This project will investigate how specific linguistic features, particularly modality and hedging, impact the accuracy and reliability of summaries of legal deposition transcripts generated by software utilizing large language models (LLMs). While LLMs are increasingly used in legal technology for tasks such as summarization and information extraction, there is a lack of research on how they handle linguistically complex input that conveys uncertainty, obligation, or indirectness. These elements are common in legal discourse and complicate both the identification of factual content to include in summaries and the selection of appropriate language to use in summarizing that content.

The study will focus on how LLM-based summarization handles epistemic and deontic modality, as well as related features such as negation, conditionals, intensifiers, and hedging language. I will curate a set of legal transcripts that reflect these patterns and analyze how their presence influences the factuality, coherence, and grounds of model-generated summaries. Evaluation will involve both qualitative and quantitative methods, including targeted prompting experiments, error analysis, and entailment or natural language inference (NLI) models, to assess consistency between the source and summary.

To support interoperability and reproducibility, the project will use a modular NLP pipeline built in Python, incorporating tools such as Hugging Face Transformers for model access and prompt testing, spaCy or NLTK for linguistic annotation, and OpenAI or Bedrock APIs for comparative model evaluation. I will also explore ways to generate confidence scores or explanation layers using attention visualization or sentence-level NLI scoring techniques. Findings will contribute to a deeper understanding of how linguistic nuance impacts summarization quality and provide practical insights for enhancing the reliability of legal AI applications.