UNIVERSITA' DEGLI STUDI DI PADOVA

Dipartimento di Medicina Molecolare

Corso di Laurea Magistrale in Scienze Infermieristiche e Ostetriche

TESI DI LAUREA

Extraction and Integration of Information between Structured Data and Free Text for the Characterization of Accidental Falls: A GPT-Based Approach

Relatore: Prof. Dario Gregori

Correlatore: Dott.ssa Giulia Lorenzoni

Laureando: Davide Rango

Matricola: 2060413

2022/2023

Index

Abstract

Background: Accidental falls in healthcare settings are a relevant problem, causing injuries, extended hospital stays, and significant societal burdens. The ongoing investigation into the causes of accidental falls, despite the preventive efforts, highlights the potential of artificial intelligence (AI) in interpreting fall patterns, identifying risk factors, and subsequently mitigating future incidents. GPT-based models show a transformative potential in AI, surpassing conventional methods in text classification, suggesting a paradigm shift in the field.

Aim: To develop and implement an automatic coding system utilizing GPT-based models to extract and categorize accidental falls data and assess the feasibility of such models in clinical risk management. Furthermore, a comparison between GPT-3.5 and GPT-4.0 models was carried out.

Methods: Our study analysed accidental falls' incident reporting records from an Italian Local Health Authority. Expert clinicians manually categorized the incident descriptions concerning fall location and injury presence, serving as the gold standard. The GPT 3.5 and 4 models were employed for free text classification and their performance was evaluated according to the gold standard.

Results: The study included 254 records with a free-text description of the fall, 187 reported information on injury location, while 93 records reported information of injury occurrence as a result of the accidental fall. GPT-4 exhibited better performance compared to GPT 3.5, accurately classifying fall locations in hospital rooms (accuracy: 0.904), bathrooms (specificity: 0.894), and hallways (accuracy: 1). For what concerns injury detection, GPT-4 identified 30 out of 34 cases with an accuracy of 0.881 (95% CI 0.817; 0.946).

Conclusion: This study has demonstrated the potential of GPT-based models in extracting and analyzing data concerning accidental falls within hospital settings. Their effectiveness in identifying patterns and circumstances surrounding accidental falls offers valuable insights for clinical risk management and fall prevention.

Introduction

Accidental falls in healthcare facilities are common events that occur at all ages and may cause loss of independence, injuries and increased length of hospitalization (1). Fall-related adverse events impose a significant social and economic burden for individuals, their families, and the healthcare system (2,3). About 30-50% of falls result in some physical injury and fractures occur in 1–3% (4); each year, about 700000 falls worldwide result in death (5).

In Europe, the incidence of this phenomenon has been nearly 2000 cases per 100000 in 2017, the prevalence has been of about 5000 cases per 100000 in 2017, and the YLL (year of life lost) has been nearly 17 million in 2017 (6).

Falls are particularly prevalent in nursing homes and hospitals, where patients, often with acute or chronic illnesses that impair judgment and mobility, are required to stay in environments that are new and unfamiliar to them. (7). In those settings, the history of falls, use of walking aids, and disability are strong predictors of future falls (8).

An essential resource to reduce the fall burden is to evaluate each patient's risk for accidental falls at admission. The patient must undergo a comprehensive evaluation, considering their previous history of falling, their frequency, characteristics and context, fall risk factors, and their physical, cognitive, psychological, and social resources (9). Stratifying the fall risk may allow for the adoption of personalized primary prevention strategies, preventing accidental falls.

Falls are a nursing-sensitive indicator (10) and a reduction of accidental falls in the hospital environment is an indicator of a strong safety culture and a good quality of care provided (11).

With this aim, in Italy, the Ministry of Health has approved Recommendation No. 13, "Prevention and management of patient falls in healthcare facilities", addressed to all healthcare settings and to all healthcare operators, in order to enhance patient safety and to establish a data collection system of in-hospital sentinel events, including fatal falls and falls resulting in severe injury.

Despite the interventions to prevent the phenomenon, the problem of falls in healthcare is still relevant. One possible reason could be that risk management interventions are not adequately implemented in the in-hospital setting. One might also consider that a gap in understanding the patient safety culture among nurses and inadequate adherence to fall risk assessment protocols could contribute to this issue (12). Furthermore, the fact that the burden of falls is still relevant could be related to the inefficacy of available interventions due to an incomplete understanding of the phenomenon of falls.

The use of artificial intelligence to evaluate accidental fall data might help to clarify the falls patterns (13), identify relevant risk factors, and, as a result, prevent and manage future falls (14).

The advent of Generative Pre-trained Transformers (GPT)-based models has the potential to instigate a paradigm shift in the field of Artificial Intelligence, outperforming previous state-of-the-art models in numerous tasks, including text classification, where they surpass traditional machine learning techniques (15).
Several papers have highlighted the promising potential of utilizing GPT-based models in healthcare data, particularly in electronic health records characterized by unstructured free-text information. These models can potentially address a wide range of tasks, including clinical and research-related activities. However, despite several publications claiming the potential and theoretically revolutionary role of such tools in various medical fields (16-18); there are still only a few preliminary healthcare applications of GPT-based models in the literature (19,20).
This study aims to develop and implement an automatic coding system utilizing GPT-based models to extract and categorize accidental falls data and assess the feasibility of such models in clinical risk management. The final purpose is to understand better patients' accidental fall patterns to enhance patient safety and prevent future falls.

Materials and Methods

Our analysis considered falls notified through the Incident Reporting system to the Risk Management Service of a Local Health Authority in Italy (name not revealed for a Non-Disclosure Agreement). This Local Health Authority includes two hospitals and three health districts.

The Incident Reporting form is a structured data collection form based on two distinct sections. The first part includes information regarding the dynamics of the fall (such as date, hour of the day, location, department involved, potential causes, etc.), including a brief description of the incident, socio-demographic information about the patient, factors contributing to the fall (intrinsic and extrinsic) and strategies employed to prevent it.

The second part concerns the outcomes of the fall. This section includes details about possible injuries resulting from the fall, investigations performed to diagnose the potential damage, therapeutic interventions, and the patient's pharmacological therapy at the time of the fall. All information is pre-coded except for the narrative description of the event.

Reference method: manual classification

Details about the location of each accidental fall and any resulting injuries were extracted from the available free-text descriptions. Two experienced clinicians manually categorized these descriptions into specific groups:

1) Location of the accidental fall: hospital bathroom, hospital room, hallway;
2) Fall-related injury (defined as a physical damage resulting from the fall): yes vs. no.

All records were anonymous and thoroughly examined to prevent any potential data disclosure. The records were in the Italian language.

GPT-based classification

OpenAI's Application Programming Interface (API) endpoints were used as the basis for the classification task. The API provides developers with access to advanced artificial intelligence models, such as GPT (Generative Pre-trained Transformer), which can be used for a wide range of applications, including Natural Language Processing (NLP), text generation, translation, summarization, and more. One of the most significant advantages of using the OpenAI API is the ability to perform few-shot learning, where the artificial intelligence model can quickly adapt to new tasks with minimal training.

GPT models were accessed through a public API in R using the OpenAI package (22). This package enables users to submit prompts to the trained model hosted in the cloud and get the model's responses. Beyond safeguarding OpenAI's intellectual property, this approach offers the notable advantage of relieving users of the huge computational costs of running such a large deep neural network (DNN) on a local machine. The cost of each API request is contingent on the number of tokens in both the prompt and response. A token represents the fundamental input processed by the model and generally corresponds to a portion of a word, approximately 3/4 characters on average (21). We utilized the GPT-3.5 variant known as "gpt-3.5-turbo-0301" and GPT-4 for our purposes.

When making requests to the GPT model, users have the option to include additional parameters that can shape the nature of the generated response. Among these parameters, “temperature” and “presence-penalty” are the most important, playing a key role in settling the level of determinism in the model's responses. In our tests, we specifically configured the “temperature” parameter to a value of 0.2 and “presence-penalty” parameter to a value of 0.8. “Temperature 0.2” represents the exploration level or randomness in the algorithm, where lower values often imply a more deterministic search. Meanwhile, “presence-penalty 0.8” indicates the significance or weight assigned to penalizing certain solutions or actions, promoting optimization towards desired outcomes. These parameter settings influence the algorithm's behavior in seeking optimal solutions within a given problem.

The choice for the parameters set was motivated by the fact that we extracted specific information for which there is an univocal answer (high determinism). At the same time, we want to obtain results in a machine-readable format suitable for direct use in statistical analysis. Creating a well-designed English prompt allows us to achieve this goal with a single query to the model for most documents. However, since the model has been trained in natural language, it is possible that responses in unexpected formats may occasionally be provided. To mitigate this problem, we opted for a lower "temperature" setting, understanding that this increases the likelihood of encountering incorrect formats in multiple attempts when repeating the request.

We asked GPT to extract information about accidental falls in healthcare facilities texts. We follow OpenAI's guidelines (21) to generate prompt messages for each of these extractions. The specific prompts provided to GPT in the completion task are reported in Table 1.

Statistical Analysis

Descriptive statistics were presented as percentages (absolute numbers) for categorical variables. The assessment of GPT's performance in the classification task involved computing accuracy, sensitivity, and specificity, along with reporting bootstrap 95% Confidence Intervals (CI). All analyses were carried out using R software version 4.3.2 and the OpenAI libraries (22).

Results

The database included 263 accidental fall records containing free-text event descriptions from 2016 to 2017. Nine records were excluded from the analysis due to duplication. The final dataset included 254 records. The patient's median age was 78 years.

Most of the accidental falls occurred when the patients were hospitalized (91.6%) while they were not alone (53.2%), either in their hospital bedroom (71.5%) or in the toilet (19.4%), especially during the night (34.8%). Most of the patients were at risk of fall (65.8%). However, only a relatively small proportion of patients (27.4%) had a personalized fall prevention plan in place.

Patients primarily fell from a standing position (45.7%) due to slipping (45.6%) often while engaged in activities such as walking (21.2%), getting out of bed (26%), or using the toilette (17.3%). The causes of accidental falls were frequently attributed to walking barefoot (43.8%) or wearing inadequate footwear (37%), and bed-related factors such as raised (33.6%) or lowered bed rails (23.4%). Half of the accidental falls did not result in any harm (50.2%), and when an injury occurred, the consequences were generally mild, such as contusions (39.7%) and excoriations (31.4%). The injuries have rarely required treatment (82%). Most patients were on more than four concomitant medications (60.6%) at the time of accidental fall (Table 2).

Classification performance

A total of one hundred eighty-seven records reported the information about the injury location in the fall description text, while 93 records reported the information about injury occurrence (Table 2).

GPT-3.5 exhibited good accuracy in identifying accidental fall locations, correctly classifying 102 out of 126 falls that occurred in hospital rooms (Table 3), resulting in an accuracy of 0.868 (95% CI 0.818; 0.92) (Table 4). It also demonstrated strong performance in detecting accidental falls in hospital bathrooms, achieving an accuracy of 0.915 (95% CI 0.872; 0.957) and a sensitivity of 0.965 (95% CI 0.965; 1). However, its sensitivity in detecting accidental falls in hallways was lower (0.586 95% CI 0.059; 1). Regarding the injuries, it identified 30 out of 34 cases of fall-related injuries (Table 3) with an accuracy of 0.752 (95% CI 0.667; 0.839).

GPT-4 generally outperformed GPT-3.5. First, for what concerns the fall location, it correctly identified all the accidental falls occurring in the hallways (Table 4). Furthermore, it correctly classified 109 out of 126 accidental falls in hospital rooms (Table 3), resulting in an accuracy of 0.904 (95% CI 0.861; 0.941) and a sensitivity of 0.865 (95% CI 0.809; 0.925). It also showed a high sensitivity (0.982 95% CI 0.94; 1) and specificity (0.894 95% CI 0.838; 0.946) in identifying accidental falls in hospital bathrooms (Table 4). In the identification of fall-related injuries, GPT-4 displayed good performance metrics (Table 4), showing an accuracy of 0.881 (95% CI 0.817; 0.946), sensitivity of 0.88 (95% CI 0.759; 0.974), and specificity of 0.881 (95% CI 0.79; 0.963).

Tables S1 and S2 present some examples of misclassified records, including the GPT output.

Discussion

This paper aimed to illustrate the feasibility and the value added by GPT-based models into the realm of clinical risk management, specifically in the surveillance of accidental falls, with the overarching goal of enhancing patient safety and preventing future falls. Furthermore, a comparison between GPT-3.5 and GPT-4.0 models was carried out.

The integration of AI models, particularly GPT 3.5 and GPT 4, showed varying accuracy, sensitivity, and specificity levels in identifying fall's location and assessing the presence or absence of fall-related injuries. GPT-3.5 exhibits strong performance in identifying all instances of injuries (sensitivity: 100%); however, it encounters substantial challenges in accurately classifying non-injuries, misclassifying over one-third of these cases as injuries (specificity: 0.612 95% CI 0.491; 0.737).

GPT-4 displayed better results in the classification task, demonstrating an accuracy rate of over 90% for determining the location of accidental falls and about 88% accuracy in detecting the presence of fall related injuries.

The model exhibited a high capability in interpreting partial information and successfully classified most of the description records accurately.

The results also indicate the potential for these artificial intelligence models to be effective in non-English-speaking contexts, underscoring their versatility in processing multilingual healthcare data.

Despite some occasional misclassifications, the study underscores the value of GPT models in augmenting clinical risk management, enhancing fall prevention strategies, and improving patient safety protocols in healthcare environments.

Present results are promising. The implementation of such automated systems, capable of analyzing and categorizing complex data from healthcare incident records, create opportunities for a new paradigm in clinical research and risk management.

A detailed analysis of the circumstances, causes, and consequences of accidental falls can provide an important knowledge base for future research in patient safety and fall prevention.

The use of GPT models can support research by providing quick and accurate access to clinical data, enabling researchers to identify patterns and correlations that might be challenging to identify manually. Effectively extracting information from large volumes of unstructured data could open new paths for early risk identification, personalized intervention protocol development, and predictive analysis in the realm of patient safety.

Ultimately, the integration of these GPT models in medical research offers an innovative and promising outlook to explore new directions in fall prevention, optimizing patient safety, and advancing public health.

Limitations

A limitation of this work is the small sample size. This paper relies on a small number Incident Reporting records from a single Risk Management Service within a Local Health Authority. Greater diversity and a larger number of data could improve the generalizability of the findings.

One limitation encountered was the occasionally inadequate information in some accidental fall description records, which could have affected the AI models' data extraction efficiency. However, it's crucial to recognize that these textual records were not specifically prepared for textual analysis. They were, instead, compiled from routine clinical practice activities. Notwithstanding this, the GPT models demonstrated impressive performance, effectively analyzing and classifying the data even though it was not optimized for GPT-based analysis (23).

Conclusion

This study has demonstrated the potential of GPT-based models in extracting and analyzing data concerning accidental falls within hospital settings. Their effectiveness in identifying patterns and circumstances surrounding accidental falls offers valuable insights for clinical risk management and fall prevention.

While the promise of Artificial Intelligence is evident, constant improvements are required, considering both technical and ethical challenges. Beyond these considerations, integrating GPT-based models in clinical research represents an innovative opportunity to advance clinical data analysis and improve patient safety and fall prevention plans.

References

1. Cameron ID, Dyer SM, Panagoda CE, Murray GR, Hill KD, Cumming RG, et al. Interventions for preventing falls in older people in care facilities and hospitals. Cochrane Bone, Joint and Muscle Trauma Group, editor. Cochrane Database of Systematic Reviews [Internet]. 2018 Sep 7 [cited 2023 Aug 24]; 2020(1). Available from: http://doi.wiley.com/10.1002/14651858.CD005465.pub4
2. Dykes PC, Curtin-Bowen M, Lipsitz S, Franz C, Adelman J, Adkison L, et al. Cost of Inpatient Falls and Cost-Benefit Analysis of Implementation of an Evidence-Based Fall Prevention Program. JAMA Health Forum. 2023 Jan 20;4(1):e225125.
3. Sherrington C, Michaleff ZA, Fairhall N, Paul SS, Tiedemann A, Whitney J, et al. Exercise to prevent falls in older adults: an updated systematic review and meta-analysis. Br J Sports Med. 2017 Dec;51(24):1750–8.
4. Oliver D, Healey F, Haines TP. Preventing Falls and Fall-Related Injuries in Hospitals. Clinics in Geriatric Medicine. 2010 Nov;26(4):645–92.
5. World Health Organization. Fact Sheet. Falls. [Internet]. 2018. Available from: https://www.who.int/news-room/fact-sheets/detail/falls
6. James SL, Lucchesi LR, Bisignano C, Castle CD, Dingels ZV, Fox JT, et al. The global burden of falls: global, regional and national estimates of morbidity and mortality from the Global Burden of Disease Study 2017. Inj Prev. 2020 Oct;26(Suppl 2):i3–11.
7. LeLaurin JH, Shorr RI. Preventing Falls in Hospitalized Patients. Clinics in Geriatric Medicine. 2019 May;35(2):273–83.
8. Deandrea S, Bravi F, Turati F, Lucenteforte E, La Vecchia C, Negri E. Risk factors for falls in older people in nursing homes and hospitals. A systematic review and meta-analysis. Archives of Gerontology and Geriatrics. 2013 May;56(3):407–15.
9. Montero-Odasso M, Van Der Velde N, Martin FC, Petrovic M, Tan MP, Ryg J, et al. World guidelines for falls prevention and management for older adults: a global initiative. Age and Ageing. 2022 Sep 2;51(9):afac205.
10. Oner B, Zengul FD, Oner N, Ivankova NV, Karadag A, Patrician PA. Nursing-sensitive indicators for nursing care: A systematic review (1997–2017). Nursing Open. 2021 May;8(3):1005–22.
11. Alanazi FK, Lapkin S, Molloy L, Sim J. The impact of safety culture, quality of care, missed care and nurse staffing on patient falls: A multisource association study. Journal of Clinical Nursing. 2023 Jun 12;jocn.16792.

12. Nadia P, Yulianty Permanasari V. Compliance of the Nurse for Fall Risk Assessment as a Procedure of Patient Safety: A Systematic Review. KLS. 2018 Dec 5;4(9):207.
13. Lindberg DS, Prosperi M, Bjarnadottir RI, Thomas J, Crane M, Chen Z, et al. Identification of important factors in an inpatient fall risk prediction model to improve the quality of care using EHR and electronic administrative data: A machine-learning approach. International Journal of Medical Informatics. 2020 Nov; 143:104272.
14. O'Connor S, Gasteiger N, Stanmore E, Wong DC, Lee JJ. Artificial intelligence for falls management in older adult care: A scoping review of nurses' role. J Nursing Management. 2022 Nov;30(8):3787–801.
15. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language Models are Few-Shot Learners. 2020 [cited 2023 Nov 7]; Available from: https://arxiv.org/abs/2005.14165
16. Arslan S. Exploring the Potential of Chat GPT in Personalized Obesity Treatment. Ann Biomed Eng. 2023 Sep;51(9):1887–8.
17. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell. 2023 May 4;6:1169595.
18. Grünebaum A, Chervenak J, Pollet SL, Katz A, Chervenak FA. The exciting potential for ChatGPT in obstetrics and gynecology. American Journal of Obstetrics and Gynecology. 2023 Jun;228(6):696–705.
19. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT. Diagnostic and Interventional Imaging. 2023 Jun;104(6):269–74.
20. Sezgin E, Sirrianni J, Linwood SL. Operationalizing and Implementing Pretrained, Large Artificial Intelligence Linguistic Models in the US Health Care System: Outlook of Generative Pretrained Transformer 3 (GPT-3) as a Service Model. JMIR Med Inform. 2022 Feb 10;10(2):e32875.
21. OpenAI [Internet]. 2023. Available from: https://beta.openai.com/
22. Rudnytskyi I. OpenAI: R Wrapper for OpenAI API [Internet]. 2023. Available from: https://CRAN.R-project.org/package=openai
23. Lorenzoni G, Rampazzo R, Buratin A, Berchialla P, Gregori D. Does the Integration of Pre-Coded Information with Narratives Improve in-Hospital Falls' Surveillance? Applied Sciences. 2021 May 13;11(10):4406.

Tables

Table 1. GPT prompts for extraction of clinical information about location of the accidental falls and falls related injuries.

Completion sequence	GPT Prompt Text
#1	"This is a data frame containing records of accidental falls in healthcare facilities. Each record corresponds to a fall that may have occurred in a specific location within the hospital."
#2	"In the first column is reported the id of the accidental fall. In the second column is reported the description of the fall."
#3	"The text is in Italian"
#4a	"Where did the fall occur? Choose from the following options: 'hospital bathroom', 'hospital room', 'hallway'. Use only the words corresponding to these categories and use only lowercase. Do not write any other output or comments."
#4b	"Did the patient sustain any injuries, either immediately or as a result of the fall? Answer with 'yes' or 'no'"

Table 2. Falls characteristics (2016-2017). Data are absolute numbers (percentage).

	N = 254
Setting: *inpatient* (%)	229 (91.6)
People witnessing the fall: *Someone* (%)	134 (53.2)
The patient was at fall risk: *Yes* (%)	160 (65.8)
Fall prevention plan (%)	
None	57 (24.8)
Standard fall prevention plan	110 (47.8)
Personalized fall prevention plan	63 (27.4)
Time of day when the fall occurred (%)	
Morning	77 (31.6)
Afternoon	43 (17.6)
Evening	39 (16.0)
Night	85 (34.8)
Location where the fall occurred (%)	
Hospital bedroom	173 (71.5)
Hospital bathroom	47 (19.4)
Hallway	8 (3.3)
Outpatient clinic	11 (4.5)
Outside	3 (1.2)
Location from which the patient fell (%)	
Standing position	111 (45.7)
Lying position	80 (32.9)
Seated position	52 (21.4)
Mechanism by which the fall occurred (%)	
Stumbling	17 (7.8)
Slipping	99 (45.6)
Losing consciousness	20 (9.2)
Off balance	28 (12.9)
Loss of strength	16 (7.4)
Not available	37 (17.1)
Action at the time of fall (%)	
Walking	49 (21.2)
Getting out of bed	60 (26.0)
Getting on bed	2 (0.9)
Using toilette	40 (17.3)
Reaching for something	18 (7.8)
Sleeping	4 (1.7)
Lying on a stretcher	1 (0.4)
Standing up/sitting down	31 (13.4)
Not available	26 (11.3)
Extrinsic cause (patient) (%)	
Walking barefoot	64 (43.8)
Inadequate footwear	54 (37.0)
Walking aids	11 (7.5)
Clothes	8 (5.5)
Medical devices (e.g., surgical drains)	9 (6.2)
Extrinsic cause (environment) (%)	
Wet floor	12 (11.2)
Confined space	11 (10.3)
Lowered bad rails	25 (23.4)
Raised bed rails	36 (33.6)
Poor lighting	8 (7.5)
Structural safety issues (e.g., uneven floor)	15 (14.0)

Number of concomitant medication (%)	
1-2 concomitant medications	59 (30.6)
3 concomitant medications	14 (7.3)
≥ 4 concomitant medications	117 (60.6)
No concomitant medication	3 (1.6)
Fall-related injury: *No* (%)	120 (50.2)
If any, injury type (%)	
Hematoma	9 (7.4)
Contusion	48 (39.7)
Sprain	2 (1.7)
Excoriation	38 (31.4)
Wound	15 (12.4)
Fracture	9 (7.4)
Injury treatment: *None* (%)	159 (82.0)
Gold standard location	N = 187
Hallway	5 (2.6)
Hospital bedroom	126 (67.4)
Hospital bathroom	56 (30.0)
Gold standard injury	N = 93
Yes	34 (36.6)
No	59 (63.4)

Table 3. Number of reports correctly classified and misclassified by GPT 3.5 and 4 according to the location and fall related injuries. Data are absolute numbers (percentage). The percentages were calculated according to the overall number of records that underwent GPT classification (187 for location and 93 for injuries)

	Hallway	Hospital bathroom	Hospital room	Presence of fall related injury
GPT 3.5				
Positive cases correctly classified	3 (1.6)	54 (28.9)	102 (54.5)	34 (36.6)
Positive cases misclassified	2 (1.1)	2 (1.1)	24 (12.8)	0 (0.0)
Negative cases correctly classified	175 (93.6)	117 (62.5)	60 (32.1)	36 (38.7)
Negative cases misclassified	7 (3.7)	14 (7.5)	1 (0.6)	23 (24.7)
GPT 4				
Positive cases correctly classified	5 (2.7)	55 (29.4)	109 (58.2)	30 (32.3)
Positive cases misclassified	0 (0.0)	1 (0.6)	17 (9.1)	4 (4.3)
Negative cases correctly classified	182 (97.3)	117 (62.5)	60 (32.1)	52 (55.9)
Negative cases misclassified	0 (0.0)	14 (7.5)	1 (0.6)	7 (7.5)

Table 4. Performance metrics according to location and fall related injuries. For each metric, point estimates (95% Confidence Interval) are reported. When the metrics are equal to 1, it means that there were no classification errors.

	Hallway	Hospital bathroom	Hospital room	Presence of fall related injuries
GPT 3.5				
Accuracy	0.953 (0.920; 0.979)	0.915 (0.872; 0.957)	0.868 (0.818; 0.92)	0.752 (0.667; 0.839)
Sensitivity	0.586 (0.059; 1.000)	0.965 (0.911; 1.000)	0.812 (0.74; 0.881)	1.000
Specificity	0.962 (0.935; 0.984)	0.894 (0.832; 0.947)	0.984 (0.945; 1.000)	0.612 (0.491; 0.737)
GPT 4				
Accuracy	1.000	0.92 (0.882; 0.957)	0.904 (0.861; 0.941)	0.881 (0.817; 0.946)
Sensitivity	1.000	0.982 (0.940; 1.000)	0.865 (0.809; 0.925)	0.88 (0.759; 0.974)
Specificity	1.000	0.894 (0.838; 0.946)	0.983 (0.946; 1.000)	0.881 (0.79; 0.963)

Supplementary Material

Table S1. Examples of records misclassified for the location by GPT-4.

Fall description record	GPT classification	Gold standard
"Ha scavalcato la sponda per scendere, si è aggrappato e si è fatto scivolare lentamente a terra. Dice di essersi aggrappato con ginocchia e spalla sinistra e non ha riportato alcun danno. Non perdita di coscienza"	The text does not provide information on where the fall occurred.	Hospital bedroom
"Paz agitato ha scavalcato le spondine e nella caduta riporta ferita lacero contusa gomito e ginocchio dx. Non appariscente trauma cranico"	The text does not provide information on where the fall occurred.	Hospital bedroom
"Ha scavalcato le spondine e si è tenuto ad esse aggrappato, poi ha perso le forze adagiandosi lentamente a terra. Già caduto a terra fratturandosi il femore destro."	The text does not provide information on where the fall occurred.	Hospital bedroom

Table S2. Examples of records misclassified for the presence or absence of injury by GPT-4.

Fall description record	GPT classification	Gold standard
"Dinamica non ricostruibile (paziente demente con allettamento cronico, non testimoni se non la badante che prestava assistenza notturna) si precisa che durante la notte la badante, come da richiesta dei familiari e da noi concessa, è stata ininterrottamente presente al letto della paziente dalle ore 20 alle ore 7. non sono segnalati in diario clinico eventi significativi. non si è in grado di risalire alla dinamica che ha determinato le suddette lesioni, che non erano presenti nella giornata del 20/01/2016."	The text does not provide information on whether the patient sustained any injuries from the fall.	Presence of fall related injury
"Si rinvia la vaccinazione 1 epa A perché il b/o è scivolato dal lettino dopo che la mamma si è allontanata improvvisamente dallo stesso abbandonandolo i due operatori impegnati uno al carrello e l'altro al computer per inserimento dei dati vaccinali come da procedura. valutata dal pediatra presente nell'altro ambulatorio vaccinale. "Posizionato ghiaccio in zona frontale con presenza bozza. B/o reattivo non sembra aver riportato alcuna lesione di rilievo. Si danno indicazioni sul comportamento da tenere a domicilio. il 2/08/2016 contattato la mamma telefonicamente: il bambino non ha presentato alcun sintomo di rilievo."	Absence of fall related injury	Presence of fall related injury
"L'infermiere sentiva un tonfo e trovava la signora a terra tra il bidet e il lavandino, al mattino. Lamentava dolore all'ipocondrio sx e arcata sopraciliare dx"	The text does not provide enough information to determine if the patient sustained any injuries from the fall.	Presence of fall related injury
"Il paziente (presente la moglie) andando in bagno è scivolato cadendo all'indietro e battendo la testa. Chiamati dalla moglie, abbiamo trovato il paziente a terra vicino al letto vigile e collaborante. Messo a letto con borsa di ghiaccio"	The text does not provide information on whether the patient sustained any injuries from the fall.	Presence of fall related injury
"Il paziente nell'uso della toilette è caduto. Riferisce perdita di coscienza non lesioni apparenti, lamenta lieve fastidio braccio sinistro"	Presence of fall related injury	Absence of fall related injury
"Ha scavalcato le spondine e si è tenuto ad esse aggrappato, poi ha perso le forze adagiandosi lentamente a terra. Già caduto a terra fratturandosi il femore destro."	Presence of fall related injury	Absence of fall related injury