Reasoning over Legal Texts Using Large Language Models and Automated Reasoning*

Verena Praher[1], Endre Szasz-Revai[2], and Wolfgang Windsteiger[3]

[1] RISC Software GmbH
verena.praher@risc-software.at
[2] Rise2Reality Software
Endre.Szasz-Revai@rise2reality.com
[3] Johannes Kepler University Linz (JKU)
Research Institute for Symbolic Computation (RISC)
Linz, Austria
Wolfgang.Windsteiger@risc.jku.at

Abstract

This presentation describes an ongoing project in the area of tax law. The goal of the project is to bring computer-support to the decision process regarding taxation for international businesses. In particular, in a first prototype, the focus lies on *transfer pricing*, which is a non-trivial price-determination and taxation process that concerns businesses with divisions distributed over different countries or economies. We describe the ideas pursued in the project, we do not yet have results nor can we report on their critical evaluation by practitioners.

1 The Problem and the Requirements

Transfer pricing is the process of calculating certain prices based on laws and regulations from various sources, e.g., EU, OECD, federal laws, but also company-internal guidelines. The decision, which combination of rules finally determines the prices, is non-trivial for human experts and should be supported by software. The main requirements for the system are:

- *Trustworthiness*: the user must rely on logically sound reasoning.
- *Explainability*: the user should be able to track from which sources a decision is derived.
- *Scalability*: it should be capable of dealing with huge datasets of legal and non-legal texts.

2 The Proposed Design of the Software Prototype

The proposed design combines Large Language Models (LLM) with Symbolic Reasoning. Regulatory texts are formulated in natural language whereas reasoning tools are based on some sort of logic language. In order to make these texts accessible to logical reasoning it requires the formalization of all norms in the language of the prover. This can be seen as a translation from one language into another, and this is where LLMs demonstrate strong capability. The target language is essentially predicate logic. It will become clearer after the first examples whether a subset will be sufficient or certain generalizations are required. In order to guarantee sound

*This work is supported by the Austrian Research Promotion Agency (FFG) in the frame of project "TaxoLogic: Explainable Hybrid Artificial Intelligence for Transfer Pricing", project no. FO999936923/66571492.

application of the norms, the reasoning itself will be purely symbolic without any assistance from an LLM. However, turning the symbolic proof into a chain of logical arguments comprehensible for a domain expert needs another layer of translation. Depending on the reasoning tools in use this might be done by an LLM as well.

3 Symbolic Methods & Machine Learning Techniques

In this approach, we do not enhance LLMs by symbolic computation nor do we empower symbolic algorithms through machine learning. We rather combine the two worlds on the "problem level", i.e., we split the real-world problem into pieces that can then be tackled by symbolic and machine learning methods separately.

Machine Learning Techniques The machine learning components are modeled as a pipeline to bridge the gap between natural and formal language. This pipeline should prepare and translate unstructured legal norms into predicate logic expressions. The LLMs have to be capable of translating legal norms into predicate logic. We will benchmark larger models with explicit reasoning capabilities and smaller models that are specifically fine-tuned on tax-related legal norms, evaluating both extraction quality and suitability for on-premise deployment to ensure data sovereignty.

In the first pipeline step we apply information extraction techniques to structure the input and try to build an ontology or knowledge graph of the terminology and relationships between concepts. We intend to use named entity recognition to identify concepts and context engineering to segment text into manageable parts. This intermediate form should reduce ambiguity and allow us to control semantics for the translation.

In a subsequent step, the LLMs translate the structured natural language into predicate logic expressions. To ensure well-formed outputs, constraints are enforced during generation, e.g., by restricting the output format to valid logical expressions using guardrails. Since this translation is inherently imperfect, the resulting formulas must be reviewed and corrected. To guarantee trustworthiness, we employ a human-in-the-loop verification. Experts review and correct the proposed expressions to ensure that the legal norms are correct before handing them over to the symbolic reasoning engine.

Symbolic Methods One of the challenges for formal reasoning based on propositions extracted from laws and guidelines is their inherent inconsistency. When building up the body of formulas we try to identify contradictions and, if possible, resolve them by consulting available court-rulings. In addition, since certain contradictions might not be detected a-priori, we need to restrict the reasoning engine such that it would not apply certain standard inference rules based on the law of contradiction, e.g., any goal is proven as soon as the knowledge base contains a contradiction.

For a prototype we base the reasoning engine on the Theorema system because Theorema by design explains all reasoning steps in a human-comprehensible form. In addition, we have access to the implementation of the inference mechanism so that we can easily adapt to special requirements coming from the application domain. Finally, another advantage is that the Theorema language is untyped and the system is not tied to a certain underlying logic. In its standard form it supports a certain variant of higher-order predicate logic, but it is possible with minimum effort to include parts of modal or deontic logic into the language and add appropriate reasoning capabilities, if this is indicated by the domain of tax law.