



Written narratives to understand the experience of individuals with pain: Could Artificial Intelligence (AI) help in integrating them in clinical practice?

Rubén Nieto¹, Andreas Kaltenbrunner², Jacopo Amidei², José Gregorio Ferreira De Sá², Klara Albajes³, Mayte Serrat⁴, Albert Feliu-Soler^{5,6}, Juan Vicente Luciano^{5,6,7}

¹eHealth Lab Research Group, Faculty of Psychology and Educational Sciences, Universitat Oberta de Catalunya, Barcelona, Spain
²AI and Data for Society, Internet Interdisciplinary Institute, Universitat Oberta de Catalunya, Barcelona, Spain
³Psyclinic, Barcelona, Spain
⁴Unitat d'Expertesa en Síndromes de Sensibilització Central, Servei de Reumatologia, Vall d'Hebron Hospital, Barcelona, Spain
⁵Department of Clinical and Health Psychology, Faculty of Psychology, Autonomous University of Barcelona, Bellaterra, Spain
⁶CIBER of Epidemiology and Public Health (CIBERESP), Madrid, Spain
⁷Teaching, Research & Innovation Unit, Parc Sanitari Sant Joan De Déu, Sant Boi De Llobregat, Spain

AIMS:

Written narratives (WN) can uncover the personal, subjective, complex, and multidimensional nature of pain, offering insights from the perspectives of those experiencing it (Nieto et al., 2020). This study explores the potential of Artificial Intelligence, specifically Large Language Models (LLMs), to replicate the results of human qualitative analyses of these narratives, focusing on content accuracy and assessments of disability and pain severity. Specifically, we aim to:

1. Evaluate the agreement between human assessments of pain narratives and those generated by GPT models.
2. Analyze the correlation between assessments by GPT models and humans in relation to various outcomes.

METHODS:

STEP 1

43 individuals with fibromyalgia were asked to freely write about their pain experiences. Additionally, standardized self-reported questionnaires were administered to assess fear-avoidance beliefs, the impact of fibromyalgia, and levels of depression and anxiety.

STEP 2

Independent experts in pain management analyzed the writings to rate on a scale from 0 to 10:

- **Pain Severity:** The perceived magnitude of fibromyalgia in terms of pain and overall suffering as conveyed in each participant's text.
- **Disability:** Defined as the perceived extent to which fibromyalgia disrupts the usual activities and life of the participants.

GPT-4 was asked to do the same task as the experts:

STEP 3

"As an expert psychologist specializing in evaluating pain in patients diagnosed with fibromyalgia, you are tasked with analyzing patient narratives about their pain and then scoring them on a scale from 0 (indicating no severity or disability) to 10 (indicating maximum severity and disability). These patients' explanations about their pain and how they feel it are all written in Spanish. The level of severity, defined as the perceived intensity of pain and overall suffering, and disability, defined as the extent to which fibromyalgia hinders patients' usual activities and quality of life, are to be rated based on your interpretations of the patients' texts. Scores should accurately reflect the severity and disability levels described in patient narratives without inflation. A holistic evaluation capturing the complexity of experiences is crucial. Pay attention to phrases indicating coping mechanisms, resilience, or mitigating factors that may reduce perceived severity or disability. Consider contextual understanding, including coping strategies, support systems, and adaptive behaviors, which may mitigate perceived severity and disability. Your role involves receiving a patient's narrative, enclosed within triple slashes, and analyzing it. You are expected to return your analysis in JSON format, with the following keys: "severity_score" providing the scores for severity ranging from 0 to 10, "disability_score" providing the scores for disability ranging from 0 to 10, "severity_explanation" providing an in English explanation for the severity score and "disability_explanation" providing an in English explanation for the disability score"

STEP 4

Two experts evaluated the explanations provided by GPT based on a seven-point scale (from strongly disagree to strongly agree):

- **Question 1:** Could the explanation have been written by a psychologist expert in fibromyalgia?
- **Question 2:** Does the explanation accurately represent the pain severity expressed in the narrative?
- **Question 3:** Would I use the pain severity score and explanation above to assist in assessing the patient's pain?

See some examples of ratings and corresponding explanations by GPT-4 here:

RESULTS:

Agreement between experts, and between experts and GPT:

	Expert 1 vs Expert 2		Expert 1 vs GPT		Expert 2 vs GPT	
	Pain severity	Disability	Pain severity	Disability	Pain severity	Disability
Percent agreement	0.29	0.31	0.36	0.23	0.32	0.35
Weighted percent agreement	0.96	0.95	0.94	0.94	0.95	0.94
Gwet's AC2	0.87	0.83	0.84	0.79	0.84	0.80
Krippendorff's alpha	0.66	0.69	0.45	0.49	0.49	0.57

(Note: in red low indexes, yellow near to adequate and in white adequate ones)

Correlations with standardized outcomes:

		FIQR				HADs		
		Total	Function	Impact	Symptoms	TSK	Anxiety	Depression
Experts	Severity	0.41*	0.18	0.45*	0.46*	0.18	0.34	0.44*
	Disability	0.44*	0.24	0.48*	0.43*	0.19	0.38**	0.46*
Mean 10 Trials Temp 1	Severity	0.43*	0.23	0.43*	0.47*	0.18	0.32**	0.41*
	Disability	0.49*	0.27	0.49*	0.51*	0.16	0.34**	0.45*

*p<0.01; **p<0.05

Experts assessments' (mean (SD)):

	Pain severity			Disability		
	Question 1	Question 2	Question 3	Question 1	Question 2	Question 3
Expert 1	5.93 (0.26)	6 (0)	5.77 (0.43)	5.93 (0.26)	6 (0)	5.77 (0.43)
Expert 2	6.84 (0.37)	6.79 (0.41)	6.65 (0.48)	6.84 (0.37)	6.79 (0.41)	6.65 (0.48)

CONCLUSIONS:

Written narratives appear to be valuable to approach the experiences of individuals with pain from their perspective, providing insights that standardized questionnaires may not always capture. LLMs could be promising in facilitating pain narratives and analyzing them automatically or semi-automatically, facilitating the tasks from professionals. In fact, our results support agreement between GPT-4 evaluations, and evaluations from experts. Also, the evaluations from GPT-4 correlated with standardized outcomes as evaluations from experts.