (%)	
Adult child	8266 (62.7)
Grandchild	2012 (15.3)
Spouse	523 (4.0)
Formal caregiver	202 (1.5)
Friend	75 (0.6)
Sibling	57 (0.4)
Unknown	2044 (15.5)
ADRD[b] stage, n (%)	
Early	1675 (12.7)
Moderate	6554 (49.7)
Late	2935 (22.3)
Unknown	2015 (15.3)
Care setting, n (%)	
Home	7145 (54.2)
Transition	891 (6.8)
Memory care	728 (5.5)
Nursing home	531 (4.0)
Assisted living	452 (3.4)
Unknown	3432 (26)
Support need, n (%)	
Informational	5843 (44.3)
Emotional	4453 (33.8)
Respite	190 (1.4)
Legal	131 (1.0)
Medical	48 (0.4)
Financial	27 (0.2)
Instrumental	18 (0.1)
No explicit support needed	2469 (18.7)
Dominant emotion, n (%)	
Sadness	3839 (29.1)
Fear	3364 (25.5)
Anger	2273 (17.2)
Bittersweet	1655 (12.6)
Hope/joy	722 (5.5)
Neutral	1326 (10.1)
Sentiment[c], mean (SD)	2.61 (1.96)
Burden level[c], mean (SD)	6.53 (2.70)

[a]LLM: large language model.
[b]ADRD: Alzheimer disease and related dementias.
[c]N=13,179.

Support needs were primarily informational or emotional. Informational support was the dominant type (5843/13,179, 44.3%), followed by emotional support (4453/13,179, 33.8%). Posts with no explicit support needed accounted for 18.7% (2469/13,179). Explicit respite (190/13,179, 1.4%), legal (131/13,179, 1.0%), medical (48/13,179, 0.4%), financial (27/13,179, 0.2%), and instrumental (18/13,179, 0.1%) support needs were much less frequent. Table 1 also shows that the corpus was mainly negative in emotional tone. Sadness was the most common dominant emotion (3839/13,179, 29.1%), followed by fear (3364/13,179, 25.5%), anger (2273/13,179, 17.2%), bittersweet emotion (1655/13,179, 12.6%), neutral tone (1326/13,179, 10.1%), and hope or joy (722/13,179, 5.5%). The mean sentiment score was 2.61 (SD 1.96) on a 0-10 scale, where higher values indicate more positive tone. The mean burden score was 6.53 (SD 2.7) on a 0-10 scale, where higher values indicate greater caregiver burden.

Sentiment and Burden by Nondemographic Groups

Table 2 presents mean sentiment and mean burden scores across all nondemographic inferred variable groups. By caregiver role, adult-child posts had the highest mean burden score of 7.02 (SD 2.29) and the lowest mean sentiment score of 2.40 (SD 1.79), spouse posts had a similar mean burden score of 7.00 (SD 2.31) and mean sentiment of 2.47 (SD 1.84), and grandchild posts had a mean burden score of 6.50 (SD 2.58) and mean sentiment of 2.56 (SD 1.95). Posts coded as formal caregivers, friends, or unknown caregiver role had lower mean burden scores (formal caregivers: 4.01, SD 3.3; friends: 4.71, SD 2.72; and unknown: 4.76, SD 3.39), respectively, while sibling posts had a mean burden score of 6.70 (SD 2.76).

Table 2. Mean inferred sentiment and burden by nondemographic inferred variable group.

Group	Number of posts	Mean sentiment (SD)	Mean burden (SD)
Caregiver role			
Adult child	8266	2.4 (1.79)	7.02 (2.29)
Grandchild	2012	2.56 (1.95)	6.5 (2.58)
Spouse	523	2.47 (1.84)	7 (2.31)
Formal caregiver	202	4.13 (2.50)	4.01 (3.30)
Friend	75	2.85 (1.64)	4.71 (2.72)
Sibling	57	2.47 (1.88)	6.7 (2.76)
Unknown	2044	3.43 (2.30)	4.76 (3.39)
ADRD[a] stage			
Early	1675	2.88 (1.74)	5.38 (2.47)
Moderate	6554	2.45 (1.79)	6.96 (2.17)
Late	2935	2.05 (1.72)	7.77 (2.16)
Unknown[b]	2015	3.76 (2.45)	4.27 (3.39)
Care setting			
Home	7145	2.36 (1.73)	7.01 (2.29)
Transition	891	2.07 (1.49)	7.82 (1.67)
Memory care	728	2.84 (2.16)	6.59 (2.48)
Nursing home	531	2.23 (1.74)	7.17 (2.31)
Assisted living	452	2.57 (1.84)	6.79 (2.29)
Unknown	3432	3.3 (2.31)	5.05 (3.22)
Support need			
Informational	5843	2.49 (1.37)	6.22 (2.43)
Emotional	4453	1.63 (1.10)	8.02 (1.60)
Respite	190	1.26 (1.28)	9.16 (0.91)
Legal	131	1.67 (1.03)	7.92 (1.67)
Medical	48	1.85 (1.09)	7.58 (1.90)
Financial	27	1.22 (0.75)	8.67 (1.14)
Instrumental	18	1.72 (1.49)	8.22 (1.86)
No explicit support need	2469	4.86 (2.58)	4.24 (3.13)
Dominant emotion			
Sadness	3839	1.49 (0.72)	7.99 (1.59)
Fear	3364	2.07 (0.82)	6.72 (2.10)

Group	Number of posts	Mean sentiment (SD)	Mean burden (SD)
Anger	2273	1.48 (0.77)	7.95 (1.63)
Bittersweet	1655	4.18 (1.94)	5.55 (2.60)
Hope	314	6.66 (1.58)	2.9 (2.39)
Joy	408	8.16 (0.96)	2.51 (2.01)
Neutral	1326	4.57 (1.28)	2.69 (2.16)

[a]ADRD: Alzheimer's disease and related dementias.
[b]Unknown denotes insufficient context and should not be read as a clinical category or absence of burden.

According to Table 2, burden increased across inferred ADRD stage. Early-stage posts had a mean burden score of 5.38 (2.47) and mean sentiment of 2.88 (SD 1.74), moderate-stage posts had a mean burden score of 6.96 (SD 2.17) and mean sentiment of 2.45 (SD 1.79), and late-stage posts had the highest mean burden score of 7.77 (SD 2.16) and lowest mean sentiment among stage groups (2.05, SD 1.72). Unknown-stage posts had lower burden (4.27, SD 3.39) and higher sentiment (3.76, SD 2.45), suggesting that posts without stage cues often contained less specific or less acute caregiving narratives. Burden also varied by care setting. Transition posts had the highest mean burden score (7.82, SD 1.67) and low mean sentiment (2.07, SD 1.49), followed by nursing home (7.17, SD 2.31), home care (7.01, SD 2.29), assisted living (6.79, SD 2.29), and memory care (6.59, SD 2.48). Posts with unknown care setting had a lower mean burden score (5.05) and higher mean sentiment (3.30). Note that posts coded as Unknown for ADRD stage (mean burden 4.27, SD 3.39) or care setting (mean burden 5.05, SD 3.22) lacked sufficient contextual evidence for those classifications. Because stage, setting, and burden were inferred from the same narrative, sparse posts may also have contained fewer explicit burden cues. These lower means may therefore partly reflect information availability or shared-method measurement and should not be interpreted as evidence that an Unknown group is absence of burden.

Support need type showed some of the largest differences in Table 2. Posts seeking respite support had the highest mean burden score (9.16, SD 0.91) and low mean sentiment (1.26, SD 1.28), followed by financial support needs (8.67, SD 1.14), instrumental support needs (8.22, SD 1.86), emotional support needs (8.02, SD 1.6), legal support needs (7.92, SD 1.67), and medical support needs (7.58, SD 1.9). Informational support posts had a lower mean burden score (6.22, SD 2.43) and higher mean sentiment (2.49, SD 1.37). Posts with no explicit support need had the lowest mean burden score among support need groups (4.24, SD 3.13) and the highest mean sentiment score (4.86, SD 2.58).

Dominant emotion was also strongly aligned with burden and sentiment. Posts coded as sadness had a mean burden score of 7.99 (SD 1.59) and mean sentiment of 1.49 (SD 0.72), and anger posts had a mean burden score of 7.95 (SD 1.63) and mean sentiment of 1.48 (SD 0.77). Fear posts had moderately high burden (6.72, SD 2.10) and low sentiment (2.07, SD 0.82). In contrast, bittersweet posts had intermediate burden (5.55, SD 2.6) and higher sentiment (4.18, SD 1.94), while hope, joy, and neutral posts had lower burden scores (2.90, SD 2.39, 2.51, SD 2.01, and 2.69, SD 2.16, respectively) and higher sentiment scores (6.66, SD 1.58, 8.16, SD 0.96, and 4.57, SD 1.28, respectively). This pattern supports the internal consistency of the inferred measures: negative emotions were associated with higher burden and lower sentiment, whereas positive or neutral emotions were associated with lower burden and higher sentiment.

Primary and Secondary Stressors

The open-vocabulary stressor extraction produced a broad set of short caregiver-burden phrases. Figure 3 presents clustered stressor topics by primary versus secondary placement, and Figure 4 presents the most common within-post primary-secondary pairings. A primary stressor was identified in 10,961 posts (83.2%), whereas a secondary stressor was identified in 7398 posts (56.1%).

Figure 3. Shared stressor topics by primary versus secondary placement.

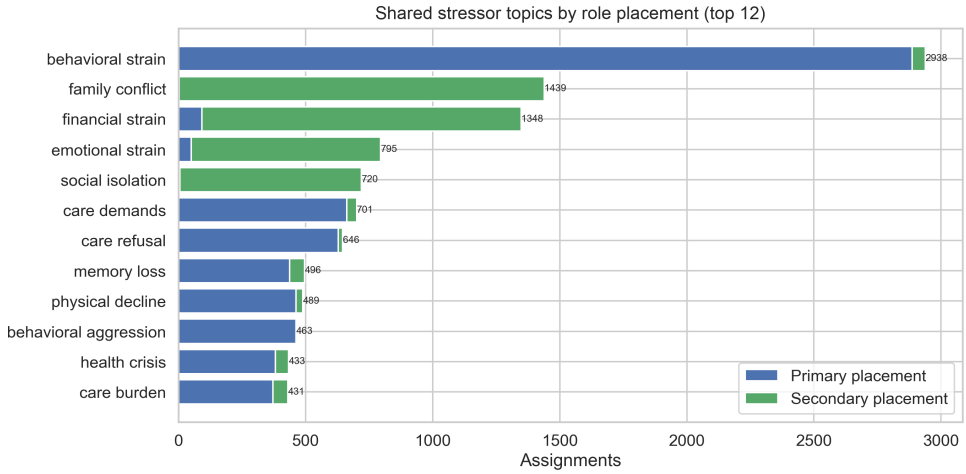
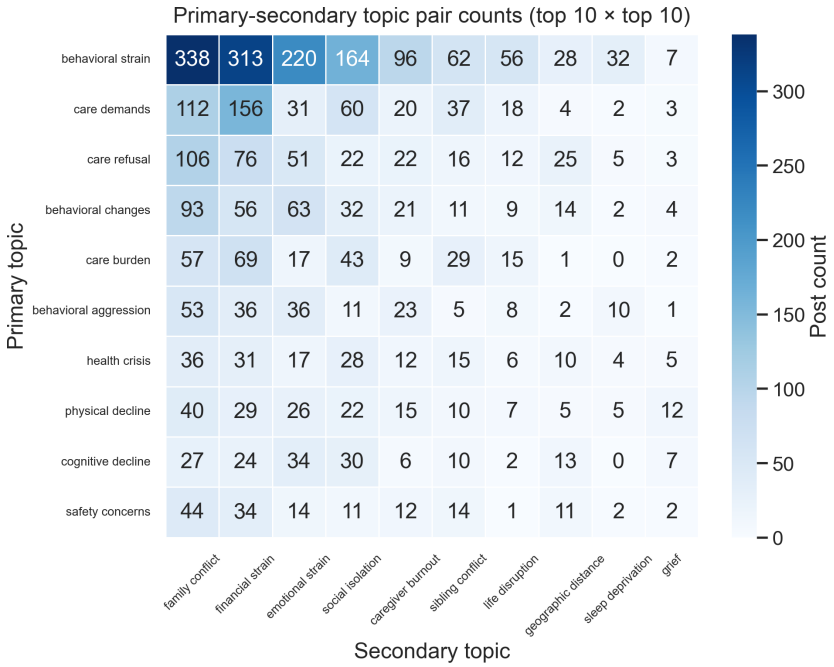


Figure 4. Within-post pairing of primary and secondary stressor topics.



Behavioral strain was the dominant clustered primary stressor theme (n=2272), followed by behavioral changes (n=709), care demands (n=663), care refusal (n=629), physical decline (n=462), memory loss (n=418), cognitive decline (n=398), health crisis (n=382), behavioral aggression (n=372), and care burden (n=370). Other clinically recognizable primary stressors included incontinence management, safety concerns, communication difficulty, disease progression, wandering, medical decisions, repetitive questioning, hallucinations, end-of-life concerns, delusional behavior, and placement transition.

Figure 3 also shows that secondary stressors were led by family conflict (n=1436 mentions), financial strain (n=1255), emotional strain (n=744), social isolation (n=714), caregiver burnout (n=354), sibling conflict (n=336), life disruption (n=213), geographic distance (n=209), sleep deprivation (n=122), grief (n=90), mental health concerns (n=73), and legal issues (n=72). The secondary stressor distribution suggests that caregivers' Reddit disclosures often move from symptom-management problems to practical and relational consequences: who pays, who helps, who refuses to help, how caregiving constrains the caregiver's own life, and how distance complicates supervision.

Figure 4 shows that within-post primary-secondary stressor pairings further clarified these pathways. Among posts with both a clustered primary and secondary stressor, the most frequent pairings were behavioral strain with family conflict (n=338), behavioral strain with financial strain (n=313), behavioral strain with emotional strain (n=220), behavioral strain with social isolation (n=164), care demands with financial strain (n=156), care demands with family conflict (n=112), and care refusal with family conflict (n=106). These pairs suggest that behavior-centered and care-demand narratives may function as central stressors linking direct dementia care demands to downstream family, financial, social, and mental-health consequences.

Temporal Patterns

Figures 5 and 6 present quarterly changes in inferred categorical fields, sentiment, and burden. Quarterly percentages represent the distribution of coded posts, and quarterly burden and sentiment values are means across posts. As noted above, these trends represent post-level patterns and should not be interpreted as unique-user prevalence or within-person change.

Figure 5. Quarterly trends in inferred categorical fields. ADRD: Alzheimer disease and related dementias; N/A: not available.

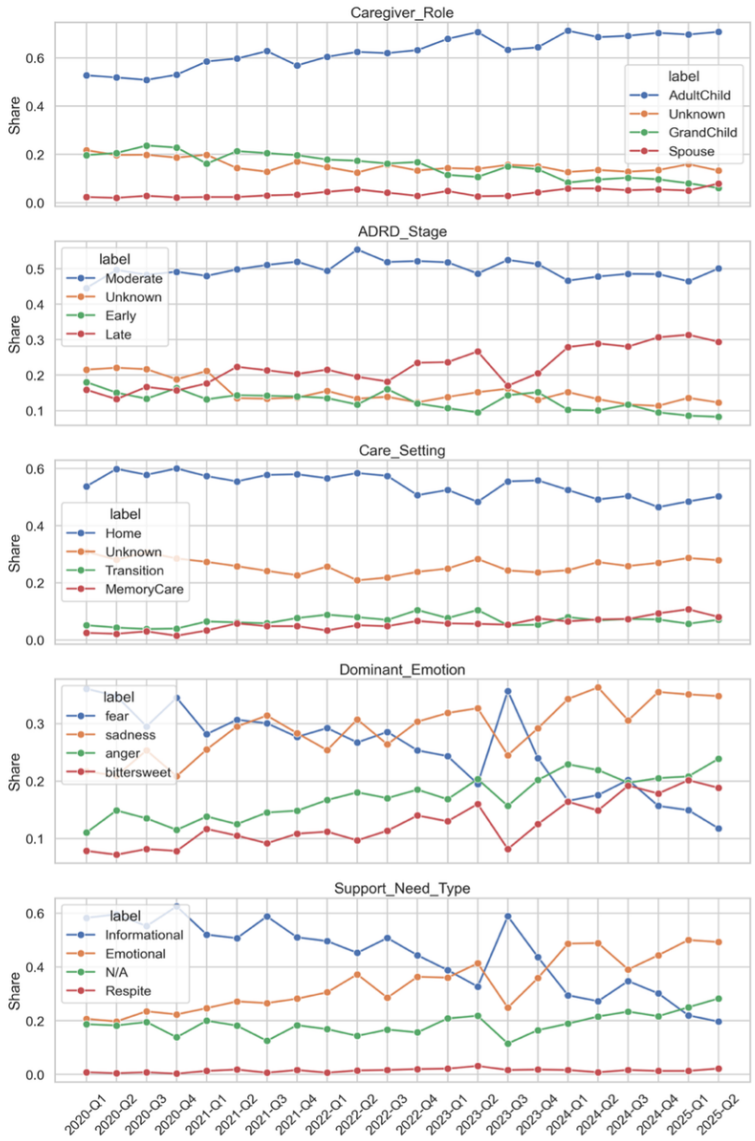


Figure 6. Quarterly trends in inferred sentiment and burden.

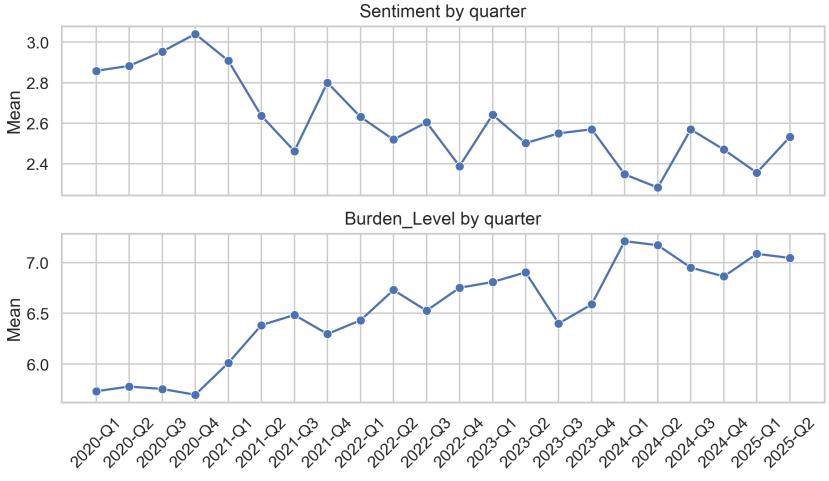


Figure 5 shows that the composition of posts changed over time. Late-stage ADRD posts increased from 15.9% (2095/13,179) in 2020 Q1 to 29.4% (3875/13,179) in 2025 Q2, while emotional-support posts increased from 20.7% (2728/13,179) to 49.2% (6484/13,179) over the same period. Informational-support posts moved in the opposite direction, decreasing from 58.3% (7683/13,179) in 2020 Q1 to 19.7% (2596/13,179) in 2025 Q2. Care-transition posts fluctuated across quarters, ranging from 3.8% (501/13,179) in 2020 Q3 to 10.5% (1384/13,179) in 2022 Q4. Figure 5 also helps interpret the temporary dip in burden observed in 2023 Q3. That quarter retained 600 posts, but it had fewer late-stage posts (2240/13,179, 17.0%) than 2023 Q2 (3519/13,179, 26.7%) and fewer care-transition posts (685/13,179, 5.2%) than 2023 Q2 (1384/13,179, 10.5%). It also had fewer emotional-support posts (3268/13,179, 24.8%) and more informational-support posts (7749/13,179, 58.8%) than adjacent quarters. Dominant emotion also shifted in 2023 Q3: fear increased to 35.7% (4705/13,179), while sadness declined to 24.5% (3229/13,179). These compositional differences are consistent with a case-mix explanation involving more advice-seeking and uncertainty-oriented posts rather than a missing-data artifact. However, because author-level data were unavailable, we cannot determine whether the dip reflected changes in author composition, repeated posting, or stochastic quarter-to-quarter variation.

Figure 6 presents the numeric burden and sentiment trends. Mean burden increased from approximately 5.7 in 2020 to above 7.0 in several later quarters, while sentiment generally declined. These descriptive trends should not be interpreted causally, but they suggest that caregiver burden narratives remained elevated in the later years of the corpus.

Subgroup Intersections and Socioclinical Implications

To identify clinically actionable subgroups, we conducted a targeted pairwise analysis of caregiver role, ADRD stage, care setting, and support need. For each pair of variables, Table 3 reports the subgroup with the highest mean burden among cells with more than 200 posts. This analysis is intended to complement the single-variable summaries in Table 2 by showing where high-burden caregiving contexts overlap.

Table 3. Highest-burden subgroup in each core variable pairing among cells with more than 200 posts.

Pairing	Highest-burden subgroup	Posts, n	Mean burden (SD)
Caregiver role ADRD[a] stage	Adult child+late stage	1892	7.96 (1.95)
Caregiver role care setting	Adult child+transition	667	7.93 (1.49)
Caregiver role support need	Adult child+emotional support	3096	8.10 (1.48)
ADRD stage care setting	Late stage+transition	299	8.40 (1.32)
ADRD stage support need	Late stage+emotional support	1355	8.42 (1.46)
Care setting support need	Transition+emotional support	359	8.45 (1.12)

[a]ADRD: Alzheimer disease and related dementias.

The pattern in Table 3 is that high burden is concentrated around late-stage ADRD, care transitions, adult-child caregiving, and emotional support needs. Adult-child caregivers appeared repeatedly in the highest-burden pairings. Adult-child posts with emotional support needs had a mean burden score of 8.10 across 3096 posts, and adult-child in care-transition setting posts had a mean burden score of 7.93 (SD 1.49) across 667 posts. Late-stage ADRD combined with emotional support need had a mean burden score of 8.42 (SD 1.46) across 1355 posts, and care-transition posts with emotional support needs had a mean burden score of 8.45 (SD 1.12) across 359 posts. Late-stage care transitions also had high burden (mean 8.40, SD 1.32, across 299 posts).

Table 4 further shows which subgroup categories had the highest burden within the most common primary and secondary stressor topics. To save space and focus on stable, clinically meaningful patterns, Table 4 reports only subgroups with more than 200 posts and mean burden above 8. Behavioral strain was most burdensome in late-stage posts (mean 8.52, SD 1.54) and emotional-support posts (mean 8.49, SD 1.24). Care demands were also most burdensome when paired with emotional support needs (mean 8.33, SD 1.14). Among secondary stressors, financial strain showed the highest burden levels: late-stage financial-strain posts had a mean burden score of 8.76 (SD 1.36), and financial-strain posts with emotional support needs had a mean burden score of 8.72 (SD 1.15). Family conflict and emotional strain also reached high burden when paired with late-stage ADRD or emotional support needs.

Table 4. High-burden groups within the most common primary and secondary stressor topics[a].

Stressor type	Stressor topic	Highest-burden group	Posts, n	Mean burden (SD)
Primary	Behavioral strain	Late stage	410	8.52 (1.54)
Primary	Behavioral strain	Emotional support need	989	8.49 (1.24)
Primary	Care demands	Emotional support need	279	8.33 (1.14)
Secondary	Family conflict	Late stage	320	8.31 (1.50)

Stressor type	Stressor topic	Highest-burden group	Posts, n	Mean burden (SD)
Secondary	Family conflict	Emotional support need	613	8.22 (1.26)
Secondary	Financial strain	Late stage	290	8.76 (1.36)
Secondary	Financial strain	Emotional support need	377	8.72 (1.15)
Secondary	Financial strain	Home care	811	8.25 (1.55)
Secondary	Financial strain	Adult child	890	8.10 (1.57)
Secondary	Emotional strain	Emotional support need	424	8.27 (1.16)
Secondary	Emotional strain	Adult child	514	8.06 (1.37)
Secondary	Emotional strain	Moderate stage	424	8.03 (1.32)

[a]Groups are shown only when the subgroup had more than 200 posts and mean burden above 8.

Discussion

Main Results

In this source-matched analysis of 13,179 LLM-coded Reddit posts from the final 13,200-post scraped corpus, the LLM-assisted workflow identified clinically interpretable ADRD caregiver constructs at scale. Reddit ADRD caregiving narratives were dominated by adult-child and grandchild perspectives; moderate-stage, home-based care was the central disclosure context; sadness, fear, and anger were common; and negative emotions aligned with higher burden. Inferred burden increased across ADRD stages, from early-stage posts to moderate- and late-stage posts. Behavioral strain emerged as the dominant primary stressor, while family conflict, financial strain, emotional strain, social isolation, and burnout were common secondary stressors. Temporal analysis showed that mean burden was relatively stable in 2020, rose through 2021 and 2022, and remained elevated from 2023 through mid-2025; sentiment generally moved in the opposite direction and was lower in later years. These findings suggest that ADRD caregiver burden in Reddit narratives is not only emotional distress but also a set of linked practical, relational, financial, and support-navigation problems.

Several results are especially important for interpreting the measurement workflow. First, caregiver role, ADRD stage, care setting, support need, sentiment, burden, and stressor type were more recoverable from Reddit narratives than sociodemographic characteristics. This pattern is expected because caregivers often describe care tasks, disease stages, emotions, and support needs directly, but rarely provide income, race, education, or marital status. Second, the burden patterns were internally consistent. Late-stage ADRD posts had higher burden and lower sentiment than early-stage posts; transition posts had the highest burden among care-setting groups; and posts expressing sadness or anger had high burden and low sentiment, while hope, joy, and neutral posts had lower burden and higher sentiment. Third, primary and secondary stressors followed the logic of the Stress Process Model: direct dementia-care problems, especially behavioral strain and care demands, often appeared alongside spillover burdens involving family systems, finances, social isolation, and caregiver burnout [8].

The human validation benchmark supports the main structured fields but also clarifies the boundaries of the corpus. Human agreement was excellent for caregiver role, ADRD stage, care setting, dominant emotion, support need, and burden level, and substantial for primary and secondary stressor categories. However, agreement on whether posts were strictly in-scope as informal caregiver narratives was poor; the findings should therefore be interpreted as descriptions of high-engagement ADRD caregiving discourse on Reddit rather than prevalence estimates from a definitively verified caregiver population.

Comparison With Prior Work

This study speaks directly to recent work using Reddit and other online platforms to study caregiver narratives. For example, Huang et al [14] analyzed Reddit posts and comments from r/AgingParents, r/eldercare, and r/elderlycare and found that caregivers often disclosed care recipients' functional problems, caregiving relationship strain, care recipients' emotional problems, and scarcity of health and social resources. Guided by the Stress Process Model, they showed that primary stressors and secondary stressors co-occurred in caregiver posts and that different stressors were associated with different forms of social support received in comments. Our findings echo this stress-process pattern in an ADRD-specific corpus. By applying an LLM-assisted measurement workflow, our study extends this work by analyzing a larger longitudinal corpus and identifying more varied ADRD-specific stressors. The most frequent within-post pairings, such as behavioral strain with family conflict, financial strain, emotional strain, and social isolation, show how dementia-specific primary stressors may connect to downstream caregiver-life stressors within the same post.

The descriptive results are also consistent with prior online caregiver research on poster roles, discussion topics, and visible forms of distress. We found that adult children accounted for 62.7% (8266/13,179) of posts and grandchildren for 15.3% (2012/13,179), which is consistent with prior online-community studies showing that family members, especially adult children and spouses, are prominent caregiver voices in dementia forums [18,20]. Compared with national dementia caregiver estimates, however, the Reddit corpus appears more heavily weighted toward adult-child and grandchild voices and less weighted toward

spouse caregivers. The 2026 Alzheimer Disease Facts and Figures report indicates that over half of dementia caregivers provide assistance to a parent or parent-in-law, whereas approximately 10% provide help to a spouse with Alzheimer disease or another dementia [1]. In our post-level corpus, spouse caregivers accounted for only 4.0% of inferred caregiver roles. This difference should not be interpreted as a population difference because Reddit posts are not a representative caregiver sample. Instead, it suggests that Reddit may be especially useful for observing adult-child and younger intergenerational caregiving narratives. The 15.3% (2012/13,179) share of grandchild role inferences is notable because grandchildren are less often examined as a distinct group in traditional caregiver research. Reddit may make younger and intergenerational caregivers more visible because of age-related differences in platform use and the accessibility of anonymous peer support. Our findings also showed that moderate-stage ADRD and home-based care were common, while late-stage posts and transition posts had higher burden and lower sentiment. These findings align with prior dementia-caregiving research showing that care demands change across disease progression and that functional decline, behavioral symptoms, care transitions, and caregiving relationship strain are important sources of burden [37-40].

The emotional and support need findings also fit prior Reddit and online-forum studies. Huang et al [15] supportive-communication study found that Reddit caregivers often used direct support seeking and expressed fear, anger, and sadness. Other ADRD-focused online narrative studies have identified themes such as family issues, care facilities, financial struggles, abuse concerns, symptoms, emotional distress, nursing homes, sleep problems, incontinence, and advice seeking [18,22-24]. Our results echo this literature: informational and emotional support were the 2 dominant support needs, and sadness, fear, and anger together accounted for most posts. Our results extend this literature by adding a burden gradient to these patterns. Respite, financial, instrumental, emotional, legal, and medical support needs all had higher mean burden than informational posts or posts with no explicit support need. Sadness and anger posts had high burden and low sentiment, whereas hope, joy, and neutral posts had lower burden and higher sentiment. This adds a post-level link between emotional expression, support need, sentiment, and inferred burden that is harder to obtain from topic prevalence alone.

This study also extends prior ADRD-specific NLP, machine learning, and LLM studies by changing the unit of measurement. Earlier work used structural topic modeling, machine learning classification, sentiment analysis, or related methods to identify aggregate themes in online posts [18-22,24]. More recent LLM applications have mapped digital caregiving functions, built retrieval-augmented caregiver support systems, or summarized large social-media corpora into aggregate themes [25-27]. These studies are valuable because they scale beyond manual qualitative review and identify clinically meaningful themes. However, topic models, supervised classifiers, and many LLM topic-modeling workflows often produce broad themes or fixed labels, and they may not preserve how several constructs appear together in the same narrative. Our study adds a different use case: structured narrative measurement. The workflow inferred post-level characteristics, caregiving contexts, confidence ratings, and open-vocabulary primary and secondary stressors. This made it possible to compare burden across stage, setting, support need, and emotion; examine stressor clusters; assess uncertainty; and describe temporal changes in both categorical composition and numeric burden. These outputs should not be treated as ground truth, but the findings show how LLMs can help transform naturalistic ADRD caregiving-related Reddit discourse into analyzable, theory-guided measures when paired with source matching, schema validation, field-level correction, confidence ratings, and human validation [29-31].

Social and Clinical Implications

The vulnerable-subgroup findings suggest several practical targets for caregiver support. Care transitions were the clearest pressure point, especially when paired with late-stage ADRD or emotional support needs. This suggests that transitions should be treated as key intervention windows. Prior work on nursing home admission and home-to-long-term-care transitions shows that institutional transitions can change but do not necessarily eliminate caregiver distress, and that caregivers often have unmet emotional and informational needs during transitions [39,40]. Placement counseling, discharge planning support, anticipatory guidance about memory care or hospice decisions, and follow-up after the move may be useful when caregivers are deciding whether, when, or how to change care settings.

Emotional support needs were consistently present in the highest-burden posts. This suggests that caregiver-support systems should not focus only on informational resources. Peer-support forums, caregiver hotlines, clinical teams, and digital tools may need pathways for validation, distress screening, and referral when caregivers describe guilt, isolation, fear, or unsustainable care. The LLM's role in this study was analytic rather than therapeutic: it classified burden-related signals in posts but did not provide support, assess risk, or make clinical decisions. Automated systems may help summarize needs at scale or direct content to trained human reviewers, but the loneliness, emotional overload, empathy, and validation reflected in high-burden posts are fundamentally relational. Any caregiver-facing application should augment rather than replace reciprocal peer support, clinician and social-worker judgment, and accountable referral pathways. Financial strain also appeared as a high-burden secondary stressor, especially in late-stage, home-care, adult-child, and emotional-support narratives. This finding is consistent with dementia-caregiver studies linking financial strain to role overload and role captivity and with interview research showing that caregivers need integrated financial, legal, functional, educational, and peer-support resources [41,42]. For aging services, this points toward financial-navigation support, benefits literacy, Medicaid and long-term-care planning, and decision aids.

Family conflict and emotional strain also emerged as intervention-relevant secondary stressors. Family conflict reached high burden when paired with late-stage ADRD or emotional support needs, suggesting that conflict often intensifies when care demands become more severe. This is consistent with prior online-community findings that family issues and caregiving relationships are central dimensions of ADRD caregiving discourse [18,20]. Interventions that focus only on symptom management may miss this social context. Caregiver support may also need family mediation, role clarification, and planning tools that help relatives divide responsibility before a crisis or care transition occurs [4]. Grandchildren accounted for 15.3% (2012/13,179) of caregiver role inferences, making them a visible group in this Reddit corpus. This does not imply population prevalence, but it suggests that online ADRD caregiver support environments may include younger and less studied caregivers whose needs may differ from those of spouses or adult children.

The temporal findings add another implication: caregiver-support needs may change in form even when the underlying burden remains high. The later years of the corpus showed higher burden levels and lower sentiment, while emotional support became more common and informational support became less dominant. This suggests that digital health systems and community platforms should be able to respond to both practical advice seeking and emotional distress, rather than treating caregiver support as a single static need. We did not identify a clear external event that could explain the 2023 Q3 pattern by itself. Therefore, the temporal patterns should be interpreted as descriptive signals of changing post composition rather than evidence of a policy or social-event effect.

Looking forward, if LLM-assisted measurement is later embedded in caregiver-facing support systems, algorithmic interpretation may itself become part of the stress process. Opaque or erroneous classifications, perceived surveillance, and mismatched recommendations could create new secondary strains, whereas transparent and well-governed systems could improve navigation and access to support resources. Future research could therefore extend established stress frameworks to examine algorithm-mediated appraisal, trust, perceived control, and support access.

Limitations

This study has several limitations. First, the analytic sample is not representative of the broader population of informal ADRD caregivers. Reddit users constitute a self-selected population, and the monthly top-score sampling strategy further prioritizes posts receiving greater community engagement. Although many users in these communities may be based in the United States, posts from other countries may also be included. Accordingly, differences across subreddits should be interpreted as reflecting variation in community function, audience, and disclosure context rather than differences in the population prevalence of caregiver experiences. In addition, limited agreement in identifying in-scope posts indicates uncertainty in distinguishing general ADRD discussions from narratives clearly authored by informal caregivers. The reported percentages and mean scores therefore characterize this selected corpus of high-engagement Reddit posts and should not be generalized to the prevalence or distribution of caregiver burden in the broader population.

Second, the unit of analysis was the post rather than a verified unique caregiver. Repeated posting by the same caregiver may influence observed distributions. ADRD-specific subreddits may also include posts from patients, clinicians, or general community members. The inference workflow attempted to measure caregiver attributes only when the post supported that framing, but noncaregiver perspectives may remain in the corpus. This limitation was visible in the human validation benchmark, where coders disagreed substantially on the in-scope field despite high agreement for most downstream constructs.

Third, the inferred variables are model-generated measurements rather than ground-truth annotations. The 150-post validation benchmark supported several major fields, but it cannot fully establish performance across all rare subgroups, stressor types, or ambiguous narratives. The source-matched analysis retained only LLM-coded records whose IDs joined to the final scraped corpus. This filtering protects corpus validity but also shows the need for strict ID validation in LLM-assisted pipelines.

Fourth, this study analyzed posts from January 2020 through June 2025 rather than a longer historical period. This time window was chosen partly to manage computational cost and partly to avoid mixing the main study period with more distant pre-COVID Reddit dynamics. The COVID-19 pandemic may still have shaped caregiver needs, care access, and posting behavior during the study period. Future work should analyze longer time spans and larger post samples to test whether the main burden, stressor, and subgroup patterns are robust across different periods.

Fifth, we did not conduct subgroup analyses by demographic attributes such as race, education, income, marital status, or age. Reddit users often do not disclose these personal characteristics, and the LLM therefore assigned low-confidence ratings to many demographic fields. Analyzing these fields in the current corpus could give a misleading impression of precision. Future studies with larger corpora may still recover enough high-confidence demographic information to support cautious subgroup comparisons and to identify vulnerable caregiver groups for more targeted intervention design.

Finally, although the strict rule- and rubric-based prompt made scoring more standardized, and agreement with human coding is necessary for validation, they may not fully explain how the model reached its judgments. The workflow shows whether model outputs match benchmark labels, not why the model selected a label in an ambiguous narrative. Borderline human cases most often involved whether sparse text justified an “other” stressor rather than “unclear” and whether limited distress cues supported low burden rather than insufficient evidence. Explicit relationship terms, functional-dependence descriptions, behavioral symptoms, named care settings,

transition language, and direct support-seeking or emotion language were the observable cues specified in the coding framework, but the study did not estimate their feature importance.

Conclusions

This study demonstrates the potential of LLM-assisted measurement for studying ADRD caregiving-related Reddit discourse at scale. The main innovation is not simply using an LLM to summarize Reddit posts, but using it as a structured measurement tool that links caregiving context, sentiment, burden, support needs, and primary and secondary stressors at the post level. The findings show clinically meaningful patterns: burden increased across inferred ADRD stage, negative emotions aligned with higher burden, behavioral strain was a central primary stressor, and family, financial, emotional, and social strains appeared as common secondary stressors. The subgroup analyses further suggest that late-stage ADRD, care transitions, adult-child caregiving, emotional support needs, and financial strain are key contexts for high burden. These patterns highlight transition navigation, emotional support, financial and benefits counseling, and family role negotiation as potential targets for caregiver-support interventions and digital health tools. Human validation supported the reliability of the main structured fields, while also showing that corpus inclusion and model-generated measures require cautious interpretation. Overall, this workflow offers a scalable and transparent approach for converting naturalistic online ADRD caregiving-related Reddit discourse into theory-guided quantitative measures that can inform clinical aging research, service planning, and future intervention design.

Generative AI Use

The authors acknowledge the use of generative AI tools (ChatGPT) to assist with language editing and manuscript revision. All substantive content, analyses, interpretations, and final wording were reviewed and approved by the authors. The use of large language models (LLMs) as part of the study workflow is described separately in the Methods section.

Funding

This research was supported by grant, P30AG066614, awarded to the Center on Aging and Population Sciences at The University of Texas at Austin by the National Institute on Aging. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The raw Reddit posts analyzed in this study are based on publicly accessible Reddit data. To protect user privacy, raw post text, usernames, and individual-level identifiers are not publicly redistributed. Aggregate tables, figures, and analysis code may be made available by the corresponding author upon reasonable request, subject to ethical and platform-policy considerations.

Authors' Contributions

Conceptualization: YZ (lead), MFZ (equal)
Data curation: MFZ
Formal analysis: MFZ (lead), YZ (supporting)
Funding acquisition: YZ
Investigation: YZ (lead), MFZ (equal)
Methodology: MFZ (lead), YZ (supporting)
Supervision: YZ
Validation: MFZ (lead), YZ (supporting)
Visualization: MFZ (lead), YZ (supporting)
Writing – original draft: YZ (lead), MFZ (supporting)
Writing – review & editing: YZ (lead), MFZ (supporting)
The author MFZ is currently not affiliated with any institution but is an Independent Researcher.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Reddit data corpus funnel for Alzheimer disease and related dementias (ADRD) caregiving post extraction.
[DOCX File (Microsoft Word File), 17 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Full Batch-Mode large language model (LLM) Inference Prompt.
[DOCX File (Microsoft Word File), 19 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Human validation agreement for the 150-post benchmark.
[DOCX File (Microsoft Word File), 17 KB-Multimedia Appendix 3]

References

1. The Alzheimer's Association. 2026 Alzheimer's disease facts and figures. Alzheimers Dement. Apr 2026;22(4):e71345. [doi: 10.1002/alz.71345]
2. Gilhooly KJ, Gilhooly MLM, Sullivan MP, et al. A meta-review of stress, coping and interventions in dementia and dementia caregiving. BMC Geriatr. May 18, 2016;16:106. [doi: 10.1186/s12877-016-0280-8] [Medline: 27193287]
3. Wiles JL, Leibing A, Guberman N, Reeve J, Allen RES. The meaning of "aging in place" to older people. Gerontologist. Jun 2012;52(3):357-366. [doi: 10.1093/geront/gnr098] [Medline: 21983126]
4. Zhang Y. Family structure and transitions into and out of homeboundness: evidence from the National Health and Aging Trends Study, 2011-2024. J Gerontol B Psychol Sci Soc Sci. Oct 1, 2026;gbag206. [doi: 10.1093/geronb/gbag206] [Medline: 42826503]
5. Caregiving in the United States 2020. AARP, National Alliance for Caregiving; 2020. [doi: 10.26419/ppi.00103.001]
6. Caregivers of a person with alzheimer's disease or a related dementia. CDC. URL: https://www.cdc.gov/caregiving/about/index.html [Accessed 2026-07-30]
7. Pearlin LI, Lieberman MA, Menaghan EG, Mullan JT. The stress process. J Health Soc Behav. Dec 1981;22(4):337-356. [doi: 10.2307/2136676] [Medline: 7320473]
8. Pearlin LI, Mullan JT, Semple SJ, Skaff MM. Caregiving and the stress process: an overview of concepts and their measures. Gerontologist. Oct 1990;30(5):583-594. [doi: 10.1093/geront/30.5.583] [Medline: 2276631]
9. Friedman EM, Shih RA, Langa KM, Hurd MD. US prevalence and predictors of informal caregiving for dementia. Health Aff (Millwood). Oct 2015;34(10):1637-1641. [doi: 10.1377/hlthaff.2015.0510] [Medline: 26438738]
10. Pristavec T, Luth EA. Informal caregiver burden, benefits, and older adult mortality: a survival analysis. J Gerontol B Psychol Sci Soc Sci. Nov 13, 2020;75(10):2193-2206. [doi: 10.1093/geronb/gbaa001] [Medline: 31903481]
11. Vasileiou K, Barnett J, Barreto M, et al. Experiences of loneliness associated with being an informal caregiver: a qualitative investigation. Front Psychol. 2017;8:585. [doi: 10.3389/fpsyg.2017.00585] [Medline: 28469589]
12. Akgun-Citak E, Attepe-Ozden S, Vaskelyte A, et al. Challenges and needs of informal caregivers in elderly care: qualitative research in four European countries, the TRACE project. Arch Gerontol Geriatr. 2020;87:103971. [doi: 10.1016/j.archger.2019.103971] [Medline: 31756568]
13. Tourangeau R, Yan T. Sensitive questions in surveys. Psychol Bull. Sep 2007;133(5):859-883. [doi: 10.1037/0033-2909.133.5.859] [Medline: 17723033]
14. Huang NM, Timothy B, Ho SS. Stressors disclosed on Reddit by caregivers of older adults and social support received: content analysis. JMIR Aging. Sep 5, 2025;8:e71452. [doi: 10.2196/71452] [Medline: 40911350]
15. Huang NM, Timothy B, Ho SS. Supportive communication patterns among informal caregivers of older adults on Reddit: a content analysis. Digit Health. 2025;11:20552076251376537. [doi: 10.1177/20552076251376537] [Medline: 41018504]
16. Amorim D, Lam K, Chary AN, Plys E, Shi S. Positive aspects of caregiving: a qualitative analysis of Reddit posts. J Am Geriatr Soc. Jun 2026;74(6):1792-1796. [doi: 10.1111/jgs.70358] [Medline: 41724689]
17. Grimmer J, Stewart BM. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. Polit Anal. 2013;21(3):267-297. [doi: 10.1093/pan/mps028]
18. Cheng K, Yin Z. "Please advise": understanding the needs of informal caregivers of people with Alzheimer's disease and related dementia from online communities through a structured topic modeling approach. AMIA Jt Summits Transl Sci Proc. 2022;2022:149-158. [Medline: 35854737]
19. Ni C, Malin BA, Song L, Jefferson AL, Commiskey P, Yin Z. "Rough day … need a hug": learning challenges and experiences of the Alzheimer's disease and related dementia caregivers on Reddit. ICWSM. 2022;16:711-722. [doi: 10.1609/icwsm.v16i1.19328]
20. Ni C, Song Q, Malin B, et al. Examining online behaviors of adult-child and spousal caregivers for people living with Alzheimer disease or related dementias: comparative study in an open online community. J Med Internet Res. Nov 17, 2023;25:e48193. [doi: 10.2196/48193] [Medline: 37976095]
21. Shoults CC, Rutherford MW, Kemp AS, et al. Analysis of caregiver burden expressed in social media discussions. Int J Environ Res Public Health. Jan 20, 2023;20(3):1933. [doi: 10.3390/ijerph20031933] [Medline: 36767297]
22. Hswen Y, Xiong J, Hurley M, Nguyen TT. Experiences of Alzheimer's disease and related dementia family caregivers on Reddit communities: a topic modeling and sentiment analysis. Artif Intell Health. 2024;1(3):127-135. [doi: 10.36922/aih.3075] [Medline: 39246419]
23. Huang NM, Wong LZ, Ho SS, Timothy B. Understanding challenges and emotions of informal caregivers of general older adults and people with Alzheimer disease and related dementia: comparative study. J Med Internet Res. Feb 28, 2025;27:e54847. [doi: 10.2196/54847] [Medline: 40053723]

24. Kim J, Cai ZR, Chen ML, et al. Mental health care needs of caregivers of people with Alzheimer's disease from online forum analysis. NPJ Ment Health Res. Nov 14, 2024;3(1):54. [doi: 10.1038/s44184-024-00100-y] [Medline: 39537826]
25. Kim S, Han DY, Bae J. Transforming Alzheimer's digital caregiving through large language models. Curr Alzheimer Res. 2024;21(7):503-516. [doi: 10.2174/0115672050301740241118044604] [Medline: 39592896]
26. Hasan WU, Zaman KT, Wang X, Li J, Xie B, Tao C. Empowering Alzheimer's caregivers with conversational AI: a novel approach for enhanced communication and personalized support. NPJ Biomed Innov. Dec 4, 2024;1(1):3. [doi: 10.1038/s44385-024-00004-8] [Medline: 42032173]
27. He W, Hou B, Zheng A, et al. Advanced topic modeling with large language models: analyzing social media content from dementia caregivers. Innov Aging. Dec 26, 2025;9(Suppl 1):S38-S47. [doi: 10.1093/geroni/igaf120] [Medline: 41458886]
28. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics; Jun 2-7, 2019:4171-4186; Minneapolis, MN. [doi: 10.18653/v1/N19-1423]
29. Gilardi F, Alizadeh M, Kubli M. ChatGPT outperforms crowd workers for text-annotation tasks. Proc Natl Acad Sci U S A. Jul 25, 2023;120(30):e2305016120. [doi: 10.1073/pnas.2305016120] [Medline: 37463210]
30. Ziems C, Held W, Shaikh O, Chen J, Zhang Z, Yang D. Can large language models transform computational social science? Comput Linguist. Mar 1, 2024;50(1):237-291. [doi: 10.1162/coli_a_00502]
31. Asirvatham H, Mokski E, Shleifer A. GPT as a measurement tool. National Bureau of Economic Research Working Paper Series, No. 34834. Feb 2026. [doi: 10.3386/w34834]
32. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The pushshift Reddit dataset. ICWSM. 2020;14:830-839. [doi: 10.1609/icwsm.v14i1.7347]
33. Tang AY, Kwak J, Xiao L, et al. Online health information wants of caregivers for persons with dementia in social media. SAGE Open. Oct 2023;13(4). [doi: 10.1177/21582440231205367]
34. Garg M, Sohn S. CareD: caregiver's experience with cognitive decline in Reddit posts. Presented at: 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI); Jun 26-29, 2023:581-587; Houston, TX. [doi: 10.1109/ICHI57859.2023.00104]
35. Google. Gemma 4: our most capable open models to date. Google Blog; 2026. URL: https://blog.google/innovation-and-ai/technology/developers-tools/gemma-4/ [Accessed 2026-07-30]
36. Google. Gemma 4 model overview. Google AI for Developers; 2026. URL: https://ai.google.dev/gemma/docs/core [Accessed 2026-07-30]
37. Huang HL, Shyu YIL, Chen MC, et al. Family caregivers' role implementation at different stages of dementia. Clin Interv Aging. 2015;10:135-146. [doi: 10.2147/CIA.S60574] [Medline: 25584022]
38. van den Kieboom R, Snaphaan L, Mark R, Bongers I. The trajectory of caregiver burden and risk factors in dementia progression: a systematic review. J Alzheimers Dis. 2020;77(3):1107-1115. [doi: 10.3233/JAD-200647] [Medline: 32804093]
39. Gaugler JE, Mittelman MS, Hepburn K, Newcomer R. Clinically significant changes in burden and depression among dementia caregivers following nursing home admission. BMC Med. Dec 17, 2010;8:85. [doi: 10.1186/1741-7015-8-85] [Medline: 21167022]
40. Lee K, Chung J, Meyer KN, Dionne-Odom JN. Unmet needs and health-related quality of life of dementia family caregivers transitioning from home to long-term care: a scoping review. Geriatr Nurs. 2022;43:254-264. [doi: 10.1016/j.gerinurse.2021.12.005] [Medline: 34953331]
41. Liu Y, Dokos M, Fauth EB, Lee YG, Zarit SH. Financial strain, employment, and role captivity and overload over time among dementia family caregivers. Gerontologist. Sep 17, 2019;59(5):e512-e520. [doi: 10.1093/geront/gnz099] [Medline: 31322654]
42. Fan Q, DuBose L, Ory MG, et al. Financial, legal, and functional challenges of providing care for people living with dementia and needs for a digital platform: interview study among family caregivers. JMIR Aging. Sep 5, 2023;6:e47577. [doi: 10.2196/47577] [Medline: 37526513]

Abbreviations

ADRD: Alzheimer disease and related dementias
BERT: Bidirectional Encoder Representations from Transformers
IRB: institutional review board
LLM: large language model
N/A: not available

NLP: natural language processing
RQ: research question

Edited by Jinjiao Wang; peer-reviewed by Eugene Yeoh, Jacqueline Guendouzi; submitted 01.Jun.2026; final revised version received 17.Sep.2026; accepted 08.Oct.2026; published 24.Nov.2026

Please cite as:
Zhang Y, Zhang M
Alzheimer Disease and Related Dementias Caregiver Burden on Reddit: Large Language Model–Assisted Measurement Study
JMIR Aging 2026;9:e103305
URL: https://aging.jmir.org/2026/1/e103305
doi: 10.2196/103305

© Yan Zhang, Mingfei Zhang. Originally published in JMIR Aging (https://aging.jmir.org), 24.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Aging, is properly cited. The complete bibliographic information, a link to the original publication on https://aging.jmir.org, as well as this copyright and license information must be included.