alexi

Understanding AI Legal Research

October 23, 2025

Katie Sharp
Product Manager, Alexi

Who am I?

- Currently, I am the **product manager** at Alexi.
- Trained as a **lawyer**, I sit at the intersection of law, AI, and product.
- Helped develop Alexi's legal research technology - which delivers **reliable answers** to complex legal questions.
- Great appreciator of the work **librarians** do

Table of Contents

01

AI Primer

Extractive AI

- Analyzes and **extracts information** from high volumes of data sources
- Each output ties back to source data

Generative AI

- Creates **content** based on direct prompts
- Best for single tasks (drafting, summarizing)
- Requires human direction for each step

Agentic AI

- Executes **multi-step workflows** autonomously
- Achieves a larger objective with less human intervention
- Requires clear goals + oversight

Large Language Models (LLMs)

A prediction machine

- Trained on patterns, not reasoning
- Generates text it predicts you want
- Not a database and not authoritative

Natural Language Processing (NLP)

How AI reads and works with language

- Converts human language into a format machines can interpret
- Identifies patterns in words, meaning, and structure
- Foundation for chatbots

Understanding Tokenization

What are Tokens?

Tokens are the building blocks of how LLMs process data.

When you interact with a LLM:

1. Your text is split into tokens.
2. The model processes these tokens.
3. The response is generated as a sequence of tokens, then converted back to text.

Helpful rules of thumb for English

1 token ≈ 4 characters

1 token ≈ ¾ of a word

100 tokens ≈ 75 words

1-2 sentences ≈ 30 tokens

1 paragraph ≈ 100 tokens

Understanding Context Windows

A context window defines the **maximum number of tokens** an LLM can process in a single request including both the input and the generated output.

Essentially, it is the LLM's **working memory.** It determines the length of conversation possible before the LLM starts to forget details from earlier in the exchange.

02

Current State of Legal AI Research: Vals Report (US)

Vals Legal AI (US)

Methodology

This study extended a previous study (VLAIR) and evaluated four AI products on legal research tasks, studying their ability with respect to a lawyer control group (the Lawyer Baseline).

The study assessed the participants' ability to respond to 200 U.S. legal research questions.

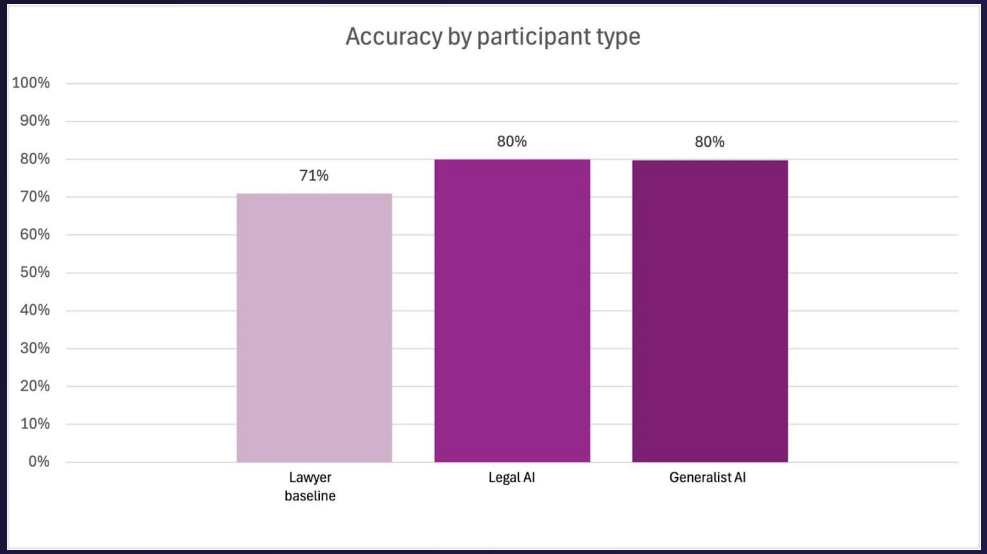
The responses were evaluated for **accuracy, authoritativeness,** and **appropriateness.**

Accuracy by Participant Type

1. Both legal AI and generalist AI can produce highly accurate answers to legal research questions

2. Currently, sources and citations are the differentiators for legal AI.

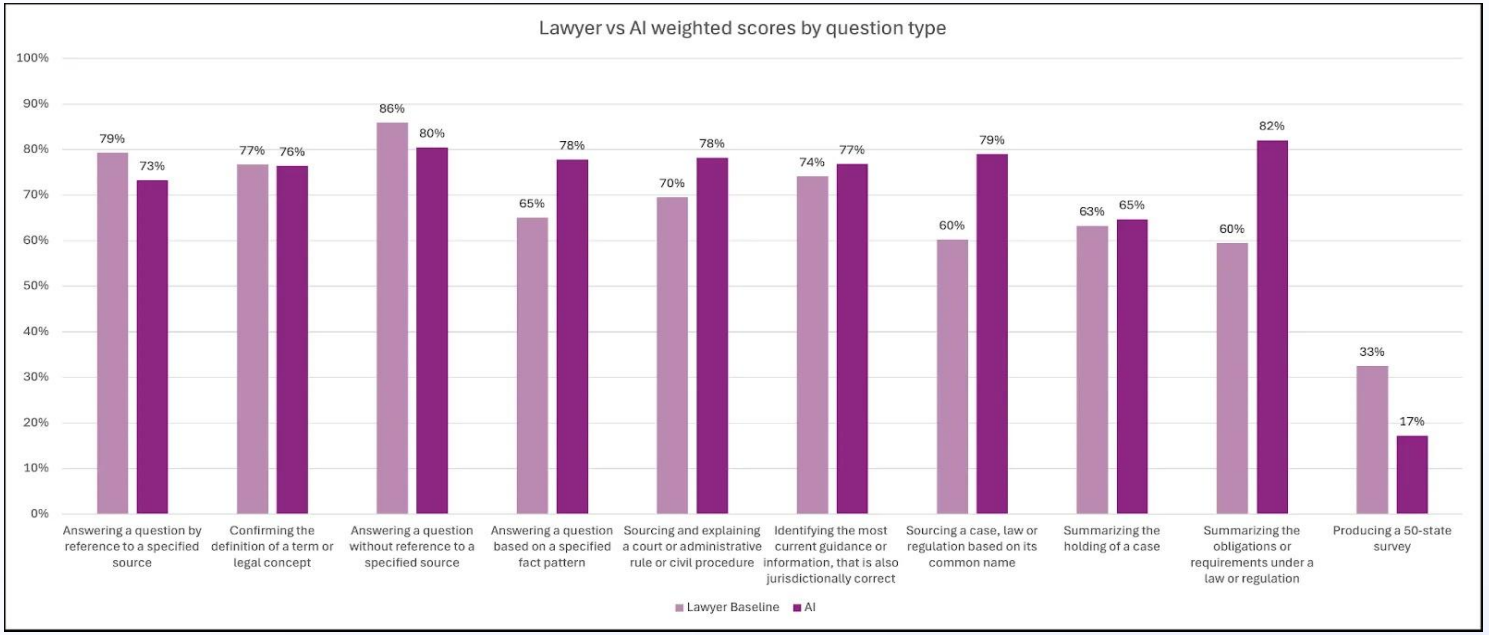
3. Specialized legal AI tools were more likely to decline to answer a question where it deemed there to be a lack of available source data. ChatGPT would rely on its knowledge to provide an answer to the question.



Source: https://www.vals.ai/industry-reports/vlair-10-14-25

Where lawyers outperform AI

There remain many **capability gaps** that lawyers are best placed to fill.



However, there is still a substantial risk of **inaccuracy** and **hallucination** with generative AI.

370

Reported instances globally of practitioners using AI to prepare court submissions that have been found to contain hallucinated citations and/or incorrect legal analysis

$100-$350+

Penalties Canadian lawyers have faced for relying on materials in Court.

Hallucinations

- Off-the-shelf LLMs hallucinate *at least 58% of the time* on legal fact questions
- Models often do not recognize their own mistakes or correct user false assumptions
- Source: Matthew Dahl et. al., “Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models (2024) arxIV:2401.01301

03

Working with Lawyers who use AI

Use of Gen AI Among Canadian Lawyers

- The number of lawyers who use AI rose from 14% to 26% from 2024 to 2025
- Another 12% said they plan to start using it within the next year
- Source: Thomson Reuters 2024 State of the Canadian Law Firm Market report

Verification Checklist

- Confirm source
- Check currency
- Check jurisdiction
- Confirm citation
- Note up

How Librarians Can Guide

- Steer tool selection
- Encourage verification
- Teach grounding principles & best practices
- Be the AI literacy anchor

04

How to Improve Results with AI: Context & Prompt Engineering

Prompt v. Context: What's the difference?

Prompt Engineering

The art of providing the right instructions to an LLM.

Often a single, well-written prompt.

Context Engineering

Controlling the system in which the well-written prompt operates.

Context includes user preferences, documents, custom-built tools, memory.

Techniques to Improve Your Prompts

- **Self-Criticism:** Ask the model to improve its own output or, if you can, ask another platform or LLM to criticize the output.

- **Provide Examples (Few Shot):** can you include in your prompt examples of narrative style or judgment calls to show it what to do?

- **Chain-of-Thought:** Ask the LLM to explain its reasoning to you.

> Context engineering is discipline of designing and building dynamic systems that provides the right information and tools, in the right format, at the right time, to give a LLM everything it needs to accomplish a task.

Phil Schmid | Senior AI Relation Engineer at Google DeepMind

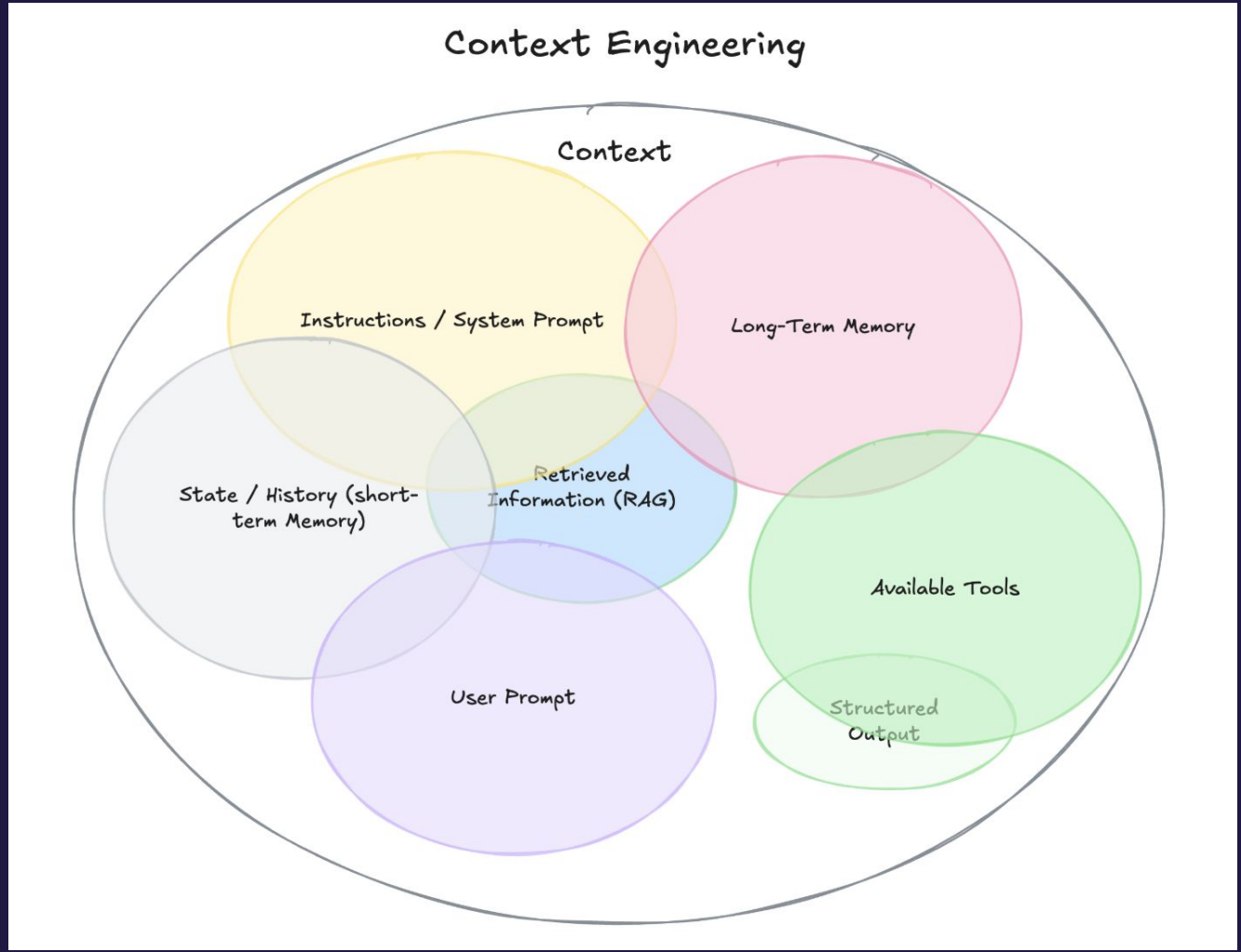


Image from philschmid.de/context-engineering

What problem does thinking about context solve?

1. Information overload
2. Relevancy
3. Consistency
4. Scalability

> Context engineering is the delicate art and science of filling the context window with just the right information for the next step

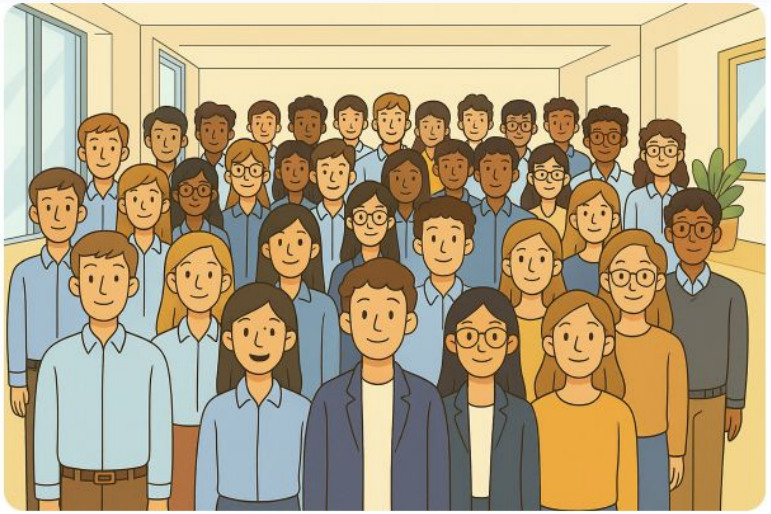
Andrej Karpathy

Best Practices for Context Engineering

- Instruct the model on what tool you want it to use (and **not** use)
- Experiment with different prompt instructions
- *Ask yourself*: can this task be broken into smaller subtasks or will that add unnecessary context/potential for conflict?
- Be specific in what you want the task to achieve
- Tell it how to handle ambiguity or missing data

Breakdown the desired output into tasks

Imagine you have a 100 employees at your beck and call. But none of them are great at listening.



alexi

Thank you

Closing statement