

Large Language Models for Rail Safety

BUILT BY FIRST ANALYTICS

OCTOBER 2024



ChatFRA

Two source FRA databases:

- Rail equipment accident/incident data (Form 54)
- Highway-Rail Grade Crossing Incident Source Data (Form 57)

DEPARTMENT OF TRANSPORTATION FEDERAL RAILROAD ADMINISTRATION (FRA)		RAIL EQUIPMENT ACCIDENT/INCIDENT REPORT		OMB No. 2130-0500	
1. Name of Reporting Railroad		1a. Alphabetic Code	1b. Railroad Accident/Incident No.		
2. Name of Other Railroad or Other Entity with Consist Involved		2a. Alphabetic Code	2b. Railroad Accident/Incident No.		
3. Name of Railroad or Other Entity Responsible for Track Maintenance (single entry)		3a. Alphabetic Code	3b. Railroad Accident/Incident No.		
4. U.S. DOT Grade Crossing Identification Number		5. Date of Accident/Incident month day year		6. Time of Accident/Incident AM <input type="checkbox"/> PM <input type="checkbox"/>	
7. Type of Accident/Incident (single entry in code box)		1. Derailment		4. Side Collision	
		2. Head on collision		5. Raking collision	
		3. Rear end collision		6. Broken train collision	
		7. Hwy-rail crossing		10. Explosion-detonation	
		8. RR grade crossing		11. Fire/violent rupture	
		9. Obstruction		12. Other impacts	
8. Cars Carrying HAZMAT		9. HAZMAT Cars Damaged/Derailed		10. Cars Releasing HAZMAT	
11. People Evacuated		12. Subdivision		13. Other (describe in narrative) Code	
13. Nearest City/Town		14. Milepost (to nearest tenth)		15. State Abbr. Code	
16. County		17. Temperature (F)		18. Visibility (single entry) Code	
19. Weather (single entry) Code		20. Type of Track Code			

DEPARTMENT OF TRANSPORTATION FEDERAL RAILROAD ADMINISTRATION (FRA)		HIGHWAY-RAIL GRADE CROSSING ACCIDENT/INCIDENT REPORT		OMB No. 2130-0500	
1. Name of Reporting Railroad		1a. Alphabetic Code	1b. Railroad Accident/Incident No.		
2. Name of Other Railroad or Other Entity Filing for Equipment Involved in Train Accident/Incident		2a. Alphabetic Code	2b. Railroad Accident/Incident No.		
3. Name of Railroad or Other Entity Responsible for Track Maintenance (single entry)		3a. Alphabetic Code	3b. Railroad Accident/Incident No.		
4. U.S. DOT Grade Crossing Identification Number		5. Date of Accident/Incident month day year		6. Time of Accident/Incident AM <input type="checkbox"/> PM <input type="checkbox"/>	
7. Nearest Railroad Station		8. Subdivision		9. County	
10. State Abbr. Code		11. City (if in a city)		12. Highway Name or Number Public <input type="checkbox"/> Private <input type="checkbox"/>	
13. Type			17. Equipment		
A. Auto			4. Car(s) (moving)		
C. Truck-trailer			5. Car(s) (standing)		
D. Pick-up truck			6. Light loco(s) (moving)		
F. Bus			7. Light loco(s) (standing)		
G. School bus			A. Train pulling - RCL		
J. Other motor vehicle			B. Train pushing - RCL		
K. Pedestrian			C. Train standing - RCL		
Code			D. EMU Locomotive(s)		

Two Modes / Features

- Use the bot as a traditional BI (SQL) query tool
- But adds capability to ask questions of unstructured data (e.g. textual incident narratives)

ChatFRA Examples

BI/SQL QUERYING AND INTERROGATION OF TEXTUAL NARRATIVES

ChatFRA

Simple BI reporting / SQL generation mode

Prompt

For 2023, provide a total of each incident type. Sort and display them in descending order by total.

Response (exact quote)

In 2023, the total number of each incident type, sorted in descending order by total, is as follows:

- Derailment: 1413 incidents
- Other impacts: 446 incidents
- Hwy-rail crossing: 244 incidents
- Other (describe in narrative): 202 incidents
- Side collision: 125 incidents
- Obstruction: 121 incidents
- Raking collision: 23 incidents
- Fire/violent rupture: 23 incidents
- Rear end collision: 8 incidents
- Head on collision: 3 incidents
- RR grade crossing: 2 incidents
- Explosion-detonation: 1 incident
- Broken train collision: 1 incident

Please note that these numbers are based on the data available in the database and the actual numbers may vary.

ChatFRA

Simple BI reporting / SQL generation mode

Prompt

How many grade crossing incidents were there in Illinois from 2020 to 2023, by year, and in total?

Response (exact quote)

Based on the data from the Highway-Rail Grade Crossing Incident Source Data, here are the number of grade crossing incidents in Illinois from 2020 to 2023, by year:

In 2021, there were 127 incidents.
In 2022, there were 148 incidents.
In 2023, there were 102 incidents.

So, in total, there were 377 grade crossing incidents in Illinois from 2020 to 2023.

The SQL query I used to get this information is:

```
SELECT year, COUNT(*) as incidents FROM Embeddings WHERE state = '17'  
AND year BETWEEN '20' AND '23' GROUP BY year
```

Complex BI reporting / SQL generation mode

Prompt: *What are the 3 grade crossings that reported the highest number of people injured in 2024? Where are they (FIPS State, county)? Can you retrieve the incident narrative and date of each incident?*

Based on the data from the Highway-Rail Grade Crossing Incident Source Data, the 3 grade crossings that reported the highest number of people injured in 2024 are as follows:

1. Grade Crossing ID: 057218E

- Location: Weld County, FIPS State Code: 8
- Incident Narrative: "GRADE CROSSING: TRAIN 6(28), OPERATING WITH ENGINES 816, 41 WITH CARS 61049, 39046, 32011, 32060, 38 024, 33031, 34032 AND 35007, STRUCK A SEMI-TRUCK AT COUNTY ROAD 63 (DOT #057218E) AT APPROXIMATELY MP 502.0, ON THE BURLINGTON NORTHERN SANTA FE BRUSH SUBDIVISION. THE ENTIRE TRAIN DERAILED. ENGINES 816 AND ENGINE 41 DERAILED ON ITS SITE. CARS 61049, 39046, 32011, 32060, 38024, 33031, 34032 AND 350 07 DERAILED UPRIGHT."
- Date of Incident: January 29, 2024
- Number of People Injured: 14

2. Grade Crossing ID: 345254U

- Location: Christian County, FIPS State Code: 21
- Incident Narrative: "I12815 STRUCK A SUBURBAN SUV THAT WENT AROUND GATES AT A CROSSING. PROTECTION ALSO AT CROSSING: PAV EMENT MARKINGS (STOP LINES & RR XING SYMBOLS)."
- Date of Incident: January 16, 2024
- Number of People Injured: 4

3. Grade Crossing ID: 069146M

- Location: Bond County, FIPS State Code: 17
- Incident Narrative: "L-CHI6551-18I TRAVELING SOUTHBOUND STRUCK A VEHICLE TRAVELING WESTBOUND ON SINGLE MAIN TRACK. NO HAZ MAT RELEASED. USER'S AGE UNKNOWN"
- Date of Incident: February 17, 2024
- Number of People Injured: 4

The SQL query I used to get this information is:

```
SELECT gxid, state, county, narrative, day, month, year, TOTINJ
FROM Embeddings
WHERE year = 24
ORDER BY TOTINJ DESC
LIMIT 3
```

ChatFRA

Queries based on non-structured data, e.g. narrative text

Prompt

How many incidents involved livestock?

Response (exact quote)

Based on the search terms I used, there have been 93 incidents involving livestock. This includes incidents where the narrative mentions "livestock", "cattle", "pigs", "goats", "sheep", or "horses".

Please note that this number might not include all incidents involving livestock if they are referred to using different terminology in the narrative. I'm fairly confident in this answer, but there's always a chance that some incidents were missed due to the limitations of text search.

ChatFRA

Queries based on non-structured data, e.g. narrative text

Prompt

Are there any mentions of human factors in derailments?

Response (exact quote)

Based on the data in the 'incidents' table, there are 165 incidents where the narrative mentions both 'human' and 'derailment'. This suggests that human factors may have been involved in these derailments.

However, please note that this is a simple text search and does not necessarily imply causation. The narratives would need to be read in detail to confirm the exact role of human factors in these incidents.

I'm fairly confident in this answer, but there could be other interpretations or ways to approach your question.