**Transcript**

July 3, 2025, 4:19PM

 **Todd Helmus** 0:03  
Work.  
Qualitative interviews and.

 **Ryan Brown** 0:10  
So.

 **Todd Helmus** 0:12  
Qualitative document reviews.  
And maybe before then, we can go into more specific literature reviews.  
So we will start with the interviews and then turn over to.  
Thinking about the process, the battle rhythm of conducting interview based research, be it focus groups or individual interviews or however whatever term you or specific methodology you use.  
What aspects of that process do you think?  
Are most.  
Have most potential for automation.

 **Brian Jackson** 0:56  
I mean.

 **Ryan Brown** 0:56  
Definitely. Oh, go ahead, Brian.

 **Brian Jackson** 0:58  
No, I I have a really basic one, but I mean like basic data cleaning and de identification.  
You know, so you know we're we're we're we're doing a bunch of interviews and you know, even if you're talking, there's a lot of umm's in there and things that you don't realize you're saying.

 **Todd Helmus** 1:07  
Oh.

 **Brian Jackson** 1:18  
And so, you know, training in AI to be like, you know, any interjection or, you know, simple thing said by this speaker, you know Brian.  
Jackson the interviewee or interviewer can be, you know, masked or eliminated, and then something that just that helped with smart de identification.  
I mean, it's a drop in the bucket, but I think that's a test that like current AI, should be able to do well.

 **Todd Helmus** 1:45  
Have you applied that?  
Have you applied it to the identification of?  
What do you use or what do you use for the umm extraction?

 **Brian Jackson** 1:51  
So we we just don't do the almonds extraction at all.  
And we I mean we.

 **Todd Helmus** 1:55  
You're just thinking it's a potential use Rajat.

 **Brian Jackson** 1:57  
Yeah, that's a potential use, but the the DE identification, I mean we're we've been recording in teams.  
So you know it becomes a global find and replace. But I mean that's just because teams ties everything to everyone's teams accounts. So in some ways in some ways that makes de identification easier.  
But you know using other tools and other recording mechanisms might not.

 **Todd Helmus** 2:19  
Caitlin.

 **Caitlin McCulloch** 2:24  
Honestly, I think that Brian and I, the task that we had with the social media posts is one.  
I don't think it's quite there yet, but I would really, really like AI to be ready for, which is finding rarer examples.  
So we had functionally 500,000 social media posts. We were trying to find violent extremists talking about financing that's going to be an incredibly small amount of this very, very large sample.  
So just the ability to do.  
A reasonably complicated, you know, comfortable jargon filled.  
Like find and search.  
In that case, is just incredibly useful, like finding a signal in the noise.  
Currently, I think that it's helpful, but you still have to do a bunch of post processing because it doesn't understand sarcasm.  
There's a lot of jargon around financing in these communities, etc.  
But like hopefully in the future, that is a task I would love to see even further automated.

 **Ryan Brown** 3:25  
Are you there?  
Up.  
You froze for a second.  
I'll jump in unless someone else is championing the bit here.  
I I really.  
I mean, I am less skeptical, I guess of current capabilities for analysis, which is the most.  
The most burdensome usually.  
Stage of qualitative analysis and the most expensive because right now to do it well, we do it with teams of people. We do simultaneous codebook development, long conferences to synthesize, we do.  
You know, manual creation of Capas for intercoder reliability.  
What I've seen with Muse.  
Is that yes, there are still some weaknesses with sometimes over interpretation of excerpts or under interpretation, inability to notice subtleties in sarcasm or phrasing.  
Is that because it's so much faster you can load?  
Way and analyze way more documents.  
So depending what you're looking for, if you're looking for sort of a gist of a pattern and a very large data set, it's just so much quicker.  
Than it used to be.  
So I think you can iterate on large data sets much more quickly and and I've seen it take essentially what would be like a 20 days of Labor task and turn it into a one day of Labor task for one individual.  
Is it as good? Probably not.  
But it's it's lot faster.

 **Todd Helmus** 5:02  
Go hit Sean and Julie.  
But what?  
What, Ryan? What? What program are you using for that?

 **Ryan Brown** 5:08  
Oh, just using Muse, which I'll admit bias because I I helped pratique a lot with the development of the sort of the user interviews and the user interface for that.  
But yeah, Muse is Muse is 100% what I'm using.

 **Todd Helmus** 5:23  
That's pretty cool.  
And that's still sort of in the sort of beta phase, right?

 **Ryan Brown** 5:27  
Yeah, it is. It is.  
But I think he's.  
I think it's got enough bandwidth and processing now that that folks who are interested can get on it, but it's you can just go to qdarand.cloud to experiment with it.

 **Todd Helmus** 5:40  
Yeah. Thank you, Sean.

 **Sean Robson** 5:44  
Yeah, I think you know from my perspective, we've used it for again like social media posts really to recruiting project. And you know I think we found it useful for like generating themes, the classification of content to those themes was less useful, would make a lot of mist.  
But I think in we didn't invest in this, but I know that it's certainly possible if you if you spend the time depending on your project you can.  
You know create.  
You can sort of follow up with just NLP.  
Do some traditional machine learning to sort of improve the accuracy of which it's going to sort of classify those?  
So it's certainly possible, but it's just I think it depends for us the scope of our project wasn't wasn't large enough that we needed to invest in that.  
So we we took the initial classification to those teams and then we just had the Ra's essentially go back and do sort of a human review of of what the language models coded.

 **Todd Helmus** 6:44  
So I'm saying I'm saying, you know, using the machine to help.  
Speed some tasks along, but then having a human interface on that is important.  
Uh, Julia?

 **Julia Kaufman** 6:58  
Yeah. I just want to follow along with with Ryan because we we use music.

 **Todd Helmus** 6:59  
And then Caitlin.

 **Julia Kaufman** 7:03  
I mean, we gave Prateek feedback on Muse because we specifically had we had a very specific use which is like when we get all this open-ended data you know thousands and thousands of rows, it's right there and you know it's it's it's short.  
It's I think it's really good for a thing like Muse because they're just like short open-ended responses that people wrote in, in a survey. And there's so many of them, right?  
So it just saves us so much time to just like feed it into something like Muse.  
And I'm just like doing that right now to to, like, remind myself with the new set of data as I'm talking to you guys. And it's just like it just gives so much grounded up information so quickly.  
But of course, you know the thing that you have to do afterwards is like verify that, yeah, these are the the codes that you're in and feed it new codes and like code up things and you have to do a lot of back and forth.  
But I think once we're good at it, it just saves us so much time. And the other thing we're doing it for to Brian's point, is just.  
Try we have to de identify.  
By the open-ended responses before we give it out to people because we want to make sure that no one can identify the respondents, so we have to take out all the location information and things like that.  
It's very good at big data on that point too, so I really appreciate that in particular.  
There are other things I could talk about, but I'm interested to see what other people say.

 **Todd Helmus** 8:16  
I'm going off camera to save a little bandwidth here, Caitlin.

 **Caitlin McCulloch** 8:23  
I think that I just wanted to tap in on that and say I think that I really see the promise of AI tools and I think it's what they are helpful in research. I think that the thing that I always want to some people is like the human.  
Interface and I think that the part that makes me worried about using it in research analysis is that I do feel like there are some folks who are not doing delayed post processing. The revision deederation for things and I think that because it's so quick and you get.

 **Julia Kaufman** 8:50  
Yeah.

 **Caitlin McCulloch** 8:50  
Such a nice analysis response.  
Like you can just sort of be like, oh, it answered my question.  
It's all done like this is great.  
I can sort of move on from this point of point.  
And I think that that's super dangerous.  
So I think like the human interface is what I really want a foot stomp and possibly having like.  
Some discussion of that honestly, even in Rand research or products and like sort of the iteration that you went through or the the things that you tweaked on the human side?  
I think would be really cool in encouraging use of these methods.  
But also sort of in encouraging use of integrity in these methods.

 **Todd Helmus** 9:33  
Brian.

 **Brian Jackson** 9:34  
Yeah, I mean, so mine mine's a little bit of a related point to that.  
I mean, so a long time ago, during the Iraq war, what somebody in the military intelligence community said to me is their complaint about Palantir was that it allowed a user who didn't understand what the tool was doing to take garbage data and produce an absolutely beautiful social.  
Network graph that resulted in somebody dying.  
That may or may not have been the right person to die.  
So I mean, So what?  
This comes down to for me about the AI tools is like I mean the data looks wonderful like at the end like, you know, oh, you know, I've gotten all of this stuff. What I think Rand needs to figure out how to do is is like, what are?  
The best.  
I mean, Caitlin talked about, you know, sort of like laying out what human checking you did. But like, what really are the best practices for validation, I mean?  
And Caitlin and and my project, you know we we sort of.  
Had models competing with each other and you know some of them were better than others. Some of them were worse in different ways.  
I mean, we could sort of do that because it was a grant that we could sort of go in different directions on. But you know, kind of that question about like what does validation mean?  
Like, what do we think the error bar is like? You know what?  
You know, you know and and maybe it's how much checking you know on your output is is like Rand practice.  
To be able to say like OK, I checked.  
This much so you can trust the results this much because you know circling into our quality assurance process like this is something that we need to explicitly define what good enough is because it will never be perfect.  
It's a black box and so like how much do we need to know and how much do we need to articulate for the result to be, you know, up to Rand's quality standard.  
That's what people.

 **Julia Kaufman** 11:36  
Yeah. I just want to add too. Yeah, totally.  
I agree with everything you guys said about what needing after you generate something checking it, but also the more information you can feed into the models. Like if you have things coded up already like 200 rows in my examples are coded up. It does so much of.  
A better job. So the more you can give it information, the better it does too.  
But the checking is also even after you give information, you just have to keep going back and forth the iteration.  
Is so important.

 **Todd Helmus** 12:06  
How much are you guys finding?  
It seems like so they're, you know, there's some complex methods. It seems to to use this technology.  
I mean what?  
And I get it that Rand has developed a tool and and maybe this is what goes to what Ryan mentioned. I think with Alina developing some sort of process, but you know is there a is there a playbook that you're using for this?  
Are you trial and error?  
Are you leaning on fellow conversations with other Rand analysts to to think about?  
How to use this? Or you just like winging it? Like I'm gonna just mess around with this AI system and see what it can generate Caitlin.

 **Caitlin McCulloch** 12:45  
No, I can't see jumping on things like clearly I'm very strong in everything and I think it's a mixture for me. So I taught a methods class for the masters and one of those was.  
Search and so did sort of a deep dive, kind of like chat engineering, you know, prompting like that. And what it does to your results and did a bunch of like toy examples of if you are shifting things in prompts, what does that look like? So like some?  
Of that, some of working off folks at Rand, we work with Qiu on who is just this great programmer.  
Are so like learning from how she has, who's done this? A bunch has handled things.  
But I don't think that there's like a formal process, and if anything, I think there's actually a lot of people basically being like, go forth, do all the AI to just like everyone at Rand, which is great and you should play in that sandbox.  
But I would actually love to see more like.  
Written frameworks or guidelines or like hey, this might be helpful for figuring out if you're having a lot of variation in your thematic you know the categories that it's offer you or something similar.

 **Todd Helmus** 13:47  
Ryan, thank you, Caitlin.

 **Ryan Brown** 13:50  
Yeah, I think it's very doable.  
I mean, so I've I've been doing qualitative work for like 20 or 30 years and I use what is considered gold standard.  
And there's, you know, some variation there. But in human fully human teams.  
Implement that with with the large language model enabled analytic platform.  
And then view the LLM. Essentially a tireless research assistant. So in a lot of cases you have one or two, you know π and then qualitatively, then you have multiple Ra's and some of them you're teaching a code for the first time.  
So you you can ask them, redirect them, redefine the codes.  
Put ad inclusion exclusion criteria.  
Run integrator reliability.  
Rerun it with the new instructions.  
That's essentially what you would be doing over weeks.  
So I kind of view you think about what's how you've staffed the project with AI and soon we'll have multiple agents to write. They'll be able to take different perspectives on the data.  
Think about what you use teams for because of exhaustion or the, you know just the need to have parallel computing power.  
Cognitive power and see what can be implemented with AI and then.  
Eventually, we're gonna need to run a large parallel experiment where you have, you know, your full gold standard human coding team.  
And then you have, you know, various AI processes and you see what you do a comparison, and then there's probably gonna need to be something in the academic literature that shows that.  
And then folks are gonna be more comfortable.  
Of course, that's not gonna test for all cases, right?  
There's different styles of data, different, you know, but, but I think it's it's quite.  
Doable to? It's just we haven't.  
Done all that testing yet?

 **Todd Helmus** 15:43  
Other observations on on how we're learning to use this, Kimberly.

 **Kimberly Jackson** 15:47  
So this is.  
Maybe. I mean you can tell me if this is beyond the scope here, but I what I've noticed is so one a lot of the projects that I'm working on.  
I think that people are approaching these tools with just in that ad hoc manner of like I'm gonna play around with this and and see what happens, which I don't.  
For the record, I don't think is necessarily a bad thing because we have a lot of people here who are. Aii wouldn't say that they distrust AI, but their AI.  
Newbies and wow, this thing's actually really cool.  
But it is related to my broader point, which is.  
I'm getting. I think that there's sort of like this. Of course, we're in this institutional curious.  
We're in this at a point institutionally where we don't know what our role is and what our future is.  
I think everybody's a little bit scared of of of what's next, right?  
And I think that's fair and it's a real legitimate concern that things that we used to put a bunch of labour behind.  
And were hallmarks of our giant doorstep, reports our literature reviews, and I think that there is some hesitation to, I mean, AI can be excellent for a lit review, right?  
It's not perfect.  
We got to check the sources got all that but.  
At a time when we are realizing that the sponsor wants less of the 30 page lit review that they wanted before or we thought that they wanted before.  
And at a time when the machine can do it, a hell of a lot faster.  
With a lot fewer resources.  
Than than humans can.  
I think that it's a.  
It just introduces an additional layer of. I don't know if it's mistrust or by using this tool.  
Am I going to?  
Is that gonna reflect back to me that I am no longer as relevant as I thought that I was?  
There's, I hear a lot of these conversations.  
I I'm the first person to tell you. I think it's not.  
Of course you have to have humans making the decisions and providing the nuance and the perspectives.  
I got it but.  
I think that that's that's that is a.  
At least a consideration, if not an impediment, for some.

 **Todd Helmus** 17:58  
Any thoughts on that observation?

 **Nathan Beauchamp-Mustafaga** 18:03  
I mean, so I think in in many ways the typical.

 **Todd Helmus** 18:07  
Vary from go ahead.

 **Nathan Beauchamp-Mustafaga** 18:09  
Sorry, the typical Chinese language work we do.  
I mean, I call it glorified lit reviews, right?  
It's the same fundamental thing. We just on the China team call that the product, right?  
It is like we found a bunch of literature. We read a bunch of literature, let us tell you what it says. And if we over perform, it's let me tell you what it means.  
Like you know, and so the litter view on our end is like, what does everybody else you know?  
I didn't see s s already tell you about this question.  
So we.  
I would like for it to do good at litter views for Chinese single stuff, but like.  
I think the the challenge, at least for us, which is a niche case, is that it's really good at like surveying a general pool of data.  
But you know the way we do it like I need to be able to identify different services within military.  
Like, how do they think about it? And so being able to like segment out different groups and identify them and then like, oh, within, you know, the Chinese Air Force, they think about it this way or the Chinese, you know?  
Do you think about it this way?  
Those are the kind of hallmarks of being able to do good, like trying to trying to military research and so.  
So far, at least, you know I haven't been able to find.  
Good tools that can quickly let you segment it out and then kind of have like stacks. And so one of the tools we're trying to build.  
Is the ability to do like custom rags on like define corpus of text which would basically be like I have a hundred articles but I wanna label them and segment out like OK.  
Here's the Chinese Navy articles.  
And then tell me what it says.  
And so, so far at least we on the China team aren't really there to actually make that realistic.  
And so thus like it's not actually very practically useful yet, which is one of the reasons why nobody on the team releases AI for their research.  
But like that would be. You know something that would be like really practical and actually speed up some of our research. I think the other part of it that I would like to get better at is oftentimes like Chinese military, for example, preformulaic.  
There's like problem statement research we did implications blah blah blah.  
And so like again being able to do bulk, but not just on the whole article, but to be able to say like.  
Oh please only read section three or four that has the word implications in it.  
Like pull that right and I don't know. So far. Maybe we're just too basic, but that is again, like a step beyond really where we're able to get so far.  
So those are like 2 components that I think we're trying to work on eventually for the China team stuff.

 **Todd Helmus** 20:55  
Caitlin then Brian, and then I'll switch gears a little bit with some different questions. Go ahead greatly.

 **Caitlin McCulloch** 21:01  
I feel like Nathan's point just jogged something for me, which is like I feel like AI currently can read, write like a mediocre RA literature review.  
Like a literature review from someone who doesn't really understand the topic that well. And like sometimes, to be fair, I feel like that's what you need.  
You don't need like you know, to be able to, you know, fluctuate out the services or know that this is, you know, particularly important piece of writing and that it's based off that piece of writing. And I feel like.  
It can definitely serve that role.  
It's just to me not to the role where I can serve like.  
And associate researcher or like a higher level policy analyst literature reviewer above. Basically like it can't give you or even like an A+ RA literature review.  
And then my secondary thought is like.  
I think you know this is slightly off everything except for the like picking like A1 chunk. Part of what we're doing with the DoD instructions and directives when we were having AI sort through them as we were looking for a very particular topic. And I think what.  
I worry about sometimes with AI is like.  
Like we did the post checking on it, it was less than 50% accurate because it hadn't identified a single keyword like there was one keyword that it had just had not attached to. The fact that it was this sort of thing.  
And like with iteration, probably could have like gotten past that, but then I would worry that there would be another keyword.  
Do you know what I mean?  
So like I just think.  
That's part of the reason why I'm so like I'm gonna say I'm gonna be kind to myself and say skeptical of AI and its current uses, but thanks.

 **Todd Helmus** 22:33  
Brian and I want to ask about, like, what are the like after Brian talks?  
It'd be helpful to go through like for for literature reviews or document reviews like what are the, what are the steps of that process that are that are relevant to the automation. Go ahead, Brian.

 **Brian Jackson** 22:52  
Yeah, I mean, I'll, I'll kind of bridge into that 'cause. We were sort of talking about the reviews already a little bit, but you know kind of echoing, you know kind of Kim's point about researcher anxiety about these tools. It kind of you know and reinforcing ca.  
'S thing like. I'm not sure how to judge the goodness of the synthesis that comes out of the AI tool. Like, you know, I had projects early in my career.  
That was, you know, taking.  
A body of work that was not really related, you know.  
Or was was related but not focused on what I was trying to understand. This was technology and you know, technology used by terrorist groups.  
And so I was trying to find the insights, you know, that poli sci people had had about technology that they might not have realized were insights about technology by reading a whole bunch of stuff.  
And then synthesizing it and like that was something that I could do. I'm sure I could ask an AI to do that, and it would produce a result.  
And like understanding.  
You know, I I might be able to understand whether that's a good or a bad result now, but like many people, don't.  
And so there there are metrics where I can look at AI output and go wow, that's amazing.  
That's good.  
I know that's quality and one of those is, you know, sort of the bridge to the lit review workflow and that's the searching part, like if the AI finds something unexpected that I look at and go, wow, that's cool.  
I never would have looked in that journal.  
I never would have, you know, used the specific keywords that.  
Would have gotten that.  
That's that's a time where I could be like, Oh yeah, definitely that's quality.  
Like AI brought goodness to this.  
That wouldn't have been there otherwise, but on on the things like the synthesis, even though it's fast and it it, it's mostly there like I have a hard time sorting out how to judge goodness and that's particularly threatening for a research organization that.  
'S entire sales pitch is goodness.

 **Todd Helmus** 24:51  
I mean, what is the when when you say you're asking the lit review to do synthesis, what is that?  
What is that?  
What does that refer to?

 **Brian Jackson** 24:57  
Well, I mean that's that's that's the you know you I mean in in some ways that's you know everything that we've been talking about about any qualitative data set you know I mean tell me what the themes are in this you know tell me what you know you.  
Know bring together these resources into a discussion that speaks to this question.  
I mean that's that's sort of synthesis. I mean, in my old old work, I was asking the question like, is it as easy for terrorist groups to get technology and use?  
Actively, you know, as most people at that point assumed that it was. And so for me that was, you know, sort of going through and like Easter egg hunting and weaving it together and trying to decide what was credible and what wasn't.  
I'm sure an AI could do that and would spit something out.  
But like I wouldn't be able to judge whether it was good or bad.

 **Todd Helmus** 25:48  
So one one piece, talk about Easter egg content.  
I mean, I know there's there's various different ways that people approach literature reviews, with some people going the route of, you know, building a spreadsheet and like, you know, these are the 50 documents that matter. You know, these are the 3000 documents reviewed based on parameters. You.  
Know these two, these 500 meta criteria for the literature review.  
Of these 300, I've now identified like 815 rows in Excel files.  
You know what is the age of the participants and what are the entry criteria for the participants or you know what were the key findings with respect to A and then B&C?  
You know, like really breaking it out.  
I don't always do literature reviews that way, but I know it's a it's a more systematic approach that some people use, and I've heard that AI can be good for some of that, but not all of that.  
Any sort of experience in in that area?

 **Julia Kaufman** 26:46  
Yeah, I I feel like Ryan and I are like Muse muse.  
We should use Muse, but I do think that Muse it doesn't like if you feed it abstracts it.  
I don't think.  
I totally think you know, I don't think we should ever expect that AI is going to give you a high quality research level.  
Output right now, like if you trained it maybe, but if you don't train it, I don't know why you would expect it to, right?  
But it I think the pieces of literature review process, if you have the right.  
Tools. It's pretty good at it.  
Like if you feed it like lots of abstracts like, I think again really big data. It does a really good job of helping you get to a smaller set of data to well, like if like.

 **Todd Helmus** 27:26  
In what way?  
How are you, Julie?  
How are using it in a way that that that that that you feel?

 **Julia Kaufman** 27:31  
Like five? Yeah, yeah. Like 500 abstracts, right.  
And you're like, here are my inclusion criteria. Here are my exclusion criteria.  
Find me the ones that fit right or even more than that, right, of course.  
Again, checking back to be like you know, let's look at fifty of those abstracts.  
How well did it do that? You know, that kind of thing. But I think it is a time saver in that way.  
But I again I totally agree.  
Like if you expect.

 **Todd Helmus** 27:54  
So giving it very clearly defined tasks narrow scope tasks.

 **Julia Kaufman** 27:56  
It.  
Yeah, very clearly defined tasks, but also the training data that it can use can get going better and better I think.

 **Todd Helmus** 28:05  
Sean.

 **Sean Robson** 28:07  
Yes, a couple other things.  
So we, I mean we've used it for like extraction tasks and it's like it's done a phenomenal job of you know for particularly in job descriptions.  
And so we're able to, you know, give it very specific instructions in terms of like extracting the job title, you know, what are the knowledge, skills, abilities and it does that we had a human check it and it's essentially 100% accurate in terms of just ext.  
The information that we need.  
The the I'll say in terms of literature review.  
John David's like been working on like ask Grand 2.0 and with Grafrag and it you know, it'll put things into different categories in terms of like methods, you know, results, findings, recommendations.  
The problem is that you know it.  
It'll tag things you know, like if if you have a Rand report that, like criticizes, like other research in terms of their findings, it puts those, it just logs those as part of the findings.  
And so you really have to sort of pay attention to to that.  
So there's still.  
I think there's still quite, it's just it kind of loses the full context when you're using that, that technology.  
I do think it's going to improve, but I just think sort of the current state, it makes it difficult to do it automated.

 **Todd Helmus** 29:26  
Ryan. Then we'll switch gears for a last few questions.

 **Ryan Brown** 29:30  
Yeah, one of one of the most labor intensive.  
Parts of a literature review in the classic gold standard way.  
Is a sometimes 10 page long list of brainstormed keywords and that deleted keywords and added keywords and different Boolean parameters.  
Lom's are fantastic at doing that.  
Without formalizing, we're doing the search right now.

 **Todd Helmus** 29:51  
I mean for doing the search for searching the documents, yeah.

 **Ryan Brown** 29:54  
I think some initial synthesis.  
I agree, Shawn, that when you get to very.  
Fine grain interpretation of articles it it can skip layers of inference.  
That's talking about, especially if you're trying to review reviews.  
So it like one of the one of the the standard techniques for trying to get something quickly is review of reviews and it will skip around whether it's talking about something at one level of abstraction or another.  
But yeah, the searching, it's searching and getting me to like a subset of articles that then I'll manually look at.  
It's been very useful.

 **Todd Helmus** 30:35  
All right.  
Last question, last two questions I have and maybe go through and you can answer one or both as you wish.  
I'm curious.  
Are there risks associated?  
What risks you see associated with adopting?  
AI for methods at Rand and what can you know what is your take on institutional adoption?  
And are there things that the institution can do to promote responsible adoption? Caitlin.

 **Caitlin McCulloch** 31:06  
I mean, obviously that's what this I think that the institution has to promote responsible adoption because otherwise we will be left behind in a tool that will make research much faster and much easier.  
So I think like AI use is inevitable in some ways, and I think that we need to think really seriously about how we want to use it and and what what ways are best for RAM to continue to use it in a way that promotes our integrity and.  
I think.  
The thing that I would lean into the most is transparency.  
For the human interface and to have iteration and revision and what the process for using it looks like on the research side, helping folks to really approach it ad hoc play use the tool, but also have sort of guidance for what you're going to need to to show.  
Up to QA with so that people will believe that you use this in a responsible way and know kind of you didn't just go. Here's a giant chunk of data AI tell me what my themes are.

 **Todd Helmus** 32:02  
OK.

 **Caitlin McCulloch** 32:05  
Give me a little output that shows.  
Where they are.  
And I think that.  
Doing so early and having those frameworks in place early would really set us up to use AI more, but also to use AI in a way that would hopefully like set standards for research outcomes.

 **Todd Helmus** 32:27  
Nathan.

 **Nathan Beauchamp-Mustafaga** 32:31  
My concerns, honey, I, you know, like everybody said, are just like the pros and cons is like it's a good RA, which is say we don't let Ras write reports.  
You know, like if I say, go find me some stuff on the Internet from that's in Chinese.  
Tell me what it says like, you know, my entire job is to have the nuance of, like, who's writing it.  
Where are they writing it?  
All these things and so.  
He's yeah, I can do that.  
But does not do that, you know immediately.  
And naturally.  
And so that's always my fear, right?  
That overlook something or miss something?  
So the the you know the ability to flatten it down can be very good, but also very.  
Miss the rich nuance of it.  
So that's I guess my concern for it.  
2nd in terms of what you know, rank can do institutionally to build on Caitlin's point. To me, it seems like there are, you know, two kind of competing, well, multiple competing priorities like we need a signal.  
That were using AI in our research to show that we're interested in experimenting with it.  
Flip side is you wanna say I just simply completed this report with AI for you, because then they're gonna be like, well, one. Why did I pay you half $1,000,000 to use, you know, AI to do this in 1/2 day or two.  
Like, well, I don't as a sponsor person, I don't believe AI. And so you know, now I don't believe anything you've told me. And so it feels like ran needs.  
Maybe some a standard disclaimers you can provide with standard?  
Levels or like gradients of like how did I use AI? My research and like have some boilerplate language that we can include in briefings and reports, and you know, like Rand has a scale of how much of my research did I outsource to AI from 1:00 to 10:00?  
This project is on a three like, you know, this project's on an 8 and we took the risk. But you also gave us $75,000 in assets to get your report in six weeks.

 **Todd Helmus** 34:24  
Good.

 **Nathan Beauchamp-Mustafaga** 34:28  
So like you agreed to that risk and that's OK. That's where we live.  
Right. Like I don't know it just like maybe that'd be one way to help define it would be good for QA. People could have to assess themselves and talk through anyway.  
That's my idea.

 **Todd Helmus** 34:42  
Feel like an upfront approach to transparency.

 **Nathan Beauchamp-Mustafaga** 34:43  
I just feel like I need a wall.

 **Todd Helmus** 34:45  
On on, on how we're gonna use it in in particular projects.

 **Nathan Beauchamp-Mustafaga** 34:49  
Yeah. And I think it would force some definitions and it's always gonna be, you know, mixed.  
But like, if we are honest with ourselves, there are reports that RA is right. 75% of it and no one asks questions and no one knows 'cause. It's a good RA. There's some reports.  
It's very obvious to, RA wrote.  
All of it. And the senior person put their name on it, right?  
So like the AI is actually just recreating some of the problems we already have.  
Question is like is it a different color of problem? You know, do we find it more naughty when people do that?

 **Todd Helmus** 35:21  
I like that question. Do we find it more naughty?  
Hi, Kimberly.

 **Kimberly Jackson** 35:27  
So I think building off of what Nathan said, I mean, even if in even if we had a requirement.  
For every project that requires a methodology section, it's just.  
A AI was used for this, this and this and and it.  
It can be even if we could never.  
I love the idea of like the 1:00 to 10:00, but that sounds like a Rand Bridge too far, even though I like the idea.  
But even if it was just like you were expected to have in like.  
A paragraph that says how AI was used in this report, if at all.  
And that's this says nothing of the the standards and how you ensure and you know all of that but.  
Even starting with that and and in terms of how do you get, how do you get the institution to incorporate it?  
One, I think you do, like Nathan talked about, you do need some kind of standard. Otherwise I think you know we would our our collective brain would explode.  
We have to have a standard, but just from.  
From the Department of Defense perspective, and this is a, this is a massive organization that is trying really hard to do 2 things.  
One adopt AI themselves to make their jobs easier.  
And they're trying to figure out how to do that.  
So to the extent that we can communicate this as we're trying to reflect the the needs and and the similar terms of our sponsors makes sense. And then the second thing is they're trying to do things a lot faster.  
I think we're seeing that a lot.  
They want.  
They want quicker turn reports.  
They want shorter reports.  
They they want the analysis at time of need, and AI is a primary way to do that quickly.  
But we have to do it with all of the other caveats we have to do it responsibly.  
But it's I I think a bit more of a mandate than a a nice to have and I've been treating it as a nice to have.  
I'm like, I'm in that camp, right?  
And I found it useful so far, but.  
Because it is not incorporated into the QA process in a meaningful way, because there's such uneven knowledge of how to use it.  
And how do I what?  
I don't know how to do which would actually be very useful is, you know even like as Brian was talking about. And Shawn was talking about, I don't know how to do some verification of this stuff. Like is it gonna be? Is it gonna be more work for?  
Me to.  
To ask for a summary of X thing and then go back like Wikipedia, right?  
Go back and check every single thing that's in there.  
And is that gonna be worth my time?  
So if we could have more guidance on the the best practices of how to ensure that the AI thing you generated is actually accurate, that would be, that would for sure facilitate my use of those tools.  
Not sure about anybody else.

 **Todd Helmus** 38:13  
OK.  
Thank you. Best practices guide, Brian.

 **Brian Jackson** 38:16  
Yeah, I mean I I related point, I mean I would say it's not so much an AI disclaimer in the final report in the kickoff and whatever, but it's the AI protocol. You know in this project we only plan to use AI to reduce our 400 page.  
Report that we will reflexively write to A50 page report that you can actually use by reading that we'll know that it did an OK job doing that versus, you know, in this project we plan to use AI to categorize and code the fundamental data. That's the found.

 **Todd Helmus** 38:33  
Yeah.

 **Brian Jackson** 38:44  
For all of the conclusions that we're drawing on.  
On this we are going to verify that by manually checking 10% of those categorizations.  
If those work, we're going to say that's good enough. If this many are wrong, we're going to check another 10% and then you'll figure out what tweaking that we have to do until, you know, we get to a standard that we can articulate.  
But you know, as as as Kimberly said, like, not everyone at Rand.  
Knows to think that way you know, or you know, knows what's good enough.  
You know, I mean, for me as a scientist, maybe doing that, you know, I check 10%. You know I found two errors out of 200.  
We're good, you know, no problem.  
But like I could see somebody else who went, oh, my gosh, there were two things wrong.  
Like what do we have to do now, you know?  
So figuring out like how the institution's gonna handle that, and like, set the expectations I think would be super important and super, I mean and and that would make a lot of people more comfortable I think about adopting because then they could articulate like.  
Oh, you know, I followed step AB and C.  
And here are the results and we're good.

 **Todd Helmus** 39:51  
Chip last quick question.  
Should should the business units require researchers to develop AI plans for the research ahead of time? We will this project.  
We're not going to use AI this project we as as a way of like promoting adoption.

 **Ryan Brown** 40:08  
That sounds like an extra layer.

 **Brian Jackson** 40:08  
It it'd be more.  
Yeah, it'd be more useful than a lot of the things we do in kick off meetings.

 **Ryan Brown** 40:10  
Oh, go ahead.

 **Nathan Beauchamp-Mustafaga** 40:10  
Yes.  
Yes, yes, I think.  
Even if people say, oh, I don't plan to use it and then it's well, we should, you know, yeah. Experiment and you'll figure it out.  
But like I mean, I'll just be honest with the China team. I I literally don't know of anybody experimenting with it.  
Maybe the RIS are right?  
They're great, but like.  
One is not functionally immediately practically useful. Sure, so it's easy.  
But like no one else is interested, right?

 **Todd Helmus** 40:40  
So there's a risk that that requiring that could push researchers to maybe adopt approaches that might not be very useful or relevant.

 **Nathan Beauchamp-Mustafaga** 40:48  
I'm very worried that in three years we're in China, work will be antiquated, like it'll be like irrelevant, right?  
I mean, are is gonna get \*\*\*\* data and turn out great reports and call it cool. And people are gonna be like, well, they gave it to me in 90 days.  
So that's all I need, right?  
Like and I'm gonna beat them on data quality because I get better sources than they do, but like.  
You know we can't.  
Structurally, we can't be minor in AI and so.  
Like you know that I think that's a huge trade off.  
Ray needs to, you know, I'm saying just for the trying to China part. But like that's my concern.

 **Todd Helmus** 41:29  
Thank you.  
Hey, y'all, it's 1:00.  
I appreciate you taking this time for the conversation.  
It was really helpful if you have any other thoughts, feel free to hit me a note.

 **Nathan Beauchamp-Mustafaga** 41:39  
Can I get one more idea that might be relevant? I think nsrd primarily, but all the divisions should go and buy programmers time and be like you're 50% your entire like 50% of your entire job is working on nsrd projects and then they match and pay.

 **Todd Helmus** 41:41  
Yeah. Thanks Julian.

 **Nathan Beauchamp-Mustafaga** 41:58  
For it on the back end with Nsrd project money, you have to do it.  
Arrow should do it, but like.  
You know, basically make sure there is time available and then for projects that are relevant.  
Big oh, you don't know much about AI.  
That's OK. Like you know this nice person over here.

 **Todd Helmus** 42:12  
You can hire a consultant you can like, bring a consultant on your project.

 **Nathan Beauchamp-Mustafaga** 42:13  
We just hired. It's gonna like, be your AI person.  
Yeah, like forced staffing.  
That will increase uptake.  
That's what I want, right?  
I want like a dedicated AI person, like a programmer on my on my project.

 **Brian Jackson** 42:27  
The other nice thing about that I've had AI people who didn't understand what the project was doing and that was frustrating and and difficult.

 **Nathan Beauchamp-Mustafaga** 42:27  
Yeah, that's fine.

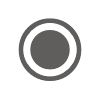
 **Brian Jackson** 42:37  
So having something like that that created a requirement for understanding I think is important.  
I mean because that feeds back into the checking.  
What checking means depends on what the project is doing.

 **Nathan Beauchamp-Mustafaga** 42:49  
Sure.  
Yeah, but you have to reserve their time ahead of time, right?  
Like the reason my project is slow is because they're all busy doing random GR things, so they can't tell me about, right?  
Like they're all extremely busy, so you have to like, buy their time now so that they actually have time when products come in anyway.  
That's my my nsrd management pitch.

 **Todd Helmus** 43:07  
That's helpful.  
Thanks. Thanks for that, Nathan.  
Alright guys. Thank you.  
Do you have any other thoughts? Send me a note or hit me up.  
But thank you for your time.  
This has been really helpful.

 **Nathan Beauchamp-Mustafaga** 43:17  
Thanks Todd.

 **Ryan Brown** 43:18  
Thank you.

 **Todd Helmus** stopped transcription