Just Letting AI Write Your Reports

Intro Music [00:00:02] Now this is recording RTI International's Practice Area presents Just Science.

Intro [00:00:12] Welcome to Just Science, a podcast for justice professionals and anyone interested in learning more about forensic science, innovative technology, current research, and actionable strategies to improve the criminal justice system. In this special release episode, Just Science sat down with Anthony Berghammer, research data scientist at RTI International, Ian Adams, assistant professor in the department of criminology and criminal justice at the University of South Carolina, and Christian Quinn, managing principal of Fulcrum Innovation, LLC. To discuss evidence-based, backed findings about the benefits and limitations of AI-assisted police report writing software. Law enforcement leaders who are considering adopting AI police report-writing software may be inundated with marketing claims about how these tools can save time and increase efficiency across their agency. In response, researchers have collected real data about the capabilities of AI report writing in the field and have identified important considerations for using AI tools in a way that is transparent and vigilant. Listen along as Anthony, Ian, and Christian describe whether AI report writing software actually saves time, the need for guard rails and human oversight and the future of AI and policing. This episode is funded by the US Department of Justice's COPS Office. Some content in this podcast may be considered sensitive and may evoke emotional responses or may not be appropriate for younger audiences. Here's your host, Carlee Ruiz.

Carlee Ruiz [00:01:34] Hello, and welcome to Just Science. I'm your host, Carlee Ruiz, with RTI International. Today's episode kicks off a new series funded by the COPS Office in exploring artificial intelligence in law enforcement. In each episode, we'll take a deep dive into a specific AI technology that's making its way into modern policing. In today's episode, we're focusing on AI-assisted police report writing software, so we'll explore the technology behind it, examine the evidence supporting or challenging its effectiveness, and discuss implications for agencies considering its use. So, joining me today are three incredible guests. We have Anthony Berghammer, a data scientist at RTI International who specializes in the technical side of AI. We have Dr. Ian Adams, a renowned policing researcher who has conducted some of the first studies on AI-assisted report writing, and Christian Quinn, a former deputy chief of Fairfax County and founder of Fulcrum Innovation, where he helps law enforcement navigate emerging technologies like AI. So thank you all for being here. So we have a lot to cover, so I'd like to jump right in. So Anthony, I think it's important to start by breaking down this technology for our audience, help us kind of unpack a bit of that black box behind AI assisted report writing. So can you kind of get into it about how it works? How does it take a transcript and spit out a pretty decent looking police report?

Anthony Berghammer [00:02:48] Absolutely, that's a great question. And I guess I'll start by saying, you know, AI is a term that's used, especially the last few years, to encompass a lot of different things. And I think specifically when people think of AI, they usually think of like Chat GPT or some of these large language models where you can input text and you get a great response out of it. But specifically with the report writing, there are a lot of different components that go into it. Um, especially in kind of the framework that you mentioned of kind of that end to end process of taking audio or body cam footage and then producing a report. So within that whole system, there are a few kinds of buckets. One is the initial data processing of taking that actual audio transcript and converting that to text. Um, same with video converting it to images and then running processing on that

information, whether it's object detection. Named entity recognition to see which names are mentioned, which car or vehicle types are mentioned locations. And then relationship, tonal expressions, you know, is there all these little components that go into it to extract information. Then that's fed into what we're thinking of as like the AI, the language model that actually takes that information and produces a report. That's obviously a super high level and I know there's a ton of details, but I think it's important to really understand that there's a lot of components that at different points could go wrong or handle data differently. So understanding each of those interworking components is really important, obviously scratching the surface, but I think that's a good starting point for AI.

Carlee Ruiz [00:04:18] Yeah, definitely. So, you know, kind of just takes all this input and then brings it together to kind of develop this report at the very end of it. And so maybe just diving a little bit into prompt engineering. So I know with a lot of these technologies, the developer who develops the prompt, are they messing with the original AI? How does that work?

Anthony Berghammer [00:04:38] Yeah, that's a great question. So each vendor is going to do it differently. And that's kind of the IP or the secret sauce that each vendor may have. So there are going to be these background prompts that might be as simple telling the AI model, you are a police expert in this area. And obviously there's going to be kind of more complicated, nuanced things that one vendor may do versus another. And then each of the little components that go into it, how do you process the data? What data do you have? To fine-tune your model to make sure it's, you know, trained towards the policing field instead of, you, know, retail marketing or whatever the case may be. So I think that's where a lot of the specific policing and report writing technology comes in is these vendors taking policing data and then tailoring these models a little bit as well as some kind of IP that goes into it to make it produce the results that it's producing.

Carlee Ruiz [00:05:29] Thank you for that kind of like technological breakdown without getting too into the weeds. You know losing some of us who are not so technical. I kind of want to jump now to Dr. Ian Adams. I want to jump into the research side of things and what your team have been finding as far as the evidence behind AI sister report writing. Is it saving agencies time like all these vendors are claiming? You know what are your thoughts and feelings on that?

Dr. Ian Adams [00:05:53] Yeah, thanks for having me along Carlee. So I've been long interested in AI and machine learning and so starting a couple of years ago when Chad GPT first came out, I sort of immediately saw like, oh no, cops are gonna definitely use this to write reports, right? It's an amazing tool. Front line officers are under a lot of time pressure. There's mounting call demand. And as most know, following the Floyd protests of 2020, we saw quite a bit of staffing pressure across the United States. And so that was a heady mix to throw this cool efficiency gaining tool into. And so I first started going up, and I still do lecture at the FBI National Academy on this topic with police executives. And the thing I heard most at the beginning was sort of like questions about what is this thing? What can it possibly do? But quickly that morphed into a lot of excitement and I shared in that excitement, right? Like, hey, if we can capture some efficiencies in the report writing part of operations, hey, maybe that means we can spend more time doing other policing stuff. And soon thereafter, in the spring of 2024, Axon launched their or announced their product draft one and with it came a bunch of marketing claims. Key amongst them that it was going to save 82% of Report writing time which was you know, implausibly precise to me Let's put it that way like when people start claiming 82% instead of 85 I was like how that is an interesting number and I thought well some time before that actually an agency

had gotten a hold of me and said hey, we're testing this product and would you like to be an evaluator of it? And so I went up there and my team ran a randomized control trial, which is sort of the highest level evidence that we can provide to agencies. And to our surprise, and to a lot of people's surprise, what we found is that it, in fact, it didn't save 82% of time. In fact, it didn't save any time. It didn't cost more time, but it didn's save any Since then, more and more agencies have sort of replicated this finding in their own use of that tool. And even more broadly, I would say, just like emerging across sectors, not just in policing, but say in code writing as well, computer coders and business more generally, this first generation of LLMs is just broadly not capturing the efficiencies that we had hoped for. And I pause there just to say like, maybe there's other benefits it is capturing, right, but like, because the marketing claims have been directed so strongly at time savings and efficiency savings, I think that naturally draws researcher attention as well. And the research in this area is pretty convincing right now that it's probably not saving time. So if time is the ultimate goal of an agency in purchasing these types of products, I think that they're likely to be disappointed.

Carlee Ruiz [00:08:50] That's a really important finding, especially like, you're not actually saving time or saving efficiency. So is there anything else or other findings out there that's maybe touching on like the quality of report or any other possible benefit?

Dr. Ian Adams [00:09:05] We have done other studies in here as to the quality of it. Maybe we could put a pin in there and come back because I think there's lots of different ways to think about what quality means in this context. The other study that we did that was recently published was going back to those same officers in our original experiment and trying to get a sense of like, how did they experience the product? And one of the interesting conundrums that we're all going to have to confront here, I think. Is that when we asked these officers, did this tool save you time? About 50%, about half of them said, yeah, this saved me time. Now, we know that is empirically not true, right? We know I can look at those same data and say, like, no, it's not true that it saved you time. But it's pretty powerful confirmation bias sometimes when we want to believe something saves us time. We really need a tool to save us time, and here comes some magical promises that say, this thing is going to save you time. It turns out that a lot of us just being human beings. Will naturally sort of believe that it's saving our time. So like one message to chiefs and other executives that I try to get across is, don't be surprised if your staff come to you and say, hey, we love this tool. It's saving us time. But you don't actually see the times. You can easily find yourself in that space right now with this generation of technology.

Carlee Ruiz [00:10:26] Kind of what I'm hearing is if you're planning to like implement this technology or do the trial pieces, maybe really have some sort of objective measures in there to see is it truly saving time and not just be going based off the feelings of your officers or at least maybe evaluating, Hey, what do you want this technology to do? Are you wanting it to save time? Or is there any other benefit you're maybe you can look at that could be good for this technology?

Dr. Ian Adams [00:10:48] I might not even be the best voice on this. One of the better thinkers about sort of goal-driven technology adoption is actually here with us and we're lucky to have him. So that's sort of Christian's area that I've heard him talk about.

Christian Quinn [00:10:58] And I come at it more from a perspective of probably operational deployment. And some of that's going to be less data driven, more anecdotal. So you kind of, you do tend to hear those things like most people are not saying like, well, we haven't realized any efficiencies because from a perception standpoint, they believe

that they have. But you know, I would submit to you, like when you started to get into the the weeds of it, some other things of anything that is met with optimism, I think we need to really look at what is the scope and utilization of it in both a field setting and more across the agency. I would submit to you that brief interactions in the field are very different than a comprehensive interview by a detective. So I'm interested in, you know, how well does it do there? I would admit to you, that someone reporting something suspicious in their neighborhood It's very different than maybe the interaction you're going to have with a member of the drug subculture or a criminal gang or something like that, that is simply going to fall into using slang and more nuanced terms and things like that. Because again, I always say this about artificial intelligence, and it's not to diminish the power and capability of the tools. But at this point, it doesn't know anything. It is. Putting text together that is semantically correct, that sounds very conversant, that sounds very good, that really is based on this notion of tokenization and reassembling characters, whether it's in words, in sentences, in paragraphs, and longer. That is based on statistical probability and some functions, right? And that is not to diminish the magic of it. Like, I don't choose magic as a term intentionally, but on a surface, that's kind of what it looks like to the end user. Like how does it generate these outputs? But really when you get into it, it's just that. It's the assembling of text. So I think the other part of that too is looking at the expectations of our end users and giving them both some semblance of training and also some semblance of responsibility, that this is not a replacement. For taking good notes. This is not a replacement for doing a thorough investigation and documenting things like the non-verbal, like the visual observations and things like that. And thinking really critically, because this is my worry with all tools. So when something's new, people are more vigilant about the outputs. People are more vigilent about the utility of it. Then they get comfy and this complacency sets in like it does with everything else. And that's where, you know, when we say automation bias, these are the things that lead us in our everyday lives to follow the directions of Siri or Google or whatever other application that we use and go the wrong way, knowing instinctively like this probably isn't the right way to go. But we trust in the tech because we think this has got to be smarter than me. It's got to know something I don't know. And the trouble in police work is that the stakes are too high for that, that we can't just put blind faith in the tools. So kind of that human in the loop sign-off is a big piece to it. And I think the last thing I would offer at this point too, before we get further down the road, is kind of that sense of choosing at the outset the extent to which we are disclosing to the communities, preparing anything that's gonna go beyond just a field report, if it's a criminal investigation, working in concert with our prosecutors' offices. In the U.S., that varies drastically from state to state, the control that the prosecutor's office has over most agencies and what they do and what tools they adopt. In some states, the prosecutor's office of New Jersey, they're driving those decisions. But in the vast majority of states, that's not the way that it is. It's they are either saddled with what the police department does after the fact. And I admit that sometimes as a profession, we're not the most collaborative partners on the front end, that we will adopt things that ultimately the prosecutors are left to deal with when it makes to court. And I would submit to you that in the age of emergent tech, particularly AI, they've got to be in at the ground floor. They've got be a partner in that conversation to talk about, like, one, the most basic question of, is this an acceptable technology to use? Because in some municipalities, prosecutors have come out and said, we won't take cases to court. In others, they said we want it disclosed on the front end that this was used to gather facts. Then just the last thing I would say, just as a discipline, we need to study more, is the notion of cognitive debt. And when you look at like some of these early studies, just in terms of the difference between using an LLM GPT to generate something versus using it to iterate on something that you created in research versus like you not using one at all. When you look some decent science on this, in terms memory and recall and that sort of imprinting in your mind the event, You really

can't refute that writing a report, making notes in the field, coming back, thinking about it, writing it down, and then reviewing your own work, that does create a more lasting imprint on your mind with respect to like the memory of the event. And I'm not saying we don't use tools of this nature. I'm just saying we'd be mindful of the trade-off. I don't mind tools helping people do their best work. I do mind sourcing the work to the tool and just having humans sign off on it as opposed to using it to augment their capacity.

Carlee Ruiz [00:16:26] Yeah, you touched on a lot of things there of, it's really good at making a report sound nice and good, but does it have the substance it needs required for that report? And then if the officer is not writing it and we get complacent on double checking that, you know, when it comes to court, eight months, maybe even years later, depending on the case, will they remember as well as if they have written themselves. And then also... Is that truly what happened in the case of like when they wrote that report wasn't fully accurate if they were being complacent and not fully double checking it before hitting submit.

Christian Quinn [00:17:03] And I think we think about like, I think about anyway, maybe I'm a pessimist is kind of the, what's the worst thing that could happen, right? To me, one of the worst things that could happen would be especially your bigger events that don't tend to go to trial for like a year where you have like a homicide or something like that. And now you have an officer on the stand who's testifying, who's being cross-examined by defense and they are asking. Did you record this? Is this your report? Is this this? The worst thing that could happen would be an officer saying, I don't know, the AI wrote that, I signed it. You know, and we I'm sure we could, you know, hypothesize like other bad things. But to me, in terms of like the utilization of this tool, that is one of the last outcomes that we want is someone disclaiming their report because they outsource that responsibility to the tech. And I think not that that's going to happen, but that as we adopt these tools, We have to be really vigilant about that and impress upon users that you are responsible for the output that this is a tool to help you record your observations and to help you provide a narrative of what you experienced, what you heard, what you observed that the AI can't possibly do for you in terms of like what you see in the nonverbal. That communication is often of greater value than what is said, especially when someone's lying. So I think we have to look at it kind of holistically.

Anthony Berghammer [00:18:21] One thing to piggyback on that too is really, we're seeing that across other industries where it's really the use of the tool and how do we put safeguards on it to make sure, you know, one, are people reading it? Because there is that disconnect of, okay, wow, this great report, it sounds good, I just need to skim it, it reads better than anything I could write, and it has those components. But then if you just skim something one time, then you have to go out and defend it six months later, you don't actually know it the same way that if you researched it or wrote it. So how can we build? These tools in a way that still has enough of the kind of human in the loop and manual review to make sure that people are, I don't want to say learning because it's still the report that they went through, but actually working with the tool instead of just the tool doing everything.

Dr. Ian Adams [00:19:03] Yeah, and I think to Christian's point, I've taught report writing to officers and I teach writing generally, you know, as a scholar, writing is the act of thinking. Like this is the actual behavior of thinking is the writing down. And we don't even, I think Christian's absolutely right that there's like pretty good science on this. But like, I don't think you need the science to just kind of intuitively know as a sergeant, if you were asked to remember a report a year after you read another officer's report versus you writing that

report, which are you more likely to know? And the answer is, of course, the one you wrote. The more that we turn these writing workflows into review workflows, it's very, very likely we're going to get less recall. It's just going to imprint on your brain later. I'm now almost 10 years out from leaving the road. I get calls all the time from friends who say, hey, remember when this happened? And I don't remember that it happened because I didn't write anything on it, right? Like I may have been there. I trust them that I was, but it doesn't mean I have an active memory for it. So I think we're in right at the beginning of a pretty big shift in rethinking these workflows. And if I were to make any kind of suggestion, it would be that we start to think about the difference between using these tools in an end-to-end manner versus a middle-to middle manner, right? I don't think it's realistic to say let's not use it. They're too good. These tools are really good, really useful, but to the degree that you can get as much front loaded work, whether that be an outline, some jumbled sort of notes that you've handwritten, your own words spoken into a recording or a microphone or something, you're taking some action to get us to the point. Where the LLM then can take that information, get it into shape, get it into like well-written English comp, and then that's not the end, that's just the middle still. Now you're going to take that piece to its final conclusion. You're going to read every sentence, you're gonna make the edits to every sentence. You're gonna do the behaviors of thinking, because it's a mistake to think that thinking and memorization is occurring just in our heads, sort of in that liminal space of our minds. It is the behavior of like writing, typing, editing. Those are the behaviors of thinking and the more that we can retain some of that in these report writing workflows I think the better off the officers will be in the cases that they work.

Christian Quinn [00:21:22] Yeah, I would just add to that. I think the act of especially if you are working investigations, writing, especially supplementing other reports, it's almost forces you to take inventory of what you know, and what the sources of those were like I spent a lot of time in Intel, and frequently you're having to challenge yourself like, why do we think that? How do we know that? Is it because a single informant with an agenda told us that, or because we factually observe that or we pulled that out of some, like, source or we did a search warrant for something that is objective that you know there's different levels of credence in terms of like what is known and putting that all together without qualification as to so-and-so said this so and you know putting some degree of like veracity to those different points as opposed to all of it is equally weighted. I think there's utility in that, and we want to make sure that we don't lose that. And I think you can, you know, to Ian's point. We're not going to turn our backs on tools in tech that are like too good, right? Cause that's foolish to stick our heads in the sand and say like, oh, it's not quite there yet or what have you. Cause people know better. But I think what we need to be is really intentional and mindful about how we roll things out and also respectful of the ability of the end users to grasp how tools work. And this is how we do other things. Like when we talk about like the breathalyzer and radar and things like that, like when you first start using it, you go through these like painful exercises and knowing things about it that might make you say like why don't you know this much about this but just having a foundational understanding of the capabilities and limitations of the things that you're using does you a lot of favors later when you actually make use of those things in a real-world setting and it also does you a lot favors later. When you're subjected to scrutiny and you have to explain how did I come to these conclusions that I did or I've represented this something to be factual, how did I get here?

Anthony Berghammer [00:23:19] Yeah, I think you made a great point. And I think learning how to use the tools is, is number one. And I, think it's, you know, we, we learned how to Google, it happens through osmosis where you just get better at kind of knowing what's going to produce the best results. And I. Obviously this is a lot more intense and

higher stakes than Google, but I think over time you learn how these tools work, what you can trust, maybe a little more than not what's higher stakes and what isn't and then. I think it's on a lot of the vendors as well to really have those conversations and do the due diligence to make sure that the tool isn't abstracting the actual users from the work, you know, that people are actually still engaged in the report writing process and understanding how the tools they're using work instead of here you go, here's the report.

Dr. Ian Adams [00:24:03] I don't know if this is what you mean by that, Anthony, but it like, it sort of got brought up too by a little bit earlier when you were talking about how this set of what we call custom instructions that go alongside these reports, usually hitting the API of the vendor right are hidden. And you referred to them as that's where the magic is or where the IP is the end of the, I assume you meant intellectual property there. I think this is a conversation that needs to be had. I haven't heard it, but. If we just think about like the creation of this police report that's going to go become a legal document and it's the first sort of like link in a chain that forms the entire criminal case, right? It's going hit every stage of the criminal case. Is it fair for a huge part of what created that report to be hidden from not just public view, but from the view of the agency and the user themselves? I understand the intellectual property argument. I feel the same way about my own custom instructions, right? So in my ChatGBT account, I have a set of custom instructions that have evolved with me since November, 2022, when that tool came out. I feel protective of it. It feels like it's something I created yet to sort of like. That into the frame of intellectual property when we're talking about police reports being created on a mass scale and perhaps weighing on the civil interests and liberties of the people that police departments serve and interact with seems to me at least debatable, right? Because three components to that LLM-generated report. One is underlying data. So this might be a transcript from a body-worn camera in the case of something like graph one. It might be set of notes or other musings by the officer in other products. Then you get the set of custom instructions. Those combined hit the foundational model, hit ChatGPT 5, 4, whatever you're using. And in the end, then we get a police report. To hide one third of, well, more than a third, because we're also not going to see the sort of black box of the foundational module. So to hide two thirds of the input that gave us a police record and say, that's intellectual property in the context of a huge public interest, you know, the right to stay free and face your accusers and all the other amendments. I mean, maybe I'm tipping my hand a little bit too much. It's more than debatable for me. Like, I think that that is obviously something that needs to be peeked inside. And even if you set aside my concern about the public interest. As the officer, don't you want to know what set of instructions went along with your own text to create that report? California just recently passed a bill that makes it mandatory now that there at least be a disclosure at the end of these reports. Frankly, every commercial vendor was already doing something like that, but it didn't take that next step. It didn't make a demand on the set of custom instructions, and I think that's something we ought to think pretty carefully about.

Anthony Berghammer [00:26:58] Just one quick thing, I think it also begs the question of, because there's so much abstracted from the foundational model with a lot of the training that goes into that, the data that goes into fine tuning for the vendor model plus the system prompts, at what point do we just say it's a tool, but regardless if the officer didn't read every line or whoever doesn't read every line and stamp it, then. Does the system matter? These are probabilistic models at the end of the day. So they're going to produce, there's a random component to it. So can we trust any of it? You know, it is, it's a, it it's kind of a weird legal question in that way of, you know, does it become this thing that was produced by a computer? And then it's on somebody else to go through every word. So that's it's completely separate.

Carlee Ruiz [00:27:40] One quick thing, Anthony, can you explain what the probabilistic part of that so people understand what you mean by that?

Anthony Berghammer [00:27:47] Yeah, yeah. So when you use something like an AI model or a large language model and you input text, it breaks that into what are called tokens, which you can think of as words. And then all that it's doing is essentially predicting the next word. So if you ask it a question, it's going to predict the answer to that question. It was token by token. Each step in that process is based on probability. This is the probability that this was the next token. This was the probability of this is the next where it takes the maximum. Probability essentially, and produces that output. So while it's trained on a ton of data and it does really good because it's trained on essentially all written text, there is a point that it is saying, hey, maybe it's 51% chance this is blue is the correct word when the reality is, oh, it was something different. Obviously, it does a good job of predicting that next token, but it is still probabilities at the end of the day.

Christian Quinn [00:28:41] And the other thing I would say on the front end before we ever even get it. That far down the road and do a deployment is I'm a big proponent of limited pilots. I think Ian really summed it up in terms of the perception of efficiencies versus the reality of efficiencies. What is an efficiency? Is it time saved? Is is effort? Is accuracy? Is it reliability? What is it that we're going for? And then if you do a limited pilot, what is it you want to measure during that? And I'm big proponents of partnering with academia or like outsourcing that to somebody who does that that will do a good job of it that's impartial and then looking at what are our kind of like go-no-go decisions on are we going to do a full implementation like anything else with body-on-cameras or what have you like These are expensive undertakings and you have to look at a lot of these things through a lens of like, what are we not going to be able to do because we chose to do this? Like at an executive level, like business decisions have to be made. Like, what, are we getting out of this and is that worth it? And then what is the metric, not just like a feeling like, Oh, I like it. I don't like it, like, What is it that we're going for in terms of like an evidence-based decision, uh, around the business aspects of running an agency in terms, of what we're realizing for return on investment.

Dr. Ian Adams [00:29:54] Yeah, Christian must have woken up trying to get me excited about this conversation because he mentioned like two of my favorite things, which is careful evaluation and balancing public money expenditures in careful ways. So on the first, we should be treating pilots as opportunities to evaluate. And you know, like I think everybody here is that's going to be a non controversial statement. RTI does incredible has for decades. I do decent work, have for a little while. We're just two examples though. There's like every region of the United States has somebody who is capable of doing good applied policing valuation. If you have questions on that, reach out to one of us. We are pretty easy to find. And then the second is, you know, back to that original study where we showed a big fat null on time savings. You know, when the time came for that agency to make a decision about deployment, they chose not to. And the costs to them were substantial, right? Like it's an agency with over a $30 million annual budget, but like every agency, the vast majority of those funds are tied up in personnel costs. And so like leftover, they have like two and a half million dollars, let's call it. Well, the quote for the full suite of the app acts on sort of evidence.com platform, which included the draft one product that I evaluated was over $500,000 equated to 23% of their post payroll funding. If you just like strip out that and say, well, what about just the draft one, just the drop one quote was around 11% of there after payroll. I think we have to have really careful talks about spending public monies when the primary goal was not met, the primary what we're

trying to achieve. And that's why, going back to Christian made an earlier point, it's really important that the chief executive and their team has a careful conversation about what it is we're tying to achieve, is it time savings on report writing? If so, then this test would have been a fail. But if it's something else, then let's evaluate that. I had another agency contact me and say it's sort of a proxy of time savings. They wanted to save, they were spending too much time on overtime related to what they perceived as a report writing problem. Officers were holding reports because the call volume was so high, they end up holding reports to the end of the shift and now they can't get all those reports done and so they're paying overtime. Well, that's measurable, right? I can get that answer for you, given the right data. So if it's overtime, is it quality? You want your sergeants to be rejecting fewer reports. Great, let's measure that thing. The point being, all of this has to happen pre-deployment. That limited rollout is a really important time to try and distinguish what's important to this agency in this context at this particular moment. And then, if it meets those goals and you decide to deploy, having that conversation again six months or a year, having it in two years and continuing to monitor because I promise you every every chief in this country, every sheriff in this county has another use for those same dollars. That $500,000 can be spent somewhere useful. And if we're spending it in a way that's not achieving our goals, I don't think we're doing ourselves or our communities any justice.

Carlee Ruiz [00:33:04] And definitely, I think also a big proponent as well that I've seen from other agencies is that they all tend to start on a trial without talking to their prosecutor office, which we've touched on a little bit about getting your prosecutor in the loop. But it's also like making sure, you know, this is very new technology and into making sure that like all your, your I's are dotted, your T's are crossed when it comes to having your prosecutor on board. You could start this whole thing and then they'd be like, nope. And then that's it. That's it for you guys. And so... You know, really making sure checking you in your IT department, whether it can kind of handle this kind of technology and making sure having a conversation, you know with your community with your mayor's office, whoever needs to be involved in those conversations as well.

Anthony Berghammer [00:33:46] I'm a big fan of pilots and we do a lot of pilots here at RTI, you know, across other, you know not just policing, but in other ways. And I think the biggest thing is kind of doing that risk analysis and understanding your KPIs beforehand. Like if it is time efficiency, great. What data do we have to measure that? How do we compare the before and after, so to speak, or the different options? Pilots especially with this new technology, you don't know what's going to bring value. I mean, and I think we see that kind of on the technical side of people saying, Oh, could we use AI for this? And a lot of times the answer is, I don't know, until we try, until we run a pilot and see, Hey, this didn't work or Hey, this actually worked a lot better than expected. So I'm a big fan of pilots and, and, but making sure that they're done right, you know, with the right metrics and the right measurements.

Dr. Ian Adams [00:34:28] And when they're done right. Like science doesn't promise you the answer you think you're going to get, right. In full disclosure, I was as shocked as anybody when we found that this initial draft one product didn't save any time. I've been lecturing on it for two or more years at that point, fully with the sort of prior expectation that this thing would save time. But that's the value of science when we pre-register our guesses, so to speak. And we show like this is the careful way I'm going to produce the highest level evidence possible, then we go with that. Maybe maybe the next outcome that we choose is slightly different. That is how we build the iterative knowledge base, though. And my fear is that we're right in the middle of probably a more rapid adoption curve than we even saw during the bodywork camera rollout phase, right? Body cameras begin

getting really rolled out at the mid-2015 in a heavy way. There were some before that. By 2019, we probably have 60% of major agencies today, more like 95% of major agencies. So it took a full decade to really achieve full penetration. Here, when we're talking about AI in general, not just LLM report writing assistance tools, I think we're way above that, right? I have a project with a team that I run a lab out of both Purdue, Utah, University of Utah, Purdue University, and the University of South Carolina. This is some NIJ funded grant work where we're trying to just describe the adoption of all kinds of advanced tooling, including AI tools. And we're already seeing like upwards of 70% of agencies are already in some phase of experimenting with drones, for example. Not just drones as first responders, but, you know, other uses for the drones. That's way higher than I think most people anticipate. If we've got nearly more than two thirds of agencies at least dipping their toes in those waters. And we don't have a good number for the vast array of artificial intelligence. Framed tools that are out there right now. IACP is coming up in a few weeks. My favorite event of the year is to walk onto the vendor floor, or what I call the magical hall of wonders without empirical proof, where there will be claim after claim after claim about magical outputs of all these different tools. And these chiefs are going to be walking in and there's a lot of pressure on chiefs to achieve efficiencies, to achieve better outcomes, do more with less. That's true today as it ever was and they're going to under a lot pressure. I would just say to the degree it's possible, take a deep breath, ask for a limited rollout, team up with someone you trust to give you some balanced answers and then make good deployment decisions because it's your budget that's at stake.

Christian Quinn [00:37:09] And I think when it comes to this issue, it's almost like going back to a basic question of what's the purpose of a police report? And I would submit to you that we want something that is as close to factually accurate as possible, that is an easily understood representation of what happened or at least how it was represented to our personnel when they got there. And we want that understandable by somebody who was not there. And we want that to be an independent representation of what happened for the purpose of even the person who was there their capacity to recall it as they perceived it and it sounds like really elementary and foolish but when you have a tool that can write a report that is narratively compelling and more grammatically correct than the writer is capable of producing there's almost this temptation like that's a better report and i would submit you that it's not it's because that's not the purpose of the report is to pass an English exam. The purpose is to have something that the person who was there can use as a means to recall and like probably another whole other podcast is the science of memory. But we certainly don't want to forfeit an opportunity to have a representation that the person who was there can rely back on when it's a year later. If something goes cold and it's four or five years later, somebody who wasn't even there has to rely on that to investigate. I think that's a very elementary question, but still a pertinent one is to you know, what is our objective in terms of like producing a written representation of of what occurred that the police department went out and investigated and what those who were there experienced and I will harp on that one last point about. The non-verbal and the visual observations and the field observations are highly pertinent. This may come as a surprise, but some people lie to us and taken on its face as to verbally what was said is not valuable. I think documenting that the person was averted in their conduct, they were sweating, they were nervous, they we're looking around. Those are pertinent facts. That accompany what was said with respect to the officer's perception of do I feel like I was being lied to and and that's to substantiate why did I sense that what were the cues whether they came from the environment or the person speaking them that belonged in the report as well.

Dr. Ian Adams [00:39:29] Maybe this is a good time to return to one I tried to pin earlier, which is, does this make for a better report? And it's kind of like, it depends on your definition of what good means, right? What is a Good Police Report? I think we can go back 100 years to the beginning of modern policing and find complaints of the quality of police reports, right. And they run the gamut from sort of substantive concerns about completeness and accuracy to more sort of English-lit 101 concerns about grammar punctuation and spelling. And I think it's pretty easy to get caught up in that, right? Like any conversation about police reports amongst police officers and those who have been in the field will be kind of like laughing a little bit about just how bad some of the spelling and grammar can be in a police report. That said, I have yet to see an officer get jammed up on cross over a misspelled word or a misplaced semicolon. That is just not the worry when I'm thinking about the quality of police reports. However, officers often are and will be jammed up on cross over bad facts, missing facts that they should have been aware of and that sort of missingness, that sort-of goodness question. And so it is awfully compelling and easy to look at the output of an LLM and say, yes, that's well written, right? It strung together five sentences into a paragraph and had a thesis sentence and used a lot of EM dashes to make it look fancy. But that is probably not a good police report. At least those aren't the metrics that I would judge makes for a good police report it can never really and no LLM can ever really provide a complete police report because of what Christian is out in here. That a police report is a collection of facts perceived and or reported by the officer. The officer's perceptions matter a lot. And to the degree that an LLM-generated report can't and doesn't include those perceptions, it's going to be a less quality report. That's just, we can't overcome that hurdle at this level of technology. So in no way, I think, I haven't even even seen vendors really claim that like this is the final replacement for police reports. Again, at its best, maybe in a middle-to-middle workflow, we can get a better produced police report when everything's going right, and we probably ought to organize our energy around that sort of workflow.

Carlee Ruiz [00:41:52] What are your guys' thoughts on, so I've seen some police report writing software that will ask questions to the officer. Did you interview any witnesses? Did you canvas the area for video surveillance? What are you thoughts around those type of softwares that are doing that kind of thing?

Dr. Ian Adams [00:42:09] Potentially useful, right? Like that's the completeness question. Maybe they can help answer. I think everybody here has probably reviewed a police report where the patrol was busy and they didn't include a contact number for the victim or a contact for a witness, right. It's a common little oversight. So to the degree that we can use some smart prompting, it's good. I try to keep my scientific hat on tightly. So I would just say like, I don't know of any evaluation of that type of product yet, but potentially useful.

Christian Quinn [00:42:36] I like it for multiple reasons, because I was a proponent of that back in the analog days. Like I remember creating a cheat sheet for auto theft reports. And the reason I did this was because upwards of at least one in 10, probably closer to one in five, auto theft report is a false report. And it's to either collect on the insurance, to mask some other thing. And I just created a sheet for patrol like if you're taking an all of that report please ask these questions and please make these notations like can the person account for the keys are they intoxicated were they recently in an accident are they behind on their payments did they recently change their insurance there are a whole bunch of what I call false report indicators that taken collectively are really helpful to know when that report comes up to a detective's desk and they're stuck with it and now the person's had like three or four days to like on their story and harden their position as opposed to like they have to commit to those facts at the beginning. I would also submit to you that We

have done that with other masts as well, like early on when we adopted Taser, we created a Taser supplement that it wasn't just, did you use it? Didn't you? What happened? It was like, where did the probe strike? What was their reaction? Did you gain compliance upon initial shock? Were additional shocks required? Like these are things and metrics and facts that we want. And I think that prompting has utility. And they don't think really, it really matters. Is it as we mature as a profession. Certain incidents, certain event types, we want to ask those things like stranger sex offenses, same thing, method of coercion, method, of entry. These kinds of things are helpful in the investigation later. And we're not discounting the competency of our patrol officers at large agencies. We're just looking at it through a lens of you don't necessarily handle these high stakes outlier events all the time. These are the things that in a perfect world, we get at the outset of a preliminary investigation.

Dr. Ian Adams [00:44:32] And to put that at a different level too, the use of checklists goes back a long ways in policing, but also in nearby critical professions. Pilots use checklists, surgeons use check lists for complex tasks. It makes everybody safer. There's been some good work over the last couple of years by Corey Haberman and his crew at the University of Cincinnati in using checklists to improve initial investigations in policing. And of course, there's like that book, The Checklist Manifesto. Gawande is the author's last name. I don't remember his first name. He shows like this, you can use these really profitably in a lot of different areas of human life. And so to the degree that we can build checklist-like behaviors into even everyday processes and then execute those through some sort of LLM in a report writing context, probably we are going to see benefits just because we've seen so many benefits in areas of human life. Like I tell my students, there's no part of the human brain labeled policing, so when we see something working for human beings everywhere, like we should probably expect that it has good effects in policing as well.

Anthony Berghammer [00:45:36] Yeah, I mean, I think the LLM serves a couple purposes in that kind of framework. I mean on one hand, it frames it as a tool instead of this automated thing. So I think it helps kind of leave the human in the loop. The officer is still actually having to answer the questions. And I think it potentially could add to checklists just because it's dynamic. So it can take the context and think of new questions to ask for like maybe an edge case, this thing that you've never seen before happens, you know, are there questions that should be thinking? Cause it does serve as a good kind of brainstorming partner, thought partner, you know in all these other fields. So maybe that dynamic checklist approach could help, you know, kind of improve that area.

Dr. Ian Adams [00:46:12] It's maybe we get some of the learning back out of these learning machines, right? Like it's the type of thing we can think of as building over time. And we can imagine a future not to get too predictive about things, but we can imagine a feature where some of these automated agents, you have some sort of unique relationship with officer individually and can say like, Hey, officer Adams, you're really bad at getting like date of birth. So I'm going to make sure that that's something I'm asking you about, or I'm checking in your reports, I'm check in your files pretty robustly. And maybe I'm good at other things, and so it doesn't have to bring up those things. It kind of learns over time. We're not that far away from, we're already, of course, seeing personalized LLM-based personalities that are taking off. The future is wide. I suspect we're going to look back in five to 10 years and think a lot about the questions and concerns we have today feel naive and simple.

Carlee Ruiz [00:47:08] We've talked a little bit, you know, if you're planning implement, you with trials and moving forward. So if you are an agency that either did a trial and went, no, this is not for us, or you're starting with that decision from the get go. What is some like

recommendations on to come to that decision to how to protect themselves? Because I keep hearing the argument for and I don't know if this is really the best argument for AI report writing software is your officers will just use Chat Sheet and they're just going to do it behind your back. What would you recommend to an agency that's deciding not to move forward with that technology, knowing that fact?

Dr. Ian Adams [00:47:43] I don't want to start with advice because I think Christian is going to have the advice, but I will say one of my first thoughts upon seeing Chad GPT, the original 2022 November version was, oh, cops are definitely going to use that. And I went out to Reddit immediately within a few weeks and there on the subreddit RLEO, which is a credential based access, right? You have to have credentials to get behind. There was immediately cops saying, hey, look at this cool thing. I'm using it to write reports. Am I going to get in trouble? Right, immediate. And so it's definitely like I don't want to dismiss that concern that you just raised because I think it's an important one We definitely do not want for a variety of reasons that everybody listening probably understands the very sensitive and personally identifiable information Traveling on the open web that is a bad bad recipe So we have we should do something whether that's policy based I'm not a big fan of the bans on these things. I think we should be finding ways to accommodate technology Policing has always been an early adopter of technology. They have to be because the first two uses for every advanced technology is crime and porn. And so cops immediately get pulled into those technological crime battles. They have be fast adopters and be creative reaction, policy, behavior, etc. As to what agencies should do, I think that's an excellent question, Christian.

Christian Quinn [00:49:05] I almost look at that through a lens of like, it's not necessarily like a binary, if this, then that, not this, than this, in terms of like a vendor-based product versus like the open tools. I think the biggest concern I have around all open models, whether it's ChatGBT or Gemini or Grok AI or what have you. And I often say this so like it won't surprise people that know me is I have far less concern with like that 20 year employee that's like tech averse. It's like, I just don't know if this is for me. Then I do that for your employee that has been using GPT since 2022 and believe they're like an AI whisperer. And those are the folks who before it all came to light that if you were using the share function on chat GPT all this time those are being leaked to the public wild and they could be found with search engines and other things like that and even though OpenAI and Google and others have moved to scrub those they exist forever in the wild and I like really cringe at the thought of like how much stuff is out there probably in terms of complex investigations that is in the domain somewhere that could be found that even if you are taking the step of scrubbing a name and scrubbing an address you know contextually you could use an LLM to figure out like hey where this happened so like someone that gets a hold of a complex burglary report or a stranger sex crime I would submit you like you know the professions having a hard enough time with trust and accountability as it is if we are going to and i often say this about. When we take reports is people don't call us when they're at the best they cause because something terrible happened to them or they are at their worst. And when we are the custodians of those stories and that information that's not the same as PII. That's not just losing somebody's name, date of birth, and social security number. That's failing to keep their secrets. So I think we need to take a look from a professional cultural stance. We need to be better about safeguarding the people we come in contact with secrets, because we can't say, hey, trust us, put your faith in the effort and our intentions as a profession. We're trying harder to be better and then do things like that and leak investigations online unintentionally because we have employees go in freelance and think, you know, everybody's looking for the hack. Everybody's looking for the shortcut. So, I agree with Ian, in that we don't... Want to take a stance like hey just don't

use anything. But we do want to educate our workforce as to like where are the vulnerabilities? Where are the limitations and liabilities? Especially in using anything that's an open source tool. They weren't designed for that and the other thing is many of these providers they don't even want us using their stuff. If you look at Anthropics terms of use, like you can't use Claude for anything law enforcement related other than back office efficiencies and analysis and accounting and things like that. Or analysis around like human trafficking and exploited persons. But like almost every other function of law enforcement is prohibited by terms of use. So I think that's like probably a lot to unpack, but I would submit to you that it's not a, if you don't use this, they're just going to go out and do whatever because that's an accountability issue. That is a knowledge and information so that they hopefully voluntarily comply and make good decisions. And also one where an agency has the wherewithal to say like, we're not using this and here's the why of that. And we're using it in this way because this is where the vulnerability lies with respect to the risk to the community.

Dr. Ian Adams [00:52:35] And we also shouldn't, like, short-change the value of policy here, right? Like, we know from decades of work that police officers are responsive to agency policy. Like, it's easy to say like, well, sure, somebody's going to violate it. They're going to do whatever they want anyway. Maybe, but that's probably not a good aggregate view of what's going on inside agencies. When officers disagree with a DUI change in the 1980s, guess what? Law and policy change officer behavior anyway. Same goes with domestic violence, mandatory arrests. Same thing goes with stop questioning frisk in New York City. Like, officers can have their own personal feelings that are conflicted with policy direction and policy wins. It's simply not true. One of my least favorite sayings is that, you know, culture eats policy for breakfast. Yeah, if your policy is shit and if your culture is shit, but like in a well-run agency, a policy is absolutely directive of officer behavior. And so to the degree, like, I'm not here to tell a chief. Here's the right policy, but I am here to say make a decision and get a policy out on it If you're if your decision is we're not going to use that tool for this that reason Please do explain it like christians obviously rooting for but like get that thing down in policy I still when I go to the fbi national academy give these lectures I take a poll every time who has an agency policy on llm's and in march 2023 it was exactly zero. By October of that year, we had like two maybe out of 40 students in each class. Today, maybe half of those agencies. But that's not enough. These tools are being used hundreds of millions of times a day. We obviously need to have agency policy that provides some guidance to the officer and some protection to the agency to say, hey, we did our best to get a good policy out there on this.

Christian Quinn [00:54:31] I would just add to that that no decision is a decision because you can't say you shouldn't have when you gave no direction whatsoever. And I think that is kind of an important point is at least attempt to get something down that provides some semblance of direction. And then ideally, you want to accompany that with an explanation if it's something that's not going to be popular or seems counterintuitive.

Anthony Berghammer [00:54:54] I think that's a good point. It's just really important to have the conversation, make a decision, have the policy, you know, train people and get ahead of it instead of behind it. Because yeah, I think you make a great point that no decision is a decision.

Carlee Ruiz [00:55:06] Well, I think we're running near the end of our time together, so I kind of wanted to give you just a last chance. Any final thoughts that you want to share with our listeners around AI police report writing software, anything like that.

Dr. Ian Adams [00:55:19] I'll go first. It's just one piece of advice, and it's, there's probably people listening who've been in policing long enough that they remember when they got their first pager or cell phone on the job. And they wouldn't have predicted here we are 25, 30 years later about where mobile technology is. We are at the pager stage of artificial intelligence and policing. We have no idea. And frankly, if somebody tells you they have a good idea what's coming in the next five, ten years, let alone two years, they're probably full of it and shouldn't be listened to. They're trying to sell you something and don't have a good basis for it. The only answer in these kinds of times is to be extremely vigilant and curious and somewhat tied to an empirical reality. Our perceptions of what's going on are very likely to be wrong. And so like, let's do our best to work together as researchers and agencies and policymakers and other interested folks, vendors, everybody really needs to be pulling the cart here to provide the best evidence about how these things are being used and what their actual effects are out in the world before we sort of hit a hundred percent penetration on these tools.

Anthony Berghammer [00:56:27] I mean, I think the biggest thing is be skeptical. I mean you can still be optimistic because I'm optimistic of everything that AI can do, but don't be afraid to be skeptical and don't believe everything that comes out because there are a lot of kind of snake oil sales people out there trying to sell the brightest thing and automate all this and that, and I think being skeptical, but also doing the research and the due diligence to see does this work. What are you hoping to improve in measuring those, doing the pilots and doing the research, but with that kind of experimental skeptic hat on, because I think it's really valuable, but you have to do your own testing and you have to see if, you know, how to use it and kind of grow with it instead of against it.

Christian Quinn [00:57:05] Yeah, and I'm also cautiously optimistic. I think I often like to say that, especially in this profession, tech is not replacing anybody. Tech is helping good people do their best work. And we're still in a window of time where we're figuring out how can the tech help our good people do their best work and how is the tech going to change because there's a lot of really candidly some neat conversations around that where LLMs are probably going to be maxed out a little bit in terms of like computational power and available data to train models but they're still going to evolve because there're going to almost like bolt-ons and add-ons to them that help offset some of their limitations. So I'm kind of eager to see what that looks like. But, you know, to both Ian and Anthony's points about, in the interim, you do need to be, like, really vigilant about, like not getting so far down the road with some of these tools that we outsource certain functions and decisions to them, that we really need to be the masters of the tools and still committed to foundationally the right things, because when we look at You know, the communities that we serve, and like I referenced, you know, getting safeguarding victim stories and things like that. It's really easy to lose that in the noise of time that we save or efficiencies that we realize and those kinds of things. And being mindful of keeping the main thing the main things.

Carlee Ruiz [00:58:29] Well, thank you all for doing this deep dive into AI tech with us and really sharing your experiences and the lessons you've learned so far. It's been a pleasure talking with all of you today.

Dr. Ian Adams [00:58:40] Thanks, Carlee, really appreciate it.

Anthony Berghammer [00:58:41] Thank you so much.

Carlee Ruiz [00:58:42] So if you enjoyed today's episode, be sure to like and follow Just Science on your platform of choice. Our next topic on this AI and policing series will be around AI and body-worn cameras. So stay tuned for that. For more information on today's topic and resources in the law enforcement field, visit rti.org or the Office of Community Oriented Policing Services. Thank you for listening today. I am Carlee Ruiz and this is another episode of Just Science.

Outro [00:59:11] Stay tuned for future Just Science episodes about the use of AI in policing. Opinions or points of views expressed in this podcast represent a consensus of the authors and do not necessarily represent the official position or policies of its funding.