HAAG Weekly Report (Simplified): Week 7 D&O

By: [redacted]

Sections to Complete

Time-Log

- What did you do this week?
  - Attended publication seminar
  - Sent out best output for manual review with the normalized saliency scores
  - Conducted Nathan's Challenge with the manual review
  - Experimented with Michael's annotation interface and offered feedback
  - Finalized Abstract
- What are you going to do next week
  - Finalize Introduction based on feedback
  - Finish Manual Review
- Blockers, things you want to flag, problems, etc.
  - Dr. Alexander is also performing a review, which has not yet been completed. We thought this would be completed by now, but she had something come in. She plans to work on it over the weekend. I cannot work on the classifier until then.

Abstracts:

Title: Table Meets LLM: Can Large Language Models Understand Structured Table Data? A Benchmark and Empirical Study

Authors: Yuan Sui, Mengyu Zhou, Mingiie Zhou, Shi Han, Dongmei Zhang

DOI: https://doi.org/10.1145/3616855.3635752

Abstract: Large language models (LLMs) are becoming attractive as few-shot reasoners to solve Natural Language (NL)-related tasks. However, there is still much to learn about how well LLMs understand structured data, such as tables. Although tables can be used as input to LLMs with serialization, there is a lack of comprehensive studies that examine whether LLMs can truly comprehend such data. In this paper, we try to understand this by

designing a benchmark to evaluate the structural understanding capabilities (SUC) of LLMs. The benchmark we create includes seven tasks, each with its own unique challenges, \eg, cell lookup, row retrieval, and size detection. We perform a series of evaluations on GPT-3.5 and GPT-4. We find that performance varied depending on several input choices, including table input format, content order, role prompting, and partition marks. Drawing from the insights gained through the benchmark evaluations, we proposeself-augmentation for effective structural prompting, such as critical value / range identification using internal knowledge of LLMs. When combined with carefully chosen input choices, these structural prompting methods lead to promising improvements in LLM performance on a variety of tabular tasks, \eg, TabFact(\uparrow2.31%), HybridQA(\uparrow2.13%), SQA(\uparrow2.72%), Feverous(\uparrow0.84%), and ToTTo(\uparrow5.68%). We believe that our open-source (please find code and data at https://github.com/microsoft/TableProvider) benchmark and proposed prompting methods can serve as a simple yet generic selection for future research.

Summary:

This paper investigates how well large language models (LLMs), like GPT-3.5 and GPT-4, understand structured data such as tables. While tables can be inputted through serialization, there has been limited analysis of LLMs' true comprehension of tabular structures. The authors create a benchmark with seven tasks (e.g., cell lookup, row retrieval, size detection) to evaluate the structural understanding capabilities (SUC) of LLMs. Their results show that performance varies significantly based on input format, content order, role prompting, and partition markers. To improve performance, they propose **self-augmentation** techniques that use LLMs' internal knowledge for structural prompting. These techniques, when combined with optimized input choices, yield measurable gains on several benchmark datasets. The benchmark and methods are open-source and intended to support broader research in this area.

What did you do and prove it

I shared the integrated results with the updated normalized saliency scores. The manual review is a painstaking process, but it is necessary to be done correctly. We have finalized the abstract. I finished my work for the Introduction but need to make the necessary revisions, based on feedback from my peers in Overleaf. I was also designated as the individual for Publication Seminar. We have also put together a project timeline, shown below:

Paper:
Workshop deadlines:
https://jurix2025.di.unito.it/

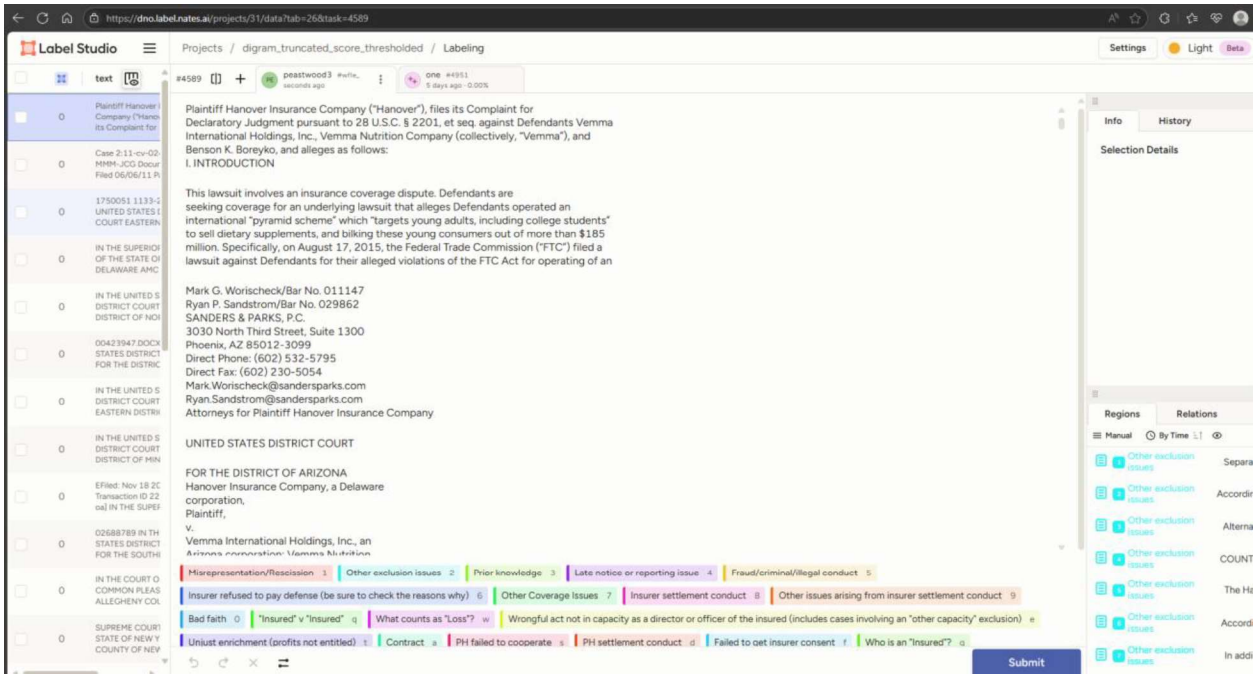
- Abstract submission deadline: August 28, 2025 (recommended)
- Paper submission deadline: September 4, 2025
- Notification of acceptance: October 8, 2025
- Camera-ready deadline: October 15, 2025
- Workshops, Tutorials, and Doctoral Consortium: December 9, 2025
- Main Conference: December 10-11, 2025

Our deadlines:

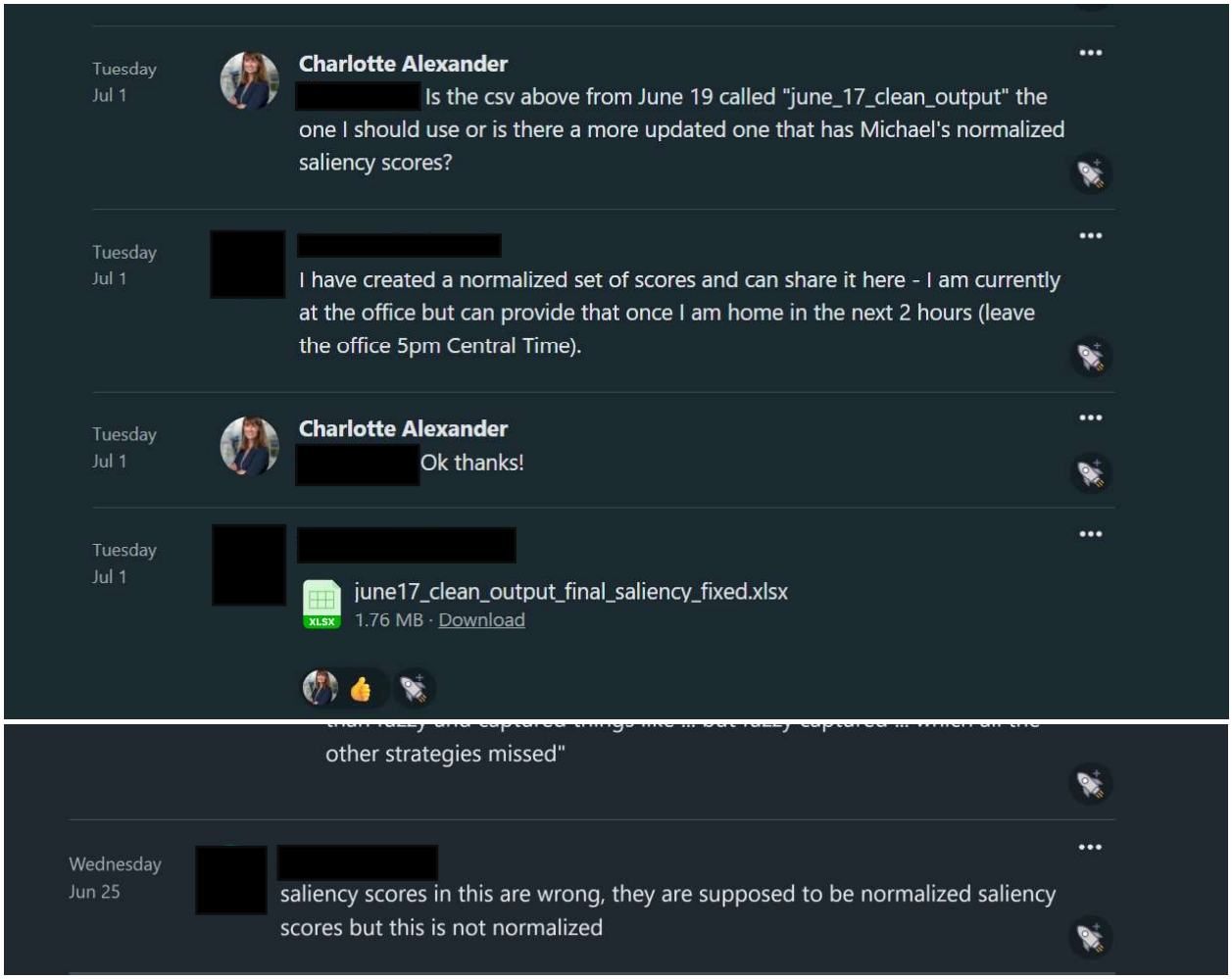
Suzanne and Evan will tag-team on continuing to update existing draft

- End of June: abstract and intro; Xin and Isidora can help
- Beginning of July: finalize manual review and weighting strategy
- First half of July: classification
- End of July: all code written; pipeline figured out; results generated
- August: writing and revising

Tested out the annotation interface, as requested:



Sent out the finalized pseudo-labeling, with the normalized saliency scores integrated:



Finished Abstract:

Automating D&O Insurance Analysis: NLP Applications in Coverage Litigation

First AUTHOR [a,1], Second AUTHOR [b] and Third AUTHOR [b]
[a] *Short Affiliation of First Author*
[b] *Short Affiliation of Second Author and Third Author*
ORCiD ID: First Author https://orcid.org/....-....-....-...., Second Author https://orcid.org/....-....-....-...., Third Author https://orcid.org/....-....-....-....

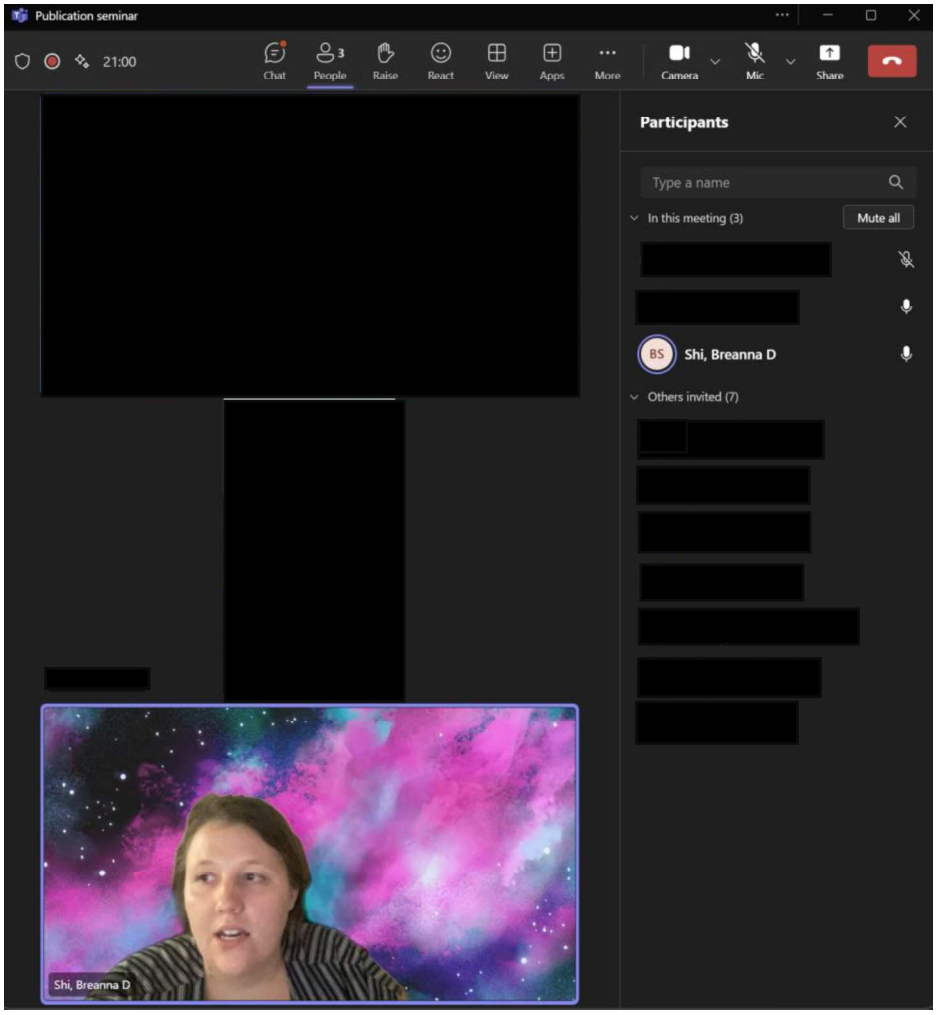
Abstract. Directors and Officers (D&O) insurance litigation addresses disputes over coverage obligations, based on complex contractual provisions. The bulk of these provisions require laborious parsing to ascertain issues associated with the cases. In an effort to streamline these efforts, this paper presents an NLP-based framework for identifying and classifying legal issues—such as "late notice" or "what is a claim"—within D&O court filings. We develop a multi-stage pipeline, combining sentence-level chunking, transformer-based semantic similarity (SBERT, ModernBERT), TF-IDF saliency scoring, and fuzzy string matching to generate pseudo-labels from a small set of annotated sentences, that would, in effect, support the training of a supervised classifier designed to generalize across litigation documents. To facilitate manual expert review, we integrate these outputs into a web-based annotation interface, in which legal reviewers can confirm or revise predicted issue-relevant sentences, thereby improving the accuracy and scalability of issue identification, even in the presence of noise or heterogeneous input. By integrating D&O litigation filings into NLP-based analysis, our findings evidence time-saving benefits with maintained precision for research and application in the field, particularly for insurer audits, legal analytics, and policyholder dispute tracking.

Keywords. D&O insurance, legal NLP, pseudo-labeling, litigation analytics

1. Introduction

Directors and Officers Insurance (D&O Insurance) is a specialized form of liability coverage designed to safeguard the personal assets of corporate directors, officers, managers, and supervisors [1]. This protection extends to liabilities arising from wrongful acts committed in the course of their professional duties, including but not limited to errors, omissions, breaches of fiduciary duty, negligence, misstatements, or misleading

Attended Publication Seminar:


Publication seminar
21:00
Chat
People
Raise
React
View
Apps
More
Camera
Mic
Share
Participants
Type a name
In this meeting (3)
Mute all
BS
Shi, Breanna D
Others invited (7)
Shi, Breanna D