Jędrzej Wydra

Can an Algorithm Speak on Behalf of Justice?
Linguistic Challenges and Perspectives of Artificial Intelligence in the Judiciary*

Let me begin by providing some academic context, which significantly influences the perspective from which I approach this topic. I am currently nearing the end of my incredible journey at a Doctoral School in the discipline of law. However, throughout my academic career, I've had the privilege of engaging in two equally important intellectual adventures: five years studying law and another five years studying mathematics. My primary area of research is forensic science, but I also dabble in formal logic and the mathematical foundations of artificial intelligence. Outside university, in my professional life, I design, build, and apply AI systems.

Until recently, the concept of language was clear to me. I was raised in the tradition of ideas passed down from my mentor (the mathematician Professor Świrydowicz), who was taught by his mentor (the lawyer Professor Ziembiński), who in turn was taught by his mentor — the philosopher Professor Ajdukiewicz, the author of the concept. I understood language as something constructed and functioning according to six rules: 1) rules defining vocabulary, 2) grammatical rules, 3) axiomatic rules, 4) inferential rules, 5) rules defining a universe, and 6) rules of denotation.**

This formal definition is convenient for lawyers, mathematicians, and logicians because it turns language into a tool for achieving specific goals and resonates with our desire to frame reality in strict structures. This view (at least for me) provided a certain sense of security, rooted in the belief that language was a clear-cut issue.

However, the rise of ChatGPT, particularly the GPT-4 model, which responds like a human, forced me to rethink my understanding of language.

GPT stands for "Generative Pre-trained Transformer". Models of this type essentially transform one text into another. That's all. They operate based on complex neural networks, which, contrary to popular belief, have very little in common with the structures in our brains. Conceptually, there are similarities, as both rely on an enormous number of connections. But that's where the similarities end. Neurons in the brain are physical entities that process electrochemical signals. Literally, we could physically count them by opening a human's skull (though I don't recommend it — microscopes are expensive). I've personally seen a neuron under a microscope during my middle school biology class (although I suspect it wasn't a human neuron). But what about artificial neural networks in language models? No one has seen them because they don't physically exist. They are virtual constructs, essentially instructions that tell a computer's processor what calculations to perform and in what order.

Computer neural networks don't process internal or external stimuli the way our brains do. They just generate the most probable sequence of symbols based on the data they were trained on (known as training data) and the input they receive at the time of use. Therefore, the processes governing biological neural networks and computer neural networks are fundamentally and qualitatively different.

Where can we best observe this difference? René Descartes proposed the concept of radical mind-body dualism. Though likely too extreme, Descartes' view offers an interesting lens through which to examine this situation. He identified two independent substances: the material

one and the mental one. The mental substance doubts, understands, imagines, asserts, denies, wills, refuses, and feels. Psychologists often refer to this substance as the soul or the mind. None of these attributes apply to language models.

Now we arrive at the question: Is what a language model generates truly language? According to Ajdukiewicz's formal definition, yes, definietly. But accepting this position impoverishes a concept of language — or even a concept of humanity — because, in Aristotelian terms, it strips matter of its form. The essence of language should be the transmission of mental states. Otherwise, we're left with something that looks like language, behaves like language, but is actually a quasi-language. This situation won't improve because AI based on neural networks will never possess a mind.

The lack of a mind and the generation of the most probable sequence of symbols pose serious risks for the judiciary. One of the most popular AI applications considered by judges is simplifying the drafting of text of reasons of legal judgments — a tedious and often disliked task.

I can't imagine this scenario working. A language model can provide reasons for anything (though the quality of those reasons is another matter). With the right prompts, we could easily generate arguments for contradictory claims. This means I could make an arbitrary decision and effortlessly obtain justifications for it from a language model. In my opinion, this vision dangerously borders on totalitarianism.

A text of reasons of legal judgments must be a linguistic reflection of the judge's mental processes. The decision must be preceded by an analysis of legal norms, an assessment of the facts, and subsumption. The reasoning behind this process must be reflected in the text of reasons, which is supposed to prevent judicial arbitrariness. What will protect us from despotism if a language model begins to create a facade by generating reasons for any statement?

Should we, then, fully entrust judicial powers to language models, thereby eliminating humans, who are prone to abuse, from the process? Absolutely not! While the quality of language models is improving, remember that these models will always remain mere generators of the most probable text.

But is the most probable answer also the right answer?

Furthermore, a functional language model must be trained on vast amounts of data, which, in significant part, comes from the Internet. What percentage of this data is legal text? What percentage is academic? A mere drop in the ocean. Handing any authority to language models would be no different from placing a celebrity with sketchy knowledge in the role of an expert. This is a one-way ticket to troubletown.

The rise of AI presents two additional threats. First, improving model quality requires training on ever-larger datasets. Soon, these datasets will include AI-generated text, which inevitably leads to a decline in quality. In essence, the AI phenomenon is like a snake eating its own tail.

Second, the increasing complexity of language models requires greater computational power. Running AI on local devices is already a challenge, so cloud infrastructure is typically used. This situation is unlikely to improve anytime soon. It means that using AI involves sending data to an

external provider and losing control over how that data is handled, creating significant personal data protection concerns — but that's a topic for another story.

I don't want to be seen as an opponent of AI, because I certainly don't believe lawyers and language models should remain in opposing camps forever. Language is both a tool and an object of study for lawyers. Not all uses of language serve to communicate our internal mental processes. Lawyers often employ language instrumentally. Take notaries***, for example. The notary's role is to protect the interests of both parties and prevent rash decisions in legally significant matters. Generating a draft notarial deed using a language model wouldn't undermine these tasks. The notary could review and approve the document instead of writing it from scratch. This could even improve the notary's service quality by speeding up the process. Especially since the current practice is similar. The first draft of a deed is often prepared by a secretary.

Looking beyond notaries, lawyers across all specialties frequently face the challenge of processing large volumes of text, including many types of documents such as contracts or legal acts. Much of this text is irrelevant to the case at hand, and searching through it is time-consuming. An AI system with semantic understanding and the ability to generate summaries or extract key fragments could significantly enhance a lawyer's efficiency.

So, could we ever see AI replace lawyers in certain tasks? Absolutely. One of lawyers' exclusive skills has been navigating the legal system and understanding the law. If ordinary people had this ability, many simple issues could be resolved without a lawyer's help. The Polish AI system Legitio.pl already offers useful assistance with simple legal matters. It can analyze a basic task and provide fairly good quality advice, which should be sufficient for the vast majority of everyday issues. However, only a human lawyer can recognize when a case description is incomplete and ask follow-up questions. This is why only a human lawyer can take full responsibility for their advice.

I don't know if AI will ever fully replace lawyers in providing legal advice, but in my opinion, this is the most endangered area of their work.

*Essay prepared for the III Polish Conference on Language and Law titled 'Legally – (In)correct' (own translation, originally 'Prawnie – (nie)poprawni'), Poznań, November 18-19, 2024. The text was translated with the support of ChatGPT. The original version can be found here.

**Rules defining vocabulary determine which words belong to a language. Grammatical rules dictate how to construct more complex expressions. Axiomatic rules identify certain sentences as true, independently of other sentences. Inferential rules indicate which sentences are true based on the truth of others. Denotational rules specify what a given expression refers to. Rules establishing the universe define the set of all objects to which the denotational rules apply.

***A notary in Poland is a legal professional who acts as a public official, drafting notarial deeds and certifying the validity of documents and signatures. Their role is to ensure that legal transactions comply with the law and to protect the interests of the parties involved.