Approaches for LLM Based Automated Categorization of Hospital Imaging Safety Event Reports

Chintan Shah, MD, MS, Neuroradiologist, Medical Director for Imaging AI, Cleveland Clinic
Brandon English, RT, HACP-PE; Shelly Horvath, RN, HACP-PE; Rekha Mody, MD; Po-Hao Chen, MD, MBA

Background/Problem Being Solved

Organizing safety event reports by patterns is critical for quality improvement. Currently, these are manually categorized into one of 40 predefined categories by trained personnel, a time intensive process. Large Language Models (LLMs) offer a promising avenue for reducing manual effort.

Intervention(s)

A dictionary of definitions and examples for 40 pre-defined categories was created by subject matter experts (SMEs). Approaches for LLM-based automated categorization were tested on 320 reports and compared with SME-assigned categories. Llama 3.1-8b was used and hosted locally via Ollama. Four approaches were evaluated:

1. Zero-Shot Categorization: A prompt was provided with all 40 categories.
2. Instructional Modification: To address frequent misclassifications involving “Near Miss,” clarification guidelines were introduced.
3. Top-3 Restriction: Top-3 most frequently used categories were identified. LLM was limited to selecting one of the top three when applicable, or "Other” when not applicable.
4. Retrieval-Augmented Categorization: Category definitions were embedded and used to retrieve the top 10 candidate categories, which were then passed to the LLM for final categorization.

Barriers/Challenges

Challenges included the large number of categories, nuanced definitions, and utility to manual workflow.

Outcome

Approach #1 resulted in 49% (158/320) correct category assignments. “Near miss” reports were incorrectly assigned (0/32). Approach #2 improved the categorization rate overall to 53% (170/320), and in the “Near miss” category to 63% (20/32). Approach #3 improved performance in each of the top 3 categories (51/52 “contrast extravasation”, 25/32 “delay in exam”, and 30/32 “near miss”). However, other reports were commonly mis-categorized into these categories, resulting in an overall lower performance of 44% correct (140/320). In approach #4 the correct category was retrieved in 89% (284/320), however, the final categorization was correct in 64% (205/320).

Conclusion/Statement of Impact/Lessons Learned

Integrating LLMs into safety event categorization workflows has highlighted the challenges above. Subsequent efforts focus on summarizing the event for trained personnel to expedite categorization.

Figure(s)

Summary of event (summarized and modified for privacy)	Actual Category (SME-assigned)	LLM – assigned Category			
		Approach 1	Approach 2	Approach 3	Approach 4
IV contrast infiltrated under patient's skin during scan, 100 cc's of Omnipaque 300 , new IV inserted in opposite arm to complete exam.	Contrast Extravasation	Contrast Extravasation	Contrast Extravasation	Contrast Extravasation	Contrast Extravasation
Pt had a CT with IV contrast after emergent prep protocol and experienced a breakthrough reaction showing symptoms including itching, hives, and low oxygen levels.	Contrast Reaction	Contrast Reaction	Contrast Reaction	Near Miss	Contrast Reaction
A CXR was ordered but the quality was poor due to significant motion artifact and it couldn't be read. A second order for imaging had to be placed instead of retaking the first one, resulting in an additional charge to the family.	Delay in Exam	Delay in Exam	Delay in Exam	Delay in Exam	Delay in Exam
MRI Dense Breast Limited WO/W IVCON studies can only be performed at Main Campus, not at other facilities. Patients must be rescheduled for these studies at Main Campus.	MRI Scheduling	MRI Scheduling	MRI Scheduling	Delay in Exam	Scheduling
An outpatient was ordered to have an X-ray of their left side for rib pain, but they had come in with complaints of right-sided rib pain.	Near Miss	Misread	Near Miss	Near Miss	Near Miss
A wrong side was ordered for an elbow, with left being ordered instead of the required right side, causing a delay.	Near Miss	Delay in Exam	Near Miss	Near Miss	WSSP
A patient was brought in for a scan, but claimed he was a victim of ID theft and didn't know why he was there. The appointment was cancelled after verifying his identity issues.	Other	Other	Informatics	Near Miss	Not Imaging/Employee
A CT order was changed to without contrast but the provider did not inform us.	Patient Safety	Delay in Exam	Clinical Engineering	Other	Delay in Exam
Transporter took 30 minutes to transport a patient upstairs, then went on break without informing staff, causing a delay in care for over an hour due to short staffing.	Transport	Transport	Transport	Delay in Exam	Transport
A patient was exposed to additional radiation due to a mix-up in X-rays, specifically a left femur X-ray being canceled and replaced with an incorrect right hip X-ray before being reordered correctly.	WSSP	Delay in Exam	Near Miss	Near Miss	WSSP

Abbreviations: LLM – Large language model; SME – Subject matter expert; WSSP – Wrong site or side procedure.

Table 1. Example event descriptions with Actual Category and LLm-Assigned categories. Descriptions summarized and modified for privacy.

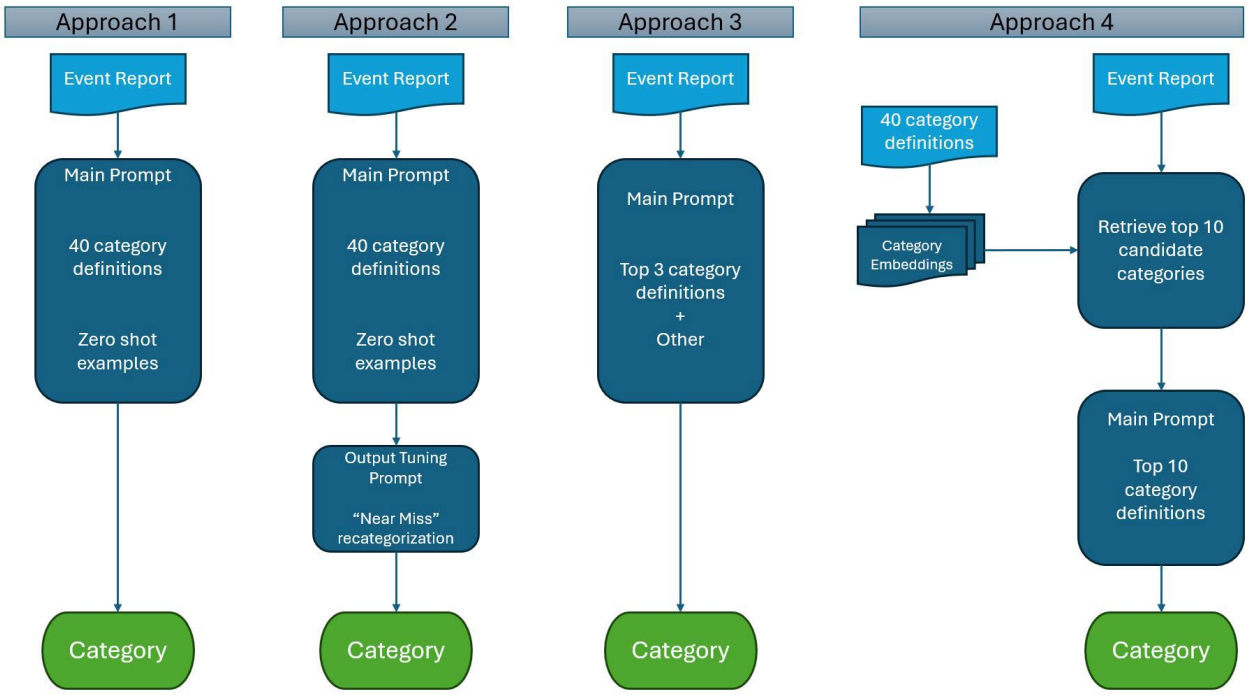

Figure 1. Four approaches used for LLM-based categorization of imaging safety event reports.

Keywords

Artificial Intelligence/Machine Learning; Quality Improvement & Quality Assurance