

AI, large language models, law and contracts

Conversation with OpenAI GPT 4.1 – August 2025



It's about artificial intelligence and it's about contracts: contract drafting and contract management, also known as CLM or Contract Lifecycle Management, by means of AI.

In this analysis, AI is intended to refer to large language models (i.e., machine learning, language modeling, machine processing of natural language) on a generic level. What is called domain-specific language models is thus definitely not subject to the ongoing analyses!

First, a personal observation: the first generation of language models widely available to the general public, say ChatGPT (3.5), Google Bard, and Claude 2.0, were, in this regard, restricted in their ability to understand and process complex legal questions.

However, with the rollout (or announcement) of advanced models (in particular GPT 4.1, but also Gemini 2.5 and Claude 4.0 upwards), things do seem to be moving forward, because a skilled handling of these models in terms of input means that these models are capable of generating enhanced contractual constructs (multilingual and even multijurisdictional, if necessary) at least - according to my observation.

A skilled handling of these AI models? Well, I mean handling them in a dual capacity—both in terms of their content and from a legal perspective.

Let's take as an example a contract for the design, manufacture, and commissioning of a series of rail vehicles between a municipal body as the purchaser and a manufacturer as the supplier of such vehicles.

This means that I must try to approach the model – as best I can – from the perspective of:

- a) an engineer in the field of rail vehicle development working for the manufacturer and
- b) a lawyer - possibly even a specialist in public procurement law - in the service of the municipal body.

Remarkably, now that I am a professional in one capacity but a 'layman with expertise' in the other, I am able to see things opening up, so to speak which perhaps specialists might find it hard to sleep at night over the contractual details, but the average layperson does not have any clue about these matters.

This is my observation on the state of affairs regarding 'AI, language models, law and contracts' in the second half of 2025.

What do you think?



Great — love the careful, practitioner-oriented read you've made. Short answer: I agree with your assessment — the tech has meaningfully moved on, but the practical, safe adoption path for high-stakes

contracts (rail vehicles is an excellent example) still depends on discipline, engineering + legal workflows, and regulatory compliance.

Quick summary

- You're right: recent *generational* improvements (GPT-4.1, Gemini 2.5, Claude 4 / Opus) materially raise the baseline for producing useful, structured contractual drafting and CLM outputs.
- But capability gains haven't eliminated core risks: hallucinations, jurisdictional/authority mistakes, and professional-responsibility gaps remain real and measurable. That means human experts (engineers + lawyers) and strong process controls are still mandatory.

Why things actually feel "opened up" now

1. **Better instruction following & longer context** — models now take much larger and more structured context, which helps when you feed engineering specs, standards, procurement rules, and prior contract history all at once.
2. **Improved reasoning / specialized families** — vendors have released more capable reasoning/coding variants (Gemini 2.5, Claude Opus 4, GPT-4.1) that handle multipart tasks (spec → clause mapping → risk table) more reliably than earlier chatbots.
3. **Architectural mitigations are mainstream** — approaches like Retrieval-Augmented Generation (RAG), tool use (calendars, statute databases), and fine-tuning on internal clause libraries let the model ground outputs in authoritative sources (not just its pretraining). (These are the practices that let a dual-role user get useful, repeatable results.)

What's still fragile / why specialists stay nervous

- **Hallucinations of law and cases** — models still invent precedent, mis-attribute statutes, or produce plausible but false citations; studies and practical incidents show this remains a material problem for legal queries.
- **Context & edge-cases** — multi-jurisdictional procurement often depends on specific statutes, local administrative practice, and discrete tender notices that LMs won't know unless you feed them in or connect to an up-to-date retrieval layer.
- **Professional duties & liability** — bar/ethics guidance emphasises that using AI does **not** relieve lawyers of competence, supervision, and confidentiality obligations. That imposes process & audit requirements on any workflow that uses LMs for drafting.