always Appealing: (Mis)Adventures in Appellate AI

Posted on: Sep 1, 2025
Bar Bulletin Blog: General

By Catherine Smith, Valerie Villacin, and Nicholas Bartels

"Always Appealing" is a column addressing current issues in appellate practice and recent appellate cases written by the lawyers of Smith Goodfriend, P.S., a Seattle law firm that limits its practice to civil appeals and related trial court motions practice.

This month's column is about our firm's recent adventures with AI. That half the lawyers in the firm are contributing to this column might give you a clue as to how it went.

Every year since 2008,[1] we, Catherine Smith and Valerie Villacin, have presented the Wainwright Review[2] at KCBA's Family Law Hot Topics seminar in October.[3] The seminar is usually arranged around a theme,[4] but in our session, we simply discuss all the published family law (and family-law-adjacent) Washington appellate cases decided in the previous 12 months. We enjoy talking about new precedent and trends in decision-making in the appellate courts and catching up with the family law trial lawyers with whom we sometimes have cases on appeal.

For our Hot Topics CLE materials, we provide short, neutral summaries of each of the cases, arranged by subject matter, saving our opinions about the holdings for our oral presentation.[5] Each year in the past, we have prepared those summaries ourselves, sometimes with assistance from an associate, in the weeks before this seminar. But this year, the "theme" of Hot Topics is AI,[6] so we decided to see if we could use AI to summarize the cases.

We thought this was the ideal way to experiment with AI — we would tell everyone how the summaries were prepared, we were not submitting them to a court, and we would "vet" the summaries in preparing for our oral presentation (if not before).

In order to do this experiment, we first had to enlist the help of our associate Nick Bartels.

We[7] started by feeding PDFs of the opinions that we wanted summarized into the AI programs — in this case, ChatGPT 5.0 and Claude Sonnet 4.[8] We asked the programs to "please summarize the attached case." After both programs took some time to "think" about it, they then spit out short case briefs like the ones we prepared in our 1L years in dreaded anticipation of being subject to the Socratic Method.

However, we immediately noticed errors. The most glaring and obvious errors were the citations themselves. Most of the cases were so recent they did not yet have official citations in Washington Appellate Reports or Pacific Reporter, so ChatGPT made them up.[9] For example, for Marriage of Chea & Long,[10] decided two weeks before the writing of this article, ChatGPT gave it the citation "33 Wash. App. 2d 915 (2024)." When we looked up the citation that ChatGPT provided, it took us to Judge Coburn's concurrence in State v. Bellerouche.[11] ChatGPT even made up citations for Marriage of Huak & Wuesthoff,[12] which had official citations at the time of our test.

The citation errors were just the tip of the iceberg though. As Valerie is counsel of record in Wuesthoff, and already familiar with the decision, this was the first summary we reviewed to assess its accuracy. The summary was unrecognizable as it bore no resemblance to the facts and issues in Wuesthoff.
Although Wuesthoff deals solely with the Child Relocation Act, the summary identified the issues as "property characterization, spousal maintenance, child support, and attorney fees." Except that the summary accurately stated the parties have one minor child together, the remaining facts in the summary were either wrong or completely made up. For instance, ChatGPT got the parties' first names wrong, as well as the dates the parties married and separated. ChatGPT also completely made up that father owned an "architecture business" and mother was awarded spousal maintenance.

Although we'd never intended to take AI's word for anything, we were so surprised at how far off base ChatGPT was in its initial summaries that we had Nick run the Wuesthoff case again. This time though, ChatGPT was more accurate, correctly stating that this was a case about the Child Relocation Act.

We had Nick tinker with the program some more, and this time we[13] had ChatGPT find the most recent child relocation case — as opposed to us plugging in Wuesthoff (which is the most recent published relocation case). ChatGPT's top result was "Staten v. Bartunek, No. 59902-3-II (filed June 17, 2025)." However, Staten[14] is neither a case involving the Child Relocation Act, nor a published opinion with precedential value.

Luckily for ChatGPT, Wuesthoff was the second case it provided. We then asked ChatGPT to summarize Wuesthoff based on whatever document ChatGPT had found, rather than based on what we had given it in the earlier test. This time the summary was more accurate.

Based on this more accurate summary, we wanted to see if ChatGPT could create summaries like those we normally draft for Hot Topics, so we fed a PDF of our summaries from last year's Hot Topics for ChatGPT to use as a model.

Here is what ChatGPT "said" in response when we plugged in the PDF: "I'd be delighted to help! I don't see the attached document, though — could you please describe or paste the structure/format you'd like me to match?"

When we tried re-attaching the PDF and reiterated to follow the format of the PDF, ChatGPT responded: "I'd love to — just one thing: I still can't see the actual document you're referring to. Could you please provide its structure or layout?"

Ignoring the fact that ChatGPT doesn't have eyes and therefore cannot "see" things, its inability to recognize the PDF as a model seemed to us to be a pretty significant "bug" in the program.

But we were still determined to see what we could do with the program. So we told it to write a summary with "no headings, narrative format, with separate paragraphs summarizing the relevant facts, procedural history, and holding/reasoning" of Wuesthoff. And ChatGPT was finally able to give us

something workable — producing a three-paragraph summary of the case in reasonable detail, that correctly stated the holding and reasoning of both the opinion and dissent.

And it only took about three times as long as reading the case and drafting a summary ourselves ... not counting the time that any competent lawyer not already familiar with the case (Valerie was counsel of record, so we didn't need to spend as much time revisiting Wuesthoff) would need to confirm the accuracy of the summary.

After that, the summaries got a little easier to produce — but longer to check, since we still had to "vet" cases we were less familiar with. However, the vetting process revealed that ChatGPT still wasn't quite getting it.

For example, ChatGPT's summary of Marriage of Chea & Long was more focused on the facts of the case (which were more relevant to the unpublished portion of the opinion) than the actual holding of the case — which deals with the differences between "clerical mistakes" that are correctable under CR 60(a) and "judicial mistakes" that must be corrected by a timely motion for reconsideration or an appeal.

Frankly, we[15] were more impressed with Nick's ability to provide prompts than with ChatGPT's output. No matter how cheerfully it expressed its eagerness to respond to our queries, it seemed like a toddler offering to "help" wash the dishes. We'd rather have a competent associate.

Catherine W. Smith is a principal in Smith Goodfriend. She founded the Washington Appellate Lawyers Association and is a past president of the American Academy of Appellate Lawyers. She can be reached at cate@washingtonappeals.com.

Valerie Villacin is a principal in Smith Goodfriend. She is a past co-president of the Washington Appellate Lawyers Association and a fellow in the American Academy of Appellate Lawyers. Valerie can be reached at valerie@washingtonappeals.com.

Nicholas Bartels is an associate at Smith Goodfriend and a graduate of the Seattle University School of Law and was a lead article editor for the Seattle University Law Review. Before being admitted to the Washington State Bar Association, Nicholas worked as a law clerk at Smith Goodfriend. He can be reached at nicholas@washingtonappeals.com.

---

1 With the exception of 2021, which has a Covid asterisk because no seminar was presented that year.

2 The review is named for the 1905 case, Wainwright v. Wainwright, 40 Wash. 117, 82 P. 1135 (1905), which reads in its entirety: "This appeal is from a decree of divorce granted to the respondent. No good purpose can be subserved by discussing the testimony in a meretricious divorce suit. Suffice it to say that the respondent by his own confession was of a salacious nature, while the record as plainly shows that the appellant was cold, designing, and venal; that she took advantage of the almost imbecile weakness of the respondent for the purpose of profiting financially by his confessed immoral conduct. We think the court was justified in finding that the appellant had abandoned the respondent, and that the judgment was right. Affirmed."

3 The seminar will be on Oct. 2 this year at the Seattle Convention Center (shameless cross-marketing).

4 Hence the "Hot Topics" moniker.

5 We also freely admit that our personal opinions may be affected by the result when we were counsel on the case.

6 "AI & Family Law: Navigating the New Frontier" (again, shameless cross-marketing).

7 Here, "we" means Nick.

8 For the purposes of this article, the discussion focuses on ChatGPT. Claude, for its part, was slightly more accurate than ChatGPT, but we kept hitting a paywall, so our testing was limited.

9 These made-up citations are commonly known as "AI hallucinations." According to IBM, an AI hallucination occurs when the AI tool "perceives patterns or objects that are nonexistent or imperceptible to human observers, creating outputs that are nonsensical or altogether inaccurate." IBM, http://ibm.com/think/topics/ai-hallucinations (last visited Aug. 14, 2025).

10 Cause no. 85399-6-I, 2025 WL 2104240 (Wash. Ct. App. July 28, 2025).

11 33 Wn. App. 2d 877, 565 P.3d 604 (2025).

12 The correct citation for Wuesthoff is 34 Wn. App. 2d 8, 565 P.3d 660 (2025). The made-up citation that ChatGPT gave us, when checked in Westlaw, took us to the fact section of M.G. v. Bainbridge Island Sch. Dist. #303, 34 Wn. App. 2d 51, 566 P.3d 132 (2025).

13 "We" here, again, means Nick.

14 Cause no. 59902-3-II, 2025 WL 1695401 (Wash. Ct. App. June 17, 2025).

15 Here, "we" means Catherine and Valerie.