MistralLarge

Model demonstrates a solid foundational ability in text analysis, successfully identifying major entities, extracting relationships, mapping plot structures, and tracking story evolution. However, a deeper evaluation across different news articles (*Delfi, LA, LSM, TVNET*) reveals both strengths and recurring limitations that highlight the system’s current capabilities and areas for improvement.

These limitations indicate that while the model provides *useful insights*, it still operates in a relatively structured, pattern-based manner rather than exhibiting comprehension.

#### **Named Entity Recognition**

Across all evaluated articles, the model demonstrates a consistent ability to identify major entities, particularly locations and actors. However, recurring misclassifications—such as treating *NATO* and the *EU* as locations rather than geopolitical organizations—highlight a need for refinement in entity categorization. While the model generally performs well in actor recognition, it occasionally overlooks key figures, such as *Saeima* in *Delfi*, or misclassifies government institutions as entities despite their indirect mention.

Event detection varies in effectiveness; the model successfully recognizes explicit events but struggles with implicit or complexly worded occurrences. Additionally, while it accurately detects direct date mentions, it has difficulty processing more nuanced temporal expressions, such as *"next April"* or *"within a few days."* This suggests that while its structured pattern recognition is reliable, it requires improvement in handling contextual time references.

#### **Relationship Extraction**

The system’s relationship extraction capabilities are functional but exhibit notable limitations in depth and accuracy. While it correctly identifies fundamental connections between entities, it lacks the interpretative ability to capture more complex relationships. Some extracted relationships are too generic, failing to specify the nature of interactions (e.g., *"Balticconnector, damage"* without context). In some cases, the model even generates inferred relationships that are not explicitly present in the text, demonstrating assumed correlations rather than precise textual evidence.

#### **Plot Discovery**

The model effectively captures essential plot elements but occasionally misprioritizes thematic focus. Instead of aligning with the primary subject matter, it tends to emphasize specific keywords or entities, such as *Balticconnector*, while neglecting broader l implications. Additionally, redundancy is present in some outputs, where key details are repeated rather than progressively built upon.

Structural coherence remains a challenge. The sequencing of extracted information does not always align with the article’s narrative flow, leading to inconsistencies in how the plot is reconstructed. This suggests that while the model excels in identifying key details, it requires improvements in recognizing how those details interconnect within a structured storyline.

#### **Story Evolution**

The model's ability to track story evolution remains one of its weaker aspects. A common issue is the misidentification of shifts within the narrative—frequently, the model treats the introduction of an article as a shift rather than recognizing genuine turning points in the storyline. This results in an incomplete or skewed representation of how a story develops over time.

Causal relationship extraction performs moderately well, with the system correctly identifying some cause-and-effect relationships. However, it tends to miss additional relevant causal links that could enhance the understanding of the article's progression.

DELFI

**Named Entity Recognition**

Model has successfully identified key entities within the text, which is generally satisfactory. However, an issue arises with the classification of "Saeima" as a location, despite its role in the article being more akin to that of an actor. While the system has performed well in recognizing actors, it incorrectly identified "Saeima" as a place rather than an institutional entity involved in decision-making.

Furthermore, the system has demonstrated strong performance in detecting events, with results closely aligning with human annotations. However, there is one instance where an entity could be interpreted as an event, though contextual analysis suggests it does not fully meet the criteria of an event. Additionally, while the system accurately recognizes explicit mentions of days and dates, it struggles with more complex timeline expressions.

**Relationship Extraction**

Models capabilities are generally adequate, as the identified relationships correspond to those present in the article. However, discrepancies remain when comparing the models output to human interpretation. Specifically, while the model successfully identifies relational links, it lacks the nuanced interpretative abilities of human annotators, who can discern more complex relationships and abstract connections. This suggests that model operates in a more structured and rigid manner, whereas human interpretation allows for greater flexibility and contextual awareness.

**Plot Discovery**

Models approach to plot discovery does not entirely align with the primary focus of the article. Instead of emphasizing the central themes—namely, gas prices and commercial actors—it has placed disproportionate emphasis on "Balticconnector." While certain extracted elements are relevant, they only partially capture the article's overarching narrative. This indicates that the system may prioritize specific keywords or topics without fully grasping the broader context of the discussion.

**Story Evolution**

Model does not effectively track story evolution but instead identifies a shift in the main storyline as its primary narrative progression. While it demonstrates competence in causal relationship extraction, some critical aspects are either omitted or lack depth. Despite this limitation, the extracted information remains relatively comprehensive.

—

LA

### **Named Entity Recognition**

Model has successfully identified all locations mentioned in the text. However, it has incorrectly classified certain actors, such as NATO, as a location. Additionally, the system failed to recognize some locations that were very similar in structure but distinct in meaning. This suggests that while the model is capable of identifying well-defined locations, it struggles with differentiating similar-sounding entities, particularly when a single shared word is present.

In terms of actor recognition, the model performed well, correctly identifying all actors and incorporating full explanations for each. However, while the model detected a greater number of events, some instances appeared repetitively. Model included the phrase “iespējams sprādziens” as an event, even though it does not explicitly appear in the text. Nonetheless, for events that were manually identified, the system successfully recognized them as well.

Timeline - model successfully identified most times . However, it failed to recognize the phrase **“nākamā gada aprīlī”** ("next April"), which may be more complex to interpret due to its relative rather than absolute nature.

### **Relationship Extraction**

The relationship extraction results are somewhat mixed. While some extracted relationships are meaningful and align with the article, others lack clarity, making it difficult to discern the type of relationship being suggested. For instance, in cases such as **[Balticconnector, bojājumi]** ("Balticconnector, damage"), the extracted relationship lacks specificity, failing to indicate whether the entity is the subject, object, or cause of the damage. This suggests that while the system identifies connections between entities, it does not always provide a clear relational structure.

### **Plot Discovery**

The system’s approach to plot discovery differs from human analysis. While it places greater emphasis on the investigative aspects of the article, it also introduces redundancy by repeating certain elements throughout the steps. Additionally, the sequencing of extracted information does not fully align with the article’s narrative structure, leading to inconsistencies in how the storyline is reconstructed. This suggests that the system may prioritize specific recurring terms over accurately mapping the logical flow of the text.

### **Story Evolution**

Models recognition of story evolution is generally acceptable. While the identified shift does not correspond to the very first event in the article, it also does not fully align with a significant turning point. Instead, it appears to capture an introductory stage rather than a clear narrative shift.

Causal relationship extraction is relatively strong, providing valuable insights into the article’s structure. One section of the extracted story is particularly well-identified, though additional refinement could enhance completeness.

LSM

### **Named Entity Recognition (NER)**

Model effectively identifies country names; however, it struggles with distinguishing other entities classified as locations. Specifically, it incorrectly categorizes actors such as **NATO** and the **EU** as locations rather than geopolitical organizations.

While the system successfully recognizes the majority of entities, it occasionally overlooks less prominent ones, particularly those that are not as easily identifiable. Additionally, although it performs well in identifying certain key elements, it fails to detect more precise details.

In terms of timeline the system does not identify explicit mentions of specific days. However, it demonstrates an improved ability to recognize more complex time expressions, which indicates progress in processing nuanced temporal references.

### **Relationship Extraction**

Models relationship extraction capabilities are highly effective, as the identified relationships generally align with the article’s content. Many of the extracted relationships are meaningful and contribute to a coherent understanding of the text.

### **Plot Discovery**

While model correctly identifies important details within the article, it struggles with accurately capturing the progression of the plot. The extracted elements do not fully reflect the logical development of the story, particularly in terms of how events unfold and where narrative peaks, such as the climax, are positioned. This suggests that while the system is effective at recognizing critical information, it lacks the ability to structure these details in a way that mirrors the natural flow of the storyline.

### **Story Evolution**

Models story evolution presents notable challenges. It fails to recognize shifts in focus within the narrative, instead treating the article’s main subject as a shift rather than acknowledging how the story evolves over time. This indicates a fundamental misunderstanding of how narratives transition between different focal points.

Despite this limitation, the system does demonstrate strength in summarization. In a single sentence, it successfully extracts the core theme of the article, which is highly effective. However, while this summary is useful, there are additional details present in the article that are not reflected in the extracted evolution. .

TVNET

### **Named Entity Recognition**

Model exhibits the same recurring issue as in previous cases, incorrectly classifying **NATO** and the **EU** as locations. However, aside from this misclassification, the recognized entities closely align with human annotations.

In terms of actor recognition, model performs at a human-equivalent level, correctly identifying all relevant entities. Additionally, while it does not recognize specific days for unknown reasons, it demonstrates the ability to process more complex timelines, indicating a preference for structured or abstract time references over simple date mentions.

### **Relationship Extraction**

Although the extracted relationships may differ slightly from those identified by human annotators, model performs well overall. The relationships it detects are meaningful and closely align with those established through manual annotation, demonstrating a high degree of accuracy in capturing entity interactions.

### **Plot Discovery**

Models approach to plot discovery mirrors its handling of story evolution, which is a positive outcome. The key elements of the narrative are well identified, and the overall structure and thematic focus align with human interpretation. However, there is a slight discrepancy in how the system assigns plot discovery points, causing a minor misalignment in its representation of narrative shifts. Despite this, the overarching storyline and conceptual understanding remain largely intact.

### **Story Evolution**

Shift detection continues to present challenges, as model does not distinguish between a genuine narrative shift and the beginning of a story. Instead of identifying a transitional point, it treats the article’s introduction as a shift, leading to inaccuracies in tracking how the narrative progresses.

While additional relationships could be extracted to provide a more comprehensive understanding of the story, the ones that have been identified are relevant and grounded in the text. However, models output does not fully capture all possible relational links, suggesting room for improvement in detecting and integrating a wider range of connections.