

Harnessing AI and social media to understand real-world patient experiences in systemic lupus erythematosus

Siqi Yang¹, Carina Hawryluk², James Liu², Niki Eckert³, Jessica Otoo², Ernest R. Vina⁴, Lixia Yao¹

1. Polygon Health Analytics, Philadelphia, PA, USA
2. Boston University, Boston, MA, USA
3. Independent Researcher, New York, NY, USA
4. Division of Rheumatology, College of Medicine-Tucson, University of Arizona, Tucson Arthritis Center, Tucson, AZ, USA

Correspondence to:

Lixia, Yao, PhD;
Polygon Health Analytics LLC, Philadelphia, PA, USA;
Email: lixia@polygonhealthanalytics.com

ABSTRACT

Objective: To apply large language models (LLMs) to Reddit posts referencing systemic lupus erythematosus (SLE) to identify patient-expressed unmet medical needs, symptom experiences, and healthcare challenges, demonstrating how AI-enabled social media listening complements traditional patient-experience research.

Methods: We extracted 4,633 posts from ten SLE-related or health-focused Reddit communities using the public Reddit API (October–November 2025). After removing duplicates, promotional content, and posts with insufficient information, 2,603 posts remained. A thematic codebook was developed through manual review of 300 posts and iteratively refined. Two LLMs (Gemini 3.0 and GPT-5.2) were evaluated for automated thematic labeling using percent agreement, Cohen's κ , and a human-annotated reference set ($n=100$). The best-performing model was used to quantify theme prevalence, followed by qualitative review of representative narratives.

Results: GPT-5.2 demonstrated higher performance ($F1=0.844$) than Gemini 3.0 ($F1=0.811$), with substantial inter-model agreement across main themes (mean $\kappa=0.71$). Posts reflected multidimensional experiences. The most frequent subtheme was *Advice Seeking* (84.1%), followed by *Emotional Coping* (55.6%). Common symptom-related themes included *Pain* (37.2%), *Other Symptom Presentations* (37.6%), *Fatigue* (24.7%), and *Acute or Worsening Flares* (30.2%). Diagnostic uncertainty was prominent, including confusion about laboratory results (24.0%) and emotional impact of uncertainty (33.0%). Qualitative review highlighted emotional distress, reliance on peer communities for interpretation of symptoms and labs, and difficulty managing complex treatment regimens.

Conclusion: LLM-enabled social media listening offers a scalable method for synthesizing large volumes of unstructured patient narratives, providing timely insights into lived experiences and unmet needs among individuals discussing lupus online. Findings align with established qualitative literature while highlighting persistent gaps in patient education, communication, and care coordination. This analytical framework can be applied across disease areas to support patient-centered care, measurement development, and evidence generation relevant to therapeutic and health-services research.

Keywords: Systemic lupus erythematosus; social media listening; large language models (LLMs); patient-reported experience; thematic analysis

What is already known on this topic

- People living with systemic lupus erythematosus (SLE) experience substantial unmet needs related to diagnostic uncertainty, symptom burden, emotional distress, medication challenges, and healthcare system barriers.
- Traditional qualitative methods (e.g., interviews, focus groups, surveys) capture valuable patient perspectives but are limited by small sample sizes, recall bias, and restricted question frameworks.
- Social media listening has emerged as a promising way to collect real-time patient insights, and recent regulatory guidance acknowledges its value as patient experience data. However, systematic, scalable analysis of large patient-generated datasets has historically been constrained by analytic burden and variability.

What this study adds

- This study is among the first to apply **state-of-the-art large language models (LLMs)** to a large corpus of SLE-related social media posts, enabling scalable thematic analysis of thousands of patient narratives.
- It provides a **validated methodological framework** for using dual-LLM agreement, human-annotated references, and performance benchmarking (precision, recall, F1) to ensure reliability in automated thematic labeling.
- Findings reveal a **multidimensional patient burden** consistent with prior studies while uncovering persistent gaps in patient education, confusion around laboratory testing, care coordination challenges, and heavy reliance on peer communities for advice.
- The approach demonstrates that LLM-enabled social media listening can generate timely, granular, patient-prioritized insights at a scale unattainable by traditional methods.

How this study might affect research, practice, or policy

- **Research:** Establishes a reproducible, scalable framework for integrating LLM-based thematic analysis into patient-focused evidence generation, accelerating insight extraction from large unstructured datasets across disease areas.
- **Clinical practice:** Highlights actionable gaps in patient education, communication, and care coordination, informing interventions to improve clinical encounters, shared decision-making, and symptom management support.
- **Policy and regulatory science:** Demonstrates how social media–derived patient experience data, when paired with rigorous quality controls, can complement formal qualitative studies and support patient-focused drug development, measurement development, and health-services planning.

INTRODUCTION

Systemic Lupus Erythematosus (SLE) is a complex, chronic autoimmune disease characterized by immune-mediated inflammation affecting multiple organ systems.¹ Globally, the incidence rate of SLE is estimated at 5.14 per 100,000 person-years²,

with approximately 90% of cases occurring in females.¹ The disease exhibits profound heterogeneity in clinical presentation, disease trajectory, and immunological features, which complicates diagnosis, delays appropriate treatment, and contributes to persistent disparities in care.^{3,4} Together, these challenges impose a substantial and multifaceted burden on individuals living with SLE.

Despite substantial treatment advances over the past several decades^{5,6}, critical gaps remain in our understanding of disease burden and treatment from the patient perspective. Physician-assessed disease activity and biomarker-based evaluations often fail to align with patient-reported health status, symptom severity, and quality of life.⁷ This discordance reflects a fundamental dilemma of current clinical assessment frameworks: they tend to prioritize objectively measurable disease activity over the lived experience of illness while the biological mechanisms of SLE are only partly understood and biomarkers may not consistently capture symptom burden. Consequently, important dimensions of unmet medical needs, such as symptom fluctuation, treatment trade-offs, psychosocial burden, and disruptions to daily functioning remain underrecognized or insufficiently characterized. Addressing this gap is essential for advancing patient-centered care and informing therapeutic development that meaningfully reflects patient priorities.⁸⁻¹⁰

Traditionally, patient experience data has been collected through structured methods such as interviews, focus groups, and surveys.^{11,12,13} While these approaches provide valuable insights, they are inherently limited by small sample sizes, recall bias, and the constraints of predefined questions.^{11,12,13} These limitations are particularly pronounced in heterogeneous conditions like SLE, where recruitment can be challenging and patient experiences vary widely across different disease phenotypes and life stages. Thus, traditional methods may only capture a partial and static view of patient burden, rather than the dynamic and evolving nature of living with SLE.

In today's digital age, social media platforms have emerged as invaluable repositories of real-time patient insights and experiences.^{14,15} They offer unprecedented access to a wealth of unfiltered, candid narratives from individuals living with chronic diseases like SLE. Through social media listening, researchers and healthcare providers gain unique opportunities to identify and address previously unrecognized patient needs and concerns. Recognizing the potential of such data, regulatory bodies like the U.S. Food and Drug Administration (FDA) have officially acknowledged social media as a legitimate source of qualitative and quantitative patient experience data, providing guidelines on its ethical use starting in 2022.¹⁶

The integration of advanced natural language processing techniques, including large language models (LLMs), has revolutionized how we interpret and utilize these vast datasets.^{17,18} Multiple studies have applied LLMs to social media data across various diseases, including evaluation of glaucoma treatment¹⁹, analyses of unmet needs and provider burnout in thyroid disease²⁰, and assessments of public perceptions of obesity treatment²¹.

In this study, we apply LLM-enabled social media listening to systematically analyze Reddit posts from lupus-related communities, aiming to generate a comprehensive, patient-prioritized map of unmet medical needs and lived experiences in SLE. This approach provides a scalable and timely complement to patient-focused evidence generation, with direct relevance to care delivery, patient education, and shared decision-making. In addition, it establishes a reproducible analytical framework for leveraging patient-generated data to inform measurement development and therapeutic strategies.

METHODS

Data collection and preprocessing

This study used the free, publicly available Reddit Application Programming Interface (API) to retrieve patient-generated content related to SLE from ten subreddits (Figure 1). Subreddits are topic-specific forums on Reddit in which users create posts and engage in threaded discussions. Two complementary data retrieval strategies were employed. First, we collected posts published from eight lupus-specific subreddits: r/lupus, r/lupussupport, r/lupusRantAndRelief, r/lupusResearch, r/lupusmicrobiome, r/ItsNeverlupus, r/menwithlupus, and r/Livingwithlupus. Second, we conducted keyword-based searches (“lupus” OR “systemic lupus erythematosus”) within two broader health-focused subreddits (r/autoimmune and r/chronicillness). Data extraction was conducted between October 14, 2025 and November 25, 2025, yielding a total of 4,633 posts.

To ensure data quality and relevance, we implemented a multi-step preprocessing pipeline. Posts that contained both empty body text and title of five or fewer words were excluded due to insufficient context to reliably reflect patient or caregiver experiences. Duplicate posts were identified and removed using an exact-match rule based on identical author identifiers and post titles. Promotional materials (e.g., advertisement or research recruitment posts) and other non-patient reported content were identified and excluded with assistance from two state-of-the-art LLMs (Google Gemini 3.0 and OpenAI GPT-5.2).

Thematic codebook development

A thematic codebook was developed to systematically classify patient experiences and unmet medical needs. Three public health researchers independently reviewed and manually annotated 300 randomly selected posts to inductively generate an initial set of themes and subthemes. To ensure comprehensive coverage of the full dataset, an “Other” category was incorporated into the preliminary codebook, which was then used for an initial round of automated labeling across all posts. This step allowed for the identification of content not adequately captured by the initial thematic structure. If no “Other” category was identified in the entire dataset, we remove “Other” category from the final results. If no posts were assigned to the “Other” category across the full dataset, the category was excluded from the final codebook.

Following a review of a sampled model outputs, the research team iteratively refined the codebook to improve definitional clarity, enhance conceptual coverage, and ensure mutual exclusivity between subthemes. The finalized codebook comprised six main themes and 43 subthemes (Table 1) and was subsequently applied for definitive automated thematic labeling of the full dataset.

Table 1. Percent agreement and Cohen's κ between Gemini 3.0 Pro and GPT-5.2

Themes	Percent Agreement	Cohen's κ
Diagnosis Journey & Uncertainty	86.86	0.74
Delayed, Difficult, or Incorrect Diagnosis	93.89	0.79
Emotional Impact of Uncertainty and Disease Progression	81.18	0.56
Testing & Results Confusion (ANA, dsDNA, complements)	93.12	0.82
Provider Trust/ Distrust	91.20	0.64
Other	92.20	0.21
Emotional & Social Support	98.27	0.42
Advice Seeking	96.27	0.86
Emotional Coping (e.g., venting, hope, resilience)	86.71	0.73
Identity and/or Future Uncertainty	89.70	0.68
Relationship Strain / Interpersonal Challenges	96.70	0.84
Bereavement / Loss	99.15	0.73
Healthcare System Barriers	89.01	0.76
Access To Care (availability, wait times, transport)	96.12	0.80
Patient Education Gaps & External Misinformation	90.43	0.50
Dismissal Or Minimization By Providers	95.70	0.85
Insurance Challenges (coverage, prior auth)	99.35	0.91
Financial Burden Beyond Insurance	98.16	0.81
Care Coordination Challenges (referrals, records)	95.01	0.65
Clinical Care Errors / Safety Events	98.85	0.62
Medication Experiences & Side Effects	94.08	0.88
Medication Effectiveness / Response	92.24	0.80
Medication Side Effects / Adverse Events	95.93	0.87
Infusion Experience (e.g., infusion centers, scheduling)	98.27	0.74
Treatment Recommendations	90.32	0.66

Supplements Or Alternative Therapies	98.12	0.83
Treatment Adherence Challenges (forgetting, intolerance)	96.24	0.63
Biologics vs Traditional Medications	96.77	0.68
Clinical Trials Or Research Participation	99.69	0.84
Other	96.39	0.16
Symptoms & Physical Experiences	92.24	0.76
Acute or Worsening Flares	88.78	0.73
Fatigue	96.77	0.91
Pain	94.43	0.88
Skin Manifestations (rashes, photosensitivity)	94.85	0.86
Cognitive Symptoms (brain fog, memory)	98.77	0.91
Other Symptom Presentation (e.g., renal, cardio)	84.48	0.69
Other	94.08	0.18
Lifestyle & Self-Management	85.79	0.72
Symptom Management Strategies	84.67	0.57
Diet and Nutrition Changes	97.77	0.85
Exercise Adaptations	98.35	0.81
Stress Management Strategies	96.39	0.53
Workplace Disclosure / Accommodations & Disability Benefits	96.62	0.80
Sun Exposure Management	97.69	0.80
Self-Tracking / Tools (journals, apps)	98.62	0.74
Activities Of Daily Living Adjustments	87.98	0.65
Body Image / Identity Adjustments	94.12	0.60
Reproductive / Fertility Treatment Considerations	98.73	0.83

GPT, Generative Pre-trained Transformer; κ , Cohen's Kappa coefficient

LLM evaluation and model selection

Two state-of-the-art LLMs, Google Gemini 3.0 and OpenAI GPT-5.2, were used for automated thematic labeling via their respective APIs. Prompt engineering was performed to optimize instructions for consistent labeling of Reddit posts to the main themes and subthemes defined in the final thematic codebook. Both models were instructed to assign all applicable main theme and subtheme labels for each post, rather than selecting a single dominant theme, to reflect the multidimensional nature of patient narratives.

To assess the reliability of automated thematic labeling, we evaluated inter-model consistency by calculating percent agreement and Cohen's kappa (κ)²² across all main themes and subthemes (Table 1). Agreement was defined as both models assigning the same binary label (Yes or No) for a given theme or subtheme within a post. Interpretation of κ followed established empirical guideline: ≤ 0 , indicating no agreement; 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, almost perfect agreement.²³ The distribution of agreement across posts for both main themes and subthemes agreement was further visualized to characterize consistency patterns between the two LLMs (Figure 2).

Additionally, three public health researchers independently annotated 100 randomly selected posts to establish a human-annotated reference dataset for performance evaluation. Discrepancies were resolved using a majority-vote consensus approach to determine the final labels for each main theme and subtheme. Using this consensus reference, model performance was evaluated using precision, recall, and F1-score, and the best-performing LLM was subsequently selected for downstream thematic analysis.

Thematic analysis and reporting

Using the best-performing LLM, we quantified the prevalence of each main theme and subtheme by calculating counts and percentages across the 2,603 patient posts (Table 2). To support interpretation of the quantitative findings, we conducted a qualitative review of representative posts within each main theme to identify salient patient experiences and unmet medical needs. Selected examples were synthesized into key phrases and illustrative summaries to contextualize the thematic patterns observed and are presented in Figure 3.

Table 2. Distribution of main themes and subthemes identified by the best-performing model (OpenAI GPT 5.2)

Main themes	Subthemes	OpenAI GPT 5.2
Diagnosis Journey & Uncertainty	Delayed, Difficult, or Incorrect Diagnosis	432 (16.6%)
	Emotional Impact of Uncertainty and Disease Progression	860 (33.0%)
	Testing & Results Confusion (ANA, dsDNA, complements)	624 (24.0%)
	Provider Trust/ Distrust	426 (16.4%)
	Other	82 (3.2%)
Emotional & Social Support	Advice Seeking	2190 (84.1%)
	Emotional Coping (e.g., venting, hope, resilience)	1448 (55.6%)
	Identity and/or Future Uncertainty	572 (22.0%)
	Relationship Strain / Interpersonal Challenges	278 (10.7%)
	Bereavement / Loss	47 (1.8%)

Healthcare System Barriers	Access To Care (availability, wait times, transport)	290 (11.1%)
	Patient Education Gaps & External Misinformation	332 (12.8%)
	Dismissal Or Minimization By Providers	424 (16.3%)
	Insurance Challenges (coverage, prior auth)	104 (4.0%)
	Financial Burden Beyond Insurance	122 (4.7%)
	Care Coordination Challenges (referrals, records)	205 (7.9%)
	Clinical Care Errors / Safety Events	46 (1.8%)
Medication Experiences & Side Effects	Medication Effectiveness / Response	656 (25.2%)
	Medication Side Effects / Adverse Events	527 (20.2%)
	Infusion Experience (e.g., infusion centers, scheduling)	82 (3.2%)
	Treatment Recommendations	435 (16.7%)
	Supplements Or Alternative Therapies	152 (5.8%)
	Treatment Adherence Challenges (forgetting, intolerance)	161 (6.2%)
	Biologics vs Traditional Medications	120 (4.6%)
	Clinical Trials Or Research Participation	23 (0.9%)
	Other	30 (1.2%)
Symptoms & Physical Experiences	Acute or Worsening Flares	785 (30.2%)
	Fatigue	642 (24.7%)
	Pain	969 (37.2%)
	Skin Manifestations (rashes, photosensitivity)	632 (24.3%)
	Cognitive Symptoms (brain fog, memory)	183 (7.0%)
	Other Symptom Presentation (e.g., renal, cardio)	980 (37.6%)
	Other	82 (3.2%)
Lifestyle & Self-Management	Symptom Management Strategies	569 (21.9%)
	Diet and Nutrition Changes	190 (7.3%)
	Exercise Adaptations	111 (4.3%)
	Stress Management Strategies	124 (4.8%)
	Workplace Disclosure / Accommodations & Disability Benefits	210 (8.1%)

	Sun Exposure Management	133 (5.1%)
	Self-Tracking / Tools (journals, apps)	71 (2.7%)
	Activities Of Daily Living Adjustments	479 (18.4%)
	Body Image / Identity Adjustments	162 (6.2%)
	Reproductive / Fertility Treatment Considerations	98 (3.8%)

GPT, Generative Pre-trained Transformer

RESULTS

Collection of SLE related posts

Among the 4633 collected posts, 2,030 (43.8%) were excluded according to the predefined data preprocessing pipeline (Figure 1). Exclusions included posts with missing author identifiers or insufficient title clarity (n = 252), duplicate entries (n = 1,013), and promotional and other non-patient reported content (n = 765) (Figure 1). The final dataset consisted of 2,603 unique posts (56.2%), of which 2,564 (98.4%) published after 2020 (Table 3).

Table 3. Number of posts by publication year

Year	Number of Posts
2011	8
2012	3
2013	9
2014	1
2015	2
2016	2
2017	3
2018	5
2019	8
2020	33
2021	33
2022	46
2023	71
2024	278
2025	2,101

LLMs consistency and performance

Inter-model agreement achieved high consistency across main themes and subthemes (Table 1 and Figure 2). Overall, percent agreement and Cohen's κ indicated substantial agreement between Gemini 3.0 and GPT-5.2 for main themes (percent agreement = 91.04%; mean κ = 0.71; range = 0.42 – 0.88) and for subthemes (percent agreement = 94.43%; mean κ = 0.71; range = 0.16 – 0.91).

At the main-theme level, almost perfect agreement was observed for *Medication Experiences & Side Effects* (percent agreement = 94.08%; κ = 0.88). Substantial agreement was also observed for *Healthcare System Barriers* (percent agreement = 89.01%; κ = 0.76), *Lifestyle & Self-Management* (percent agreement = 85.79%; κ = 0.71), *Diagnosis Journey & Uncertainty* (percent agreement = 94.08%; κ = 0.74), and *Symptoms & Physical Experiences* (percent agreement = 92.24%; κ = 0.76). In contrast, *Emotional & Social Support* demonstrated moderate agreement despite high percent agreement (percent agreement = 98.27%; κ = 0.42).

Across the 43 subthemes, 15 subthemes demonstrate almost perfect agreement (κ range = 0.81 – 0.91). The highest agreement was observed for *Insurance Challenges* (e.g. *Coverage*, *Prior Authorization*), *Fatigue*, and *Cognitive Symptoms* (e.g. *Brain Fog*, *Memory*), each achieving κ = 0.91 with percent agreement of 99.35%, 96.77%, and 98.77%, respectively. Among the remaining subthemes, 17 demonstrated substantial agreement, four demonstrated moderate agreement, one demonstrated fair agreement and two demonstrated slight agreement (Table 1).

Post-level agreement analyses further supported high inter-model consistency. Across the six main themes, 75% of posts exhibited greater than 80% agreement between the two LLMs (Figure 2a). *Healthcare System Barriers* showed the highest consistency, with more than 75% of classified posts achieving 100% agreement. For the remaining main themes, more than half of posts demonstrated 100% agreement at the subtheme level (Figure 2b).

Compared with the consensus reference labels derived from 100 randomly selected posts, GPT-5.2 achieved higher overall performance than Gemini 3.0 (F1: 0.844 vs 0.811), driven by higher precision (0.743 vs 0.694) with similarly high recall (0.977 vs 0.975). Across the 36 subthemes that have sampled posts in our human evaluation, GPT-5.2 outperformed Gemini 3.0 in ten subthemes, Gemini 3.0 outperformed GPT-5.2 in five subthemes, and the models performed comparably in 21 subthemes. Both models demonstrated very high recall, indicating strong sensitivity for detecting relevant main themes and subthemes, while GPT 5.2 showed higher precision, resulting in a higher F1 score and a better balance between sensitivity and false-positive control. We therefore chose GPT-5.2 as the best-performing model and conducted the following thematic analysis using it.

Thematic distribution from the best-performing model

Table 3 summarizes the distribution of main themes and subthemes identified by GPT-5.2. Of the 2,603 analyzed posts, each post received at least one subtheme label, with a mean of approximately six subthemes assigned per post.

The most frequent subtheme was *Advice Seeking* (84.1%), followed by *Emotional Coping* (e.g., *venting*, *hope*, *resilience*) (55.6%) within *Emotional & Social Support*. Discussions within *Symptoms & Physical Experiences* were also common, including *Other Symptom Presentation* (e.g., *renal*, *cardio*, 37.6%), *Pain* (37.2%), and *Acute or Worsening Flares* (30.2%). Within the *Diagnosis Journey & Uncertainty*, *Emotional Impact of Uncertainty and Disease Progression* was frequently labeled (33.0%). Treatment-related content was prevalent as well, notably *Medication Effectiveness/Response* (25.2%).

Qualitative review of patient narratives

To contextualize the quantitative distribution, we conducted a qualitative review of representative patient narratives (Figure 3). Several narratives described difficulty navigating referrals, records, and care coordination. For example, one patient shared an uncertain diagnosis (*"it could be lupus...without no follow-up plan"*), and another portrayed the path to diagnosis as a prolonged *"puzzle,"* as the journey spanning multiple years (*"battling for 6 years just to be diagnosed."*).

Many posts reflected reliance on peer communities for validation and guidance, particularly after perceived dismissal in clinical encounters or during delays in accessing specialty care. For example, one patient wrote, *"Please let me know y'all's opinions"* regarding whether beta-blocker use could be related to drug-induced lupus. Another solicited community opinions on interpreting laboratory results (*"low, total immunoglobulin E... what it means in relation to what I have going on?"*) while a long awaiting clinical appointment (*"Don't see the doctor for another 2 months"*).

Beyond informational needs, numerous posts captured emotional distress and identity challenges, including self-doubt and feelings of invalidation (e.g., *"I'm not sure what's wrong with me"*), as well as guilt related to role functioning when symptoms interfered with caregiving or work responsibilities (*"Feel like a failure of a mother...can't even hold myself up most days."*).

Narratives emphasizing multisystem involvement and functional impairment were common. One patient described that lupus *"doesn't pick one system; it attacks them all,"* alongside accounts of symptoms being minimized as *"just a rash."* Accounts of severe, diffuse pain and heightened sensitivity appeared frequently (*"Everything hurt."*). In discussions of acute or worsening flares, posters described abrupt escalation of symptoms and extended recovery periods, including delayed sequelae (e.g., *"hair loss lagging 2–3 months after severe systemic flare."*).

Posts also conveyed anxiety driven by ambiguous diagnostic pathways and fragmented care planning. For instance, one patient described cross-specialty deferral during

evaluation of acute foot pain, stating that the podiatrist *“refused to treat my foot pain”* and told the patient, *“I don’t treat lupus.”* Finally, posts reflected uncertainty regarding whether treatments were working and how quickly improvement should occur, illustrated by posts such as *“Monthly ER visits since June...each triggered by new infections due to immunosuppression”* and *“Torn between fear of flares if dose drops vs. fear of toxicity if it stays high”*, which demonstrated current unmet medical needs of SLE patients.

DISCUSSION

This study applied LLMs to analyze lupus-related Reddit posts and identify patient-expressed unmet medical needs and lived experiences in SLE. Our finding indicated that patients described a multidimensional spanning from diagnosis uncertainty, emotional and social needs, healthcare system barriers, treatment experiences, symptom burden, and self-management challenges. Advice seeking and emotional coping were the most prominent narrative patterns, suggesting substantial reliance on peer networks particularly when patients feel unheard in clinical settings. These psychosocial needs align with prior qualitative studies describing ongoing gaps in communication, validation, and daily functioning among individuals living with SLE.^{24,25}

Specific symptom experiences shared by patients corresponded with established qualitative literature emphasizing patient-centered voices on pain²⁶, fatigue^{27,28}, acute or worsening flares²⁹. Cognitive symptoms, including brain fog and memory difficulties, were also commonly discussed, consistent with previous qualitative research on lupus-related cognitive impairment.³⁰ Notably, Other symptom presentations was one of the most prevalent subthemes, reinforcing that SLE rarely exists in isolation with its impact often amplified by coexisting chronic conditions. These observations were supported by a prior study focused on describing the impact of comorbidities on quality of life in SLE³¹ and underscore the need to approach disease burden as prioritized by patients, rather than focusing on a narrow set of symptoms.

Treatment experiences emphasized unmet medical needs in medication decision-making and long-term treatment management, including uncertainty about effectiveness, concerns about adverse effects, and adherence challenges. Prior studies have described medication concerns and adherence challenges driven by the fear of side effects, forgetting, disbelief in medication,³² while identifying trust in rheumatologists and structured follow-up as key facilitators of sustained adherence.³³ Similarly, patient narratives in our results reflected uncertainty regarding medication effectiveness and drug-drug interactions, as well as barriers to adherence, suggesting that SLE management often involves medication regimens that may seem as complex as the disease itself with limited clarity on long-term impacts.

Moreover, unmet medical needs related to uncertainty and concerns about disease progression were also frequently observed within the diagnosis journey, consistent with prior findings that uncertainty may persist even after diagnosis.³⁴ Provider trust and distrust emerged as an important contextual factor for previously described medication adherence facilitators³³, as early clinical interactions may shape patients confidence in treatment recommendations and engagement with longitudinal care. Delayed diagnosis

or misdiagnosis were commonly reported, often in the context of symptom minimization or misattribution, aligning with qualitative studies across diverse geographic settings that have examined barriers and facilitators associated with timely diagnosis and appropriate treatment in SLE.^{35–38}

While diagnostic uncertainty³⁴ and challenges in patient–provider dynamics^{32,33,39} have been well documented in previous qualitative studies of SLE patient experience, confusion related to serologic testing of patient experience during the diagnostic process has received comparatively limited attention in prior qualitative and social media-based studies. In this study, patient confusion surrounding interpretation of laboratory test results, such as anti-nuclear antibody (ANA) and complement levels, appeared as a frequent unmet medical need. This knowledge gap often drove patients to seek peer interpretations of laboratory values and discordance between symptoms and test results.

Taken together, our findings were largely consistent with existing literature from multiple main themes on unmet medical needs and patient-reported experience in SLE while also highlighting persistent gaps in patient education and communication between clinician and patients that warrant greater attention in patient-centered care. The consistency of major unmet medical need themes with prior literature also supports the credibility of LLM based thematic analysis as a complementary approach for synthesizing large scale unstructured patient narratives from social media.

This study had several limitations. First, the analysis is subject to self-selection bias inherent to social media research. Users who post in lupus-related subreddits may not represent the broader SLE population and may have higher symptom burden, greater unmet needs, or stronger engagement with online peer communities than those who do not participate in online forums.⁴⁰ Consequently, the narratives captured may overrepresent more persistent, disruptive, or emotionally salient experiences. Second, the anonymity of the Reddit platform precluded the verification of users' clinical diagnoses and restricted access to sociodemographic information.⁴¹ As a result, our analysis relied on self-identified SLE status, which could not be independently confirmed and may have included individuals with other lupus-related disorders. This limitation is inherent to social media research and should be considered when interpreting the generalizability of our findings. Furthermore, subgroup analyses by age, sex, race or ethnicity, or geographic were not feasible, and variation in unmet needs across social determinants of health could not be assessed.

Third, the analysis was restricted to a single social media platform. Although additional lupus-related platforms were explored, data access restrictions and usage policies limited compliant large-scale extraction. Finally, while LLMs offer scalable analysis of unstructured data, their outputs may vary across repeated runs.⁴² We observed approximately 10% variation in classification during automated identification of promotional and non-patient reported posts, consistent with prior evidence of stochastic variability in LLM outputs, analogous to inter-rater variability in human annotation.

CONCLUSION

Our study demonstrates that large-scale social media listening, enabled by state-of-the-art LLMs, can provide timely, patient-centered insights into unmet medical needs and lived experiences in SLE. Consistent with prior qualitative and survey-based studies, patient narratives reflected a multidimensional burden spanning diagnostic uncertainty, symptom and treatment challenges, emotional distress, and gaps in clinical support and communication. By systematically analyzing thousands of patient-generated narratives, this work illustrates an efficient and scalable approach to synthesizing unstructured data that complements established patient-focused methods. More broadly, the analytical framework presented here can be extended to other disease areas to support rapid identification of patient priorities, inform patient-centered care strategies, and guide evidence generation relevant to therapeutic development and health services research.

Contributions:

SY and LY conceived and designed the study and drafted the manuscript. LY supervised the study. CH, NE, and JO conducted qualitative review and manual annotations. JL conducted data collection, preprocessing, and LLM-based thematic classification. SY conducted the data analysis. ER provided clinical review and critical revisions. All authors reviewed and approved the final version.

Data availability statement

The data analyzed in this study were derived from publicly available Reddit posts accessed via the Reddit API. Processed datasets and analysis codes may be made available by the corresponding author upon reasonable request, subject to platform terms and privacy considerations.

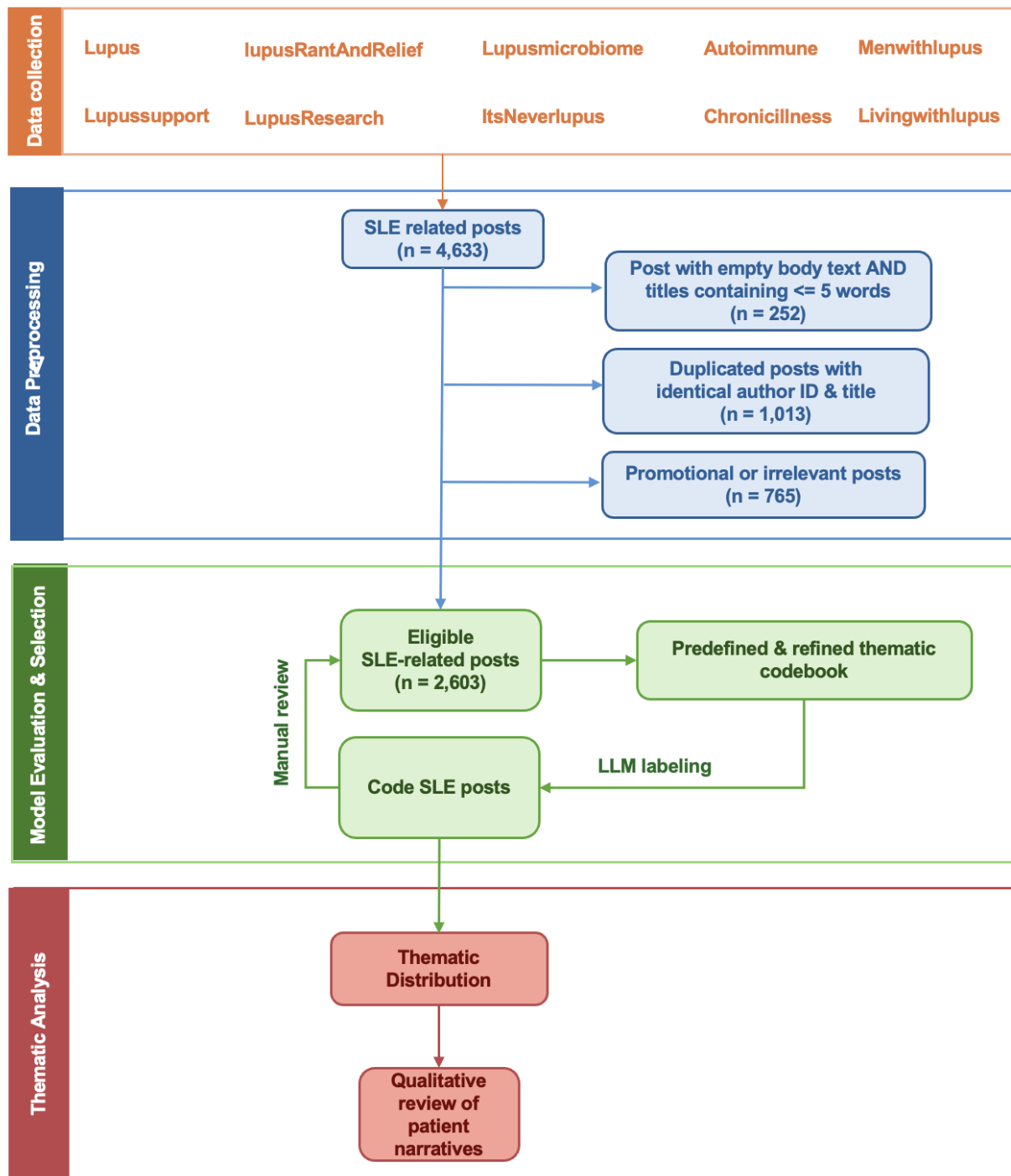
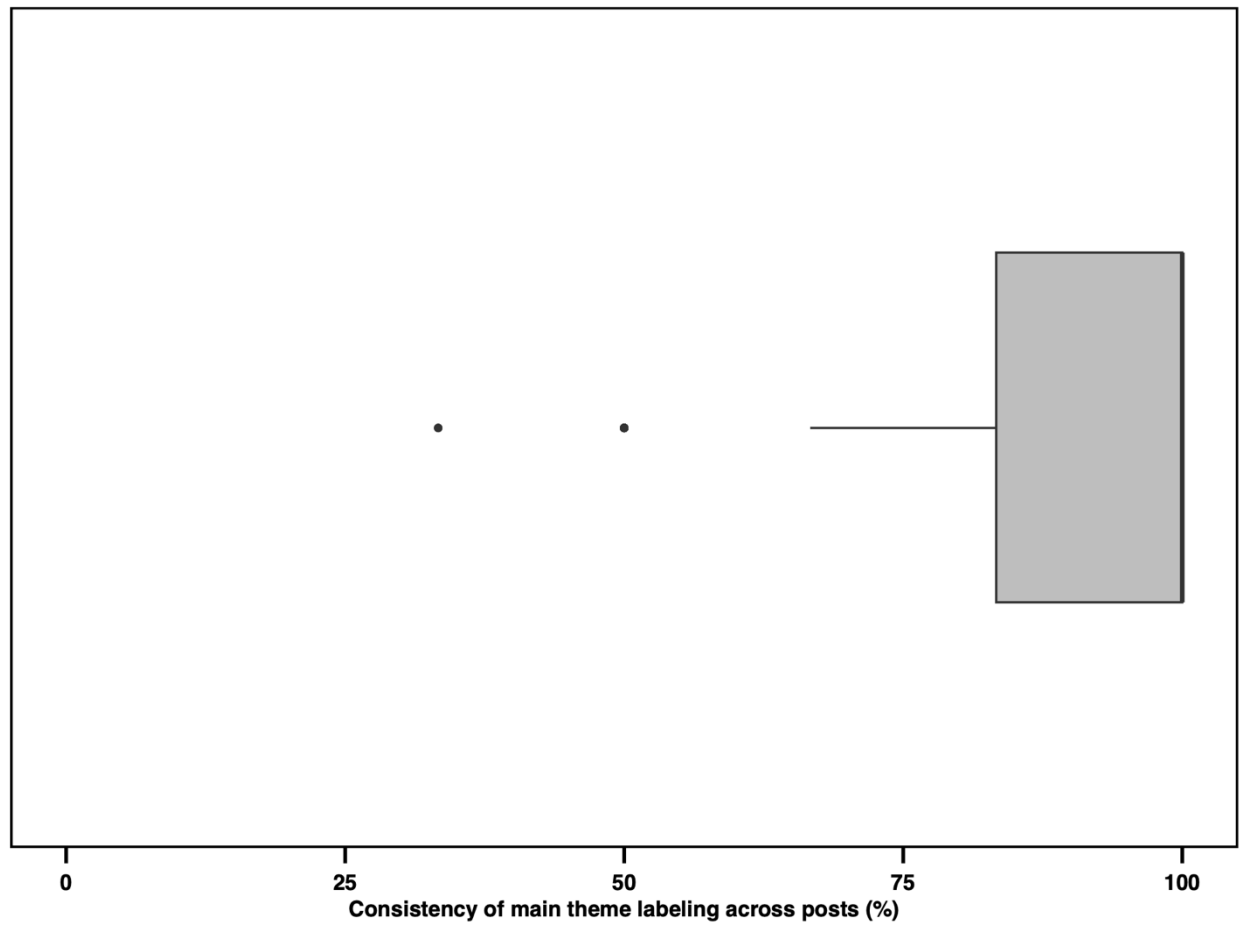


Figure 1. Workflow diagram for data collection, preprocessing, and LLM-based thematic analysis

(A) Main themes consistency distribution



(B) Subthemes consistency distribution

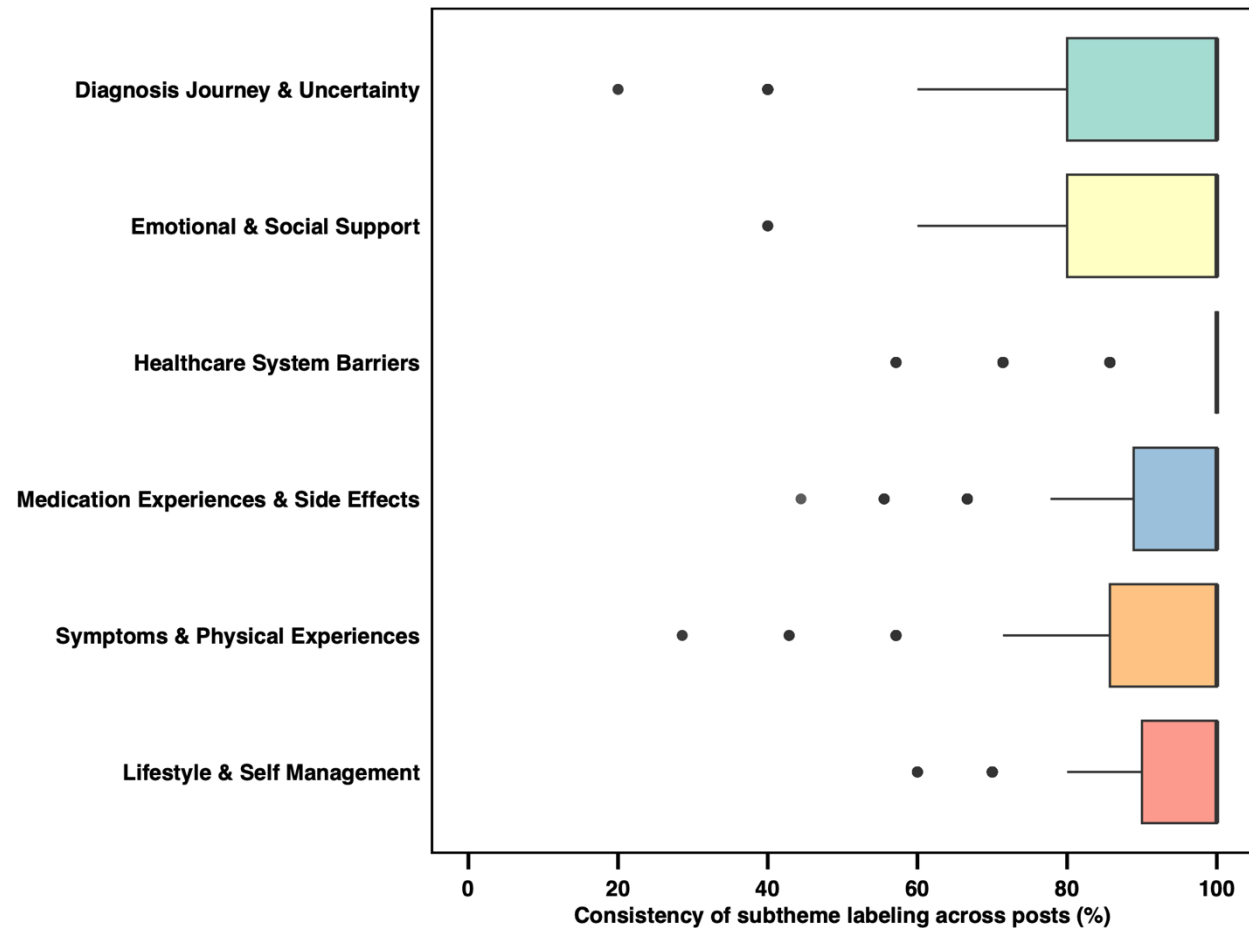


Figure 2. Thematic consistency of LLMs.

(A) Boxplot showing the distribution of inter-model consistency for main theme labeling across posts. Consistency is defined as the percentage of posts assigned the same main theme label by Gemini 3.0 and GPT-5.2. The box denotes the interquartile range (25th–75th percentiles), and the median is shown as a line (overlapping the upper quartile in this panel). Outliers are shown as individual points.

(A) Boxplots showing the distribution of inter-model consistency for subtheme labeling across posts, stratified by main theme. Consistency is defined as the percentage of posts assigned the same subtheme label by Gemini 3.0 and GPT-5.2. Boxes denote the interquartile range (25th–75th percentiles), and the median line overlaps the upper quartile in all panels. For Healthcare System Barrier, 25th is also overlapped with median and 75th percentile. Outliers are shown as individual points.



Figure 3. Unmet Medical Needs and Patient Experiences Illustrated by Selected Patient Narratives

References

1. Siegel CH, Sammaritano LR. Systemic Lupus Erythematosus: A Review. *JAMA*. 2024;331(17):1480-1491. doi:10.1001/jama.2024.2315
2. Barber MRW, Drenkard C, Falasinnu T, et al. Global epidemiology of systemic lupus erythematosus. *Nat Rev Rheumatol*. 2021;17(9):515-532. doi:10.1038/s41584-021-00668-1
3. Scopes J, Md Yusof MY. Challenges in the diagnosis of systemic lupus erythematosus. *Clin Med*. 2025;25(5):100509. doi:10.1016/j.clinme.2025.100509
4. Pons-Estel GJ, Ramírez-Flores MF, Quintana R, et al. Addressing the challenge of global delays in diagnosis and treatment of systemic lupus erythematosus. *Nat Rev Rheumatol*. 2025;21(9):566-574. doi:10.1038/s41584-025-01277-y
5. Scherlinger M, Kolios AGA, Kyttaris VC, Tsokos GC. Advances in the treatment of systemic lupus erythematosus. *Nat Rev Drug Discov*. 2025;24(12):926-944. doi:10.1038/s41573-025-01242-0
6. Zarei-Behjani Z, Hosseinpouri A, Fotoohi M, Shafiee A, Asadi D. Beyond conventional treatment: Novel cell therapies for systemic lupus erythematosus. *J Transl Autoimmun*. 2025;11:100308. doi:10.1016/j.jtauto.2025.100308
7. Golder V, Ooi JJY, Antony AS, et al. Discordance of patient and physician health status concerns in systemic lupus erythematosus. *Lupus*. 2018;27(3):501-506. doi:10.1177/0961203317722412
8. Manrique de Lara A, Peláez Ballestas I del P, Mendoza-Pinto C, et al. Unmet Needs in the Care of Systemic Lupus Erythematosus: Insights and Recommendations From Patients and Rheumatologists in Latin America. *JCR J Clin Rheumatol*. 2026;32(1S):S58. doi:10.1097/RHU.0000000000002304
9. Moroni L, De Marchi G, Pelissero R, et al. Italian Systemic Lupus Erythematosus (SLE) Patients: Overview of Their Quality of Life and Unmet Needs. *J Clin Med*. 2025;14(23):8498. doi:10.3390/jcm14238498
10. Cornet A, Karakikla Mitsakou Z, Andersen J, et al. Experiences and unmet needs of persons living with systemic lupus erythematosus in Europe: Lupus Europe's 2024 Swiss knife survey. *Autoimmun Rev*. 2025;24(8):103838. doi:10.1016/j.autrev.2025.103838
11. Rupert DJ, Poehlman JA, Hayes JJ, Ray SE, Moultrie RR. Virtual Versus In-Person Focus Groups: Comparison of Costs, Recruitment, and Participant Logistics. *J Med Internet Res*. 2017;19(3):e80. doi:10.2196/jmir.6980
12. Tausch AP, Menold N. Methodological Aspects of Focus Groups in Health Research. *Glob Qual Nurs Res*. 2016;3:2333393616630466. doi:10.1177/2333393616630466
13. Ponto J. Understanding and Evaluating Survey Research. *J Adv Pract Oncol*. 2015;6(2):168-171.
14. Chen J, Wang Y. Social Media Use for Health Purposes: Systematic Review. *J Med Internet Res*. 2021;23(5):e17917. doi:10.2196/17917
15. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. *Am J Public Health*. 2017;107(1):e1-e8. doi:10.2105/AJPH.2016.303512
16. Research C for DE and. Patient-Focused Drug Development: Methods to Identify What Is Important to Patients. January 31, 2025. Accessed January 12, 2026.

- <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-methods-identify-what-important-patients>
17. Sarker A, Zhang R, Wang Y, et al. Natural Language Processing for Digital Health in the Era of Large Language Models. *Yearb Med Inform*. 2025;33(1):229-240. doi:10.1055/s-0044-1800750
18. Deiner MS, Honcharov V, Li J, Mackey TK, Porco TC, Sarkar U. Large Language Models Can Enable Inductive Thematic Analysis of a Social Media Corpus in a Single Prompt: Human Validation Study. *JMIR Infodemiology*. 2024;4:e59641. doi:10.2196/59641
19. Bulusu A, Cotran PR, Alwreikat AM, et al. Evaluation of Glaucoma Treatment Information on Social Media Using Large Language Models. *J Glaucoma*. Published online December 11, 2025. doi:10.1097/IJG.0000000000002673
20. Raghunathan R, Jacobs AR, Sant VR, et al. Can large language models address unmet patient information needs and reduce provider burnout in the management of thyroid disease? *Surgery*. 2025;177:108859. doi:10.1016/j.surg.2024.06.075
21. Somani S, Jain SS, Sarraju A, Sandhu AT, Hernandez-Boussard T, Rodriguez F. Using large language models to assess public perceptions around glucagon-like peptide-1 receptor agonists on social media. *Commun Med*. 2024;4(1):137. doi:10.1038/s43856-024-00566-z
22. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
23. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-282.
24. Brennan KAM, Creaven AM. Living with invisible illness: social support experiences of individuals with systemic lupus erythematosus. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2016;25(5):1227-1235. doi:10.1007/s11136-015-1151-z
25. Leung J, Ra J, Baker EA, Kim AHJ. "...Not Having the Real Support That We Need": Patients' Experiences With Ambiguity of Systemic Lupus Erythematosus and Erosion of Social Support. *Acr Open Rheumatol*. 2019;1(3):135-144. doi:10.1002/acr2.1020
26. Waldheim E, Welin E, Bergman S, Pettersson S. The experience and implications of pain in systemic lupus erythematosus: A qualitative interview study focusing on the patient's perspective. *Lupus*. 2024;33(13):1435-1445. doi:10.1177/09612033241284093
27. Cleanthous S, Strzok S, Haier B, Cano S, Morel T. The Patient Experience of Fatigue in Systemic Lupus Erythematosus: A Conceptual Model. *Rheumatol Ther*. 2022;9(1):95-108. doi:10.1007/s40744-021-00374-0
28. Raymond K, Park J, Joshi AV, White MK. Patient Experience With Fatigue and Qualitative Interview-Based Evidence of Content Validation of The FACIT-Fatigue in Systemic Lupus Erythematosus. *Rheumatol Ther*. 2021;8(1):541-554. doi:10.1007/s40744-021-00292-1
29. Rogers JL, Clowse MEB, McKenna K, et al. Patient and Physician Perspectives of Systemic Lupus Erythematosus Flare: A Qualitative Study. *J Rheumatol*. 2024;51(5):488-494. doi:10.3899/jrheum.2023-0721
30. Barraclough M, Howe A, Soberanis A, et al. The Effects of Systemic Lupus-Related Cognitive Impairments on Activities of Daily Living and Life Role Participation: A

- Qualitative Framework Study. *ACR Open Rheumatol*. 2024;6(1):21-30. doi:10.1002/acr2.11624
31. Aljohani RA, Aljanobi GA, Alderaan K, Omair MA. Exploring the quality of life and comorbidity impact among patients with systemic lupus erythematosus in Saudi Arabia. *Saudi Med J*. 2024;45(10):1071-1079. doi:10.15537/smj.2024.45.10.20240432
32. Emamikia S, Gentline C, Enman Y, Parodis I. How Can We Enhance Adherence to Medications in Patients with Systemic Lupus Erythematosus? Results from a Qualitative Study. *J Clin Med*. 2022;11(7):1857. doi:10.3390/jcm11071857
33. Farinha F, Freitas F, Águeda A, Cunha I, Barcelos A. Concerns of patients with systemic lupus erythematosus and adherence to therapy - a qualitative study. *Patient Prefer Adherence*. 2017;11:1213-1219. doi:10.2147/PPA.S137544
34. Mazhariazad F, Dianati M, Taghadosi M, Zamani B. Uncertainty and its related coping strategies in systemic lupus erythematosus. Life in the fog. *J Educ Health Promot*. 2023;12:233. doi:10.4103/jehp.jehp_1080_22
35. Colmenares-Roa T, Espinosa-Escobar C, Manrique de Lara A, et al. Barriers and Facilitators Associated With Delays in the Diagnosis and Treatment of Lupus: A Qualitative Study of Patients and Rheumatologists in the Central American and Caribbean Region. *J Clin Rheumatol Pract Rep Rheum Musculoskelet Dis*. 2026;32(1S Suppl 1):S21-S27. doi:10.1097/RHU.0000000000002290
36. Manrique de Lara A, Colmenares-Roa T, Athié L, et al. Barriers and facilitators associated with diagnostic and treatment delays in lupus in the Global South and North: a systematic review of qualitative and mixed methods studies. *Rheumatol Int*. 2025;45(5):121. doi:10.1007/s00296-025-05875-7
37. Ospina-Caicedo A, Fuentes-Silva Y, Escobar-Jimenez K, et al. Barriers and Facilitators Associated With Delays in the Diagnosis and Treatment of Lupus: Insights From Patients and Rheumatologists From the Andean/Bolivarian Region. *J Clin Rheumatol Pract Rep Rheum Musculoskelet Dis*. 2026;32(1S Suppl 1):S28-S35. doi:10.1097/RHU.0000000000002301
38. Peláez-Ballestas IDP, Pons-Estel G, Pons-Estel BA. Delays in the Diagnosis and Treatment of Systemic Lupus Erythematosus in Latin America: A Multinational Qualitative Study From the GLADEL/PANLAR Network. *J Clin Rheumatol Pract Rep Rheum Musculoskelet Dis*. 2026;32(1S Suppl 1):S1-S4. doi:10.1097/RHU.0000000000002302
39. Sloan M, Lever E, Gordon C, et al. Medication decision-making and adherence in lupus: patient-physician discordance and the impact of previous "adverse medical experiences." *Rheumatol Oxf Engl*. 2022;61(4):1417-1429. doi:10.1093/rheumatology/keab534
40. Lohmann S, Zagheni E. Diversity of social media use: Self-selection explains associations between using many platforms and well-being. *PLOS Digit Health*. 2023;2(7):e0000292. doi:10.1371/journal.pdig.0000292
41. Proferes N, Jones N, Gilbert S, Fiesler C, Zimmer M. Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Soc Media Soc*. 2021;7(2):20563051211019004. doi:10.1177/20563051211019004

42. Funk PF, Hoch CC, Knoedler S, et al. ChatGPT's Response Consistency: A Study on Repeated Queries of Medical Examination Questions. *Eur J Investig Health Psychol Educ*. 2024;14(3):657-668. doi:10.3390/ejihpe14030043