

Using Novel Phrase-Level Explanations in a Softmax-Linked Additive Explainability Model for Transformers

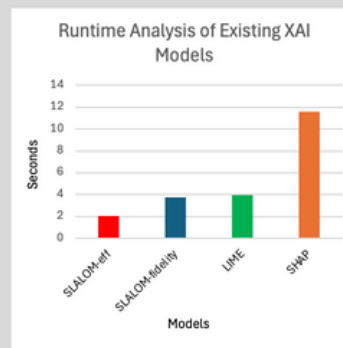
Neil Gupta, Mass Academy

Problem: Transformer models being black boxes states the need for clear explanations that accounts for the structure of language.

Objective: Develop an explainability model for transformers that accounts for phrase-level explanations and ...

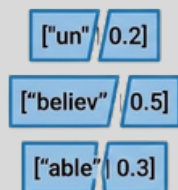


Looking at phrase-level explanations **increases** understanding

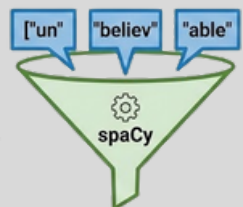


The runtime analysis demonstrates that SLALOM-eff provides a significant computational advantage by generating explanations in 2.03 seconds, effectively addressing the latency bottlenecks found in traditional feature attribution methods.

Token Decomposition
(The Math Layer)



Hierarchical Aggregation
(The Bridge Layer)



Phrase-Level Explanation
(The Human Layer)

UNBELIEVABLE

By including hierarchical phrase-level explanations, this project bridges the gap between uninterpretable transformer attribution and human understanding. The resulting model maintains the accuracy required for "Trustworthy AI" while significantly reducing the cognitive load of fragmented token-level data.

QUAD Chart w/ Main Takeaway Guidelines

POST-December FAIR STEM Website Directions
Mass Academy of Math and Science

adapted from Presentation Materials for Regeneron ISEF 2021

A “quad chart” is a single page divided into four quadrants providing a high-level summary of the project. It is intended to be more visual than detailed, quickly introducing your website visitors to what is important about your project. This exercise provides an opportunity to develop your visual abstract and methodology infographics for your Feb Fair Poster.

1. You must use a wide-screen page format no larger than the American Legal standard 8½”x11” and arranged in **Landscape orientation**.
2. The **minimum allowable font size is 14 pt.** Exception: You may use a smaller font size, down to 9 pt., for figure captions or photo credits.
3. **When considering color and font**, ensure that all text is clearly readable. Example: Blue text on red background is NOT readable.
4. **All four quadrants of your Quad Chart should each be the same size** with a single border line delimiting each, as in the example. The Title section should be only as tall as necessary to include your project title and other identifying information (see section on Quad Chart Title). The Main Takeaway should be centered, as shown in the template. It should be clearly visible and consist of no more than one sentence/phrase.
5. The Quad Chart should not include a bibliography, references, or acknowledgments.
6. All ISEF **Display & Safety rules** must be followed.
- 7: The Quad Chart will be placed on your STEM Webpage prior to Feb Fair. The week prior to Feb Fair, you will be adding additional content (namely, your poster and corresponding data) to your website for judge inspection.

