How AI is already improving traditional competition analysis
Putting theory into practice

Enrico Alemani, Ivy Kocanova, Adelle Arbo and Toby Howard[1]
25 February 2025

AI is expected to transform the way we work, but exactly how remains unclear. In this article, **Enrico Alemani**, **Ivy Kocanova**, **Adelle Arbo** and **Toby Howard** examine how AI is currently used in competition analysis, outlining the specific ways it has been applied effectively and where human oversight remains essential. They argue that while AI-powered tools can enhance efficiency, they require careful selection and supervision to ensure reliability.

Introduction

AI will transform ways of working, both in the near term and the long term. However, precisely *how* it will transform our work, and to what extent, is not yet clear. In this article, Compass Lexecon's **Data Science team** sets out their experience using AI in practice.

We use four case studies to illustrate when AI is an appropriate tool to use, how to select and tailor the best type of AI tool for the specific problem at hand, and how to ensure and demonstrate that its results are reliable. It requires not only specialised AI models, but expert human supervision.

Infinite interns with typewriters: when is it appropriate to use an AI tool?

Promises that AI will revolutionise ways of working are widespread – and in many industries that may well be a plausible path. However, in industries that require nuanced interpretation and judgement, widespread dependence on AI has not taken place – at least, not yet.

Here, we reflect on our experience of marrying the theoretical promise of AI with its practical deployment in competition law cases to explain where AI is best placed to help.

The appeal and risk of AI "assistants"

Currently, most people are familiar with AI-powered "chatbots" – similar to ChatGPT – that can be used to research and summarise documents and improve drafts.[2] These AI powered tools have an obvious appeal. They have already processed vast amounts of information – whether text or data – and produce articulate and seemingly compelling responses to questions about it comparatively quickly.

However, teething problems with these applications have damaged confidence. They can produce nonsensical answers and hallucinations to fill their knowledge gaps, misleadingly presenting their invented responses as facts.[3] And few of these tools can be given additional confidential information to supplement knowledge gaps.

These risks can deter people from using AI, or from trusting its results, and lead to an understandable tendency to revert to the "tried and tested" ways of doing things. **Figure 1** shows, for instance, that even use of ChatGPT is still relatively limited.

Figure 1: Use of ChatGPT by country, May 2024
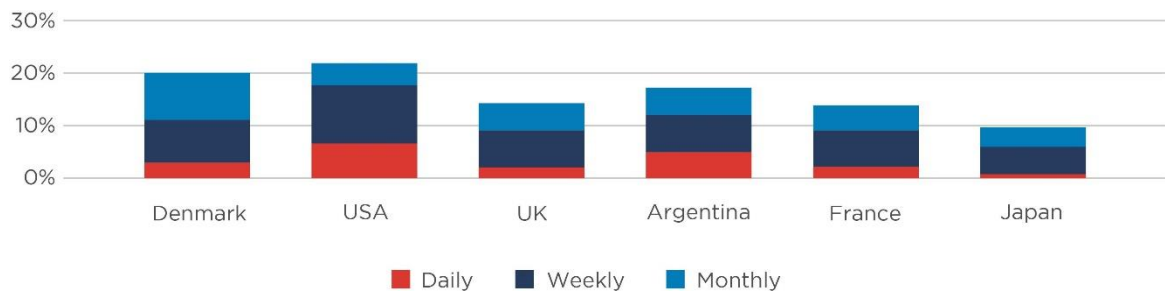
Source: Compass Lexecon analysis based on data from R. Fletcher and R. Kleis Nielsen, "*What Does the Public in Six Countries Think of Generative AI in News*?", Figure 2.

However, the broader advantages of AI are too significant to ignore. It can go beyond enhancing existing processes or fulfilling peripheral tasks; it can tackle new challenges or reimagine solutions to existing problems. Its problems are surmountable if the right tool is selected and supervised in the right way. After all, no system has yet been designed that is perfectly infallible. *Humans* can also be fallible – they have varying degrees of specialist knowledge and varying communication styles; some may even hallucinate or bluff. But in that context, everyone has learnt to manage those risks effectively, by deploying the right skills in the right place, and supervising those teams effectively.

The same principle applies here. There is no single AI tool we can use to reliably solve all problems. Instead, there are many tools, each with their own trade-offs. We have found that, much like human beings, each has a different balance of capabilities, making each one appropriate for different types of tasks. And just like any other type of assistant, how these tools are prepared and supervised for the job at hand by human experience and intelligence is crucial. If one understands the strengths and weaknesses of each tool, outcomes can be drastically improved by deploying them correctly.

What is AI (currently) good at?

The "AI-based tools" we look at in this article primarily consist of **large language models ("LLMs")**.[4] These models excel at handling high-volume, well-defined tasks that follow clear rules and patterns. They are less-effective for nuanced analysis, particularly where context, interpretation, and judgement is required.

The key distinction is that LLMs as currently designed do not use logic in the same way humans do.[5] They are statistical models trained on vast datasets to identify and replicate patterns in text. This enables them to generate human-like responses to prompts and questions without true comprehension. So, a LLM can summarise content because it can identify and replicate the patterns that are common in language, not because it *understands* what the question or its answer is actually saying in the same way a human does.

Hence, AI had been described metaphorically as providing "*infinite interns*".[6] The analogy is intended to get across the concept that LLMs function like a team of very efficient, diligent and inexperienced analysts, with supervision from a "manager". They can contribute where relatively common and simple assessments are required, particularly at scale. They can also contribute where errors are easier to spot and correct (such as coding). However, they require more careful attention and handling where errors are difficult to spot and are important.

Picking from the AI menu: which tool is best suited to the problem?

There are various AI tools available. However, the options vary in three important respects:

- **The number of parameters:** Models trained on billions of parameters are considered "Large" LLMs,[7] meaning they offer broader knowledge and can generate sophisticated, context-aware responses to a wide array of queries. However, they require significant computational resources, are typically slower and are more expensive to run. "Small" LLMs are quicker and cheaper to run, but their capabilities are restricted in their sophistication or breadth of their knowledge.

- **Flexibility:** LLMs can be categorized as proprietary or open source. Proprietary models, such as ChatGPT, tend to be powerful but are inflexible and cannot be customised. In most cases, users cannot access training data, cannot inject additional knowledge and have little to no control over how the model works. In contrast, open source LLMs typically provide greater flexibility, enabling users to access the model's architecture, introduce new data, and fine-tune parameters so that the models can be customised for a specific task.

- **Deployment:** "Off-the-shelf" LLM services are immediately accessible and provide managed infrastructure, simplifying deployment and reducing operational overhead. However, these solutions pose challenges when handling confidential data (as well as limiting flexibility). "Local deployment" offers complete control over data and customisation, but it needs significant technical infrastructure, specialised expertise and ongoing maintenance.

These three dimensions often overlap, so generally, we can refer to two broad types of LLM:

- **General-purpose LLMs:** These are large-scale models, typically proprietary, off-the-shelf solutions. They are designed for broad applications and can handle diverse queries with strong generalization capabilities. They are what most people

have in mind when they think of AI at the moment.

- **Specialised LLMs:** These models are usually small, customisable and locally deployed. They are trained for specific tasks and deployed within your own organisation, making them good for handling confidential information. However, they need technical experts to set up and maintain, have more limited capabilities compared to general purpose LLMs, and may not perform as well on tasks outside their specialised domain.

The general-purpose LLMs, like ChatGPT, are the best known and most widely adopted.[8] However, as with many other tools, the generalist tool is often not the best tool for a specific task. A well-trained and supervised specialist model is usually better than a powerful "allrounder".

Below, we describe four case studies to show how we leveraged specialised LLMs, combined with expert human supervision, to provide evidence that otherwise would not have been possible.

Case study 1: Analysing confidential documents

On a recent merger case, our team had access to thousands of confidential meeting notes from managers, which were used to assess the competitive dynamics of the industry and the closeness of competition.

Without AI, reviewing this information manually would have been prohibitively time consuming and costly. In addition, the subjective classification of results would have been difficult to audit.

Using a general-purpose LLM would also be problematic due to confidentiality reasons. As they are typically proprietary models, we could not upload any confidential data or adjust its training to ensure it performed the task well.

Instead, we built a specialised LLM that could scan these notes to provide both anecdotal

evidence and systematic patterns. Each of its features provided tangible benefits.

- **The model's small scale ensured the analysis was efficient.** It only needed to perform this specific task well; it did not need to be a sophisticated generalist. By reducing the number of parameters, we maintained strong performance, while significantly decreasing the power, costs and processing time.

- **Its flexibility was crucial to ensuring the accuracy of its analysis.** Customising the model's training meant the team could identify specific examples to ensure the LLM could perform the specific task we required. The model would then systematically assess all notes and provide all relevant references for further review. Importantly, here, AI offers an advantage that exceeds what even "infinite interns" could achieve. The results are auditable, verifiable and reproducible, which allows for iterative improvement. We could review samples of the model's judgements and assess how well it was performing. We could identify both "types" of error: the statements it wrongly identified as relevant and irrelevant to the task at hand. We could then rerun the analysis, with the model learning from its previous mistakes.

- **It was also crucial that our model was locally deployed.** That allowed us to review confidential notes without any privacy concerns, as the model was not connected to the internet and the data remained confidentially stored on our internal systems without third parties access.

These models could be equally applied to other sources of free text data, such as bidding datasets or client feedback data.

## Case Study 2: Bespoke AI using specialist knowledge

Retrieval-Augmented Generation (RAG) models are LLMs enhanced by advanced information retrieval techniques. Instead of relying solely on pre-trained knowledge, they dynamically fetch relevant information from external data sources—such as databases, document repositories, or APIs—before generating a response. A key advantage of RAG models is their ability to provide inline citations, allowing users to trace responses back to their source material. This transparency enables users to cross-check information, significantly reducing the risk of hallucinations. By grounding responses in verifiable data, RAG models enhance reliability, making them particularly valuable for applications requiring high factual accuracy, such as research or legal analysis.

We have been integrating this feature to upgrade our Decision tool – a platform which allows users to search through EC and CMA merger decisions for specific terms or concept – with a virtual AI assistant.[9] Now, by simply asking questions to our virtual assistant, users can efficiently search through documents, summarise complex information, and pull out key information from EC/CMA decisions.

The RAG model as shown below (in **Figure 2**) ensures that all information reflects the most up-to-date decisions or regulations.

The AI tool now acts as a chatbot on top of the underlying decision tool. Unlike ChatGPT or other general LLMs, its knowledge is not informed by a broad universe of data but only by the specific technical data provided. This significantly reduces the chances of hallucination, which occurs when a model relies on inappropriate or irrelevant data points. In our model, those irrelevant data points are excluded from the universe of information the model interprets, while source references support that its results are based on factual knowledge.

Figure 2: Schematic of RAG model used to upgrade our Decision tool
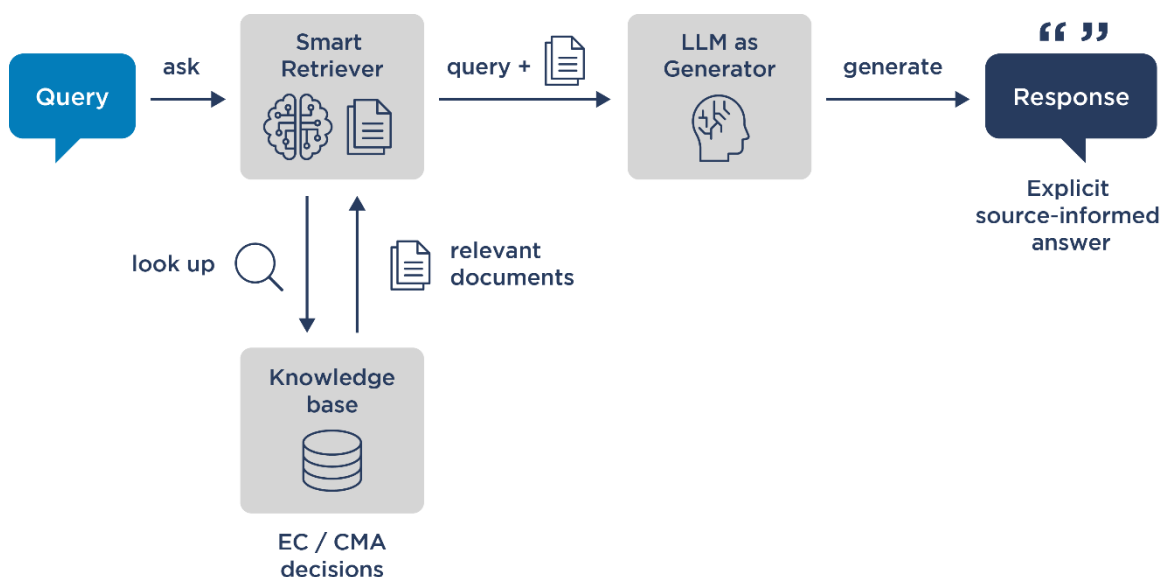
Source: Compass Lexecon

Case study 3: Bespoke AI for specific tasks

We recently prototyped an anonymisation tool: a LLM that would help economists working with confidential client data and legal privilege. The tool can be given a simple instruction, such as "Redact all names, locations, and dates", and it will anonymise the relevant information in the following way in each document.

- **Original**: "John Smith lives in New York and was born on 01/01/1990"

- **Redacted**: "[NAME] lives in [LOCATION] and was born on [DATE].

This tool automates the anonymisation process in three steps:

- **Leverage a Named-Entity-Recognition (NER) model:** An NER identifies general entity types such as names, locations and dates. This type of model is superior to basic text analysis as it helps computers understand important information instead of processing words blindly.

- **Customise the model to meet our specific needs:** An NER model is customisable, and to meet our needs we

train the model to recognise specific concepts, such as company names or market share data.

- **Build an anonymisation/redaction tool**: Each user can tailor the parameters and labels as required for their particular task.

In principle, the task could be done manually, but it would be time intensive and prone to human error. Our tool is comparatively rapid, and its results are straightforward to review and iteratively improve.

Case study 4: Training AI to ensure effectiveness

In a recent high-profile case, an AI model was used to analyse millions of posts on Twitter/X to understand how people perceived an event. Our role was to evaluate the validity of the model's responses.

During our examination, we uncovered significant limitations in how the LLM had been applied. A key issue was that the model's results were highly sensitive to even small adjustments in the instructions it received. We discovered that the model could be easily manipulated to provide favourable responses. Specifically, by altering the example cases in the instructions—used to

illustrate the task and its requirements—the model's outputs changed dramatically.

This volatility in results demonstrates a critical pitfall in LLM deployment: treating models as self-sufficient analysis tools and assuming consistent performance without proper stress testing and rigorous evaluation. Expert human supervision is still essential to ensure reliable outcomes.

In response to these challenges, we developed a rigorous framework to ensure reliable LLM use in legal proceedings. In broad terms, it involved several key phases as shown in **Table 1** below.

Central to this framework is the principle of continuous evaluation and transparency. This involves regular assessment of the model's inner workings, comprehensive documentation of decision processes, and maintenance of clear audit trails. In this sense it is like other analysis used in legal proceedings: it is vital to include sensitivities and to understand the robustness of results to changes in specifications.

By keeping human experts integral to the process and maintaining rigorous evaluation standards, the framework ensures that LLM applications in legal proceedings remain both powerful and accountable.

The future

The future of AI in competition cases is not about full scale automation or relying on a single, all-encompassing general AI. However, it's also clear that ignoring the potential of AI and sticking solely to what is "tried and tested" is no longer an option. The potential of AI is simply too transformative to overlook.

The future lies in harnessing smaller-scale customisable language models paired with rigorous evaluation frameworks. Every new technology comes with trade-offs - there is no one size fits all solution. They each require careful selection, careful training and constant supervision and tinkering. However, when handled with expertise, they unlock possibilities for compelling and novel evidence that would be impossible to access otherwise.

Table 1: Key phases involved in our framework (to ensure reliable LLM use in legal proceedings)

Phase	Core requirement
Problem definition	Clear articulation of the legal task, its constraints and success criteria.
Model selection	Careful matching of LLM capabilities with task requirements, considering domain expertise.
Knowledge injection	Tailored adaptation of the model through fine-tuning on relevant data to ensure the model understands nuances and specific terminology.
Evaluation with human in the loop	Development of comprehensive validation procedures with expert oversight. Systematically test various prompt variations to assess how different inputs influence the consistency of the model's responses. Review answers for potential biases and prejudices.

Source: Compass Lexecon

1   Enrico Alemani is a Principal Data Scientist, Ivy Kocanova is a Senior Data Scientist, Adelle Arbo is an Associate Data Scientist and Toby Howard is a Senior Analyst at Compass Lexecon. The authors gratefully acknowledge the contribution of Mark Robinson and Deiana Hristov. The authors thank the Compass Lexecon EMEA Research Team for their comments. The views expressed in this article are the views of the authors only and do not necessarily represent the views of Compass Lexecon, its management, its subsidiaries, its affiliates, its employees or its clients.

2   https://chatgpt.com/

3   BBC News, "Airline held liable for its chatbot giving passenger bad advice - what this means for travellers", 23rd February 2024. See https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know

4   These models are presented with such a vast amount of text that they can predict what the next word in a sentence is likely to be based on the surrounding context. In turn, when asked a question, they can predict what the next word in the sentence should be and hence construct coherent answers.

5   Although recent iterations of OpenAI models increasingly deploy a "chain of reasoning" approach to answers which seeks to mimic human reasoning. See https://openai.com/index/learning-to-reason-with-llms/

6   Forbes, "Why AI Is Like 'Infinite Interns'", Nov 30, 2023, https://www.forbes.com/sites/meganpoinski/2023/11/30/why-ai-is-like-infinite-interns/

7   A rough rule of thumb at the time of this article is that seven billion parameters currently constitutes a large LLM.

8   R. Fletcher and R. Kleis Nielsen, "*What Does the Public in Six Countries Think of Generative AI in News?*", Figure 1.

9   https://cl-decision-tool.com/