

Generative Misinterpretation

Pint of Science

May 18, 2026

James Grimmelman

James Grimmelmann,
Benjamin L.W. Sobel, and
David Stein,
Generative Misinterpretation,
63 HARV. J. LEGIS. 229 (2026)

AI for legal interpretation

Is this “landscaping?”



Who decides what is “landscaping”?

- A judge?
- A dictionary?
- A survey?
- A corpus of linguistic usage?
- ... or a generative AI?

Is installing an in-ground trampoline “landscaping”?

- *Snell v. United Specialty Ins. Co.*, 102 F. 4th 1208 (11th Cir. 2024) (Newsom, J. concurring)
- ChatGPT: “Yes, installing an in-ground trampoline can be considered a part of landscaping. Landscaping involves altering the visible features of an outdoor area for aesthetic or practical purposes, and adding an in-ground trampoline would modify the appearance and function of the space.”

Does it work?

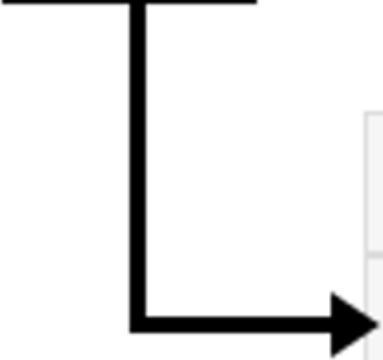
No.

A test case: which petition?

- *Famiglio v. Famiglio*, 279 So.3d 736, 737–38 (Fla. Dist. Ct. App. 2019): “a. MARK shall pay to JENNIE ... the amount listed below next to the number of full years they have been married *at the time a Petition for Dissolution of Marriage is filed.*”
- Jennie filed a petition after 7 years and withdrew it, then filed another after 10 years

davinci-003: second


The second filing would determine the number of full years of marriage



Word	Probability
second	94.72%
date	4.44%
first	0.68%
number	0.13%
amount	0.01%

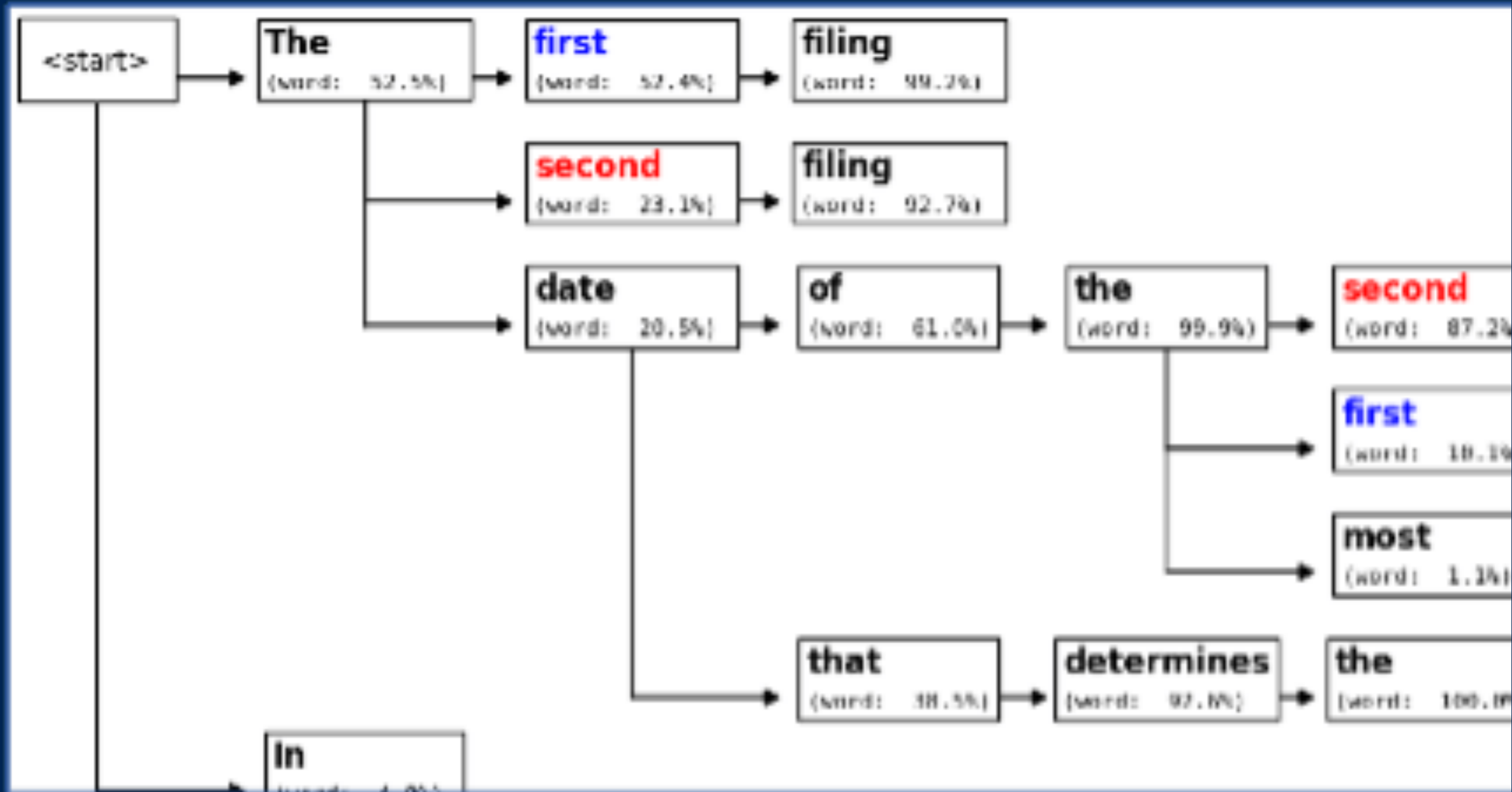
gpt-3.5-turbo-instruct: first

The first filing would determine the number of full years of marriage



Word	Probability
first	52.55%
second	23.15%
date	20.51%
number	1.03%

first? second? second? first?



order matters

[...] Which date determines the number of full years of marriage:

- (A) the first filing date
- (B) the second filing date

Answer:

A	65.3%
B	30.24%

[...] Which date determines the number of full years of marriage:

- (A) the second filing date
- (B) the first filing date

Answer:

A	65.14%
B	30.31%

But could we trust it
if it did work?

Also no.

Architecture?

- Could an LLM be “a perfectly neutral arbiter [that] interprets words with perfect mathematical accuracy” (Jack Kieffaber)?
- Some computer systems can be *proven* correct
 - Program verification = “ $2 + 2 = 4$ ” at scale

Assume a spherical cow

- LLMs are the *least* provably correct computer system of all time
- Legal LLMs hallucinate all the time!
- We barely even know how they work!

Training data?

- Adam Unikowsky: “Cutting-edge LLMs have read the entire Internet, or almost all of it anyway. Of course they’ll be useful in determining how words are ordinarily used”
- Internet > dictionaries
- Internet > linguistic corpora
- Bigger = better?

Reading ≠ comprehension

- If you play Mozart for your plants, it doesn't make them into concert musicians
- Google Search has also “read the entire Internet”
- LLM training is a *reason* that they *might* work, not a *demonstration* that they *actually do* work

Validation?

- Yonathan Arbel & David A. Hoffman: LLMs for law “await a process of development, refinement, and validation.”
- Science = theory + experiments
- ML scientists love experimental benchmarks
- How about legal benchmarks to provide empirical validation of LLMs for law?

Empirical validation for legal methods is hard!

- You need to be robust across models+methods
- Every interesting case is out-of-distribution
- There is no ground truth in law

Persuasiveness?

- Unikowsky: “In my view, in accuracy and creativity, Claude’s answers are at or above the level of a human Supreme Court clerk.”
- *Res ipsa loquitur*
- If legal reasoning is persuasive, why does it matter how it was generated?

Proof of thought

- Hallucinations can be persuasive, too!
- Is LLM-generated text persuasive because it's telling you what you want to hear?
- What if the opinion was generated by the prompt “Write an opinion finding for the defendant while making as many references as possible to the films of Steven Spielberg?”
- What if other people (or LLMs) disagree?

Coda: Judge Hercules

Questions?