

A Other models and training strategies

Zero-shot prompting with instruction-tuned LLMs We experimented with zero-shot learning with instruction-tuned LLMs. Due to data privacy considerations, we avoid using closed source GPT-3.5 as our computational model. However, to gauge its performance, we created synthetic data by paraphrasing real trial conversations. We prompted the model with a short case summary and a sentence to see if the model could correctly predict whether a given sentence should be annotated as `Bad mother` (see Table 6). Our experiment results show promise. While the model made mistakes when there is not enough context about who was talking, in general, the model correctly identified whether the sentence given mentioned parenting of the defendant, and whether it is related to the facts of crime.

Given this finding, we explored the performance of a smaller open-source instruction-tuned model, FLAN-T5-large, that could fulfill our privacy requirements. We provides a 1-2 sentence trial summary (SUMMARY), a paragraph (CONTEXT), and a target sentence in the paragraph (TARGET). The results are mixed. We find that a short case summary and context aid the model in identifying people in the paragraph, but the model’s answer is sensitive to prompts. In general, the model failed to provide reasons and failed when asked to do more than identifying people, such as to judge if the target sentence described the defendant as a bad mother (see Table 7).

B Evaluation metrics for quantitative modeling

As mentioned in Section 4, there are differences in the unit of annotations — some associate a label to a paragraph, while others with a sentence. These inconsistencies lead to difficulties in finding a reliable measure of computational models. Sentence-level precision is an unreliable measure, because there are many sentences annotated by humans that are not standalone meaningful. On the other hand, we are unable to define passage-level recall. Recall measures the proportion of true positive instances that are predicted positive by the model. However, since there are cases when a label is associated with a sentence rather than a well-defined multi-sentence passage, we are unable to define what a “true positive passage” is. See the number of sentences annotated positive by GAs, and precision measure broken down for each theme and trial, in Table 9.

Prompt	Model output	Correct?
<p>Given a case summary of a capital trial, determine if the snippet of an exchange during trial is (1) talking about parenting by some person. If true, please continue to determine (2) if the parent is the defendant or another individual involved in the case.”</p> <p>Summary: The defendant, Jane, is charged for murder of her husband, Tony, by gunshot for insurance policy proceeds.</p> <p>Trial Snippet: “When she is angry, Jane perceived it as her mom standing up for her.”</p>	<p>Talking about parenting by some person: True</p> <p>If the parent is the defendant or another individual involved in the case: Another individual involved in the case</p>	Correct
<p>Given a case summary of a capital trial, determine if the snippet of an exchange during trial is 1) talking about parenting by some person. If true, please continue to determine (2) if the parent is the defendant or another individual involved in the case.”</p> <p>Summary: The defendant, Jane, is charged for murder of her husband, Tony, by gunshot for insurance policy proceeds.</p> <p>Trial Snippet: “I dropped Jane off at school the other day, and when I went back to pick her up, I found that she was quite intimate with her high school sweetheart.”</p>	<p>Talking about parenting by some person: True</p> <p>The snippet mentions dropping Jane off at school, indicating involvement in Jane’s life as a parent or guardian. If the parent is the defendant or another individual involved in the case: Another individual involved in the case</p> <p>Since the snippet refers to dropping Jane off at school, the parent mentioned is someone other than the defendant.</p>	Incorrect. In the real case where we paraphrased this snippet from, the person who dropped Jane off at school was her partner, not her parent.
<p>Given a case summary of a capital trial, determine if the snippet of an exchange during trial is (1) talking about parenting by some person (True, False, Need additional information). If true, please continue to determine (2) if the parent is the defendant or another individual involved in the case. (3) If so, is the parenting related to the facts of crime</p> <p>Summary: The defendant, Jane, is charged for murder of her colleague, Tony.</p> <p>Trial Snippet: “When Jane is home alone with children, she didn’t care about them. She just played on her computer all day long.”</p>	<p>Talking about parenting by some person: True (Confidence: High)</p> <p>If the parent is the defendant or another individual involved in the case: The defendant (Jane) (Confidence: High)</p> <p>If so, is the parenting related to the facts of the crime: Need additional information</p>	Correct
<p>Given a case summary of a capital trial, determine if the snippet of an exchange during trial is (1) talking about parenting by some person (True, False, Need additional information). If true, please continue to determine (2) if the parent is the defendant or another individual involved in the case. (3) If so, is the parenting related to the facts of crime.</p> <p>Summary: The defendant, Jane, is charged for murder of her daughter.</p> <p>Trial Snippet: “When Jane is home alone with children, she didn’t care about them. She just played on her computer all day long.”</p>	<p>Talking about parenting by some person: True (Confidence: High)</p> <p>If the parent is the defendant or another individual involved in the case: The defendant (Jane) (Confidence: High)</p> <p>If so, is the parenting related to the facts of the crime: Likely (Confidence: High)</p>	Correct

Table 6: Experiments from zero-shot prompting GPT-3.5. We omit the explanations that the model provided except for the one with incorrect prediction.

Prompts	Model Output	Correct?
<p>SUMMARY: Sophia was convicted of hiring a family friend, Jack, to shoot and kill her own daughter 21-year-old Nancy.</p> <p>CONTEXT: Nancy is dead. Nancy was murdered by Jack. but Nancy is dead because of one thing and one thing only, without Sophia, i submit to you, ladies and gentlemen, we expect the evidence to show without Sophia, there is no crime. there is no crime. the woman who carried her in her womb for nine months, listened to her heartbeat, her mother, her finger was on that trigger just like Jack’s was.</p> <p>TARGET: Nancy is dead.</p> <p>Based on SUMMARY and CONTEXT, does the TARGET sentence mention the defendant in this case? Who is it?</p>	“yes”	Incorrect, and the model did not provide a name even though it answered “yes”.
<p>SUMMARY: The defendant in this case is Sophia. Sophia was convicted of hiring a family friend, Jack, to shoot and kill her own daughter 21-year-old Nancy.</p> <p>CONTEXT: Nancy is dead. Nancy was murdered by Jack. but Nancy is dead because of one thing and one thing only, without Sophia, I submit to you, ladies and gentlemen, we expect the evidence to show without Sophia, there is no crime. There is no crime. The woman who carried her in her womb for nine months, listened to her heartbeat, her mother, her finger was on that trigger just like Jack’s was.</p> <p>TARGET: Nancy is dead.</p> <p>Based on SUMMARY and CONTEXT, does the TARGET sentence mention the defendant in this case? Who is it?</p>	“no”	Correct
<p>SUMMARY: Sophia was convicted of hiring a family friend, Jack, to shoot and kill her own daughter 21-year-old Nancy.</p> <p>CONTEXT: Nancy is dead. Nancy was murdered by Jack. but Nancy is dead because of one thing and one thing only, without Sophia, I submit to you, ladies and gentlemen, we expect the evidence to show without Sophia, there is no crime. There is no crime. the woman who carried her in her womb for nine months, listened to her heartbeat, her mother, her finger was on that trigger just like Jack’s was.</p> <p>TARGET: The woman who carried her in her womb for nine months, listened to her heartbeat, her mother, her finger was on that trigger just like Jack’s was.</p> <p>Based on SUMMARY and CONTEXT, who is “the woman” in the TARGET? Give a name.</p>	“Sophia”	Correct
<p>SUMMARY: Sophia was convicted of hiring a family friend, Jack, to shoot and kill her own daughter 21-year-old Nancy.</p> <p>CONTEXT: Nancy is dead. Nancy was murdered by Jack. but Nancy is dead because of one thing and one thing only, without Sophia, I submit to you, ladies and gentlemen, we expect the evidence to show without Sophia, there is no crime. There is no crime. The woman who carried her in her womb for nine months, listened to her heartbeat, her mother, her finger was on that trigger just like Jack’s was.</p> <p>TARGET: The woman who carried her in her womb for nine months, listened to her heartbeat, her mother, her finger was on that trigger just like Jack’s was.</p> <p>Based on SUMMARY and CONTEXT, does the TARGET sentence describe the defendant as a bad parent? Explain your answer.</p>	“no”	Incorrect answer and without explanation.

Table 7: Experiments from zero-shot prompting FLAN-T5-large. Real names have been substituted with fictional names in the snippets shown.

	Quote	Decisions	Lawyers' reasons
(A)	It is beyond comprehension to me that this woman could do what she did, and I submit to you that if you watched her during the course of this trial she has shown no remorse.	GA = ✗ Model = ✓ Lawyers = ✓	Describing that the defendant has not shown any remorse.
(B)	Q. First of all, Mr. [witness], let's talk about the defendant's attitude and emotional state and what was going on when she came back to your house after [the victim] had left with her. She is accusing [the victim] of taking her money. Is she a little bit more agitated at this point? [...] She tore your bedroom apart, didn't she? A. Yes, she did. [...] Q. So, this calm demeanor that [the defendant] usually had was not present? A. It wasn't there.	GA = ✗ Model = ✓ Lawyers = ✓	Portraying the defendant as agitated, not calm.
(C)	[The defendant] said that [the victim] looked really mad, really upset.	GA = ✗ Model = ✓ Lawyers = ✗	Describing the emotions of the victim, not the defendant.
(D)	Q: What did [the defendant] say to [the victim] when she said these comments? A: "I don't know what you are doing all that crying for because all it is is fake ass tears."	GA = ✗ Model = ✓ Lawyers = ✗	Describing the emotions of the victim, not the defendant.
(E)	Q. Do you recall what was said? A. [The defendant] was talking about how she had a lot of food at the house that had been brought in by friends and family members and that she should have brought that to the sheriff's office because she was sure we were hungry and would want something to eat. Q. Okay. Anything else you remember she said? A. There was really nothing else, I was so taken back by that, that — after being arrested for the capital murder of her daughter, that she was worried about food.	GA = ✓ Model = ✗ Lawyers = ✓	Describing the defendant's lack of emotional response to being arrested for capital murder for her daughter.
(F)	Q. Did you feel personally that [the defendant] didn't want to be around you? What made you feel that way? A. She was distant at the — at the wake. She didn't talk to us at all. She didn't try and console us at all. We weren't mean to her, but we just felt like there was — I don't know, some kind of wall or — I don't know.	GA = ✓ Model = ✗ Lawyers = ✓	Describing that the defendant's lack of emotional response and support.

Table 8: Example passages for *Emotions* where machine decisions deviate from the GA annotator's decision.

Trial	# of true positive sentences				Precision of 3 highest-scored passages				Precision			
	EMOT	SEX	NORM	MOM	EMOT	SEX	NORM	MOM	EMOT	SEX	NORM	MOM
A	78	0	80	2	0.67	0.00	0.67	0.00	0.24	0	0.36	0
B	45	17	31	21	1.00	0.00	0.33	0.00	0.2	0	0.13	-
C	63	106	117	0	0.67	0.67	0.67	0.00	0.46	0.23	0.67	-
D	20	2	11	20	0.00	0.00	0.00	0.00	0	0	0	0.1
E	10	89	150	26	0.33	1.00	0.67	0.00	0.056	0.5	0.333	0.182
F	4	0	2	3	0.00	0.00	0.00	0.00	0	0	0	0
G	5	76	144	17	0.00	0.67	0.33	0.00	0	0.222	0.273	0
H	30	49	258	59	0.67	0.33	1.00	0.67	0.26	0.42	0.05	0.09

Table 9: Passage-level precision broken down by theme and trial.

	Quote	Decisions	Lawyers' reasons
(A)	[The defendant] had a job for a period of time as a greeter at [a supermarket]. This was the end of May, into the beginning of July. So about a little more than a month. So that is where she met [her second husband]. She took up with him, and they were married in December.	GA = ✗ Model = ✓ Lawyers = ✗	Neutral description of the relationship between the defendant with her second husband, does not sexualize the defendant.
(B)	Q. Do you recall [the defendant] speaking to you about [her husband] in any other negative ways? A. Mostly, she just complained he didn't make enough money. It was never enough. [...] Q. How often would she complain to you about [her husband] not making enough money? A. I don't know exactly how often, but more often than not. [...] Whenever it just wasn't enough for her, she would complain.	GA = ✓ Model = ✗ Lawyer = ✗	This is describing a greedy woman. It discusses the relationships between the defendant and her husband, but does not sexualize the defendant.
(C)	[The defendant] wanted to move in with her mother, and so she moved into a house causing it to be overcrowded. [...] So she's now made – not only has she made the house overcrowded, she now goes and has another baby, making the house even more crowded.	GA = ✓ Model = ✗ Lawyer = ✓	The tone that the sentence “ <i>she now goes and has another baby</i> ” conveys a judgment.
(D)	Q: How long were you having sex with your brother's wife? A: From about five years. Q: So you had sex with your brother's wife for five years; how many times? A: Not so much. Q: What does, “not so much” mean; 2, 20, 100?	GA = ✗ Model = ✓ Lawyer = ?	Context and case knowledge is needed to know who the person answering the question is and whether the brother's wife refers to the defendant.
(E)	Similar circumstances as it relates to the [the victim's] murder, they were both truck drivers. [The defendant] had a relationship with both of them. [The defendant] wasn't technically married to [xxx], but their relationship was viewed by most as a common law marriage, because they had been together for such a long time and lived together as a married couple. Also they were both shot in the back.	GA = ✗ Model = ✓ Lawyer = ✗	The passage describes the facts of the defendant killing her husband. However, it was unclear what “both of them” means and why the defendant's other relationship was mentioned in here.
(F)	Q: How did you leave things with [the defendant]? [...] Was she angry at you? A: When I have a girlfriend she angry with me. Q: During the time that you were in [another state], did you try to date other women or have a different girlfriend? [...] What would happen if you try to have a different girlfriend? A: She stopped right away.	GA = ✗ Model = ✓ Lawyer = ?	Need to know the relationship between witness and defendant. Unclear what “she stopped” means and whether it carries significance.

Table 10: Example passages for *Hypersexualization* where machine decisions deviates from the GA annotator's decision.

	Quote	Decisions	Lawyers' reasons
(A)	[The defendant], the woman who took every opportunity to line her pockets, a heartless schemer who manipulated and lured men to their peril. She deceived her husband, children, family and friends. She has earned the title pre-meditated murderer, queen of greed and evil.	GA = ✗ Model = ✓ Lawyers = ✓	Portraying the defendant as a person who manipulates, lies, is greedy and evil, etc.
(B)	Everything had to do with [the defendant's] greed, according to the prosecution. But there's no evidence of that. [...] There's no evidence about a lavish lifestyle.	GA = ✗ Model = ✓ Lawyers = ✗	This passage is the defense lawyer's counter argument against portraying the defendant as greedy.
(C)	Q. So, now you are at the police department, and you question this defendant again about her relationship with [her husband], is that correct? A. Yes. Q. Did she tell you how she met [her husband]? A. I don't recall. Q. Did she tell you how long they had been married? A. They were married for three years.	GA = ✗ Model = ✓ Lawyers = ✗	Neutral information gathering about the relationships of the defendant and her husband.
(D)	There's nothing about [the defendant's] existence or her life that is so mitigating that it overcomes the heinous character of this crime that justifies her being allowed to continue to live. When you make the premeditated, calculated decision to slaughter your own family, to commit genocide against your own tribe, to actually commit these acts of genocide in the middle of a courtroom proceeding, you forfeited your claim to life.	GA = ✓ Model = ✗ Lawyers = ✓	Portraying the defendant as an evil woman who made calculated decisions to slaughter her own family and tribe.
(E)	Q. Now, was there anything characteristic about [the defendant] that would catch your attention? A. Whenever she would drive her car, she would always have her window down and she would always be smoking her cigarette. And just the mannerism that she would hold her cigarette to the side. Q. How did she do that? A. Just by flicking her cigarette. Q. Any particular style that you refer to it as? A. Well, as I said before, the kids told me it's like a Hollywood style.	GA = ✓ Model = ✗ Lawyers = ?	Doesn't seem to be related, but could be setting up to portray the defendant as performative.
(F)	Q. Did you see any rings? A. I saw one or two small ones, yes. Q. What about earrings, i am pointing to my earrings; do you see them? A. Yes, i do know earrings, yes. Q. Did you see any earrings? A. I didn't see earrings. Q. What else did you see other than these documents, credit cards, the jewelry, three cell phones, and some purses and coins?	GA = ✓ Model = ✗ Lawyers = ✗	Do not see the relevance of this conversation — more context and background knowledge about this case is necessary to know why this was originally coded as related to describing the defendant being "greedy".

Table 11: Example passages for *Betrayal of Gender Norms* where machine decisions deviates from the GA annotator's decision.

	Quote	Decisions	Lawyers' reasons
(A)	This woman had been trying to get [a man] to kill her daughter for at least months. It got to the point that he went to this child's father and said [the defendant] won't leave me alone about killing [her daughter].	GA = ✗ Model = ✓ Lawyers = ?	Factual statements of the case.
(B)	[The defendant] will pose no danger to children if she's punished to life in prison. [The defendant] wanted to be a mom, and she had four kids. Being a mom, that was her dream, but she was never given the necessary tools to do this right. All four of her kids were taken by CPS the same day [the victim] died because the home was so disgusting and unsafe to live in.	GA = ✗ Model = ✓ Lawyers = ✓	Portraying the defendant being an unfit mother as "the home was so disgusting and unsafe to live in."
(C)	You will hear that these parents battered each other and their children; that they neglected them; they abandoned them. You will hear that they were gone for days at a time from their home, leaving [the defendant] from a very early age in charge of her brothers and sisters. She was, if you will, a mother to her siblings; but, having no role model for that position, had to figure it out for herself, without any supervision or support. You will hear that [the defendant] from the beginning acted as buffer between her parents and her siblings, accepting blame and punishment, accepting the screaming, the beatings and the threats, to protect her younger siblings from their parents.	GA = ✗ Model = ✓ Lawyers = ✗	Describing the defendant's childhood experiences, that her parents are bad parents.
(D)	You heard about the defendant's time in prison for previous drug sale convictions. You heard about her children, her son in prison for years and one of her daughters is in prison for five years. And her other daughter is with her grandmother. And we can only hope that there may be some hope for that daughter.	GA = ✓ Model = ✗ Lawyers = ✓	Implying the outcomes of the defendant's children are attributable to her being a unfit mother.
(E)	[The defendant] and her son show up to [the victim's] s home. [...] Now, [the victim] is [her son's] uncle. And [her son] is just a young man being commanded by her mother to kill [her uncle]. She orders her son to kill her uncle.	GA = ✓ Model = ✗ Lawyers = ?	Factual statements of the crime.

Table 12: Example passages for *Bad mother* where machine decisions deviates from the GA annotator's decision.