Capturing Patients' Lived Experiences with Chronic Pain through Motivational Interviewing and Information Extraction

Hadeel Elyazori[1,*], **Rusul Abdulrazzaq**[1], **Hana Al Shawi**[1], **Isaac Paki Elom Amouzou**[1], **Patrick King**[1], **Syleah Manns**[1], **Mahdia Popal**[1], **Zarna Narsihbhai Patel**[1], **Secili DeStefano**[2], **Jay Shah**[3], **Lynn H. Gerber**[1,4], **Siddhartha Sikdar**[1], **Seiyon Lee**[1], **Samuel Acuña**[1,†], **Kevin Lybarger**[1,†]

[1]George Mason University, [2]Optimal Motion Physical Therapy,
[3]National Institutes of Health Clinical Center, [4]INOVA Health System

[*]Corresponding author helyazor@gmu.edu

[†]Contributed equally to this work as senior authors

Abstract

Chronic pain affects millions, yet traditional assessments often fail to capture patients' lived experiences comprehensively. In this study, we used a Motivational Interviewing framework to conduct semi-structured interviews with eleven adults experiencing chronic pain and then applied Natural Language Processing (NLP) to their narratives. We developed an annotation schema that integrates the International Classification of Functioning, Disability, and Health (ICF) with Aspect-Based Sentiment Analysis (ABSA) to convert unstructured narratives into structured representations of key patient experience dimensions. Furthermore, we evaluated whether Large Language Models (LLMs) can automatically extract information using this schema. Our findings advance scalable, patient-centered approaches to chronic pain assessment, paving the way for more effective, data-driven management strategies.

1 Introduction

Chronic pain affects millions worldwide, diminishing quality of life and straining healthcare systems (Goldberg and McGee, 2011). In 2023, an estimated 24.3% of U.S. adults (~51.6 million individuals) experienced chronic pain (Lucas and Sohi, 2024). Beyond physical discomfort, it impacts work productivity, personal relationships, social interactions, sleep quality, and mental health (Hadi et al., 2019; Dueñas et al., 2016). Managing chronic pain remains challenging due to its multidimensional and highly individualized nature. Each patient's experience is shaped by genetics, early life events, psychological state, coexisting medical conditions, and environmental influences (Institute of Medicine, 2011; Fillingim, 2017). Many individuals experience debilitating pain without clear pathology (Fine, 2011; Dueñas et al., 2016). Over time, the persistent stress of chronic pain contributes to *allostatic load*—physiological strain that exacerbates pain severity and accelerates health decline (McCaffery et al., 2012). Consequently, understanding chronic pain requires a holistic approach that extends beyond physical symptoms.

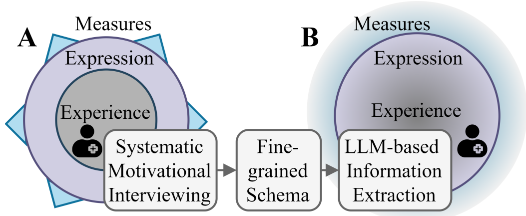

Figure 1: Integration of MI and IE to capture patients' pain experience, building on Wideman et al. (2019)

Traditional pain assessment methods rely heavily on clinical history and standardized measures, which often fail to capture the complexity of pain experience (Wideman et al., 2019; Radnovich et al., 2014; Gordon, 2015). This limitation stems from the fragmented conceptualization of pain, as shown in Figure 1A. Wideman et al. (2019) divide pain into: 1) pain experience—the subjective, intangible nature of pain that is difficult to observe; 2) pain expression—how pain is communicated verbally and non-verbally; and 3) pain measures—standardized assessments that translate expressions into numerical or categorical values. While pain measures provide objective data, they oversimplify patients' lived experiences, failing to capture the multifaceted and interconnected nature of pain. Consequently, critical aspects of pain remain poorly understood. In contrast, Figure 1B illustrates an integrated framework that our work aims to realize, where pain experience is central but is more comprehensively expressed and measured through a combination of subjective narratives and quantifiable metrics. According to the National Center for Complementary and Integrative Health (2024), adopting the "whole person" approach can lead to more comprehensive, nuanced, and effective pain assessment and treatment paradigms.

Proceedings of the Second Workshop on Patient-Oriented Language Processing (CL4Health), pages 321–330
May 4, 2025 ©2025 Association for Computational Linguistics

This study addresses these challenges by integrating Motivational Interviewing (MI), a patient-centered communication technique emphasizing empathy and active listening (Miller and Rollnick, 2013), with Natural Language Processing (NLP). We conducted semi-structured interviews using an MI protocol specifically developed to elicit nuanced, multidimensional patient narratives about pain experience. We developed a novel annotation schema to transform these unstructured narratives into structured representations by combining the International Classification of Functioning, Disability, and Health (ICF) framework (World Health Organization, 2001) with Aspect-Based Sentiment Analysis (ABSA) (Hua et al., 2024). This schema captures emotional and contextual dimensions of patient experiences, providing deeper insight into the multifaceted impacts of chronic pain. To address limitations associated with the small dataset size, we used Large Language Models (LLMs) to generate synthetic interview transcripts, supplementing real-world data for information extraction model development. Finally, we explored the feasibility of using LLMs to automatically extract the annotation schema dimensions. The contributions of this work include: 1) developing an interview protocol to elicit comprehensive patient narratives of lived experiences, 2) creating an annotation schema to systematically characterize these experiences using established frameworks, and 3) evaluating the feasibility of automating this schema using LLMs. The annotation guidelines and code are publicly available to the research community.[1]

2 Related Work

Patient narratives are important to chronic pain assessment and management, as traditional quantitative measures often fail to capture pain complexity (Georgiadis and Johnson, 2023; Robinson-Papp et al., 2015). van Rysewyk et al. (2023) found that patient narratives capture the complex interactions between physical symptoms, psychological impacts, and social consequences of chronic pain, which standardized assessments often overlook. This perspective aligns with the Multimodal Assessment Model of Pain (Wideman et al., 2019), which emphasizes moving beyond traditional measures and advocates for integrating subjective pain experiences into research and clinical practice. Recognizing their value, researchers have examined patient narratives in various clinical settings. For example, Aymerich et al. (2022) showed that narratives in a physiotherapy program informed by Acceptance and Commitment Therapy reveal both physical and psychological recovery dimensions. However, manual analysis of such narratives is time-consuming and subjective, underscoring the need for automated methods to extract meaningful insights at scale.

Early NLP research in chronic pain primarily focused on extracting and classifying symptoms from semi-structured clinical text using rule-based and machine learning methods (Rajwal, 2024). More recently, transformer-based models have advanced symptom extraction from clinical notes (Luo et al., 2022), and sentiment analysis has been used to quantify emotional distress in patient narratives (Vandenbussche et al., 2022; Nunes et al., 2023). For instance, Vandenbussche et al. (2022) systematically analyzed large-scale migraine and cluster headache narratives, identifying diagnostic patterns with unstructured text. However, the limited availability of annotated datasets restricts supervised learning approaches, particularly for analyzing unstructured patient-generated narratives. To address this challenge, recent studies have leveraged LLMs for scalable analysis of pain narratives without task-specific training. LLMs have been used to distinguish chronic pain conditions (Venerito and Iannone, 2024), extract structured insights from patient narratives (Bouzoubaa et al., 2024), and analyze sentiment in large-scale patient-reported data (Alkhnbashi et al., 2024).

This work builds on prior research by utilizing zero-shot prompting with LLMs in conjunction with a structured annotation framework to analyze chronic pain narratives. This approach enables automated pain assessment without relying on extensive labeled datasets. In contrast to previous studies that primarily focus on symptom identification and named-entity recognition, this study introduces a comprehensive annotation schema combining the ICF and ABSA to comprehensively capture biopsychosocial dimensions of pain experiences.

Even in a prompting paradigm where training data is not required, limited real-world data presents challenges in crafting effective prompts that generalize well. To address this, we generated synthetic pain narratives using LLMs to supplement real-world data and refine prompts for improved zero-shot performance. This approach

[1] https://github.com/hadeelelyazori/chronic-pain-narratives

aimed to enhance the model's ability to extract meaningful patterns without relying on extensive manual annotation or large labeled datasets.

3 Methods

3.1 Data

Semi-structured interviews (~30-60 minutes) were conducted with eleven adults reporting chronic pain (mean age: 29.5 ± 11.52 years), generating transcripts with an average of 5,500 words per interview. These interviews explored participants' lived experiences, focusing on the factors shaping pain expression and management. Eligible participants were 18 years or older and currently experiencing chronic pain. Although MI is traditionally used to facilitate behavior change, it was adapted to focus on understanding participants' experiences without influencing their behaviors. To provide some standardization, an MI protocol that emphasized engaging and focusing while excluding evoking and planning was developed. The semi-structured questions were designed to capture a broad range of factors and were informed by the National Institute on Minority Health and Health Disparities (NIMHD) Research Framework (National Institute on Minority Health and Health Disparities, 2017). This framework examines how the physical environment, behavioral patterns, cultural identity, and family and peer networks influence health. The resulting patient narratives provide a detailed, multifaceted view of chronic pain experiences.

Interviews were conducted by a team of six undergraduate researchers, with two present for each session—one led the discussion while the other documented interviewer-interviewee interactions. The interviewers had diverse academic backgrounds, including biology, forensic science, kinesiology, applied statistics, bioengineering, and healthcare research, providing a multidisciplinary perspective on patient-provider interactions. Prior to engaging with participants, researchers were trained in the interview protocol and conducted practice interviews to ensure consistency and quality. Their expertise in clinical research, physical therapy, patient communication, and data-driven healthcare analysis enriched the interview process by ensuring a contextually informed and empathetic approach. Interviews were audio recorded and transcribed using OpenAI's Whisper model (OpenAI, 2022), with speaker roles (researcher vs. participant) identified using Segmentation-3.0 (Bredin et al., 2020). Both models were run locally on a HIPAA-compliant server. The transcripts were automatically de-identified to remove protected health information (PHI) using a rule-based system (Radhakrishnan et al., 2023). A manual review was then conducted to correct transcription errors and remove any remaining PHI. All annotation and LLM experimentation utilized these de-identified records, which were securely stored on restricted servers accessible only to authorized personnel. All study procedures were approved by the Institutional Review Board (IRB).

3.2 Annotation

A comprehensive annotation protocol was developed, drawing on the concept of allostatic load, which accounts for the cumulative physiological and psychological stressors experienced by individuals with chronic pain. Allostatic load helps explain both the immediate effects of chronic pain and its long-term health impacts (Liang and Booker, 2024). This protocol was collaboratively designed by the multidisciplinary research team, whose expertise spans multiple domains. Key contributors brought specialized expertise: KL specializes in NLP annotation protocols for health informatics; SD, JS, and LHG have extensive clinical expertise in pain assessment and patient-centered care; and SS and SA bring experience in biomedical engineering, rehabilitation science, and health informatics. The collective expertise informed the development of a structured framework that integrates the ICF, a biopsychosocial framework from the World Health Organization that categorizes human functioning across body functions, body structures, activities, participation, environmental, and personal factors (World Health Organization, 2001). Since ICF does not define subcategories for personal factors, categories proposed by Geyh et al. (2019) were adopted. By incorporating both pain-related impairments and adaptation strategies, the ICF enables nuanced analysis of chronic pain experiences. To complement the ICF, ABSA was integrated to characterize implicit or explicit patient sentiments towards expressed ICF concepts, labeling them as *positive*, *negative*, or *neutral*. Figure 2 illustrates this dual-layer approach, enabling a holistic analysis of pain narratives and their perceived impact on patient experience.

The ICF includes over 1,400 hierarchically arranged concepts. Table 1 summarizes the ICF concepts used in the annotation schema, with expanded

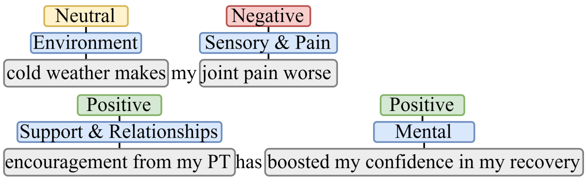

Figure 2: Annotation examples

definitions and examples provided in Appendix A. The annotation guidelines featured synthetic text examples modeled after real-world patient narratives. After training on the guidelines, four annotators—two undergraduate students (RA, a junior Biology major; HA, a senior Forensic Science major) and two graduate students (HE, a PhD student in Information Technology specializing in NLP for healthcare; ZP, a Master's student in Health Informatics)—labeled the transcripts using a local instance of Doccano[2]. Each transcript was independently annotated by two annotators, and disagreements were adjudicated.

Label	Description
Mental Fxn, *b1*	Memory, attention, emotion, ...
Sensory & Pain, *b2*	Sensing and pain experience
NMS & Movement, *b7*	Muscles, joint, ...
Tasks & Demands, *d2*	Manage tasks & routines, ...
Mobility, *d4*	Movement, transportation, ...
Self-Care, *d5*	Personal hygiene, eating, ...
Social Interactions, *d7*	Engage w/ friends, family, ...
Life Areas, *d8*	Education, work, & finances
Products & Tech, *e1*	Assistive tools and systems
Environment, *e2*	Physical environment
Support, *e3*	Physical and emotional support
Services & Policies, *e5*	Systems providing benefits.
Socio-demo, *i1*	Age, gender, education, ...
Positions, *i2*	Roles in social networks
History & Bio, *i3*	Influential life events
Feelings, *i4*	Emotional states
Thoughts & Beliefs, *i5*	Attitudes & perceptions
Motives, *i6*	Goals, needs, or aspirations
Patterns, *i7*	Habits and behaviors

Table 1: Annotation summary. Abbreviations: Functions (Fxn), Socio-demographics (Socio-demo)

3.3 Information Extraction

We used Meta's Llama family of LLMs and OpenAI's GPT-4 in an in-context learning, prompt-based setting for experimentation (AI@Meta, 2024; OpenAI, 2023).

3.3.1 Synthetic Data Generation

To supplement the limited dataset and refine information extraction prompts, we generated 20 synthetic interview transcripts, each consisting of interviewer questions and patient responses. First, *GPT-4-Turbo* was used to create 20 diverse patient profiles by combining personas from a large-scale curated dataset with Big Five personality traits (Ge et al., 2024; McCrae and John, 1992). This approach enhanced variability in emotional expression and coping styles. Using these profiles, *Llama-3.1-405B-Instruct* simulated doctor-patient interviews guided by the MI protocol used in the real interviews, producing narratives of chronic pain experiences. To ensure coherence while maintaining variability, decoding was performed with temperature of 0.6 and top-p of 0.8. These synthetic conversations were designed to mimic the structure and complexity of real-world patient descriptions. Finally, the synthetic transcripts were automatically labeled with the annotation schema using *Llama-3.1-405B-Instruct*, applying a low temperature of 0.1 for deterministic labeling. The annotation prompt included detailed instructions mirroring the annotation guidelines. An example from a synthetic transcript is provided in Appendix C.

3.3.2 LLM-Based Annotation of Transcripts

After refining the prompts, *Llama-3.3-70B-Instruct* was used in a zero-shot setting to generate ICF and sentiment label predictions for the 11 real-world patient transcripts, which comprised the test set. To ensure deterministic and controlled outputs, inference was conducted with a temperature of 0.1, top-p of 0.8, and maximum token limit of 4096. To prevent data leakage and ensure an unbiased evaluation, these *real* transcripts were excluded from the synthetic data used in prompt tuning. The prompt is provided in Appendix B.

3.4 Evaluation

Inter-Annotator Agreement (IAA) was evaluated using Cohen's Kappa to measure inter-annotator reliability and F1-score to enable direct comparison with LLM performance. Information extraction performance was assessed using precision, recall, and F1-score. Rather than evaluating individual text spans, evaluation was conducted at the conversational turn level, treating each turn as a multi-label classification instance. This turn-level evaluation aligns with the conversational nature of patient narratives, reducing sensitivity to minor variations in span selection while ensuring that extracted information retains its intended meaning.

[2]https://github.com/doccano/doccano

4 Results

4.1 Annotation

Cohen's Kappa was computed to evaluate IAA, yielding 0.52 for ICF categories and 0.43 for sentiment. To compare with LLM-extracted labels, the micro-averaged F1-score was also calculated for IAA, resulting in 0.54 for ICF categories and 0.67 for sentiment. The slightly higher F1-score compared to Kappa suggests that while there was some level of agreement on labels, discrepancies were present, particularly in sentiment annotation, where unlabeled instances from different annotators contributed to the lower Kappa. The nuanced and overlapping ICF categories introduced ambiguity, contributing to divergence among annotators. Additionally, the small dataset size limited annotators' ability to establish common patterns, increasing variability. While these IAA scores highlight challenges, they reflect the preliminary exploration of the annotation schema. Planned refinement of the annotation schema and training processes will aim to improve consistency and reliability in future iterations, as described in Section 5.

4.2 Information Extraction

Llama-3.3-70B-Instruct achieved a micro-averaged score of 0.31 F1 for ICF categories and 0.53 F1 for sentiment labels, as summarized in Table 2. While the overall performance indicates substantial room for improvement, the scores align with the observed IAA variability, reflecting the complexity of the task. Despite these challenges, the model successfully extracted some structured elements from the patient narratives, demonstrating potential for automating narrative analysis; however, performance gaps need to be addressed if actionable insights are going to be derived.

Label	**P**	**R**	**F1**	**Sup.**
Mental Fxn	0.42	0.20	0.27	25
Sensory & Pain	0.43	0.44	0.43	112
NMS & Movement	0.45	0.51	0.48	57
Tasks & Demands	0.24	0.36	0.29	33
Mobility	0.39	0.32	0.35	41
Self-Care	0.56	0.22	0.31	46
Social Interactions	0.38	0.10	0.16	51
Life Areas	0.15	0.07	0.09	30
Products & Tech	0.33	0.26	0.30	34
Environment	0.00	0.00	0.00	5
Support	0.55	0.45	0.49	94
Services & Policies	0.58	0.18	0.28	82
Socio-demo	0.00	0.00	0.00	8
Positions	0.00	0.00	0.00	2
History & Bio	0.37	0.28	0.32	46
Feelings	0.20	0.11	0.14	120
Thoughts & Beliefs	0.33	0.23	0.27	92
Motives	1.00	0.25	0.40	4
Patterns	0.00	0.00	0.00	19
Micro Averaged ICF	**0.39**	**0.27**	**0.31**	**901**
Positive	0.62	0.49	0.54	338
Negative	0.77	0.40	0.53	224
Micro Averaged Sentiment	**0.69**	**0.43**	**0.53**	**562**

Table 2: Llama 3.3 performance across ICF categories and sentiment labels

5 Discussion and Conclusions

This work presents a novel annotation schema for capturing chronic pain experiences, integrating the ICF with well-established NLP techniques, like ABSA. By structuring patient narratives within a biopsychosocial framework, this approach extends beyond traditional pain assessment methods.

Preliminary results reveal challenges in annotation consistency and automated extraction, with lower IAA suggesting ambiguities in applying ICF categories. To improve clarity and reproducibility, the schema is being refined to focus on identifying symptoms and the associated interactions. The hypothesis is that symptoms and their contextual interactions can be more reliably annotated, providing a structured basis for integrating ICF concepts at an appropriate level. Future iterations will refine the ICF label set, reassess existing data, and expand data collection to build a more diverse and robust dataset.

Zero-shot extraction experiments showed limited performance due to task complexity and annotation inconsistencies. Refining the schema should improve IAA and extraction performance. Task-specific fine-tuning may be necessary to achieve human-level performance. Incorporating realistic synthetic transcripts into fine-tuning could expand the training set, enhancing model robustness and generalization for information extraction in low-resource settings.

This preliminary study establishes an important foundation for leveraging NLP to support scalable, patient-centered chronic pain assessment. Our approach enables more nuanced and comprehensive representations of patients' lived experiences. Future work will systematically explore the avenues mentioned to improve extraction accuracy, ensuring the clinical relevance and actionable nature of the insights derived. Ultimately, this research aims to bridge qualitative patient narratives and computational methodologies, contributing meaningfully to personalized, data-driven chronic pain management and improved patient outcomes.

6 Limitations

This study has several limitations. The sample size of eleven participants limits the generalizability of findings, and the resulting annotation dataset is small, impacting both IAA and the performance of information extraction models. Additionally, the complexity and subjective nature of patient narratives introduce variability that is difficult to consistently annotate. The current zero-shot LLM-based extraction approach, while demonstrating feasibility, yields performance that may be insufficient for clinical decision-making without further refinement. Future work will involve expanding the dataset, refining annotation guidelines, and exploring fine-tuning of LLMs to improve extraction accuracy and reliability.

Acknowledgments

This research was supported by the Undergraduate Research Scholars Program of the Office of Student Creative Activities and Research at George Mason University.

References

AI@Meta. 2024. Llama 3 model card.

Omer S. Alkhnbashi, Rasheed Mohammad, and Mohammad Hammoudeh. 2024. Aspect-based sentiment analysis of patient feedback using large language models. *Big Data and Cognitive Computing*, 8(12).

Katy Aymerich, Angelika Wilczek, Soravis Ratanachatchuchai, Helen R. Gilpin, Nicolas Spahr, Clair Jacobs, and Whitney Scott. 2022. "living more and struggling less": A qualitative descriptive study of patient experiences of physiotherapy informed by acceptance and commitment therapy within a multidisciplinary pain management programme. *Physiotherapy*, 116:33–41.

Layla Bouzoubaa, Elham Aghakhani, Max Song, Quang Trinh, and Shadi Rezapour. 2024. Decoding the narratives: Analyzing personal drug experiences shared on Reddit. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 6131–6148. Association for Computational Linguistics.

Hervé Bredin, Ruiqing Yin, Juan Manuel Coria, Gregory Gelly, Pavel Korshunov, Marvin Lavechin, Diego Fustes, Hadrien Titeux, Wassim Bouaziz, and Marie-Philippe Gill. 2020. Pyannote.audio: Neural building blocks for speaker diarization. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7124–7128.

María Dueñas, Begoña Ojeda, Alejandro Salazar, Juan A. Mico, and Inmaculada Failde. 2016. A review of chronic pain impact on patients, their social environment and the health care system. *Journal of Pain Research*, 9:457–467.

Roger B. Fillingim. 2017. Individual differences in pain: Understanding the mosaic that makes pain personal. *PAIN*, 158(Supplement):S11–S18.

Perry G. Fine. 2011. Long-term consequences of chronic pain: Mounting evidence for pain as a neurological disease and parallels with other chronic disease states. *Pain Medicine*, 12(7):996–1004.

Tao Ge, Xin Chan, Xiaoyang Wang, Dian Yu, Haitao Mi, and Dong Yu. 2024. Scaling synthetic data creation with 1,000,000,000 personas. *arXiv*, 2406.20094.

Emmanouil Georgiadis and Mark I. Johnson. 2023. Incorporating personal narratives in positive psychology interventions to manage chronic pain. *Frontiers in Pain Research*, 4.

Szilvia Geyh, Urban Schwegler, Claudio Peter, and Rachel Müller. 2019. Representing and organizing information to describe the lived experience of health from a personal factors perspective in the light of the international classification of functioning, disability and health (icf): a discussion paper. *Disability and Rehabilitation*, 41(14):1727–1738.

Daniel S. Goldberg and Summer J. McGee. 2011. Pain as a global public health priority. *BMC Public Health*, 11:770.

Deborah B. Gordon. 2015. Acute pain assessment tools: Let us move beyond simple pain ratings. *Current Opinion in Anaesthesiology*, 28(5):565–569.

Muhammad A. Hadi, Gretl A. McHugh, and S. José Closs. 2019. Impact of chronic pain on patients' quality of life: A comparative mixed-methods study. *Journal of Patient Experience*, 6(2):133–141.

Yan Cathy Hua, Paul Denny, Jörg Wicker, and Katerina Taskova. 2024. A systematic review of aspect-based sentiment analysis: Domains, methods, and trends. *Artificial Intelligence Review*, 57:296.

Institute of Medicine. 2011. *Relieving pain in America: A blueprint for transforming prevention, care, education, and research*. The National Academies Press, Washington, DC.

Yunlong Liang and Cara Booker. 2024. Allostatic load and chronic pain: A prospective finding from the national survey of midlife development in the united states, 2004–2014. *BMC Public Health*, 24:416.

Jacqueline W. Lucas and Inderbir Sohi. 2024. Chronic pain and high-impact chronic pain in u.s. adults, 2023. NCHS Data Brief 518, National Center for Health Statistics, Hyattsville, MD.

Xiao Luo, Priyanka Gandhi, Susan Storey, and Kun Huang. 2022. A deep language model for symptom extraction from clinical text and its application to extract covid-19 symptoms from social media. *IEEE Journal of Biomedical and Health Informatics*, 26(4):1737–1748.

Jeanne M. McCaffery, Anna L. Marsland, Kelley Strohacker, Matthew F. Muldoon, and Stephen B. Manuck. 2012. Factor structure underlying components of allostatic load. *PLoS ONE*, 7(10).

Robert R. McCrae and Oliver P. John. 1992. An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):175–215.

William R. Miller and Stephen Rollnick. 2013. *Motivational interviewing: Helping people change*, 3rd edition. Guilford Press.

National Center for Complementary and Integrative Health. 2024. The promise of applying a whole person framework to pain.

National Institute on Minority Health and Health Disparities. 2017. Nimhd research framework. Retrieved from `https://nimhd.nih.gov/researchFramework`.

Diogo A. P. Nunes, Joana Ferreira-Gomes, Daniela Oliveira, Carlos Vaz, Sofia Pimenta, Fani Neto, and David Martins de Matos. 2023. Chronic pain patient narratives allow for the estimation of current pain intensity. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 716–719.

OpenAI. 2022. Whisper.

OpenAI. 2023. Gpt-4 technical report.

Lakshmi Radhakrishnan, Gundolf Schenk, Kathleen Muenzen, Boris Oskotsky, Habibeh Ashouri Choshali, Thomas Plunkett, Sharat Israni, and Atul J. Butte. 2023. A certified de-identification system for all clinical text documents for information extraction at scale. *JAMIA Open*, 6(3):ooad045.

Richard Radnovich, C. Richard Chapman, Jeffrey A. Gudin, Sanjay J. Panchal, Lynn R. Webster, and Jr Pergolizzi, Joseph V. 2014. Acute pain: Effective management requires comprehensive assessment. *Postgraduate Medicine*, 126(4):59–72.

Swati Rajwal. 2024. Decade of natural nanguage processing in chronic pain: A systematic review. *arXiv*.

Jessica Robinson-Papp, Mary Catherine George, David Dorfman, and David M. Simpson. 2015. Barriers to chronic pain measurement: A qualitative study of patient perspectives. *Pain Medicine*, 16(7):1256–1264.

Simon van Rysewyk, Renée Blomkvist, Antony Chuter, Rhea Crighton, Fiona Hodson, David Roomes, Blair H. Smith, and Francine Toye. 2023. Understanding the lived experience of chronic pain: A systematic review and synthesis of qualitative evidence syntheses. *British Journal of Pain*, 17(6):592–605.

Nicolas Vandenbussche, Cynthia Van Hee, Véronique Hoste, and Koen Paemeleire. 2022. Using natural language processing to automatically classify written self-reported narratives by patients with migraine or cluster headache. *Journal of Headache and Pain*, 23:129.

Vincenzo Venerito and Florenzo Iannone. 2024. Large language model-driven sentiment analysis for facilitating fibromyalgia diagnosis. *RMD Open*, 10(2):e004367.

Timothy H. Wideman, Robert R. Edwards, David M. Walton, Marc O. Martel, Anne Hudon, and David A. Seminowicz. 2019. The multimodal assessment model of pain: A novel framework for further integrating the subjective pain experience within research and practice. *The Clinical Journal of Pain*, 35(3):212–221.

World Health Organization. 2001. International classification of functioning, disability, and health (icf).

A Annotation Guidelines

Label	Description	Examples
Mental Functions, *b1*	Brain functions essential for daily life, including memory, attention, emotion, sleep disturbances, etc.	(P) “I can concentrate better since I started exercising.” (N) “I can’t remember things like I used to.”
Sensory & Pain, *b2*	Sensory abilities and perception of pain.	(P) “My pain has reduced to a manageable level.” (N) “The pain is a constant 8 out of 10.”
Neuromusculoskeletal & Movement, *b7*	Mobility, muscle strength, reflexes, and joint stability.	(P) “After months of physical therapy, my muscle strength has improved.” (N) “The stiffness in my knees has gotten worse.”
Tasks & Demands, *d2*	Managing tasks, routines, and psychological stress.	(P) “Deep breathing exercises help me stay calm.” (N) “I often skip my physical therapy homework.”
Mobility, *d4*	Movement-related activities such as walking and climbing stairs.	(P) “I’ve started taking short walks daily.” (N) “I can’t climb stairs without intense pain.”
Self-Care, *d5*	Personal hygiene, grooming, and maintaining health.	(P) “I maintain my hygiene routine despite the pain.” (N) “I often skip meals due to the pain.”
Social Interactions, *d7*	Engaging socially in appropriate ways.	(P) “Joining a support group gave me practical advice.” (N) “I don’t go out anymore because of the pain.”
Life Areas, *d8*	Tasks related to education, work, and economic activities.	(P) “I’m able to afford the best treatments.” (N) “I worry about losing my job due to pain.”
Products & Tech, *e1*	Tools designed to improve functioning.	(P) “My wheelchair allows me independence.” (N) “The outdated software at work hinders my tasks.”
Environment, *e2*	Physical environment impacting functioning.	(P) “Sunny weather helps reduce my pain.” (N) “Cold weather makes my pain worse.”
Support & Relationships, *e3*	Support from people or animals.	(P) “My family supports me a lot.” (N) “I feel isolated because my friends don’t understand.”
Services & Policies, *e5*	Governance and service systems.	(P) “The nearby clinic makes care easier.” (N) “Long wait times disrupt my therapy schedule.”
Socio-demographics, *i1*	Observable characteristics like age, education, etc.	(P) “Being financially secure helps me access healthcare.” (N) “I can’t afford transportation to appointments.”
Positions, *i2*	Roles in social and living environments.	(P) “As the youngest in my family, they all encourage me to keep up with therapy.” (N) “Because of all the responsibilities I have as an chairperson, it all affects my recovery.”
History & Bio, *i3*	Life events shaping current functioning.	(P) “Overcoming past challenges makes me resilient.” (N) “Childhood trauma makes trusting providers hard.”
Feelings, *i4*	Emotional states influencing responses.	(P) “I feel optimistic about managing my pain.” (N) “I feel anxious about my condition.”
Thoughts & Beliefs, *i5*	Attitudes about self and environment.	(P) “I believe therapy is helping me recover.” (N) “I doubt the effectiveness of my treatment.”
Motives, *i6*	Goals and aspirations driving behavior.	(P) “My goal to play with my kids motivates me.” (N) “Progress feels slow, so I’m not motivated to continue.”
Patterns, *i7*	Behavioral and cognitive tendencies.	(P) “I follow a structured medication routine.” (N) “I procrastinate on health goals.”

Table 3: Expanded annotation guidelines with examples. Parentheses indicate sentiment labels, where (P) denotes a positive sentiment and (N) denotes a negative sentiment

B Zero-shot Experimentation Prompt

To facilitate structured extraction of patient experiences, we designed a standardized annotation prompt that guides the LLM through our annotation schema. The prompt ensures consistency in identifying relevant text spans, assign ICF labels from a predefined set, and determine sentiment polarity.

It provides strict formatting guidelines, enforcing JSON output to support automation with LLMs. This structured approach enhances reproducibility and enables scalable NLP-based analysis of chronic pain narratives. The annotation prompt used in our study is presented below.

```
You are a highly skilled annotator specializing in chronic pain patient
responses, using the **ICF classification system** and **Aspect-Based
Sentiment Analysis**
### Task Overview:
Your goal is to:
1. **Identify** relevant text spans aligning with the provided ICF labels.
2. **Assign** the correct ICF label (**ONLY** from the provided list).
3. **Determine** sentiment:
    - Positive
    - Negative
    - Neutral
---
### **Labeling Rules**
- Use **ONLY** the provided ICF labels (no modifications or new labels).
- Each ICF-labeled span **must** also have a sentiment label.
- A span can be labeled with an ICF label with no sentiment label.
- A span can NOT be labeled with a sentiment label with no ICF label.
- If a span **does not match an ICF label**, exclude it.
---
### **ICF Labels (Use only these - No external labels)**
{json.dumps(icf_labels, indent=2)}
---
### **Output Format**
Return a **valid JSON object**:
{
"id": <text_id>,
"label": [
    ["<TEXT_SPAN_1>", "ICF_LABEL"],
    ["<TEXT_SPAN_1>", "SENTIMENT_LABEL"],
    ["<TEXT_SPAN_2>", "ICF_LABEL"],
    ["<TEXT_SPAN_2>", "SENTIMENT_LABEL"]
]
}
- **No explanations, no missing labels.**
- **If a span is unlabeled, exclude it.**
---
### **Text to Annotate:**
{transcript}
```

C Synthetic Data

To supplement the limited dataset, we generated 20 synthetic patient narratives using a structured pipeline. The goal was to simulate realistic patient transcripts, automatically annotate them using an LLM, and utilize them for prompt tuning. The pipeline was designed to closely mirror real-world chronic pain experiences while ensuring diversity in patient characteristics. The process consisted of three main steps:: 1) profile generation, 2) conversation simulation, and 3) automatic annotation.

C.1 Profile Generation

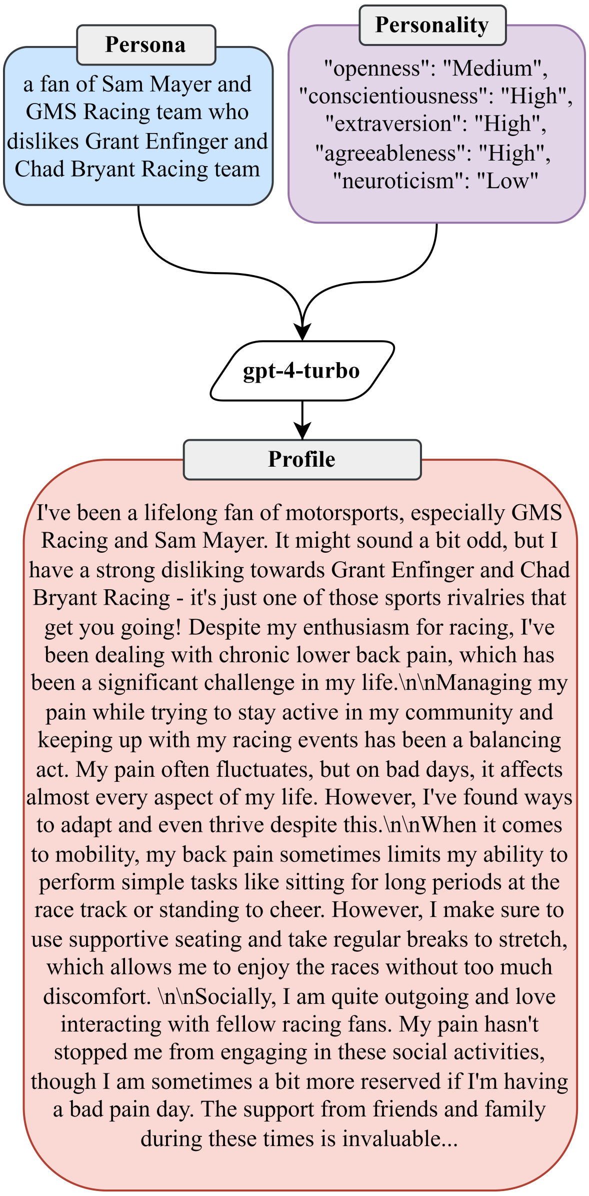

Figure 3: Profile synthetic generation example using Personas (Ge et al., 2024), the Big Five model (McCrae and John, 1992) and GPT

Figure 4 illustrates the synthetic profile generation process, which integrates personas, Big Five personality traits, and text generation. A persona is used for the demographic attributes it has (e.g., occupation, interests) and the Big Five model is used for its psychological traits (e.g., openness, neuroticism). These details are then passed through the GPT4 model, which generates a first-person narrative. The resulting profile provides a patient background, ensuring diverse and realistic chronic pain experiences for the conversation generation.

C.2 Coversation Simulation

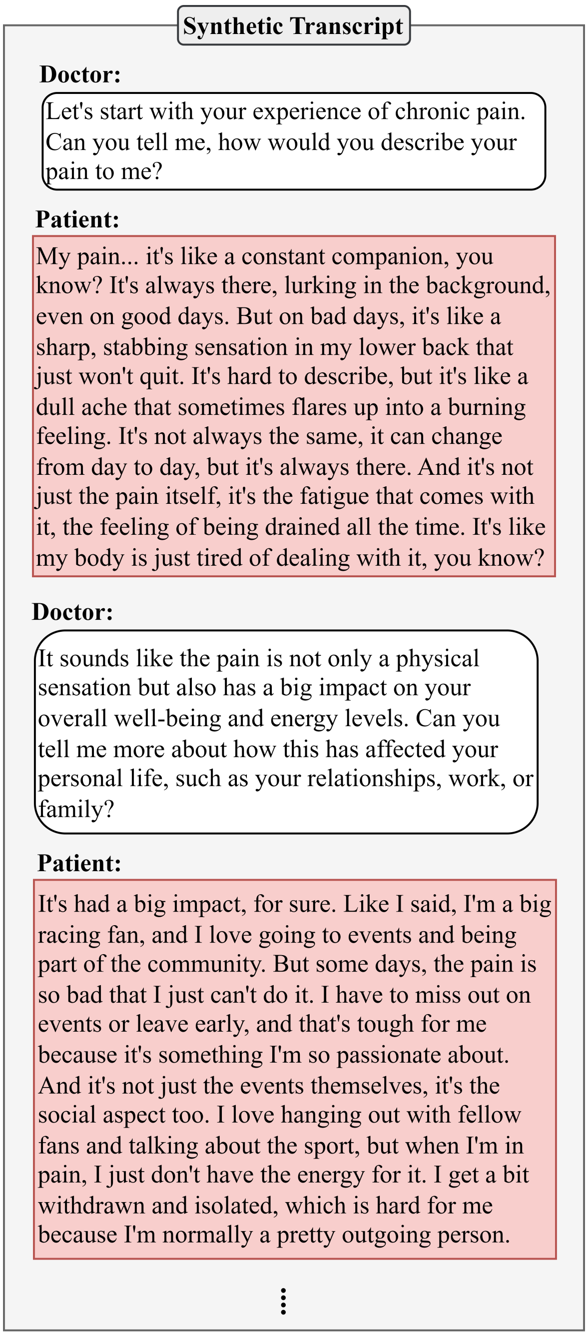

Figure 4: Synthetic transcript example