Official Transcript of Proceedings

NUCLEAR REGULATORY COMMISSION

Title:	713th Meeting - Advisory Committee on Reactor Safeguards
Docket Number:	(n/a)
Location:	Headquarters and Teleconference
Date:	Wednesday, March 6, 2024
Work Order No.:	NRC-2750

Pages 1-148

NEAL R. GROSS AND CO., INC.
Court Reporters and Transcribers
1716 14th Street, N.W.
Washington, D.C. 20009
(202) 234-4433

DISCLAIMER

UNITED STATES NUCLEAR REGULATORY COMMISSION'S
ADVISORY COMMITTEE ON REACTOR SAFEGUARDS

The contents of this transcript of the proceeding of the United States Nuclear Regulatory Commission Advisory Committee on Reactor Safeguards, as reported herein, is a record of the discussions recorded at the meeting.

This transcript has not been reviewed, corrected, and edited, and it may contain inaccuracies.

UNITED STATES OF AMERICA

NUCLEAR REGULATORY COMMISSION

+ + + + +

713TH MEETING

ADVISORY COMMITTEE ON REACTOR SAFEGUARDS

(ACRS)

+ + + + +

OPEN SESSION

+ + + + +

WEDNESDAY

MARCH 6, 2024

+ + + + +

The Advisory Committee met via hybrid In-Person and Video-Teleconference, at 8:30 a.m. EST, Walter Kirchner, Chairman, presiding.

COMMITTEE MEMBERS:

WALTER L. KIRCHNER, Chair

GREGORY H. HALNON, Vice Chair

DAVID A. PETTI, Member-at-Large

RONALD BALLINGER, Member*

CHARLES H. BROWN, JR., Member

VICKI M. BIER, Member

VESNA B. DIMITRIJEVIC, Member*

JOSE MARCH-LEUBA, Member

ROBERT P. MARTIN, Member
THOMAS E. ROBERTS, Member
MATTHEW SUNSERI, Member

ACRS CONSULTANT:
DENNIS BLEY
MYRON HECHT
STEPHEN SCHULTZ

DESIGNATED FEDERAL OFFICIAL:
CHRISTINA ANTONESCU

ALSO PRESENT:
STEVEN ALFERINK, NRR
LUIS BETANCOURT, RES*
NORBERT CARTE, NRR
STEPHEN E. CUMBLIDGE, NRR
SAMIR DARBALI, NRR
MATTHEW DENNIS, RES*
JARED GILLESPIE, PNNL*
RICHARD JACOB, PNNL*
IAN JUNG, NRR
CAROL A. NOVE, RES
JASON PAIGE, NRR
PRADEEP RAMUHALLI, ORNL

STEVE RUFFIN, RES
MICHELE SAMPSON, RES
SUNIL WEERAKKODY, NRR

* present via video-teleconference

CONTENTS

P-R-O-C-E-E-D-I-N-G-S

8:31 a.m.

CHAIR KIRCHNER: The meeting will now come to order. This is the first day of the 713th meeting of the Advisory Committee on Reactor Safeguards. I'm Walt Kirchner, Chair of the ACRS.

Other members in attendance are Ron Ballinger, Vicki Bier, Charles Brown, Vesna Dimitrijevic, Greg Halnon, Jose March-Leuba, Robert Martin, David Petti, Thomas Roberts. And I believe Matt Sunseri will join us shortly.

MEMBER SUNSERI: I'm online.

CHAIR KIRCHNER: Thank you, Matt. Our consultants, Myron Hecht and Stephen Schultz, are also joining us today. And I expect Dennis Bley to join us at some point as well. I know we have a quorum

Today the committee is meeting in person and virtually. The ACRS was established by the Atomic Energy Act and discovered by the Federal Advisory Committee Act, FACA. The ACRS section of the U.S. NRC public website provides information about the history of this committee and documents such as our charter, bylaws, Federal Register notices for meetings, letter reports, and transcripts of full and subcommittee meetings, including all slides presented at the meetings.

The committee provides its advice on safety matters to the Commission to its publically available letter reports. The Federal Register notice announcing this meeting was published on February 16th, 2024. This announcement provided a meeting agenda as well as instructions for interested parties to submit written documents or request opportunities to address the committee.

The designated federal officer for today's meeting is Ms. Christina Antonescu. A communications channel has been opened to allow members of the public to monitor the open portions of the meeting. The ACRS is inviting members of the public to use the MS Teams link to view slides and other discussion materials during these open sessions.

The MS Teams links information was placed in the agenda on the ACRS public website. Periodically, the meeting will be open to accept comments from members of the public listening to our meetings. Written comments may be forwarded to Ms. Christina Antonescu, today's designated federal officer.

The transcript of the presentation portions of the meeting is being kept. And it is requested that speakers identify themselves and speak with sufficient clarity and volume so that they can be readily heard. Additionally, participants and members of the public should mute themselves when not speaking and also silence any electronic devices or cell phones.

During today's meeting, the committee will consider the following topics. The draft final Branch Technical Position, BTP 7-19, Revision 9, Guidance for Evaluation of Diversity and Defense-in-Depth to Address Common Cause Failure Due to Latent Design Defects in Digital Computer-Based Instrumentation and Control Systems. And second, we'll take up later this afternoon, review of part of our triennial review of the NRC's research program.

We will hear about artificial intelligence and machine learning in non-destructive examination and in service inspection activities. At this time, I'd like to ask other members if they have any additional remarks or opening comments. Seeing None, I will now turn to Member Brown to lead us in our first topic for today's meeting. Charlie, the floor is yours.

MEMBER BROWN: Okay. Thank you, Walt. This morning -- you've already announced what the purpose of the meeting is for the BTP 7-19. Jason, would you like to go ahead and give your opening remarks?

MR. PAIGE: Yes, thank you. Good morning. My name is Jason Paige. I'm the Branch Chief of the Long Term Operations and Modernization Branch. And my branch is responsible for implementing the Commission direction in SRM-SECY-222-0076 on expanding the use of risk informed approaches in addressing digital I&C common cause failures or CCF.

Thank you for this opportunity to present to you the staff's implementing guidance which is being incorporated in Branch Technical Position or BTP 7-19. This has been a collaborative effort by our I&C and risk staff in NRR with support from the I&C staff and research. On February 22nd, 2023, the staff briefed the ACRS subcommittee on draft BTP 7-19, Revision 9.

During that briefing, the staff received comments from ACRS members that are related to the Commission direction in the SRM and associated BTP revision. We also received broader comments associated with the staff's long-term plans for the NRC's I&C regulatory infrastructure or comments that are beyond the scope of implementing the Commission direction. During today's briefing, our presentation will focus on implementing the Commission direction in SRM-SECY-22-0076 and addressing the comments received during the February 2023 ACRS subcommittee meeting on draft BTP 7-19.

After the presentation, we are prepared to discuss the comments that are beyond the scope of the Commission direction. However, as a reminder, we are briefing the ACRS on June 27th on all things digital I&C, including digital I&C licensing actions that are expected and that are in house, modernization of the NRC I&C regulatory infrastructure, and digital I&C advanced reactor activities. During the June 27th meeting, we will provide a holistic view on the staff's short term and long term I&C activities.

Thank you again for your comments, and we look forward to our continued interactions with the ACRS. Before turning the presentation back to Member Brown, I would like to emphasize two points made during the February 22nd ACRS subcommittee meeting. First, the staff's approach for addressing the Commission direction on the expanded CCF policy is summarized in SECY-23-0092 which is the staff's annual updated to the Commission on activities to modernize the agency's instrumentation and controls regulatory infrastructure.

In summary, for light water reactors, the staff is updating BTP 7-19. And for advanced non-light water reactors, the staff is utilizing the licensing modernization project which is endorsed by Reg Guide 1.233 and the design review guide or DRG. The staff's approach will be discussed during today's presentation.

The second point is that the Commission direction gave the staff one year to develop and complete the implementing guidance. And we appreciate the committee's flexibility on this issue. It is our understanding that the committee will be drafting a letter related to BTP 7-19, Revision 9.

And we very much appreciate getting the committee's letter or feedback as soon as possible to incorporate in the BTP to meet our one-year deadline. This concludes my remarks. And I turn it back over to Member Brown.

CHAIR KIRCHNER: May I ask Jason a question?

MEMBER BROWN: Yeah, go ahead. Fire away. You're the chairman.

CHAIR KIRCHNER: Jason, you mentioned the upcoming briefing to the committee in June. And you used the word, holistic, summary, I suppose, of where you are in terms of the overall I&C modernization. Do you envision that you would bring all this guidance together in a Reg Guide that addressed I&C in a holistic manner or in other words sweep up all these different positions that have been established over time and bring in a more coordinated set of guidance for the applicants and the staff going forward?

MR. PAIGE: So my intent of using the holistic language was really just to stay that during that briefing, we will provide a status of ongoing licensing actions that are in house. We'll provide an update on actions -- licensing actions that are expected in the future. We'll also talk about our modernization activities to the I&C regulatory infrastructure.

So we'll talk about regulatory guides that we plan on updating. And then we'll also talk about I&C activities as it relates to advanced reactors. So that's what I meant by holistic.

It wasn't necessarily stating that we plan on consolidating guidance. So I apologize for the confusion. But it was really just to provide the big picture of our activities.

CHAIR KIRCHNER: Thank you.

MEMBER BROWN: Okay. I don't have any additional things. We've got plenty of time to finish your presentation and any what I would call robust discussion that may result as we proceed through the Q&A. We've had some of it already in the subcommittee meeting.

And I would be interested to hearing your all's responses to -- I know you took some notes to our questions at that time. And hopefully that'll ease our letter preparations. So Samir, if you'd like to go ahead.

MR. DARBALI: Thank you, Member Brown. And good morning, everybody. My name is Samir Darbali. I'm an electronics engineer in the Office of Nuclear Reactor Regulation.

I'm joined today by my colleagues, Norbert Carte, also an I&C senior electronics engineer, and Mr. Steven Alferink, a risk analysis also in NRR. So we're on slide 3 which is our outline for the presentation. Today first we'll provide some background information by going over the activities that led to the development of Revision 9 of BTP 7-19, including the direction from the Commission and the staff's response.

We will then provide a summary of the

changes from Revision 8 to Revision 9. And we'll go over the changes made to the BTP since the briefing we provided the committee back in September. And we'll finish with some key messages and the next steps for revising the BTP. Next slide.

Here's a timeline of the main activities related to the development of Revision 9 of BTP 7-19. Revision 8 of the BTP was issued in January 2021. Later that year, the staff began the process to develop a SECY to recommend the Commission expand its digital I&C CCF policy and allow the use of risk informed approaches to demonstrate the appropriate level of defense-in-depth for high safety significant systems. And in August of 2022, SECY-22-0076 was issued.

The staff provided a supplement to the SECY in January 2023 to clarify the importance of point 4 of the policy. And in May of 2023, the Commission approved the staff's recommendation with some edits and directed the staff to develop implementing guidance within one year. The staff began drafting Revision 9 of the BTP in the summer of 2023 and briefed the committee in September of last year.

The public comment period started in

October and closed in November. And after that, the staff addressed public comments and went through concurrence reviews. As Jason mentioned, we briefed the digital I&C subcommittee on February 22nd. And that leads us to today's briefing.

And finally, we're expecting to issue the final BTP in May. Next slide. As mentioned earlier, the Commission provided edits to the four points in the SECY and directed the staff to clarify in the implementing guidance the new policy is independent of the licensing pathway. And so the expanded policy is technology inclusive and applies to all reactor types and includes operating light water reactors, new light water reactors, small module reactors, and non-light water reactors.

The Commission also directed staff to complete the final implementing guidance within a year. And it is important to know that if SRM-SECY-22-0076 did not modify any parts of SRM-SECY-93-087, then those parts of the original SRM are still applicable. Next slide. So as Jason mentioned, even the Commission direction, the staff has a path for addressing the Commission direction for light water reactors and for non-light water reactors.

For light water reactors, the staff's

response to the Commission direction is to revise the guidance in BTP 7-19 or the review of risk informed approaches which may result in the use of design techniques other than diversity. Because of the one-year metric to final implementing guidance, the staff has focused the edits mostly to incorporating the expanded policy. And we have also made changes to address feedback received during the September committee briefing and in response to public comments. Next slide.

For non-light water reactors, the staff provided in SECY-23-0092 an approach for addressing the expanded policy. As mentioned earlier, the staff is using the guidance in the DRG and Reg Guide 1.233 which taken together provide guidance for addressing digital I&C CCF. Reg Guide 1.233 is risk informed and includes guidance on the adequacy of defense-in-depth, and the DRG is aligned with the Reg Guide. The staff is using pre-application meetings with non-light water reactor applicants to discuss the use of the expanded policy and will also communicate this to stakeholders during advance reactor I&C public workshops. The next workshop is taking place next week on March 13.

VICE CHAIR HALNON: Samir, this is Greg. Quick question on -- I understand that this for new applications and maybe design. So for the existing fleet, common cause failure through operating experience was determined. What's the mechanism to get that taken care of with the existing fleet?

MR. DARBALI: So is your question how are common cause failures addressed for the --

VICE CHAIR HALNON: Yeah, I mean, this is -- the BTP and everything we're talking about is for future, the things that are going to be occurring in the future, either modifications or new applications.

MR. DARBALI: Right.

VICE CHAIR HALNON: So since we're learning a lot more about common cause failures and the use of digital, there has been some digital modifications already put into plants.

MR. DARBALI: Right.

VICE CHAIR HALNON: Maybe not extensively safety systems. But is it the inspection process and the operating experience process that could potentially cause either a backfit for some other mechanism to force plants to address it? And would they address it through the use of this guidance?

MR. DARBALI: Right. I mean, we're not looking at backfit considerations. So as you said, if a plant -- an operating plant wants to comes in for a license amendment for a new digital upgrade, then part of that calls looking or performing a D3 assessment to see if a common cause failure of the digital system, if it can result in the loss of safety functions. And so BTP 7-19 provides a number of ways in which applicants determine that the CCF can be prevented or considered that it can be excluded from consideration. And then if that cannot happen, then they can follow either a deterministic path to mitigate the CCF or a new risk informed path to use different design technique.

VICE CHAIR HALNON: So is that same logic that you just laid out what a plan would do if they found that the past modification suddenly through operating experience has a common cause failure vulnerability?

MR. DARBALI: So as far as operating experience, even though, yes, we have approved several digital designs, each application is different. We're talking about different plant configuration, different platforms being used. There's different development processes being applied.

So I don't think there have been enough digital upgrades that we can say -- that we can take it all together and say CCFs were not considered. If a plan was to determine, oh, plants need to consider that a CCF happens. That's what the policy as of this point says.

So they have -- regardless of what the CCF is, they have to take that into account and see how the plant can mitigate or address that CCF. If they determine that the CCF cannot be successfully mitigated and can put the plant in an analyzed condition, then I don't believe we've encountered that. Inspection activities that could capture that would be during factor acceptance testing, set acceptance testing or --

(Simultaneous speaking.)

VICE CHAIR HALNON: Or a response.

MR. DARBALI: Right. And so -- right. That inspection result would be taken against what the licensing approval --

VICE CHAIR HALNON: I think somewhere in there, you answered my question. I think this is the best guidance we have for addressing CCFs. And certainly if one happened to an existing digital upgrade that you approve years ago, this would be the first approach that someone would try to use to mitigate it, I would assume.

MR. DARBALI: Well, the approach in BTP,

again, its' staff guidance.

VICE CHAIR HALNON: Okay.

MR. DARBALI: But it lays out the plan for the staff to approve how an applicant is addressing CCF. So --

VICE CHAIR HALNON: I think I got my answer.

MR. DARBALI: Okay.

VICE CHAIR HALNON: Rather than continue on, I think I understand.

MR. DARBALI: Okay.

VICE CHAIR HALNON: Thank you.

MEMBER BROWN: Can I make one observation on your question? Based on what we've done over the last 60 years, we've had four or five new designs plus Diablo Canyon plus there was one in Florida.

MR. DARBALI: Waterford was approved.

MEMBER BROWN: Yeah, and so was Diablo Canyon.

MR. DARBALI: Right. Hope Creek was approved.

MEMBER BROWN: And there was one -- what's the one in Florida?

MR. DARBALI: Turkey Point was --

MEMBER BROWN: No, no, no. It's another

one. It's got a --

MR. DARBALI: St. Lucia.

MEMBER BROWN: No, not St. Lucia. Crystal River.

VICE CHAIR HALNON: That's the only one that's out there is Crystal River. But it's shut down and decommissioned, so --

MEMBER BROWN: Oconee. Oconee is -- yeah, I'm sorry.

VICE CHAIR HALNON: That's north Florida.

MEMBER BROWN: So I lost a couple of states in the meantime here. And then when we did the new designs from the time at least -- from the time of AP 1000 on, part of our review at least was done by me was utilizing how did they apply not just what they said in the licensing application relative to how you design these things. But the BTP was also factored into our reviews at that time.

And there's been four AP 1000, APR 1400. APWR, we got that far. But then it disappeared. And then NuScale and my brain just disappeared on the last one.

Anyway, so I think the application, the earlier versions of BTP 7-19 have been -- we were aware of those and they were at least looked at to see how things were implemented in the actual plant design. And so my understanding based on that conversation to discussions with the staff was that they had considered that as part of their review process. So this is an expansion in reality fundamentally to the risk informed approach being added most of the rest of the stuff other than the SRM for 0076 is pretty much similar to what we've had in the past.

And I see Samir shaking his head up and down. I haven't gone too far off the reservation yet. So just an observation for the committee members. I'm done.

MEMBER ROBERTS: The thing about Greg's question, in operating experience, I think -- and Samir, you can correct me if I'm wrong. But the real genesis of BTP 7-19 was more of a, I want to say, hypothesized concern about software, common cause failures, that's there's something in the system where all the redundant channels have the same software or similar software. They got tickled by the same stimulus and went into some state that wasn't predicted.

And if they had one of those in actuality, I would think the effort would be more in the software equality because the BTP is based on mitigating those hypothetical types of situations. And I don't know what more you could do besides not have them which with operating experience will come in is what the root cause was of that particular software. But then the BTP got expanded recently into hardware.

I think hardware, the experience resides in standards like IEEE 352 where you do an assessment of what the common cause failure potential is and decide what you can reasonably do to provide or mitigate. So I know we're going to take about that. I guess we talk about that at the meeting the end of June, and that's probably a good conversation to raise again then.

MR. DARBALI: So we can --

MEMBER ROBERTS: I had one other question. This slide is probably going to -- at least I can ask it. Is Reg Guide 1.233 -- the defense-in-depth model is what I'll call a plant, defense-in-depth model. I think there's actually two.

There's also kind of a regulatory or procedural defense-in-depth that's also in there. But what the BTPs have in IT leverage was this NUREG 6303 which created a model specific to I&C. It had four echelons of defense where the things that you're essentially assessed against in the whole diversity assessment process, it's pretty well still it's called out in the BTP.

So it seems to me the translation is potentially missing from Reg Guide 1.233 to a digital I&C assessment. And maybe now it's time to talk about that or maybe sometime later. But the overall process of how you assess defense-in-depth in an I&C context is something I'm not real clear on and it may require more thought.

CHAIR KIRCHNER: If I could amplify on Tom's comments, Samir, how does this loop back eventually? You're just one part of the NRR organization that continues propagate non-LWR kind of guidance and so on. But it seems to me at least -- and this is not my main area.

In I&C from a functional standpoint, the I&C system doesn't recognize what the reactor is so to speak. And it doesn't recognize what coolant the reactor is using. You still have the fundamental functions to perform controlling reactivity, controlling heat, controlling your fission product boundaries.

So it seems to me to elaborate on Tom's point, 1.233 is really at a real high level. Your DRG and your Branch Technical Position is really focused on kind of the nuts and bolts of an I&C system and hence much more useful and applicable for a non-LWR applicant. So maybe if you can loop back at the end. Or I don't want to make a major diversion now. But it seems to me this needs to come back together at some point --

MR. DARBALI: Right.

CHAIR KIRCHNER: -- within the agency.

MR. DARBALI: So you're right. The guidance in the BTP is for the I&C system itself, if you could take that box and put in whatever type of reactor. The reason that the focus of the BTP is on light water reactors is simply because the BTP is part of the standard review plan --

CHAIR KIRCHNER: Right.

MR. DARBALI: -- in NUREG-0800 which is for those large light water reactors. And also the SRP can be used for small module light water reactors. But that's really why there's a separation between the BTP and the DRG. The DRG developed to be used with the LMP and Reg Guide 1.233 for those non-light water reactors is more performance based, more risk informed technology neutral.

The staff can use -- if the staff is doing a review of non-light water reactor I&C design, they can look at the guidance in the BTP. There is that flexibility. So there's nothing really preventing the staff from using the DRG for light water reactors or the BTP for non-light water reactors. The documents themselves really are held on to those structures based on reactor technology.

MEMBER BROWN: Okay. I'm going to help out a little bit since Tom lit the fuse. We will have subsequent discussions on this. We included some items in our letter so we would get involved and discuss that.

I wanted to amplify Tom's comment relative to the software advice, limitations relative to software that the staff is faced with. In the early days, if you go back 40 years when I first started doing this in my old program, when you're dealing with a Z-80, okay, and lines of code to do your processing of roughly several thousands, if not hundreds of thousands of lines of code, there was some ability to at least do some type of inspection and say, we understand the lines of code. Once you get up into hundreds of thousands of lines of code, the staff cannot to code inspections.

You literally have to depend on how you

develop the overall architecture of the system and then foster the idea with the licensees whether new design or already existing to implement or utilize with their vendors a very high quality software development program where they track, make comments, or you can literally go in and look at the code as it's developed and why they did certain things. It's at least documented in their code development process. That's a system they can say, yeah, they're utilizing a good process.

But then you say, well, that's really not good enough. Here's the backup things we need to do and the architecture of the overall design because that's where the rubber hits the road. That's the only place they can see it right up front.

And you don't have to examine the guts of a processor or a memory chip or how every little line or line of code goes off and does something else. You don't have to do that. Now is that 100 percent perfect?

There's nothing 100 percent perfect. But that's, to me, the focus on the DRG and the modernization project which also came from MPower which was not necessarily a modernization project when we started it. It was to try to capture this approach in those documents as well as the BTP and the other Reg Guides that we work with to have them recognize this architecture approach.

And we did that with the cyber stuff as well finally. It only took three or four letters to do that. But we finally got there. Recognize that you have to think about that from the design of the system, not from the programmatic standpoint.

So I'm just trying to provide some context overall of what are the staff limitations and how do they ensure that the stuff we're putting into the commercial world is satisfactory. And that's why BTP, the DRG which is really kind of a conglomeration of a system spec in a way or advances in terms of helping them get on with this process. So that's enough of my soliloquizing here. But that's just some perspective.

CHAIR KIRCHNER: Now that we've got the architecture out of the way, we can proceed.

MEMBER BROWN: Yes, Samir. You're on again. Thank you.

MR. DARBALI: Okay. So we're now on slide 8. Thank you. So now we're going to be covering those substantive changes that were made from Revision 8 which was issued in 2021 to Revision 9.

First, on Section B.1.1, we revised it to update the language of the four points to reflect the language in the points from SRM-SECY-22-0076. Again, the Revision 8 uses the points from SRM-SECY-93-087. So we're updating those points. On Section B.1.2, we revised it to clarify the term, critical safety function and to clarify that the identification of such functions may be risk informed.

Section B.3.1.3 was revised to provide acceptance criteria for the use of alternative approaches not previously endorsed or approved. Originally, B.3.1.3 only provided acceptance criteria for approaches that were already approved or endorsed by the staff. And Section B.3.4 was added for the evaluation of a risk informed D3 assessment. This is the major change in Revision 9 to incorporate the direction in the SRM.

VICE CHAIR HALNON: Samir, this is Greg. On 3.4 and I don't know if we want to get into detail here or later. And if it's later, your presentation is fine.

There's a portion of it in the SECY and your BTP that talk about it that's a risk significant system issue. The SECY talks about if it's not risk significant, you do X, Y, Z. But the BTP doesn't go that direction, only if it's risk significant.

Was that an intentional omission from the SECY to not talk about in the BTP? BTP says if it's risk significant, you do this, this, and this. But it doesn't do the same -- it doesn't have that last portion of the SECY.

MR. DARBALI: So we do have a few slides coming up on B.3.4. But Steve, if you want to address it now or wait until later.

MR. ALFERINK: This is Steve Alferink. I was planning to address it here in my --

VICE CHAIR HALNON: Okay.

MR. ALFERINK: -- slides. But I can also answer it.

MR. ALFERINK: I'll listen for it. Thanks, Steve.

MR. ALFERINK: Thank you.

MR. DARBALI: We revised Section B.3.4 to include guidance for the evaluation of different approaches in point 4 and we'll see later that was based on a change that was to made to the point in the SRM. We also added five flow charts to facilitate the use of the BTP by the staff performing licensing reviews. And we also added language from Regulatory Guide 1.152 regarding communication independence and control of access. So again, these are the substantive changes from Revision 8 to Revision 9. Next slide.

So this is a figure that we included in the BTP and provide a review of the structure in Revision 9 and of how the sections in the BTP are organized to implement the four points in SRM-SECY-22-0076. In the next few slides, we're going to cover, I guess, in a little bit more detail on B.3.4 and also on the changes made to Section B.4. So you don't have any questions, I will turn it over to Steve.

MR. ALFERINK: Thank you, sir. Good morning, everyone. My name is Steven Alferink, and I will discuss the review guidance for risk informed --

(Simultaneous speaking.)

MEMBER BROWN: -- microphone.

MR. ALFERINK: Is this better? I'll discuss the review guidance for risk informed D3 assessment, the new Section B.3.4. We're now on slide 10.

This slide illustrates how the staff envisions a risk informed approach fitting into the overall D3 assessment process. The D3 assessment process starts by identifying each postulated CCF. Once CCF is identified, it can be addressed deterministically or by justifying alternative approaches.

These options were shown in the two boxes in the middle. If the CCF is not addressed using either of these two options and they can be addressed using a risk informed approach which is shown in the colored box on the right. A review for the risk informed D3 assessment is broken down into four steps, each of which is covered in corresponding subsections of Section B.3.4.

I'll cover each of these steps at a high level in the following slides. Next slide, please. We're now on slide 11. This slide covers the first two steps of a review of a risk informed D3 assessment.

The first step is to determine consistency with NRC policy and guidance on risk informed decision making. In this step, the reviewer will review an application that uses a risk informed approach for consistency with established NRC policy and guidance on the risk informed decision making as reported by point 2 of SRM-SECY 22-0076. For light water reactors that will be reviewed using BTP 7-19, the established NRC policy and guidance on risk informed decision making includes Reg Guide 1.174 and Reg Guide 1.21.

The second step is to review how the CCF is modeled in the PRA. In this step, the reviewer will first determine if the base PRA meets the PRA acceptability guidance in Reg Guide 1.2 or equivalent guidance for new reactors. It reflects the plant or design at the time of application. A reviewer will then evaluate how the CCF is modeled in the PRA and the justification that the modeling adequately captures the impact of the CCF on the plant. In general --

MEMBER PETTI: Question, as I understand it, PRAs today don't model the digital I&C system. So this is a pretty high bar. This is a big scope change to PRA, right, to include this for all the digital I&C.

MR. ALFERINK: The next one I think will address your --

MEMBER PETTI: Okay.

MR. ALFERINK: -- comment. So in general, CCF can be modeled in the PRA either through detailed modeling of the digital I&C system or the use of surrogate events which in the existing basic events --

CHAIR KIRCHNER: The use of?

MR. ALFERINK: Surrogate events which in the existing basic events in the PRA or new basic events added to the PRA, they capture the impact in the CCF. So you're correct, unless operating plants do not have PRA modeling of the digital I&C system in the PRA.

MEMBER BIER: And if I can follow up, my sense is in addition to the fact that it's mostly not modeled right now, I suspect that if you tried to model it down to the level of individual components and common cause failure between those components, that might not yield very accurate results. It's one thing to model kind of a black box of I&C failure and another thing to model all the components and logic inside that. So have there been examples that have done that or what's your sense how feasible that would be?

MR. ALFERINK: At the moment, my sense is that the plant most likely used the surrogate events to capture the impact, the CCF on plant. At the moment, I am unaware of doing the detailed visual system modeling.

MEMBER BIER: So in other words, it would be kind of a black box if you postulate that you had this type of failure, here's what would happen afterwards?

MR. ALFERINK: Yes, that is correct.

MEMBER BIER: Okay, thanks.

MEMBER PETTI: And that's sort of at the event level, not the fault level. There's a lot of fault, right?

MR. ALFERINK: What I'd expect them to do is if they identified which components the CCF would impact. You could potentially just fail that sort of components. So I would say it's more like the basic event level. You go through, fail those, look at what the change in risk would be.

MR. WEERAKKODY: Sunil Weerakkody, Senior Level Advisor, PRA for NRR. With respect to modeling challenges, I would hate to characterize that as a high bar unless licensee says, well, to capture the impact of the CCF, I'm going to go and model everything in detail. And that's one way to do it. That's a very difficult way to do it.

Second way that is more realistic is to carefully study the impact of the CCF and find what we call a surrogate demand. And to do that, in fact, I think the industry has recognized that as a big challenge in terms of how to do that. And that's something discussed at the -- when we go to the oldest group, the risk management committees, we talk about that.

And the oldest group committees are

already thinking about that. But it is an important point that we need to get ahead of. Thank you.

MEMBER BROWN: So your point is even doing surrogates is not necessarily well understood at this time. And to me, when I listen to both, I'm just trying to integrate both comments here.

MR. WEERAKKODY: I --

MEMBER BROWN: Once you get down either -- let me just finish --

(Simultaneous speaking.)

MEMBER BROWN: -- my thought so you can beat me to death. Doing components, you can try to do that with hardware analog systems today. Try to analyze all the systems and do a PRA on it. We've never done that.

It's just too hard with several thousand components that you could look at and then do the connectivity between division to division to division. How does one really propagate into the other? What I'm hearing from you even stepping it to a higher level a surrogate approach is not necessarily well understood at this time.

People are thinking about it. How do we go do that? But what surrogates would be useful? Maybe that's the way I would phrase it.

(Simultaneous speaking.)

MR. WEERAKKODY: No, I think if you leave it at that, I think that's -- I think you misunderstood what I was trying to convey because it is not uncommon for us in PRA as practitioners to come up with surrogate in different areas. We do this in other areas like when we use PRA to risk inform material issues. It's not difficult for experienced PRA practitioners to understand the systems and do that. But what we do recognize is in the I&C area when you identify that surrogate, you need an I&C engineer and a PRA engineer in the room to do that accurately. So as long as you had those people in the room, it's something that can be accomplished in my opinion relatively easy.

MEMBER BROWN: In order to calibrate -- thank you. I am obviously a major skeptic of trying to apply PRA to the development of the digital I&C systems. And I know it's not universally -- nobody is necessarily on the same frame I am.

And based on experience, like, 50 years of experience of developing these things since 1980, you really make it difficult -- you've really got to look at the overall design. And what we've depended upon is an architecture approach that minimizes the potential impact. And to me, the risk informed nature of that approach is we're confident to a certain level.

I don't know how to calibrate that level that if you have some type of diversity in the design, it will mitigate that somewhat also. But you're never going to be 100 percent on anything. And by the time you try to go through a detailed PRA analysis, I'm concerned about the cost to develop these systems.

And it would even limit even farther the backfit of what I call much superior systems, I&C systems into the commercial fleet as it exists today. It just concerns we that piling more design stuff on that you have to deal with up front like that is -- that's just my thought process. And I think the staff has to be cautious about how we approach those.

We want plants to upgrade their systems to these systems that are more accurate, don't drift, are more responsive. You can depend on them doing what they're supposed to do, a little bit more confidently than with the analog systems. So I mean, you've really reduced the variables that could provide variability with an analog system when you transition over to the digital systems.

They provide some really positive

benefits. So anyway, that's just another part of my speech making for the mission here. I'm not objecting to doing it. Obviously, we're putting it in because that's the direction we've got. I just think you are -- and I have to be very cautious as you implement these reviews and how you start going down this path to make sure we don't really compromise our ability to upgrade the plants.

MEMBER PETTI: So just when we see new ideas and new approaches, sometimes things are table topped, right? I mean, I don't have any sense of the confidence of this option. I mean, is it one that industry is going to use and going to want to use because this confidence that you can come up with good surrogates and you can do this?

Or does it look like it's too much of an effort? It's a branch here that people won't take because it looks like it's a bridge too far. That's what I don't get a sense on at this point.

MR. ALFERINK: I haven't heard feedback myself. But we put the flexibility in there, so it's up to an applicant. Should they choose to do the detail modeling, they could.

They wouldn't have guidance in there. Based on the current state of practice, I expect they would use a surrogate. I haven't heard any indications that would be a particularly challenging approach, at least to figure out what surrogates are and to use that approach. But like I said, we'll have to find out here in the future.

MEMBER DIMITRIJEVIC: Hi, this is Vesna Dimitrijevic. I agree that there is absolutely not really -- shouldn't be difficulty to choose the surrogates based on the critical functions which are affected. And also I want to say we use the surrogates -- we use the surrogates in the new design too because of the diversity between systems.

You can never really model this on component level. So this is nothing strange for the digital I&C. You have to choose the right surrogates because of diversity.

This is not the common cause among the systems. So you have to -- among the single system. Then with the multiple systems and therefore the usual component level is not applicable. So using surrogates is nothing strange and shouldn't present problem. That's just my experience with it.

MEMBER BROWN: Are you done, Vesna?

MEMBER DIMITRIJEVIC: Yes.

MEMBER BROWN: Okay. Thank you.

MR. ALFERINK: Slide 12, please. And we are now on slide 12. The third step is to determine the risk significance and CCF. The risk significance of a CCF can be obtained by calculating the increase in the risk from the CCF using either a bounding sensitivity analysis that assumes the CCF occurs or a sensitivity analysis that uses a conservative value less than one for the probability of the CCF which we loosely call a conservative sensitivity analysis. Next slide.

If the increase in the risk is calculated using a conservative sensitivity analysis, a reviewer will evaluate a technical basis with a conservative probability of the CCF. The impact of this assumption on PRA uncertainty and whether it is considered a key assumption and the impact of this assumption on the key principles of risk informed decision making. A reviewer --

CHAIR KIRCHNER: Steve, can I interrupt you here? Just could you explain for the public why your second bullet says due to a bounding sensitivity analysis assuming CCF occurs? Why didn't that a conservative analysis? The next bullet says assume a probability less than one. That doesn't seem conservative to me. That seems to increase the uncertainty.

MR. ALFERINK: I would argue they're both intended to be conservative. I use the term bounding. Say we use 1.0. It's guaranteed to occur.

CHAIR KIRCHNER: Exactly.

MR. ALFERINK: In our first ACRS meeting, we know that we're open to applicants using numbers less than one. For example, if they use 0.5, I think it's safe to say that the probability of the CCF occurring is certainly less than 50 percent. But if they wanted to use that, then we're trying to provide that flexibility. At the same time, it is incumbent upon them to provide the technical justification for that.

CHAIR KIRCHNER: I just struggle over the conservative in quotation marks in that bullet. I understand doing a sensitivity analysis that assumes a probability less than one. But then they have to do what you ask for next. They have to have the technical basis for it and then see what the impact on the system in question is. But I just don't see how that's conservative versus Bullet No. 2.

MR. ALFERINK: It would be less conservative than the bounding sensitivity analysis. But with respect to a 0.5, I still think it's safe to say that's a conservative number.

CHAIR KIRCHNER: All right. It's more the viewgraph keys aspect of this than the substance. Okay. Thank you.

MEMBER ROBERTS: Yeah, Walt. I think what you just said. Those second and third bullets probably should be indented because that's an or. You don't do both, right? Because the first bullet says you do one, the bounding or the conservative. And the second one, I just wanted to make sure I understand, one of the NEI comments that they provided indicated that they wouldn't expect to use a CCF probability less than one in the near term because of lack of methods to justify it.

And that kind of goes with the discussion we had about surrogate events where surrogate event would, I think, clearly mapped the bounding sensitive because you would assume the common cause happened in mild consequences as opposed to the actual equivalent. So I think probably the real message out of this is the bounding sensitivity is all we're going to see in the near term. And any further work to have a CCN less than one will require technologies that don't exist yet or techniques that doesn't exist yet. Is that fair?

MR. ALFERINK: I think that's a fair characterization.

(Simultaneous speaking.)

CHAIR KIRCHNER: -- its practical application. I mean, where the value of the whole PRA is, is improving your design. So if you went through this exercised and you did see that you had a problem because of a CCF that made a significant change, we can quibble about the next chart that's coming.

Then as a designer, I would want to go back and say, okay, I don't have enough diversity or I don't have enough redundancy or I don't have enough independence such that in my D3 assessment, it's n to satisfactory. So I go back and I change. I make an actual hardware change, whether that's via software or not. That's a different discussion.

But a component change or something to take that off the table. I mean, that's the real value of the exercise here if you're using PRA. That's my opinion.

MEMBER MARCH-LEUBA: I agree with you. I wanted to say that in this field a factor of 2 doesn't make no difference. So that probability of one has to be 10 to the minus N with any such number.

And that will be difficult to justify.

That will come along when you truly have diversity. The diversity is not a complete analog system, but it's another digital system which has a probability of having a common cause failure, but it's pretty low. But it has to be 10 to the minus 2, 10 to the minus 4, not 0.5.

MEMBER BIER: Another question is in terms of applicability of this criterion, conceivably there could be new reactors with very low core damage frequencies and 10 to the minus 6 might actually look pretty significant for some of those if core damage frequency is low enough. So what's the thinking on whether this would be applicable in that situation?

MR. ALFERINK: Certainly. We'd be using the same thresholds for both the operating fleet and new light water reactors. They'll be consistent with SECY-10-0121.

MEMBER MARCH-LEUBA: Since we're wasting time, let me put a concept out there. It's going to seem completely different to you but it's not. And it starts with a concept, a real life concept.

I know you all have heard this. When it rains, it pours. Say, for example, I'm driving and I have a flat tire. Okay. It's an event. But when you have a flat tire, and this could be a county road.

It's at night and it's raining.

And when you go pick up the spare tire, it doesn't have power on it, pressure. And then when you go into the compartment to get the AAA card to call them to help you and you remember you left on top of the piano. Those are seven different independent events which if you calculate it, it comes to 10 to the minus 25.

But when it rains, it pours. When you're talking CCF, the probability that it happened, when it happens because you're having a bad day. And we've all had those bad days. Okay? So, a little bit in there is a wise comment. We cannot assume independence for those things. It doesn't happen.

MR. ALFERINK: So the reviewer will evaluate the risk significance of CCF by comparing the increase in the risk obtained from the sensitivity analysis, the thresholds for CDF. The reviewer will determine the CCF is not risk significant. The increase in the CDF is less than one times 10 to the minus 6 per year and the increase in LERF is less than one times 10 to the minus 7.

VICE CHAIR HALNON: Steve, I completely blew my question because it was actually a conflation of two questions and it came out as nonsense. Let me restate it. And this gets back to the BTP, the guidance in it.

The BTP says if it's risk significant, and it meets the acceptance criteria which is A, B, C, D or whatever, three items, you do that. It doesn't go into if it's risk significant and it doesn't meet the acceptance criteria, what do you do at that point?

So that's the actual question. Whatever I said before, just strike that from the record in your brain and say that he didn't know what he was talking about because I just conflated two questions. This one, I was really concerned about. I didn't see a direct correlation between the SECY and the BTP.

MR. ALFERINK: I'll address that on the next slide in slide 13.

VICE CHAIR HALNON: Yeah, I saw that coming up. And I didn't know if you were hoping that last bullet was going to satisfy me or not given my --

(Simultaneous speaking.)

MR. ALFERINK: No, it's a statement to hopefully satisfy.

VICE CHAIR HALNON: Okay.

MR. ALFERINK: Getting back to slide 12. So just a few more comments on this slide. So first it is important to note that there's a fundamental different between the intent of risk evaluations performed or risk informed applications involving BTP 7-19 and those that do not involve BTP 7-19.

Evaluations performed for risk informed applications that do not involve BTP 7-19 are intended to calculate the change in risk due to a proposed licencing action and therefore reflect the as built and as operated or as to be operated. As such, those licensing actions that result in an increased risk above one times 10 to the minus 5 per year are normally not considered as discussed in Reg Guide 1.174. Evaluations performed for risk informed applications involving BTP 7-19 are only intended to determine the risk significance of the postulated CCF.

These evaluations are not intended to calculate the change in risk due to the introduction of the digital I&C system nor the baseline risk with the digital I&C system installed. Therefore, these evaluations do not reflect the as built and as operator or as to be operated. Based on the discussions during the recent subcommittee meeting, I'd like to clarify the intent of the second sub-bullet associated with a conservative sensitivity analysis.

To be clear, the staff will review all

risk informed applications to determine consistency with NRC policy and guidance on risk informed decision making, including the five principles of risk informed decision making in Reg Guide 1.174. In addition, if the increase in the risk is calculated using a conservative sensitivity analysis, the reviewer will evaluate the impact of this assumption on the key principles of risk informed decision making. The intent of the acceptance criteria is to ensure a risk informed application that calculates the increase in the risk using a conservative sensitivity analysis, addresses the impact of this assumption on the key principles of risk informed decision making in addition to the broader discussion of the application's consistency with NRC policy and guidance on risk informed decision making. Next slide.

So we are now on slide 13. The fourth step is to determine appropriate means to address the CCF. This slide illustrates a graded approach for the review based on the risk significance of the CCF. The rick significance of the CCF is characterized by mapping its increase in risk to the regions in figures 4 and 5 of Reg Guide 1.174.

This figure illustrates this mapping based on CDF. A similar figure would illustrate this process based on LERF. If the CCF is not risk significant meaning that the increase in risk falls in Region 3 of both figures, a reviewer should concluded that standard design and verification and validation processes are sufficient to address the CCF.

If the CCF is risk significant, meaning that the increase in the risk falls in Regions 1 or 2 of either figure, a reviewer will evaluate the CCF against the acceptance criteria with the level of technical justification commensurate with the risk significance of the CCF. Based on the discussions during the recent subcommittee meeting, I'd like to elaborate on two points. First, the statement that the review should conclude that standard design and verification and validation processes are sufficient to address the CCF if the CCF is not risk significant assumes that the reviewer has already concluded that the application is consistent with the established NRC policy and applicable guidance on risk informed decision making.

Second, point 3 of SRM-SECY-22-0076 states that a diverse means must be provided if a postulated CCF is risk significant and the assessment does not demonstrate the adequacy of other design techniques, prevention measures, or mitigation measures. BTP is written to provide review guidance to the staff. It is not written to provide implementing guidance for applicants. As such, the acceptance criteria in Section B.3.4.4 do not specify that an applicant must provide diverse means if the assessment does not demonstrate the adequacy of other design techniques, prevention measures, or mitigation measures. If an applicant determines --

MEMBER BROWN: Could you repeat that again, please?

MR. ALFERINK: The acceptance criteria in Section B.3.4.4 do not specify that an applicant must provide diverse means if the assessment does not demonstrate the adequacy of other design techniques, prevention measures, or mitigation measures.

MEMBER BROWN: So you're effectively saying they don't have to do anything even if it doesn't meet the requirements?

MR. ALFERINK: I'm saying the way the acceptance criteria are written does not provide that kickback we're discussing.

MR. DARBALI: And the point Steve was making before that was that the BTP revision for the NRC staff's review --

MEMBER BROWN: No, I understand that. But

I mean, it would not be very -- I mean, licensees are going to be designing their stuff recognizing that it's going to be reviewed by the staff. Where do you go to get stuff that the staff is going to be looking for? They go in the BTP.

So I don't put much credence in the fact that it's for the reviewers and not the licensee. If I was a licensee, I'd be looking at everything the staff is going to be reviewing my equipment against and making sure I didn't have any pinholes in it or giant gaps, one way or the other. So I think you'd really to walk down that path of this is just for reviewers because it's not.

To me, I mean, I look at your box and I'm not a -- PRA guys have said before. If I'm 10 to the minus 6 in CDM, I've got to be somewhere over here on the left-hand side of the corner of that box that you're up in the 10 to the minus 6 or 7 range. And therefore, a change in CDF is just blacked off after 10 to the minus -- because it's all black which means bad to me.

So if it's 10 to the minus 4 for delta CDF, is that necessarily bad if you're at 10 to the minus 7 on CDF. To me that's not necessarily all that bad. It's a pretty small change. So there doesn't seem to be a gradation in here.

This is pretty much a couple of blocks. I don't have any problem with that. But you need to be able to address in my own mind a licensee's approach where he says, hey, look, stuff has got virtually no -- or the plant is no change in real damage frequency.

Therefore, we have some more crummy stuff, I hate to use that word, but it's not quite as upscale. It doesn't have quite as much diversity, et cetera. But there's got to be some flexibility. If I was a staff member, I wouldn't -- if I had a CDF that was really, really low, then you really need to use your head and don't pound a guy to death for some increased design features.

VICE CHAIR HALNON: So Charlie, rather than working with numbers with too many zeros in them, just a direct question. Why not just include that last portion in the guidance? So if the reviewer knows a full picture, in other words, you've got -- if it passes this acceptance criteria, you're thumbs up.

If it doesn't pass it, it's a silent. But the SECY isn't. It's just one sentence more, one or two more sentences. Recycle back and have them change the design. Why not include that?

MR. ALFERINK: I got two more points I think will answer the question. The short answer is because they would use review guidance in a different section. This is specific to B.3.4.4 I'm referring to.

So the point I was going to make is if an applicant determines design techniques, prevention measures, or mitigation measures other than diversity are adequate with the CCF, the reviewer will evaluate the application using the guidance in B.3.4.4. If that point determines that diverse means are requires for CCF or that point in the policy, the reviewer will evaluate the application using the guidance in Section B.3.2 because that would be the appropriate section for reviewing diverse means. So this particular section is written for reviewing it when they've determined diverse means are not required.

VICE CHAIR HALNON: Okay. It just makes it more difficult to follow the SECY point by point.

MR. ALFERINK: I understand.

VICE CHAIR HALNON: And there's no recycle back that says even in the BTP. If it doesn't meet it, go back to 3.2.

MR. ALFERINK: Correct.

VICE CHAIR HALNON: So it's just a comment that disconnect isn't necessary in my mind. If it's intentional, I don't see how intentionally it works well. But I understand it's guidance for the staff. And they'll get used to using it.

MR. ALFERINK: It was intentional in the fact that we're writing this as review guidance, recognizing that applicants may very well use this to help inform how they either process the language we use. The structure was for review guidance to be consistent with that. And yes, I acknowledge it looks like it creates a disconnect there. That was it for slide 13. So I'll now hand the presentation back to Samir, unless we have more questions.

MR. WEERAKKODY: May I say something? Again, this Sunil Weerakkody. So you have a -- I'm sorry. Your point is a valid one, yes. This is staff review guidance.

But the industry looks at this. One of the things I want to share with you is that this BTP especially when you are at this new part, it's not going to be standing alone. There will be other communications.

There will be other guidance, in fact. That's why I keep going back to when we met with the owners' groups every three or four months. This is a point that we bring and discuss imagined guidance. So I think if you don't get everything from here, there will be other supporting documents. And I'm not making any commitments that we capture that.

VICE CHAIR HALNON: So I appreciate that. And I think the staff will work well with this without a problem. However, in the big picture of things, road maps work really well because you kind of know where you're starting and you know where you're going.

And over communication is not a bad thing when you're trying to come in cold and try to figure out how I'm going to do this assessment -- D3 assessment. What happens if this happens? What happens if that doesn't make it?

This who discussion would've been off if you just put one more bullet that said, and if it isn't, go there. That's the only point. So if that goes against your principles, I got it. But I can identify some good regulatory practices that would make that a decent thing to do.

MEMBER ROBERTS: Yeah, one observation. I'm not completely sure if this is editorial or whether this is a technical kind of question. But figure 7-19-4 which is the flow chart at the back of the BTP I think portrays the SRM position pretty well.

So it's got more information than the text of the BTP which is unusual.

It gives you a flow chart. It gives you a summary report if you read the document to figure out what the summary report is. Here, there's actually more information in the flow chart. So assuming the flow chart is what you intended, again, it may be worth taking another look and at least make the text as descriptive as the flow chart is.

MEMBER BROWN: Are you talking about the point 2 flow chart?

MEMBER ROBERTS: Point 3.

MEMBER BROWN: Point? Okay.

MEMBER ROBERTS: 7-19-4.

MEMBER BROWN: Oh, yeah. I see your point.

MR. ALFERINK: Thank you. Back to Samir.

MR. DARBALI: Thank you, Steve. Now we will go talk about the changes made to Section B.4 regarding point 4 of the policy. For SRM-SECY-93-087 and SRM-SECY-22-0076, the independent and diverse displays and manual controls that are called for by point 4 are not required to be safety grade or hardwired.

But they do have to be of sufficient

quality. For the review of an application that implements independent and diverse manual control room displays and controls for manual actuation of critical safety functions, Section B.4 of the BTP provides six acceptance criteria items. The acceptance criteria calls for displays and controls that are independent and diverse from the equipment performing the same functions within the proposed related digital I&C systems.

SRM-SECY-22-0076 includes a new sentence that allows applicants to propose a different approach if the plan design has a commensurate level of safety. We added review guidance to Section B.4. For the review of applications that do propose that different approach that does not meet all of the acceptance criteria. Next slide.

MEMBER BROWN: Backtrack just a second. You were on 24, right?

MR. DARBALI: Yes.

MEMBER BROWN: The difference -- and I'm just giving a little credit for this. I'm not going to try to beat you up or anything. SECY-93-087 was modified by the SRM, initially BTP or initially SECY that you wrote.

You used the word, shall, relative to

hardwired. And the Commission came back and said, that's too strong. It should be considered. I think that's -- I forgot the exact words.

MR. DARBALI: Considered as guidance.

MEMBER BROWN: Considered as guidance, right? And the way you phrased it here, even though they didn't even deal with point 4 -- I mean, the fourth paragraph of point 4. You rephrased it to be they do not have to be either safety related or hardwired which is a more declarative sentence. And so far, I'm giving you the benefit of the doubt. I think that's -- the wording is not totally consistent with the 93-087 which considered guidance as one thing.

But there's another point that says, you don't need to do it, period. So saying that these are similar is not exactly right. ut it doesn't preclude observing that they ought to be hardwired by something else depending upon the design that's provided as long as the pressure to turn everything into glass in software is not dictated to you for some reason. So anyway, I'm just making that point that after going back and forth like we did in the subcommittee meeting that they are not totally consistent with each other.

MR. DARBALI: Understood. I think most of the BTP is written in terms of acceptance criteria and maybe not so much as far as best practices. And I think that's -- your point is that even though they're not a requirement per the original SRM, the use of safety grade or hardwired technology is a best practice.

MEMBER BROWN: The best practice approach is not clear -- it was more clear in 087 than it is now.

MR. DARBALI: Correct.

MEMBER BROWN: You all have fuzzed it up, even though the Commission didn't say anything about it. I'm sorry I didn't have my microphone on. Court reporter, did you hear me?

Okay. Even though my mic wasn't on, you got all my transient thought processes. Thank you.

MR. DARBALI: Okay. Thank you. Now on slide 15. Okay. So now we're going to focus on those changes to the BTP since the September briefing. What we were talking about were the major changes from Rev. 8 to Rev. 9, now the changes the September briefing.

So basically, we made clarifications throughout the BTP to address several of the discussions that were held during the September briefing as well as some comments from Member Brown and Member Roberts that were provided as an attachment to the transcript also public comments. We believe that these comments helped improve the quality and clarity of the BTP. And so we appreciate the time taken in preparing these comments.

Regarding the public comments, we received a total of 35 public comments which were all provided by NEI. And None of the public comments that were received were related to applications for non-light water reactors or the DRG. We also --

MEMBER ROBERTS: Can you clarify, Samir, what that means? Does that means that None of them specifically address the DRG? Or that you assess and None of them would apply to the DRG or to non-LWRs?

MR. DARBALI: We didn't have any comments specific to the DRG or to non-light water reactors. They were strictly comments on the language in the BTP itself.

MEMBER ROBERTS: Did you conclude that None of them would affect the DRG or folks using the DRG with the pre-application engagements with the applicants to figure out whether their approaches are consistent with principle? Because it seems like there are a lot of comments from the NEI and from us for that matters. And saying they apply only to the

BTP and not to non-light water reactors or the DRG seems like a pretty strong statement. It may not literally apply because they weren't to text. But they may be attacking -- or applicable to principles.

MR. DARBALI: I think some can be extrapolated and discussed in pre-application meetings with non-light water reactor applicants. In that sense, when it's a technical matter discussing in the comment, sure, it could be applied to a non-light water reactor application.

MEMBER ROBERTS: So I'm wondering how staff is using that. At least the documentation of the common resolutions we got were very specific to the BTP. Are they being factored into the thought process while reviewing the new reactor applications or applying the DRG? Is there some flexibility?

MR. DARBALI: Yeah, I think we similar staff involved in both pre-application engagements and looking at those public comment responses. But again, the focus being on completing our milestones for the BTP, maybe the comment could've been more clear. Basically, we didn't have a comment saying, do this for the DRG or do this a non-light water reactor. Or how would this apply to a non-light water reactor? Those very explicit questions, we did not get.

MEMBER ROBERTS: Okay, thanks.

MR. DARBALI: Thank you.

MEMBER BROWN: I'm going to amplify Tom's comment in that -- and I've read this BTP 7-19 at least three or four times over the last number of years as we've gone through several -- a couple of revisions to it. And for the life of me, I&C is I&C regardless of whether it's light water or non-light water. You're going to have systems.

You got to shut it down. You got to cool the plant. You got to do X, Y, and Z. And you could literally say the BTP is useful for review of non-light water reactors. Put the rubber stamp on it and it would work just fine with no changes.

And then as you found areas that you needed changes, you could then implement those if there was some particular characteristic. But to put this aside totally and only -- the DRG is not as extensive relative to CCF type stuff if I remember correctly. Tom, am I correct? That's my memory. It's gone back a while.

MEMBER ROBERTS: Yeah, it's roughly three pages in the DRG, the 48 pages of the BTP.

MEMBER BROWN: Right. And so this is a far more definitive document, and it's very general relative to how you address various instrumentation and control systems, whether they're normal controls, shut down safeguards, whatever across the board. So I'm just throwing that out as ease of staff efforts over the years. You could go through it.

If you wanted to make a revision, you make this Rev. 10 and say it's good for non-light water reactors and you could issue it without change, that's the only point. I thought I'd get that on the record anyway. Okay. Now you can go.

MR. PAIGE: This is Jason. I just want to add a comment. I didn't say through my opening remarks during today's meeting.

But during the subcommittee meeting, I did mention that part of our approach for the DRG, we are hosting workshops in the advanced reactor community. And the main purpose of those workshops is to get feedback from industry, external stakeholders, to get those lessons learned and to understand their needs so that we can better understand the approach or what updates are needed to the DRG to meet their needs. So I just want to emphasize that.

MEMBER BROWN: My point being is that DRG, just implement the BTP and the DRG or whatever. I mean, I&C is I&C. It doesn't make any difference what you're cooling it with.

You're going to have to shut down the reactor. You're going to have to maintain cooling. And you've got to maintain confinement or whatever the containment philosophy is. And oh, there's Ian. Ian is a DRG --

(Simultaneous speaking.)

MR. JUNG: Ian Jung, Senior Reliability and Risk Analyst, previously a branch chief for I&C area that initiated the -- I was involved in most of the new reactor reviews and so on. But now I'm working on the Reg Guide 1.233 implementation within the division of defense reactors. When we developed the DRG, Charlie, it was intentional to develop a DRG from SRP.

SRP was largely focused on light water reactors. It was 500 pages long, which speak up? Okay. Okay, of course.

So we did mapping of SRP and with the non-LWR designs. We wanted to streamline staff's review, focusing on what's really important for the staff.

So those architectural descriptions we highlighted tremendously. And the key elements of the BTP were introduced into DRG so the staff can really focus on it. And many areas despite -- we understand that it's a common I&C technology.

But if you look at the BTP closely, there are a lot of languages and descriptions and guidance that are really specific to light water reactors. When you talk about even safety significance and risk significance, those are newly defined in Reg Guide 1.233 from a risk informed perspective. For regarding this topic of risk informing, by the time you get to what PRA standard you're going to use, the non-LWRs have a hits on standard -- ASME standards for PRA which we have different risk metrics, different definitions, and so on.

So I think there are some intention of being different. But at the same time, we understand the comment. And me as well as some of the staff members who are involved in TRG and non-light water reactor designs, we are fully familiar with what's going on with BTP.

We have that in our pocket. We review this new non-LWR designs. We'll utilize that. But I think the point is, I think, given the time, we are really busy working with X-energy and that part of those applicants and events.

Right now, we are really dealing with dealing with the I&C design. We are talking about defense-in-depth adequacy of the Reg Guide 1.233. They are actually translating those guidance of the LNP and TRG to see how defense-in-depth adequacy is being handled, including potential common cause failures, not just for digital but all other systems as I'm dealing with uncertainties, lack of operating experience, clip actually affects the sensitivity analysis.

All those are being part of the equation. And there's an independent, integrated decision-making panel that includes PRA, plant operation, engineering, and others who are making more conscious decisions on the adequacy of defense-in-depth. So we want to just keep going with that for now.

Next week, we have a workshop on next Thursday on, actually, this particular topic that we can share with industry and we can hear from them on the subject. So we want to have a experience and update. Future goal is eventually we know that DRG will have to be revised at some point. So we want to go there.

I just want to share this. This could be a topic at the June meeting as well. So we can elaborate in more detail at the time. Thanks.

MEMBER BROWN: That's okay. If you

remember back -- and I'm sure you do, back when we first started with the DRG approach coming out of the entire project, the comment was made that the DRG was actually a light water reactor development initially. And then it was incorporated into the LMP in a different -- under now with just non stuck in front of it.

But yet it was roughly the same. It's been expanded to include some additional information and we went from MPower to the point being is it is a more integrated standard, if you want to call it that than what we had in the past. If you look at -- remember the roadmap that was developed for the I&C world with the little arrows going.

They all get integrated. The DRG kind of encompasses that thought process into one document as opposed to 22 documents which it's difficult for licensees to get their hand around. So the DRG as you've got it is fundamentally we made the comment then, why are you calling it non-light water when in reality it's perfectly useful for light water design development as well.

So I'm just throwing that out to jog everybody's memory that this was not an isolated thing. And the proliferation of documents in the old days made it more difficult for the licensees to get stuff done and an integrated document. One suspects that they have to go to or they could rely on. It's easier for them to deal with. That's the only point I'm trying to make as I disappear into the sunset at some point.

MR. JUNG: Thanks, Charlie. Just one thing, it's a long history with the MPower and NuScale application of these basic concepts. Really it made a tremendous difference in efficiency and effectiveness, our review. And NuScale has been advertising this is one of the tremendous success area of the review by the staff.

MEMBER BROWN: As they use the DRG approach.

MR. JUNG: And also just a reminder, the ACRS wrote a letter to the Commission and presented in a Commission meeting that TRG was a significant improvement in the review of digital I&C. But the reason that it could be said is during my review of ESBWR and APWR and EPR, I&C was a critical path. And it took thousands and thousands of hours on very prescriptive reviews of areas that are not necessary safety significance.

So it's tremendous progress. And we are

still learning with a new design and new process, new framework. I think that our proposal is practiced in some of the newer term reviews. That's what we are suggesting.

MEMBER BROWN: Thanks, Ian.

MEMBER ROBERTS: I think what Ian just explained checks pretty well with what I've talked about a few minutes ago about defense-in-depth models. And it might be a good thing if we can put it in to the June 27th meeting to give an update based on the experience of the workshop, whatever you have in the last few months. But I think you made an important point which is that the -- and restating it from what I said at the outset is the existing defense-in-depth and diversity model derives from the light water reactor, NUREG 6303 where there was -- based on the defense-in-depth model for light water reactor with confinement from fuel cladding and reactor pressure vessel and containment, the four echelons of defense were derived.

And then the techniques used to assess the four echelons of defense for light water reactor, when you shift to a completely different technology, whether that defense-in-depth model even applies becomes a question, which means any techniques derived from that defense model often come into question. We talk about things like cliff edge effects and looking at what the context of defense-in-depth is for that particular plant design. Now I'm glad to hear that's what you're thinking about because it may lead to similar techniques but applied either more extensively or less extensively depending what the model is.

But it may be that the approach based on light water reactors could be insufficient or overkill, depending on what the overall defense-in-depth model is. And so the techniques will be similar. The number of systems you apply them to might change depending on the model. And again, it sounds like you're thinking the same way and very interested in hearing more.

MR. JUNG: Yes, I think the June meeting we'll prepare and have some discussion. Some of this conversation, actually, it covers both TRG and Reg Guide 1.233. In one sentence, I think that trend level model of defense-in-depth adequacy covers both not just plant level but it should go down to individual layers and individual systems, their contribution to those layers. And not only the diversity but other programs, risk insights, and all those things come into play as in totality, not just the single element. So I think we can provide an explanation on how that model works, how we are dealing with that through some of the near term applicants.

MEMBER ROBERTS: Thanks.

MEMBER BROWN: Thanks, Ian.

MR. DARBALI: And that last bullet on slide 15, key point we want to make is that regarding public comments and changes to be made since the September briefing, we did not make any substitute changes through the analysis methodologies or the acceptance criteria in the BTP that we had shared in September. Next side. So here on slide 16 are the notable changes we made to the BTP since September. We revised the BTP to consistently use the term digital I&C system which is the term used in SRM-SECY-22-0076.

We clarified that the BTP is intended to provide review guidance to the NRC staff for ensuring an application meets the policy and applicable regulation. And it is not intended as guidance to applicants for developing the D3 assessment. We do recognize applicant's look at the BTP to understand what the staff is going to be looking for.

We removed the pointers between Section

B.3.1.3 for alternative approaches and B.3.4.4 for appropriate means to address the CCF. We provided a well designed watchdog timer that is now dependent on the platform software and puts the actuators in a safe state. As an example of an alternative approach, that may address certain vulnerabilities. And we added a sentence that states that credited manual controls should be connected downstream of the equipment that can be affected by a CCF.

MEMBER BROWN: Even if they're software-based? The manual controls -- yeah, because you've allowed software-based manual controls. But I'm just pointing that out. There's a bit of an inconsistency. You're going to have to deal with that when you finally see these alternative approaches.

MR. DARBALI: Right, right. The main criteria is that they are --

MEMBER BROWN: They're pushed downstream. That's the key point.

MR. DARBALI: Yes.

MEMBER BROWN: Because eventually you've got to collect something hard that's going to apply power or take it away.

MEMBER ROBERTS: Question for Norbert on the fourth bullet. That subcommittee, I question whether the example of the well designed watchdog timer is properly characterized. And I think you would agree that the text seemed to give it more credit than you intended. And did you find a change to the example to clarify that?

MR. CARTE: Sorry. Norbert Carte, I&C Technical Reviewer. So I would say the problem is a little bit the slide. The slide only has a partial quote of the BTP.

And the full quote of the BTP says can eliminate some types -- may eliminate certain types or some types of CCF. So it's not -- so the example as is stated is, I think, adequate because it says that a watchdog timer can address certain types of CCF, not necessarily all types of CCF. So we don't plan a revision of that text.

MEMBER ROBERTS: Okay. So you would never use a watchdog timer as the only rationale to justify not considering common cause failure. Because that's the way that it reads and where it is in the text.

MR. CARTE: Well, what it says, for example, a watchdog timer not dependent on the I&C system software puts the actuators at a safe state may address certain CCF vulnerabilities. So it isn't an absolute. The door is open that some things can be addressed that way. The assumption is that other CCF concerns would be addressed by other techniques.

(Simultaneous speaking.)

MEMBER ROBERTS: Okay. I understand.

MEMBER MARCH-LEUBA: So you will have to rely on that famous PRA to tell you that reduces the probability of failure sufficiently?

MR. CARTE: Well, this is in a deterministic section. So we wouldn't be using --

MEMBER MARCH-LEUBA: It will still apply. I can see why an I&C system that is not really safety grade but is important. When you run a watchdog, you make it reliable that it may be acceptable. Something like the reactor scram, it will never be acceptable.

But there are things in the reactor that are not as important. And maybe a watchdog makes it work sufficiently well. It's a type of diversity.

MR. CARTE: Yes, it's a type of diversity that helps. But the problem is that there are different designs to watchdog timers. And one design that I'm aware of uses the sort of fault bit on the processor.

So anytime a processor locks up, it raises a fault bit. But would that happen if the clock fails? That's another issue.

So there are certain watch processor faults which then generate a watchdog timer. There's other applications of the watchdog timer where it's drove by the application program at the end of the application program to make sure the application program goes all the way through. So you know that you're not stuck in an infinite loop.

So the first watchdog timer doesn't catch you on an infinite loop. But the second one does because it ensures that you go through the application all the way to the end each time. So there are different designs of watchdog timers.

They address different concerns. And so they can be used as one of the means. We don't anticipate they would be the only means.

MEMBER MARCH-LEUBA: It would apply for watchdog timers. I design instrumentation and I can sleep much more comfortable if there is an independent watchdog timer on the computer because I know, if I made a mistake, it still caught me. I'm just making an advertisement for watchdog timers. Include them in. They're not that expensive.

MEMBER BROWN: I'm also a proponent of hardware-based watchdog timers such that if they're appropriately utilized, at least from flag or shutdown or provide a safety signal for any independent voting unit if, in fact, it locks up. And the platforms, I've looked at three of the platforms that you all have had to take approval of, Common Q, this and whatever they are. And they all incorporate multiple types of watchdogs.

Some of them are software-based within the software to make sure this thing got done or that thing got done. But the hardware-based one at the end of the whole cycle provides an additional level of -- some level of certainty that you're going to catch a lock up if it gets in there. The point being is there's not a cure all. But they certainly help out the process, particularly the hardware independent ones.

MEMBER MARCH-LEUBA: And from the point of view of marketing, if I'm a designer and you don't give me credit for it, I won't bother putting it. So there are two sides of every coin. It has to give the credit it deserves with program analysts.

MEMBER BROWN: Agreed. The good news, you put it in. It's in the paper now. It wasn't in there before. It wasn't in any of the other documents before. So the improvements in the Reg Guides and the DRGs and BTP and there's another I wanted to call off with the architectures, watchdog timers, et cetera, et cetera, are significant improvements in terms of the guidance that's being provided out to the licenses as to what to be expected to see their designs. So I'll quit again for a minute.

MR. DARBALI: Okay. Now on slide 17.

CHAIR KIRCHNER: You address manual control. Yeah, just to address your last bullet, we interrupted you.

MR. DARBALI: Right. So we added a sentence that states that manual controls should be downstream of the equipment that can be affected by CCF. On slide 17, so to summarize, the staff revised BTP 7-19 to incorporate the policy in SRM-SECY-22-0076. We made changes after the September briefing in response to public comments and feedback received from ACRS members.

We also made clarifications throughout the document. And most importantly, there were no substantive changes made to the analysis methodologies for the acceptance criteria and the BTP. Next slide. And as we have said, we are working on issuing the final BTP in May. And as mentioned a few times, separately from the work of the BTP, we are also scheduled to brief the digital I&C subcommittee in

June on a broader range of digital I&C activities. Thank you.

MEMBER BROWN: Thank you, Samir. Do any members have any other comments that they would like to make?

Hearing none, public comments. Is there anybody on the public line that would like to make a comment relative to this briefing?

Hearing none, I will turn it back over to you. Well, thank you very much, Samir, Norbert, and Steve.

CHAIR KIRCHNER: I echo that thanks. I think we're coming up to taking a break. But before we do that, I would like to extend the opportunity to first I'll start with Vesna. Vesna, have you any comments you wish to make?

Hearing none, I --

MEMBER DIMITRIJEVIC: Sorry. I couldn't find my microphone. Nothing this moment. Maybe now discussion I would like --

CHAIR KIRCHNER: Okay.

MEMBER DIMITRIJEVIC: -- to come back to some. All right.

CHAIR KIRCHNER: Great, thank you. And also let me extend that opportunity to our consultants. Dennis Bley and Myron Hecht and Steve Schultz, any comments?

MR. BLEY: This is Dennis. No, I think the members asked the key questions and staff is had done a good job on this one.

MR. SCHULTZ: This is Steve Schultz. I agree with Dennis in terms of the overall presentation. The discussion this morning was a good follow-up to what was presented and discussed at the subcommittee.

And key points were made that reflected both that interaction and also resolved some issues here. So appreciate the discussion this morning and the presentation. Very nicely done.

CHAIR KIRCHNER: Also my error, Matt, I skipped over you. Have you any comments?

MEMBER SUNSERI: I don't have any comments, Walt. Thank you.

CHAIR KIRCHNER: Thank you. Okay. With that, members, any immediate comments? We'll come back after a break and we will take up our letter writing. And we can start that off with a discussion amongst ourselves as to key points.

And then we'll proceed with the draft letter that has been prepared. Hearing None right now, let's take a break for at least long enough to get coffee and take care of other business. So we will take a break until -- it's 10:15. We'll reconvene at 10:30. So we are in recess.

(Whereupon, the above-entitled matter went off the record at 10:15 a.m. and resumed at 1:00 p.m.)

CHAIR KIRCHNER: Okay. We are back in session. We are going to hear from research. Dave, would you like to make the introduction on Ron's behalf?

MEMBER PETTI: Sure. So members, this is part of our continuing briefings on different aspects of the research portfolio for RES. Ron's not here, so I'm filling in. He's still responsible for an item, though. He just doesn't know it yet. So first, we'll hear from management. Steve.

CHAIR KIRCHNER: Okay. Go ahead.

MR. RUFFIN: Good afternoon to you all and thank you for having us today. I'm Steve Ruffin.

MEMBER PETTI: Closer to the mic.

MR. RUFFIN: I'm Steve Ruffin. Good afternoon to you all. Thank you for having us to present today. I am the branch chief for the materials engineering branch in the division of engineering in the Office of Research. This is my division director Michele Sampson here.

MEMBER PETTI: You've got to talk straight into the microphone.

MR. RUFFIN: Today -- how about now?

MEMBER PETTI: Yeah, that's good.

MR. RUFFIN: Today, you will hear from our experts from research and from NRR as we present the review of our NRC research program, artificial intelligence and machine learning and nondestructive examination and in-service inspection activities. The presentation and subsequent discussion will demonstrate how research is prepared in the agency for the future through our work on new technologies and developments related to the use of artificial intelligence and machine learning for NDE and ISI and how this research is being used by the agency to meet our licensing and oversight mission objectives and how our investment in NDE research allows the agency to use research as mantra to be ready to meet the moment as the number of qualified NDE inspectors decline and as the nuclear industry is looking to take advantage of the advances in automation to enhance inspection capabilities.

Our presenters at the table today are Carol Nove from research, Dr. Stephen Cumblidge from

NRR, and also supported our excellent support, Dr. Pradeep Ramuhalli from ORNL on my left here. On the phone, we have Dr. Richard Jacobs and Jared Gillespie from PNNL. And with that, I'll turn it over to Carol. Thank you.

MR. CUMBLIDGE: She'll return. So next slide, please. This is Stephen Cumblidge from NRR, Piping and Head Penetration Branch, DNRL. Okay. First of all, let's quickly go over the introduction background and then I'll turn it over to Carol to describe the research program, talking about the commercially available automated data analysis packages and also some of the evaluations that have been ongoing of the machine learning for the use and ultrasonic NDE. And then at the end, we'll switch back to me and I'll talk about some of the outcomes and path forward for the research.

I'm not going to go over these acronyms. But for completion, you have them. Okay. So why is the NRC interested in the (audio interference).

CHAIR KIRCHNER: We're getting feedback. Those people listening in, please mute yourself so that we don't have feedback on the Teams link. Thank you.

MR. CUMBLIDGE: Okay. These seems to be

better. All right. So why is the NRC interested in this. Okay. One of the main issues, nondestructive examinations of nuclear power plants are basically we've incorporated ASME Section 11 and Section 3 include code in 10 CFR 50.55(a)(b).

So we have incorporated -- I'm sorry. Okay. And the industry uses NDE to find in-service and flaws and also for Section 3, pre-service flaws. Now the plants are getting older. And as they get older, we're going to start seeing degradation in new places.

And we expected to see more degradation as time goes on. And so we really need accurate and good NDE to deal with the aging of the power plants. Now the issue is, you can see this is a diagram.

We have an upper head inspection and the black areas, the inspected area of the tube. But the Rorschach test you see below that is what you get when you do an ultrasonic scan of one of these upper head nozzles. And it takes a lot of training and skill to be able to make sense of that.

And keep in mind in that scan, a couple pixels out of place would should an indication of a flaw. So analyzing that data takes skill and takes a lot of work. And AI might be able to help with this process.

And essentially, the industry, they would like to reduce how much time does it take to do these inspections. They would like to reduce radiation exposure during the inspections and also how many of them they have to do. Next slide, please.

CHAIR KIRCHNER: So you've piqued us, though. Before you move on now, this is the danger of presenting to the committee, especially --

MR. CUMBLIDGE: Yes.

CHAIR KIRCHNER: -- research material. So does that picture show a flaw and where's the flaw in the picture if it does?

MR. CUMBLIDGE: It shows --

CHAIR KIRCHNER: Or are there a few pixels out of whack? Or is this a good picture?

(Simultaneous speaking.)

MS. NOVE: So I believe this data -- this right here would be called a leak path which is not what the AI systems at the moment are looking at. But this would indicate that there's a leak path from -- this would be your nozzle wall and this would be a leak path between the nozzle wall and the vessel inside. So --

CHAIR KIRCHNER: On the other side, is

that another potential leak path growing?

MS. NOVE: That would be one. But I believe this is the one in this particular image is identified as a leak path as a through wall flaw.

CHAIR KIRCHNER: Okay. Thank you. And secondly, since you use the word -- and we have one of our experts here. It's not in your acronym list. How would you define AI as when you use it?

You just use AI. It's quite the buzz word. It's coming to be like risk informed in the agency. I think everyone is going to be AI. So when you use AI, could you just give us a context for what you mean?

MR. CUMBLIDGE: We have a slide later.

MS. NOVE: When we get into the research, I have a slide specifically on that. So are we good to --

CHAIR KIRCHNER: Okay. All right. I'll wait.

MS. NOVE: Okay.

MR. CUMBLIDGE: Okay. So as I said, it takes a lot of skill and training to be able to look through these images and find the relevant indications. How do you prove that a human being can do this? How do you prove that a human being has a talent to do this?

Well, we use what's called a performance demonstration. You essentially given the person a series of scans. And they have to be able to go through them and find the flaws that are there and not make too many false calls.

And the rules for that are described in Section 1, Appendix 8. So that requires you to find and size and do all that. And the procedure has to be able to find all the flaws.

So we have a really strict regime in place for the people. The problem is, is that the industry is projecting a shortage of the number of people who can do this in the future. So they are looking maybe using AI to help meet their future needs for doing these analyses.

Like, for this upper head inspection, you need two people to it. You need essentially a primary reviewer and a secondary reviewer because it's so challenging. So in place the upper head inspections, these are one -- we'll talk about this one more because it's one of the primary use cases for the AI or automated data analysis.

And it requires multiple qualified inspectors. It takes a long time. Also the person needs to be very focused the entire time they're doing the work.

A momentary lapse in attention can result in a missed call which has happened. So there's a lot of human factors related to these inspections. Fatigue, you're in a difficult physical environment for a long period of time while doing a complicated cognitive task.

And so AI could -- or artificial intelligence or automated data analysis could help make this easier for the inspector. Next slide. What we're seeing now is that you don't have to be an expert in computer programming necessarily or have to develop machine learning to use it at this point. There are a lot of open source pools.

If you google open source machine learning tool, you'll find a lot of these logos will pop up in your screen. And these tools which are open source, publicly available. They're becoming more powerful and easier to use as the years go on.

So again, this is -- it does not take a major effort in learning how to make hundreds of thousands of lines of code to use machine learning, artificial intelligence. You have to be an AI expert. That is not as energy intensive as it used to be. And the nuclear industry, EPRI and others, are funding work to learn how to use these tools for automated data analysis or some of the more challenging or monotonous problems in the NDE world.

MEMBER MARCH-LEUBA: These tools are not related to the most recent large language model, LLM, that we hear on ChatGTP and this talk?

MR. CUMBLIDGE: They're largely image recognition or image analysis tools.

MEMBER MARCH-LEUBA: Is there like 10-year-old -- actually 2-year-old -- 2023 second quarter?

MR. RAMUHALLI: This is Pradeep Ramuhalli. So you're right. These are not the large language model type tools. The LLMs. are similar tools. Similar methods use the underlying tools that PyTorch or TensaFlow, for example, might provide. But what the industry is looking at to our knowledge at this time is really using image recognition or classical data analysis type of approaches and the LLMs.

VICE CHAIR HALNON: So how do you make sure that it is a pure AI algorithm or whatever they call it and it's not adulterated like we saw with the recent experience with the problems that we saw on the news recently? So how do you ensure that? I mean, you got so many there. Are they all validated in some way? Or do you plan on endorsing certain ones for this?

MR. CUMBLIDGE: Skip ahead. Performance demonstration that we use for human beings were used for the algorithms are they're developed essentially. And now we're going to make sure that everything works in the end and how when we use them that they'll remain good in future.

VICE CHAIR HALNON: Okay. And to be clear, you're not looking at replacing the inspector human. You're looking at enhancing the throughput quality and maybe reducing the numbers that are required on each inspection. But there will still be a human on each of the inspections. Is that correct?

MR. CUMBLIDGE: The short -- the short term and long term. Short term, yes, there'll be people. Long term, I think there's a thought industry would like to get rid of the people entirely.

VICE CHAIR HALNON: You see a path there. It's potential.

MR. CUMBLIDGE: That's a much more challenging path if you have a person involved. But industry is interested in getting rid of the inspector entirely. But we'll talk about the path forward and how to make sure you keep things going well.

VICE CHAIR HALNON: Okay, good. Thanks.

MR. CUMBLIDGE: Okay. Next slide. So again, for the near term, it involves kind of the analysis of encoded data -- recorded data, not a person taking, like, with a probe and a pipe looking at an oscilloscope, trying to find indications on the scope. The data is recorded and you can look at the entire image at once. Also in the short term, we're talking about screening and identifying regions that are indication free or classifying a region as having a possible indication.

So kind of using the screening not as the final call. Also can be used for quality control for NDE exams. Whenever you do these exams, there's a lot of numbers you can put in.

I use this frequency and it was calibrated and making sure the person is putting in the right numbers and using the right numbers as they're doing it. It's a human factors error trap when you do -- it can be a compliance issue where they do an exam and they write the wrong things down on the forms. And then the regional inspectors see it.

I think this helps keep things going smoothly. In the longer term, data compression or less extraneous data being produced because you're screening a lot out. Generating NDE reports, essentially you have the large -- you might have a very large number of indications. You want to characterize them well. Maybe you can have the machine write the report for you when you double check it.

But it would do that for you. And also, they're working on real time analysis of un-encoded data. That is where the person would -- essentially when they were scanning with the hand holding up the probe running over and they see there's an indication of crack.

I might draw a box saying, oh, look at this area more carefully. Stop here. Also, for visual testing, right now if you have perfect lighting, perfect angles, perfect everything, they can see cracks. And then you have to draw boxes around cracks if you're scanning a camera over a cracked surface.

That doesn't necessarily work if you're at funny angles and the lighting is wrong. But for now , they would like to do that in the future. So they're working on that.

And also generating fake cracks or

training people. Essentially, they make fake indications raised on previous -- instead of writing a fake essay for someone, it can make note a new inspection for someone to practice on. So that's working on that for a longer term. Next slide.

So getting back to your question, okay, how can you use this automated data analysis? The first way and the way we focus on the most here today is an assisted examination. There is where you have a fully qualified inspector.

Someone already could've done the inspection without the automated data analysis. But there being an assisted by, and the examination is being facilitated by the AI or algorithm. And also, the qualified human being makes the final calls, not the machine.

And the second way is fully automated. The people are basically there to operate the equipment and the algorithm makes all the decisions. And that's much further off. We'll see what comes of that. Next slide.

So I think you've seen some of these images from EPRI and we're taking this from an EPRI report. But the simplest way of doing the automated data analysis is you basically screen. You run over the large amount of data.

And most of it, there are very few flaws out in the field. So most of it will say nothing here, nothing here, nothing here. Oh, here's an area of interest. Here's an area of interest, kind of flag the areas of interest.

And then the inspector largely goes over the areas of interest that have been flagged by the algorithm. And if you're doing this, you can tune the algorithm to be paranoid and to call much more than you would in the field that would never pass a performance demonstration test because it's calling too many things. But then the person would go through all the flagged areas and decide what was a real flaw and what was not a real flaw.

And this is kind of a leap forward in the thought of how to use these automated processed because the human being is still making the final calls. But if you look at one of those upper head exams, they take an incredible amount of data. You can see before you use the AI screening, there's, like, 4.4 miles of scans to go through. And after the AI, take 463 feet to go through. So it condenses things down and makes it much faster and less strenuous task for the analyst.

MEMBER BIER: Excuse me. Just for clarification again, the qualifications is the set fictitious flaws that were put in. Or they are a subset of actual flaws that were found?

MR. CUMBLIDGE: For qualification, you'd only use real flaws. And then if you're training, you'd probably want to use real flaws. And then for the qualification, you would certainly only use real flaws.

And then the inspector would only be getting areas that are flagged that might have -- you might be looking at some piece of geometry that might be a flaw. And it would be up to the inspector to determine is it geometry that it's catching? Is it some material pickup that it's catching, or is it an actual crack that is causing the indication that the AI found and flagged.

VICE CHAIR HALNON: Does it have an ability to rate the quality of the probe handling? I mean, because that's a big variable in itself, plus the coupling to the pipe itself too.

MR. CUMBLIDGE: I think the -- is the scan high enough quality to be put in? I think that'd be up to the programmer. It's not currently a requirement.

MS. NOVE: I'll jump in for a second and say that right now this is all encoded. So while we're in the encoded realm, the probe handling is --

VICE CHAIR HALNON: It is what it is.

(Simultaneous speaking.)

MS. NOVE: It's simply -- yeah, exactly.

VICE CHAIR HALNON: I mean, clearly -- well, the phased array. I mean, you do the phased array, I mean, you're looking at a bunch. So you could probably figure that out because you're looking at a bunch of different angles and stuff.

But if it's just a straight use of probe, I mean, someone is just wiggling their fingers a little bit can cause -- I mean, you can see that -- a human can see that. I mean, that's -- but will AI flag that as a flaw or -- I mean, it's a training thing. I mean, you have to figure that out, I guess, as you go.

MR. CUMBLIDGE: Even a pre-screen, do you have enough coupling? Basically, if you do the scan and it picks up from somewhere. You can get all sorts of weird artifacts from mishandling the probe.

VICE CHAIR HALNON: Right.

(Simultaneous speaking.)

MR. CUMBLIDGE: And usually what happens

--

VICE CHAIR HALNON: -- not clean or something does not --

(Simultaneous speaking.)

MR. CUMBLIDGE: Yeah, the surface is --

VICE CHAIR HALNON: AI is only limited to looking at the data that it gets accepted. So --

MR. CUMBLIDGE: Usually the encoded data, the analyst looks at it and says, retake that. That wasn't taken properly.

VICE CHAIR HALNON: Right, yeah.

MR. CUMBLIDGE: And then they send it back. And that would still be in the loop of a person would have to accept the scan saying it's okay. You could in theory.

VICE CHAIR HALNON: Yeah, I was just kind of probing. Where could the human be taken out of it? And it seems like there's some subjectivity up front on whether or not they're even going to accept as a good scan.

MR. CUMBLIDGE: And the human -- at this point, the human would have to say if that was a good enough scan to even put into the AI. Like whenever it's sent over from the plant to the trailer where the person is working, it would be up to them to then take that data and load it in and do all the work. If they didn't look up front, they probably wouldn't put it in the AI --

VICE CHAIR HALNON: Okay.

MR. CUMBLIDGE: -- for analysis. But that's me speculating. I don't actually know. Next slide, please. So what are the benefits? So the one thing about -- we did a lot of work on human factors of NDE a few years ago.

And what's a good thing that we do not have that many cracks or large flaws in the nuclear industry . Largely, if it's pretty robust, haven't had that much degradation which from a safety standpoint is great. From keeping people vigilant is not.

So you have a case where you don't have that many flaws. You can spend much of your career not ever finding an in-service crack which is great from safety but again not for training. Computers don't have this problem. They don't lose vigilance.

They're trained and you put them on the task. They're not going to get board. They don't have boredom in them.

Now that said, computers do make mistakes. But they're different than the ones that humans make.

Like, the computer is not going to have a lapse of attention.

But they will only make the same mistake 100 times in a row through a program to do that. So you can't expect a human being to be given the same problem many times in a row. We'll figure it out. The computer will not.

But if you have a trained analyst who really knows what they're doing paired with a well turned algorithm, you have the best of both worlds. And you pick out a lot of the cognitive problems that you get with human analysis, if you have a good AI with a good person. And also, if we do ever get to the manual, it'll be very good at reducing what goes to inspectors. It doesn't really speed up the scans in the field. Any questions?

CHAIR KIRCHNER: Make sure you don't program in an advanced team's beliefs, one of the ones that you think you're eliminating with machine learning.

MR. CUMBLIDGE: Advanced team belief is I get the head of the team says this is good and the lower people on the team go, oh, well, I defer to him. And the computer doesn't know what status is and doesn't care. Okay. Now the problems of the automated data analysis is that we're adding a layer of complexity on top of an already very completed process.

And that leads to a series of other issues. So for one thing, if everybody in the fleet -- let's say the one automated data analysis algorithm becomes really popular and everyone uses it. And then ten years later, we find out it misses something all the time.

CHAIR KIRCHNER: By a team belief.

MR. CUMBLIDGE: Yeah. It'll make the same mistake over and over again. What happens if you find out that it's been making that same mistake for ten years across the fleet. I don't want that day to happen. I'm proactively lazy.

I'm making sure that I will not be dealing with that because I don't want to have to try and go through every single scan that's been done for the past ten years across the fleet. And then also it's really complicated. People have a lot of questions.

What happens if licensees don't really understand what it can and can't do when they use it improperly and articulate to the licencees and the regional staff to try and figure out if they're doing it right. That's not good. Also, in the performance demonstration testing, what happens is that people will go through with the algorithm assisting them.

You could never pass without the algorithm. So you start off right now we have highly qualified people. In ten years, the people are going through learning with the algorithm.

Maybe there's weaning entirely on the algorithm and they don't actually have the skills to do the work independently. And that terrifies a lot of the vendors. Then also these algorithms can be very challenging to train and to retrain the source of the machine learning versions. And if you retrain them improperly, you can start with a good algorithm that is working very beautifully to retrain it for a site specific item and then it doesn't work at all. And how do you catch that?

VICE CHAIR HALNON: Isn't that reliance completely on AI where we just talked about where industry wants to go with this total reliance, not on human interaction?

MR. CUMBLIDGE: They would like to do that and make sure they have the proper infrastructure in place to make sure that wouldn't be stopped if it was going to not work.

VICE CHAIR HALNON: So you'd have to have some kind of, for lack of a better term, quality checks in that that maybe put a poison pill in something to make sure.

MR. CUMBLIDGE: We can talk about things to do later. But also one thing is if you -- let's say you would have the assisted analysis. But the person is almost useless. It becomes by default accidentally fully automated.

VICE CHAIR HALNON: Okay.

MR. CUMBLIDGE: So that, again, we don't want -- we need the people to be trained and qualified.

VICE CHAIR HALNON: And you get more confidence, then you can get more complacent --

MR. CUMBLIDGE: Right.

VICE CHAIR HALNON: -- relative to what your true duty is supposed to be if that's what you're doing.

MEMBER BIER: If I can expand on that point, you may also be familiar with this. But some of the literature on AI assistance in medical care, like, reading mammograms, what they find in experiments is that you can get better results from having a physician and a computer system than from having two physicians because the physicians kind of have common biases. They see the same things and whatever.

But I feel like as things go forward, we're going to need a new word for kind of the computer equivalent of social loafing, right? We know already, like, okay, one shift may overlook something because they figure the next shift will deal with it. But you may also overlook things, flaws that the computer didn't see because you think, oh, well, they're fine. They already did it, and now I don't need to do it.

So it's not just do they lose capacity to do it, but are they incentivized somehow to still do a really thorough job. I mean, you could even have cases where the computer system deliberately withholds a certain number of flaws to see does the human catch most of them and then throws them back in later. It's not like it's going to throw them out. But you don't catch at least three out of these five, then you haven't really done your job thoroughly. Go look harder kind of thing. So --

CHAIR KIRCHNER: Dennis, you have your hand up?

MR. BLEY: Yeah, I do. And it's kind of the opposite question from where Vicki was. As I understood the way the intent for the systems to work is the computer paths will identify potential areas where there could be flaws.

But then when a human comes in, like someone said early, it went from 400 miles to 400 feet. The human is only looking at the 400 feet. So anything the computer missed, nobody is going to look at.

MR. CUMBLIDGE: That's correct.

MR. BLEY: Yeah, so if there's any errors of omission by the computer will never find them until something breaks later or the flaw gets bigger and it gets identified. So it seems the only function of the human here is to say whether this potential flaw is really a flaw or not. So it isn't checking the computer in any way. It's reducing a number of things you have to look at after the computer has found a basic set. That seems like a gap unless you're really convinced all potentials are found by the computer which seems to be where you are.

MR. CUMBLIDGE: Yeah, you have to tune the computer algorithm to be very paranoid essentially and catch everything.

MR. BLEY: And assume that you have seen everything in the past so you can do that.

MR. CUMBLIDGE: I think we have a slide on that as well.

MEMBER BIER: Yeah, but also the paranoia runs the opposite risk, I mean, for radiation alarms. We've seen cases where people say, oh, yeah, that's just always going off. I'm just going to ignore it. And again, there needs to be some incentive to make sure that somebody don't just say, oh, yeah, this always finds way too many flaws. I'm not going to bother looking at these.

MR. CUMBLIDGE: It has some flagged areas of interest but not make so many false calls and so helping the entire area. And the person being -- there's never any flaws. I've been doing this for 20 years. I've never seen a flaw. So it helps with that.

MEMBER BIER: Thanks.

MR. CUMBLIDGE: Also, the last thing on this and I'd mentioned. If people want to do this right, if they want to use these algorithms in the field and use them, it's going to take a new class of experts. Because it takes to use these properly, you need ultrasonic experts and it takes AI experts or machine learning experts to use them.

So it's another person in the room who has to be highly qualified to do the work. So that's another area where it's a hazard. Again, we're taking an already complicated process. We're doing the exams. And we're adding another layer of expertise which with the associated issues. Next slide.

MR. BLEY: Before you go on --

MR. CUMBLIDGE: Mm-hmm.

MR. BLEY: My apologies. I was trapped in the cone of silence. When we first began this talk, they introduced all the people who were here from the staff. But I don't know anything about all you folks. On that group of people, you have one or several who have deep experience in actually looking for flaws themselves.

MS. NOVE: Part of our team -- this Carol Nove from research. Part of our team at PNNL includes a qualified ultrasonic Level 3 who worked in the field in the industry for 38 years. So he had extremely a lot of experience.

MR. BLEY: I'm glad you have at least one. I'm not familiar with the qualification levels. I'm assuming from the way you said that, Level 3 is the highest level?

MS. NOVE: Yes, it is.

MR. BLEY: Okay.

MR. CUMBLIDGE: And Carol, I've worked in the laboratories or in the field for many years.

MEMBER DIMITRIJEVIC: Hi, this is Vesna Dimitrijevic. I have experience because I work in risk informing in-service inspection part. And we look in all of this probability of the missing false and false identification. Did you guys have any -- do you have any results which will be defining probability of detection with the probability of missed calls for this combination or this was just human?

MR. CUMBLIDGE: Later on, we have extensive --

MS. NOVE: Looking at how machine learning impacts the probability of detection is something that we haven't addressed yet in our research program. But that is part of where we're going.

MEMBER DIMITRIJEVIC: Because you have certain -- curious: are you improving on probability of detection, or on probability of a missed call? You know, I was wondering, does this result in more misses or more false calls? That's -- you know, I don't really have to know exact probability of detection but I'm just curious. Do you have more misses or more false calls in this combination?

MR. RAMUHALLI: This is Pradeep. With the smaller data sets that we have been using for to date, the results are mixed in that in some instances for certain types of flaws, certain locations, certain other factors that play a role in detection. You do have fewer misses in other cases, particularly in flaws that may be extremely small and challenging to detect anyway. The missed call rate does seem to go up.

MEMBER DIMITRIJEVIC: So it was also curious when you say that you turn the AI paranoid. Does that mean you're sort of reducing the size of the flaws which need to be detected?

MR. RAMUHALLI: When Stephen was talking about training the AI or machine learning to be paranoid, what he's referring to is training the algorithm to detect or to call a much larger fraction of the regions than otherwise. So not only would ineffectively flaws be called or caught but it might flag regions of geometry, fabrication flaws that may exist, for example. And so even noise might -- the eye may be trained even with examples of noise to say, hey, if you see something like this, flag it as something that needs to be viewed by the analyst. So that's what that paranoid means.

MEMBER DIMITRIJEVIC: All right. So it's much more wide than, than just the size of the flaws?

MR. RAMUHALLI: That's right.

MEMBER DIMITRIJEVIC: Okay, thanks.

MR. BLEY: And one -- this is really the last question for me at this point. I hear what seems to be a contradiction. You've talked about the computer doesn't learn. It makes the same mistake over and over again. But then we call these systems machine learning. And the machine learning systems, I've been familiar with either through mathematical algorithms or some other means, actually learn from their experience. And they don't make the same mistakes over and over again which is what here?

MR. CUMBLIDGE: If you retrain them. If you do not retrain them -- so you train the algorithm. You fix it. That would make the same mistake over and over again. If you then retrain it with new information, then it would learn.

But it takes the retraining for the machine algorithm to learn. You generally would not use it in a way that it would learn as you go. And we'll talk about that, how you --

(Simultaneous speaking.)

MR. BLEY: But does that somehow inform

that it just made an error so it can correct itself? I think I cut off my mic. You give it a new training set. Okay.

MEMBER BIER: And also when it's retrained, it may make new errors that it wouldn't have made before the retraining?

MR. CUMBLIDGE: That is correct. Yes, we have to deal with that. We're working on a process on how to make sure that, one, it's catching the flaws that you need and that retraining is done in such a way that it, one, keeps finding the flaws that you need and also it does not make too many false calls and raise other issues. So we're working on how to do that and how to make sure -- unless you retrained, it will not keep -- what you do not want as algorithm that changes during the inspection.

You don't want to start off with something that passes qualification, does a great job, and as you're doing the inspection, starts changing on you. That would no longer be a qualified procedure. So then it starts making the same mistake over and over again. That is, if you don't retrain it, it will.

MEMBER DIMITRIJEVIC: Well, the best way to do that is after every non-distracting examination, you do distracting examination and tell him learn the algorithm to understand when was the right and when was wrong. And you do that a million times, and I will be perfect.

MR. CUMBLIDGE: I think the licensees would object to that.

MEMBER DIMITRIJEVIC: Yeah, I know.

MR. CUMBLIDGE: So the one problem, though, with the automated data analysis is that as the plants get older and as new reactors are designed, new things are going to happen. We're going to see new things or flaws will appear in places they haven't found. Like, what happened in France, they were looking for the thermal fatigue flaws and they found stress corrosion cracking.

That wasn't expected, but it certainly happened. And these methods can be very good. If you have a known problem they're very well trained on, very well tuned for, they can be very good at that. But the new things you'll guarantee that even the best trained, best tuned algorithm will be able to find the new things or the new places.

And that never happens there, so there's no indication. So we have to be careful as we go forward understanding the limitations of even the best algorithm will have weaknesses. And whereas a human being might go, that's weird. Let's take a look at this more. The computer won't. It'll just say, no, I'm not programmed to call that. I'm not going to call it.

MEMBER BALLINGER: This is Ron Ballinger. I'm sorry for being late. I was literally trapped in a cone of silence. These algorithms also can be -- you're saying that it could missing things that are new.

But you can program them. I don't like to use the word, training. It sounds like there's some human involved. But if there's an anomaly that occurs, these programs can be programmed to flag an anomaly which is outside of the training set but which is different.

That alerts the inspector that, oh, there's something going on here that it's not trained to see. But it's different from the validation set that we use. So it's not completely in the dark if you do it right.

You'd have to do it right, though. It's theoretically possible, but you'd have to do it. Make sure that was done that way.

MEMBER MARCH-LEUBA: One central flaw is that your training said you have a lot of small defects. You have a lot of those. You have only one or two big defects. It doesn't learn about the big ones. It identifies the small ones. Oh, okay. I don't know. Never seen that one.

MR. CUMBLIDGE: How you make your training set is vital. We've learned that.

MEMBER MARCH-LEUBA: You need millions of data points to say something is -- sufficient number of weights under the network to have some reasonable chance of working.

MR. RAMUHALLI: This is Pradeep again. To answer the earlier question on anomaly detection, that is an approach that we are investigating at the moment to see what the capabilities of that are and how far can you go. What source of aspects or issues of fact might play a role in the ability of the machine learning to detect indications of anomalies that I may not have seen in its training set before. But if it encounters it, a new form of degradation essentially.

MEMBER BALLINGER: Can the algorithms estimate the probability of detection for each flaw that it sees so that the analyst that's looking at the data gets a feeling for how good the program is doing. Maybe I'm not using the right words. But it's all probability of detection.

MR. RAMUHALLI: Can I take that? So in principle, yes. It's not probability -- I would not call it probability of detection. But I would call it, like, an uncertainty estimate.

MEMBER BALLINGER: Okay. Uncertainty, that's fine too.

MR. RAMUHALLI: In principle, yes. In practice, not all algorithms are set up to do that from the get go. And that -- how one sets that up is something that will need to be investigated or examined.

MR. CUMBLIDGE: And with that, we switch over to Carol.

MS. NOVE: So good afternoon, everyone. I'm Carol Nove. I'm a senior materials engineer in the Office of Research. And I am the lead for the NDE of vessels and piping program that we have the national labs.

And Stephen was kind enough to write a user need request, NRR 2022-07, which included a task on automated data analysis. And the task asks research to provide a technical basis describing the current capabilities of machine learning and automatic data analysis for NDE. And the way research is handling that UNR request is twofold.

One, we're doing an evaluation of commercially available systems. And that work is led by PNNL. The other is evaluating machine learning for ultrasonic examinations, and that work is being led by actually Pradeep Ramuhalli who's here with us today from Oak Ridge National Labs.

So this slide goes to the question that was asked earlier about how we define artificial intelligence. Our work is looking into kind of two arenas. There's rule-based where decisions are based off of explicit rules and it's easy to determine why specific decisions are made.

And then there's learning-based which would get into the artificial intelligence machine learning, deep learning, and so forth. And the decisions are based off training data. And it's difficult to determine why specific decisions are made.

In terms of the analysis mode, there's two types that we're considering. One, where ADA provides the analyst with a flag data set which is what we've talked a lot abut so far today. And the other would be automated with no analyst involved.

CHAIR KIRCHNER: Can you put that figure back up? I can't make sense of that figure. You've got things nested in a way that I wouldn't consider artificial intelligence. I would put it perhaps -- you've got algorithms.

You have machine learning. You've got deep learning which is just more horsepower on machine learning in most cases like deep blue or something. And then you have artificial intelligence. This is something completely different. So I don't get this diagram.

MEMBER BIER: Well --

CHAIR KIRCHNER: It doesn't make any sense to me.

MEMBER BIER: -- I think it's consistent with how things are currently used.

(Simultaneous speaking.)

MEMBER BIER: That's okay. Artificial intelligence these days is mostly synonymous with machine learning and deep learning. But historically, if you were looking at artificial intelligence in 1990, it would've been all rule-based. So that's why there's a part of that deal that's outside. There's other approaches. So I'm fine with the diagram.

CHAIR KIRCHNER: So it works for you?

MEMBER BIER: It works for me. But we can discuss if you want.

(Laughter.)

MS. NOVE: Thank you. So --

CHAIR KIRCHNER: But let me interrupt you, though. What would you consider AI? I mean, on a more serious note, we had the good fortune of some presentations on artificial intelligence, the committee has. And those were very interesting because mainly what I'm seeing is programming.

I'm not seeing something that goes much beyond that. So when the agency starts bending about AI just like risk informing everything, is that just jargon because that's the current buzzword? Or we want to look cool? Or we have really intent behind using these words?

That's my concern. Everything now is risk informed. Okay, fine. So for the last five or so years in the agency, that's been quite the movement. Now we're picking up some new jargon including AI. So what's your -- when you're using that, what do you really mean?

Because if you took the AI off this diagram, I would be perfectly satisfied personally. But you don't have to please me. I'm just curious as to what does it mean in the agency to be throwing these words around.

MR. CUMBLIDGE: We're doing confirmatory research of what industry is proposing. And so we aren't -- essentially, we're looking at what they're doing. And --

CHAIR KIRCHNER: So you're repeating their words?

MR. CUMBLIDGE: Essentially. And also, they said we're going to make this diagram. We got it from elsewhere. So we're doing confirmatory research to make sure that what industry does is going to be effective in the future. And we're using their terms, but it's confirmatory work.

VICE CHAIR HALNON: In the past presentations we've had with both research and industry, we've asked where is AI going to be applied to? And the answer is, well, we're not going to apply it anywhere right now from a safety perspective. It's all going to be business systems and other stuff. You see this as the first application of AI that is being approached that could potentially affect safety?

MS. NOVE: It's one of the three identified applications that the agency has identified for use initially in the field. And this actually has been used. There's been at least four demonstrations in plants of this technology, three for upper head exams and one for dissimilar metal weld exams.

And right now, industry is using it in a way where they still have their two analysts doing independent review of the data. Then the data goes to licensee for licensee oversight to review the data. And then it goes to the system for review. In the not too distant future, they would like to replace at least one of those analysts with this type of technology.

VICE CHAIR HALNON: The reason I ask, though, is I'm wondering if we're dealing with two subsets of people here with folks that are innovating in their company. We're going to use AI for corrective action. We're going to use it for our business processes.

And then you get this other faction over here that's not talking to these guys that are actually talking about using it. Because when you ask this on one hand where you're going to use it, oh, it's not going to be safety. Don't worry about.

You're going to learn a lot about it first. And now we're getting this AI, really active watching what they're doing. And we never got that really from our AI presentations. So are we dealing with the same people here?

MEMBER MARCH-LEUBA: Yeah, but there is a big paradigm change. I'm using big words.

VICE CHAIR HALNON: Paradigm is not a big word.

MEMBER MARCH-LEUBA: Up to now, the ringleaders have insisted in knowing that you know what you're doing. When you talk about analysis codes, I can ask you, as a regulator, what is this line of code doing and why is it there? And you should be able to answer that or you don't get a license.

We are now going to a completely different mode which we call AI in which the applicant comes and says, I have a black box here. No idea how it works, but it works, and I want to use it.

That's AI. And it's a big paradigm change for the regulator. And need to be read to understand what is going to be done and how are we going to accept these black boxes?

MS. NOVE: And that's why Stephen wrote the user need that he did.

MEMBER MARCH-LEUBA: It's crucial.

MEMBER BALLINGER: This problem is basically a pattern recognition problem. And what you're trying to do is to prevent false negatives.

That's the goal.

You don't want to have a crack that you don't see that cause problems. And so the role of AI, and I hate to use that word again, is the analyst in the rules-based program only recognizes what you program in it. When you hit it with what we call artificial intelligence, whatever the heck that means, what they need to do is to ensure that whatever it does, it affects your ability to make -- to eliminate false negatives.

So the variables that AI can handle compared to what a human can handle is much bigger. And so there's an opportunity to take the data that's whatever the data is and to identify kind of nuances in the data that the analyst or the rules don't see. So that's the way I look at it because if it doesn't help you with false negatives, you're done.

VICE CHAIR HALNON: So back to my question because it was -- my colleagues went to a different spot. Am I talking to two different groups of people here?

CHAIR KIRCHNER: I don't know.

MEMBER BIER: Yes.

VICE CHAIR HALNON: That's something that we got to figure out because this just -- I mean, Vicki, what's your opinion?

MEMBER BIER: Well, I don't have a good read of what the nuclear part of the industry is doing. But I think we have to be careful about using the word, industry, because there is also presumably a whole ecosystem of AI companies or NDE companies out there. And historically, I think the software companies are much more risk tolerant shall we say, right?

The nuclear companies, at least the established utilities mostly know, okay, I don't want to do anything crazy. I don't want to get in trouble with the regulators. I don't want to have a costly flaw that goes undetected and then I'm shut down for two years or whatever.

But the software industry is completely different incentives. It's, like, if we don't ship this in the next six months, it's going to be worthless because the next company is going to have a better version. And we want to capture all the users before our software becomes obsolete. So I think we also have to be careful to distinguish nuclear industry versus software industry or whatever you want to call it.

VICE CHAIR HALNON: It looks like you have a couple friends that want to --

MS. NOVE: Yeah, I was going to say we have friends chiming in online. Matthew and Luis?

CHAIR KIRCHNER: Yes, Matthew. Go ahead.

MR. DENNIS: Hi, this is Matt Dennis, NRC staff from the Office of Research. And I was waiting until if needed. But I can try to address the question that was just brought up about are we all talking on the same page. Luis also raised his hand to assist with this question.

But Luis and I are the two that have presented to the ACRS subcommittee a couple times now on the AI strategic plan and the implementation plan and the all things AI at the agency. So I'll just say that, yes, we are all on the same page. I think there is some slight amount of confusion about what industry says.

And I say this as saying the nuclear industry says they are going to use these in non-safety applications. But I do agree with what Carol stated is that there are a handful of examples, this being one of them, the NDE example, where it may be moving into that component or that line that we have in the AI strategic plan that says NRC regulated activities. So it's not AI in the control room.

It's not AI in a safety system. But it is in an area that is NRC regulated. And so this one and couple other examples are ones where it hasn't happened yet. But as you're seeing from the presentation today and as Carol mentioned, it is something that we're monitoring quite closely and is consistent with all the other activities we're doing because we see it as being potentially near term.

VICE CHAIR HALNON: Thank you, Matt. That helps a lot. Appreciate it. Luis, did you want to get on?

MR. BETANCOURT: Yes, just to piggyback on what Matt was basically saying. Like, I guess the comment that I want to make and reemphasize, we have been monitoring this very closely with Carol's team as well as the other use cases that went across the nuclear industry. I think you guys are aware that the majority of the stuff that industry is doing is to use it for improving operational performance and mitigate risk.

But it's now slowing moving from that as well as from the research into the regulated area. This would be one of the first areas that we believe is going to be touching upon that. And like Matt mentioned, we haven't seen anything that's going to be controlling the power plant.

It's going to be mostly like made in the decision making. AI is only being used for informing those decisions and you still need to have a human component to be able to verify the decision making still done by the human. But we are coordinating across the agency all of the use cases like this one. And there's other ones that we know that they are frontrunners. But we believe that this is going to be one of the first ones coming on our doorstep.

VICE CHAIR HALNON: Thank you, Luis. You can go on. I'll have a question at the end about more stuff. But --

MS. NOVE: Okay.

VICE CHAIR HALNON: -- we're done with this one. Thanks.

MS. NOVE: So the objectives of the research program are to assess the current capabilities of automated data analysis and provide the technical basis to support regulatory decisions and code actions. And we have four primary expected outcomes to identify capabilities and limitations of automated data analysis. And we're specifically focused on ultrasonic NDE applications, identifying factors influencing the performance and their impact on NDE reliability, recommending verification and validation approaches and methods for qualifying machine learning, nuclear power plant NDE, and identifying gaps and existing codes and standards.

The first part of the program I want to talk about is we took a look at rule-based automated data analysis. And we started that process by doing a literature review and found that almost all of the recent publications are dealing with learning based analysis. So the literature -- the recent literature really wasn't addressing rule based automated data analysis.

And rule based ADA is usually used for flaw detection and signal processing. And typically an amplitude threshold would be used to identify flaw signals about the noise floor. And then there's some signal processing that can help improve the signal to noise ratio.

So, it will be -- it can achieve a high detection rate. But it also can -- that high detection rate comes with typically high false call rates because it's not able to consistently distinguish between geometric responses and flaw responses. And you can see that in the graphic on the right where amplitude threshold has been set and the software is identifying geometric responses as well as a flaw response.

So in this work, we looked at two commercially available systems. One was UltraVision and the other VeriPhase. We used carbon steel and rod stainless steel mockups in the work. And the qualified Level 3 UT analyst also analyzed all the data.

And then there was a statistical analysis conducted of the results. And those software analysis systems were optimized to have similar performance with the goals shown in the green box in the graphic being a false call probability less than 20 percent and a detection rate greater than 80 percent. And the inspections for carbon steel shown in blue achieved greater than 80 percent detection and less than 20 percent false call probability. And the inspections on the wrought stainless steel, the false call probability was too high. And then the performance of the UT --

CHAIR KIRCHNER: Carol, sorry to interrupt.

MS. NOVE: It's all right.

CHAIR KIRCHNER: Again, you're making us think and it's after lunch. So I'm a bit slow picking up what you just said. Why stainless not as detectable --

MS. NOVE: Oh, grain structure. There's a lot more noise sources in the grain structure of austenitic steel.

CHAIR KIRCHNER: That's the physical reason.

MS. NOVE: Yeah, carbon steel is clean. It's like looking through butter.

CHAIR KIRCHNER: Yeah, so does that suggest once you know that, then you tune your detectors accordingly? Seems to me I could readjust.

MS. NOVE: Not for this rule-based.

CHAIR KIRCHNER: Okay.

MS. NOVE: Because again it's only looking at an amplitude threshold. And the austenitic stainless steel and things like dissimilar metal loads are going to have a high noise threshold no matter what you do, if you're simply looking at amplitude threshold.

CHAIR KIRCHNER: So that suggests maybe a different technique?

MS. NOVE: It suggests that this --

CHAIR KIRCHNER: Because we have a lot of stainless steel?

MS. NOVE: Right, this --

CHAIR KIRCHNER: I know we have a lot of carbon steel and a variance for --

(Simultaneous speaking.)

MS. NOVE: So it suggests for these rule-based analysis that they're too kind of simple minded to work in the kinds of materials that we have in plants and that they are fine if you had a lot of carbon steel piping. A lot of this is being used in the oil and gas industry where a vast majority of their systems are carbon steel. And that's what these were really developed for.

But we wanted to understand could these be used in nuclear applications. And our bottom line was not really. But they could be potentially used alongside learning based systems and together make some use of them. Move on.

MEMBER MARTIN: Actually, could you explain the graph? I'm trying to understand that.

MS. NOVE: So the graph, the x-axis is false call probability. And the y-axis is detection rate. And you want your false call probability to be less than 20 percent and your detection rate to be greater than 20 percent for anything that you deploy in the field.

And then in this graph, the triangles -- the blue triangle represents the performance of the qualified inspector. And there's no curve for the qualified inspector because that analysis wasn't done based on an amplitude threshold whereas the blue lines and the orange lines were tied to an amplitude threshold that were used to optimize the performance. So we're seeing best case performance for these different analyses.

MEMBER MARTIN: The blue and the yellow, this is the software basically?

MS. NOVE: Yes, so the blue is carbon steel and it's the two different software programs. One is UltraVision. The dashed is UltraVision. The solid is VeriPhase. And the orange lines, it's UltraVision and VeriPhase. And then you can sort of make out the qualified inspector's performance in the orange triangle.

MEMBER BIER: So you want to be in the upper left, correct?

MS. NOVE: Yeah, you want your performance in this raw curve shown in this box is where you ultimately want to get to. So it showing that for carbon steel, the performance is adequate. But we don't do a lot of carbon steel exams --

MEMBER MARCH-LEUBA: What do you mean adequate?

MS. NOVE: What's that?

MEMBER MARCH-LEUBA: I'm hearing these guys from the autonomous driving training that we only killed 20 people this month. We're doing good.

MS. NOVE: Yeah.

MEMBER MARCH-LEUBA: I mean, if you fail to identify 20 percent of the flaws in the material, stop the program. We're not going to accept it. I mean, if your frame of mind is at 80 percent detection rate is good, you're working in the wrong industry.

MS. NOVE: I was going to say this is in terms of the reactor. This is just one aspect that you would look at. I mean, we do qualification --

MEMBER MARCH-LEUBA: You have it on here two different industrial/commercial systems. Neither of them work.

MEMBER BIER: Jose, just to chime in for some perspective, probably most medical tests you will ever have are also in that green box.

MEMBER MARCH-LEUBA: Yes.

MEMBER BIER: So --

MEMBER MARCH-LEUBA: And you get false positive. But that's not -- we don't live in that universe.

MS. NOVE: We also don't work in a world where our inspectors going out in the field are qualified to -- they pass qualifications. But they don't have to find every single flaw. There's a qualification where they have to find so many and they're allowed so many false calls depending on the qualification that they're looking to pass. So it's not a case of we're expecting an inspector to get 100 percent detection.

MEMBER MARCH-LEUBA: Yeah, but 20 percent failure rate on your best case?

MS. NOVE: This is a small data set. This is --

MEMBER MARCH-LEUBA: I made my displeasure clear, I think.

MS. NOVE: I understand. But again, we're looking to say is this something that we could even consider in the nuclear industry. And is there a technical basis because again we're driving it, developing the technical basis. And our answer was these two particular commercial systems would not be adequate for nuclear applications. That's the bottom line there.

MEMBER MARTIN: This is a small part of a much bigger picture in the sense that you're saying, well, the inspector didn't find these flaws. The flaws, this whole system has got to be -- it's part of -- you're saying to yourself, what size flaw can I tolerate? I need to be able to detect those flaws that I can't tolerate.

What are the ones I can't tolerate? Well, they're the ones that'll grow between inspection intervals and get me in trouble. So the fact that an inspector doesn't see a flaw X length, if X length is not problematic between inspections and this probability of detection.

The next time where it's a much bigger flaw is 100 percent. This all fits together with how this system has to work. It doesn't have to be -- it has to be 100 percent effective at some point, and that point has to be before you ended up with an unstable crack. Am I preaching here? I mean, that's really what we're interested in.

MR. CUMBLIDGE: I'm sorry. I got sampling statistics and whatnot. In general, the gray box would be considered on par with what you'd expect a human being in the field to be able to do.

(Simultaneous speaking.)

MEMBER DIMITRIJEVIC: -- these three

triangles. The blue triangle and yellow or orange triangle what you say, the qualified inspector. Is that what your expectation is in qualification? Like, blue shows 100 percent detection and zero false rate.

And for the WSS shows 90 percent or 10 percent. So what does the triangles on the graph means, the orange triangle mean? That means that expectations in qualification?

MS. NOVE: No, that's the results of the qualified inspectors' analysis of the data that was used in this research program. But in terms --

(Simultaneous speaking.)

MEMBER DIMITRIJEVIC: That'll give us comparison between human and ADA, right?

MS. NOVE: Yes.

(Simultaneous speaking.)

MEMBER DIMITRIJEVIC: -- you must perform actually much better.

MS. NOVE: In this particular case, yes.

MEMBER DIMITRIJEVIC: In both cases, both for the CS and for WSS, right?

MS. NOVE: Right.

MEMBER DIMITRIJEVIC: Okay.

MS. NOVE: So moving into machine lining, the scope of NDE, machine lining for NDE is pretty large. And so to focus the work, we're limiting our work to machine learning for ultrasonic NDE. And even that is a fairly large scope covering different materials, types, and inspection procedures.

So the work is being conducted sequentially starting with simpler materials. We started with austenitic welds and some dissimilar metal welds and conventional probe data to quickly identify key factors and to help establish machine learning evaluation pipeline before we move to more complex materials, flaw types, and inspection procedures. So the empirical evaluation is ongoing with the first phase of the work focused on capabilities and limitations of machine learning for NDE and building toward the rest of the aspects listed here, including identifying factors influencing the applicability to other inspections, meaning the inspections of other materials such as cast austenitic stainless steel, dissimilar metal welds, and reactor vessel upper heads.

We're assessing the effects of data augmentation, including using simulated data, establishing methods to quantify confidence in the machine learning results, and assessing the capabilities for flaw size quantification from ultrasonic data. So the generic workload for the research program, and this part of the program is the empirical aspects of it, the data collection has been done at PNNL. And the machine learning aspects are being done at Oak Ridge.

And so initially, we've collected ultrasonic NDE data from a variety of materials with multiple probe designs, frequencies, and wave modes. Pre-processing the data to remove noise and outliers, training the machine learning algorithm on the pre-process data, and then using the trained algorithm to analyze new ultrasonic data and assessing the results using multiple metrics. The graph on the right shows the flaw distribution for the six specimens used in the initial phase of the work.

Specimens were four stainless steel specimens and two dissimilar metal welds. They included plates, pipe segments, and whole pipes. The flaws were thermal fatigue cracks, saw cuts, and EDM notches.

And the data was collected with conventional and phased array probes using both sheer and longitudinal wave modes and multiple sonification angles. For the data collected on these flaws represented a range of materials and flaw conditions.

And the data set continues to be expanded as the two national lab teams review and process available data.

And the goal is to eventually publish the data, enabling researchers to use an extensive and common NDE data set to advance the machine learning for the NDE area. So this slide shows a high level overview of the workflow used for classification by the Oak Ridge team. The work in this phase of the program was done using ultrasonic data that is naturally represented in the form of two dimensional B scan images where the travel time of the ultrasound is displayed in the image along the vertical axis and the linear position of the probe is displayed on the horizontal axis.

The amplitude of the echo from anomaly such as defects correlates to a color scale which is not shown here. Using 2D images allows the work to leverage generic advances and image analysis and helps keep the work focused on our NDE specific advances being pursued by the nuclear industry. And in other words, or confirmatory research is focused on the application to NDE instead of focusing on developing new analysis techniques.

So it's along the lines of what Vicki was remarking on earlier. The first step in the processing of the B scan data is pre-processing it to a consistent shape and size of the model input. Then the classification is performed using deep networks such as convolution networks as they are widely used for analyzing images.

Once trained and fine tuned, the machine learning model is used with a blind test data set to assess performance. Multiple experiments were conducted by varying different factors, for instance, data size, flaw types and location, inspection frequencies, and wave modes. In the lower right of the slide, we show some of the metrics used by the Oak Ridge team to assess the model's classification performance.

They include classification accuracy, confusion matrices, two false positive rates, and receiver operating characteristic curves. Lastly, shown on the lower left is a graphical representation where test results are plotted against the original ultrasonic scan data. The orange area represents the B scans classified as flaws and the gray area show the B scans that were classified as non-flaws. And the bold lines are the true positions of the flaws. I'll pause her for a second if there's any questions?

VICE CHAIR HALNON: Is the overlap

anything -- because in some areas, there's some bold blue in the gray area. Is that --

MS. NOVE: Yes, the edges.

VICE CHAIR HALNON: It's the edges of it?

MS. NOVE: Yeah, and edges, we've learned through the work that there's -- labeling the training set is very important, especially around the edges.

VICE CHAIR HALNON: Okay.

MS. NOVE: I don't know if Pradeep wants to --

MR. RAMUHALLI: This Pradeep. So that is correct, what Carol was saying. As we work this research, we have learned that labeling the data so that the ML method can learn from that data properly is important. And that actually goes back to something that Stephen mentioned earlier about making the machine learning algorithm paranoid so that it actually calls more indication -- it calls more regions for review. Labeling the data, the training data properly is going to be one of the critical aspects in the process.

VICE CHAIR HALNON: Can you address the splotches on the bottom left? There's a flaw.

MS. NOVE: This right here?

VICE CHAIR HALNON: Yeah, then there's

some blue, dark blue. And then it looks like it intensifies a little bit and it's called a flaw. Can you just interest that issue between the two flaws?

MS. NOVE: I would say that is one flaw. But flaw reflection -- flaws won't reflect uniformly back to the probe. So if you're just a little bit off, you can lose amplitude. But you're recalling that in the field, it would call from --

VICE CHAIR HALNON: One whole flaw?

MS. NOVE: Yeah, here to here. And if you had something like this, there are proximity rules that would tell you, do you need to call this as a single flaw or do you need to call this whole thing as one flaw? That goes into what an inspector would do, and we're not getting into that process.

VICE CHAIR HALNON: Is that horizontal alignment -- is that reflective of the depth? The one on the right is higher in the blue than the one on the left.

MR. RAMUHALLI: So in this case, the data is actually showing like a plan view or a top view of the --

VICE CHAIR HALNON: It's just different.

MR. RAMUHALLI: It's just different, locations relative to the center line.

VICE CHAIR HALNON: Okay.

MEMBER BALLINGER: Can we go back the previous slide? Who decides what's an outlier and how do you decide?

MS. NOVE: Do you mean like an outlier, like --

MEMBER BALLINGER: Well, it says here pre-process the data to remove noise and outliers. I understand what noise is. Outliers, I'm always skeptical of a priori getting rid of what you call an outlier when it might not be an outlier. It might be the most important piece of data.

MR. RAMUHALLI: So in this case, the -- all of the -- in this particular analysis, all of the outliers that we are referring to her are essentially noise. And that is reviewed by -- that is based on review by the experts on the team who have reviewed the data and said, this is noise. This is really the region of the data that corresponds to a flaw. This is geometry, et cetera. So when it says noise and outliers, we are really only talking about noise here and kind of cleaning up the noise.

MEMBER BALLINGER: These outliers might be an indication of something unusual that you called an outlier.

MR. RAMUHALLI: That is correct.

MEMBER BALLINGER: A statistical two sigma or something like that, you just threw it out whereas there may be information contained in those outliers.

MR. RAMUHALLI: That is correct. And so eliminating based on -- eliminating data based on, like, a statistical two sigma, three sigma, whatever the criteria is. My opinion is if you're selecting data for training, it's probably not the path to follow. What you do want to do is to eliminate clear examples of noise in the data so that the algorithm is given consistent data to train with, to learn from.

MEMBER MARCH-LEUBA: Yeah, but not in the field. You're going to get some outliers too.

MR. RAMUHALLI: That is correct.

MEMBER MARCH-LEUBA: Your algorithm should be able to know what is good and what is bad.

MR. RAMUHALLI: Yes.

MEMBER MARCH-LEUBA: If you only feed it nice data, I'm very skeptical.

MR. RAMUHALLI: So in general, there are two ways that we've been looking at this data. And there's more data, as Carol pointed out, that's being added to this every day. One approach is really looking at the examples of both flaws and not flaws.

Some flaws here might be geometry, might be just plain noise as well from gray noise in the data. And seeing if the algorithms, if the machine learning can learn to differentiate between those two cases. Another scenario goes back to the anomaly detection approach that was brought up earlier where you're training the algorithms with the data that you have.

But you're also telling it that this is not what's normal. This is what's not normal that we know of. But this is what we certainly know is normal.

And then having the algorithm figure out if anything anomalous crops up, and whether it has seen it or not, whether it's part of the training data set or not, to still be able to flag that. How well can some of these algorithms perform in those cases. So both of those are approaches that we are investigating.

One of them is still in process and is something that we are still working on. But to your point is there are going to be outliers in the real world and the field. And you want the machine learning algorithm to be able to flag those and say that is something that someone needs to go take a look at.

MEMBER MARCH-LEUBA: But the algorithms should be really, really, really good at is to find something normal.

MR. RAMUHALLI: Yes.

MEMBER MARCH-LEUBA: That's normal. That is not normal. I don't know if it's the format, but it's not normal.

MR. RAMUHALLI: Correct.

MEMBER MARCH-LEUBA: And if you remove the things that are not normal in the process of training, you're --

MEMBER BALLINGER: You're biasing the training data.

MEMBER MARCH-LEUBA: Yeah.

MEMBER BALLINGER: You're biasing the training.

MEMBER MARCH-LEUBA: It's like imagine we're looking at a weather map radar. If I see a line of yellow with red coming towards me, I know something is bad. If I see everything clear and nothing, it's good. Very simple. You don't have to complicate it with too many letters. It's the same thing here, yeah.

MEMBER BALLINGER: Is there any learning

that can be had by going across the hall and talking to the Eddy Current people?

MS. NOVE: We do talk to the Eddy Current people.

MEMBER BALLINGER: Okay. I was being a little tongue in cheek here. But --

MS. NOVE: They're mostly using rule-based.

MEMBER BALLINGER: Yeah, because there's a lot of data there.

MS. NOVE: There's a lot of data.

MEMBER BALLINGER: A lot of data there.

MR. CUMBLIDGE: It was previously worked on.

MR. RAMUHALLI: So that's something that we have been having those discussions with. Also have been using rule-based approaches for the current algorithms. As Carol pointed out, that is the current state of the art.

There seems to be rule-based approaches. And that works really well for Eddy Current. The kinds of rules that are used by analysts to flag indications, those are easy to capture. Those are very, very defined. And the data tends to have clear evidence of those signal behaviors that reflect those rules. And so in that case, rule-based works really well.

VICE CHAIR HALNON: I had a question. Of course you've already jumped ahead. And help me understand what the current practice is.

But you have these scans on there. This is already part of the practice. There's scanning equipment. An inspector will look at scans like this. That's the kind of normal practice, right? And then the software is coming in on top of this, right?

MS. NOVE: So the normal practice, so the kind of scans you see here, these B scans, that would be fairly normal type of scan that you would get with a reactor vessel upper head exam. That's the type of date they collect. For pipe weld exams they're not using, this is time of flight data. They're using -- they used array data and piping exams. But both of them are image data. And so, yes, so we're used to collecting data in the form of images, and --

VICE CHAIR HALNON: There are multiple kinds of data streams that are also being fed to the software. Is that what you're implying there? I mean, this scan, that scan.

MS. NOVE: That's one of the things that we'll talk about in a few minutes is that the data that you put into these systems have to be representative of the data that you trained on in order for it to be -- for the system to be applicable to your test data or your field data. So it had to be very consistent. If you change your procedure, that may change your qualified angles or your frequencies.

You can't just automatically use your machine learning tool without retraining and requalifying it. So it's procedure specific. It's very tied together.

So going back to the program that we have ongoing at the national lab. This slide shows a detailed example of results from one of the initial experiments where data from Specimen A was used to train and data from Specimens B, C, and D were used to test the performance. And the types of flaws that are in each of those mockups is listed in this upper right-hand corner.

So for the model train on Specimen A, it showed a high classification accuracy with the test data on Specimens B and D. And this outcome demonstrates that the machine learning models efficacy is notably enhanced when the test data closely parallels the characteristics of the training data set. It's kind of along the lines of what you were just asking. We got to start somewhere.

So when the model is tested on Specimen C which contained relatively small thermal fatigue cracks that were close to the weld center line, the true positive rate fell significantly. This poor performance is not unexpected as lower detection rates for flaws on the far side of the welds is expected in the coarse grain materials used in nuclear power plants. So the simple test and others like it highlighted the need to use data for training that is similar to that expected to be encountered in the test. And more specifically, data that covers the range of flaw types, sizes and locations and essential variables such as frequency and wave noted in the inspection procedures. The data needs can be significant if a single model is used for all inspection scenarios and procedures and were evaluating whether multiple smaller data sets and associated models may end up being the better option.

CHAIR KIRCHNER: Why did you get -- did you retrain after you did C for a round pipe and then apply it to D and get a better answer?

MS. NOVE: I'm going to go on to the next slide.

CHAIR KIRCHNER: All right. Keep going.

MS. NOVE: Okay. So we're going to talk about transfer learning. So given the potential lack of large data sets, approaches that rely on fine tuning a previously trained model may also work. In this slide, we show what happened when a diverse data set including both thermal fatigue cracks and saw cuts was used to retrain the model.

The results show that the retrained model had superior performance over both model trains slowly on thermal fatigue cracks or saw cuts. And so for fuel applications, fine tuning the qualified machine learning model using site specific data could be done. And frankly, it's being done to ensure that the model's applicability to any unique characteristics that may be present in a particular site's inspection data.

Of course, such retraining brings up the question of how one would requalify the model. Again, this assessment of transfer learning and requalification is ongoing. And we expect to gain more insights from the work over the next several months.

CHAIR KIRCHNER: Make a note that you might talk to Joshua Kaiser and your -- well, he's over at NRR. But this is done often in the CHF area basically where you are training your model. You get a better, more accurate heat flux correlation as you kind of work with new data sets. Train the model and improve the -- tighten the uncertainties.

MEMBER BIER: I guess another kind of issue is shift over time, even within a single plant. I mean, obviously larger flaws are going to be easier to detect than small flaws. So if the only difference is just flaw growth, then I believe the performance will not deteriorate. But if you have some new type of degradation mechanism that doesn't show up till after 20 years, something or other, then you could have a model that you thought was well calibrated that starts performing badly eventually.

MEMBER BALLINGER: You know, the elephant in the room here is the stress corrosion crack. These are not stress corrosion cracks. They're saw cuts. They're fatigue cracks.

Is there a thought to looking at how these things do with real stress corrosion cracks?

MS. NOVE: Well, we are actually in the process of building more mockups and expanding what we're looking at, collecting more data. This is very data intensive, and we made a lot of mockups, a lot of specimens. And we won't end up having stress corrosion cracks but thermal fatigue cracks, more of those.

MEMBER BALLINGER: But stress corrosion cracks usually have an oxide layer between the two halves. And that complicates things. So I would expect -- I think the behavior is quite a bit degraded when you have an actual stress corrosion crack.

MR. CUMBLIDGE: Fortunately and unfortunately depending on what you look at it, they actually have some amount of experimental data on stress corrosion cracks on the upper head tubes.

MEMBER BALLINGER: Is EPRI in this?

MR. CUMBLIDGE: They're one of the people -- they're one of the primary people working on this with the upper head is with some of the data they've gathered as they're training.

MEMBER BALLINGER: Is EPRI involved in your program?

MS. NOVE: They're not involved in our program. But we do interact with them regularly. We actually -- the whole team of us went out to a nuclear power plant when they were running one of the demos of this.

And we spent a few days with them seeing how they were deploying this technology in the field.

So we interface with EPRI very regularly. But we are not working with them on this program. We exchange a lot of mockups all the time. So that is one area that we --

(Simultaneous speaking.)

MEMBER BALLINGER: They produce the mockups and they do the training, at least on the nuclear side.

MS. NOVE: As of right now, we have four other dissimilar metal weld mockups in house that we're collecting data on. So we do work with them. But we are doing an independent evaluation of this technology.

MR. RAMUHALLI: Can I just add one more comment? This is Pradeep again. So to your question about stress corrosion cracking and how these methods might do that is Carol's point about something that's down the line.

But having looked at similar algorithms, so machine learning as a technology itself is not new. I mean, there's been various iterations of this over the years. And certainly, there is plenty evidence in the literature and in past work on the user previous situations of machine learning for patent recognition, particularly for ultrasonic NDE data from stress corrosion cracks.

And appears that many of the things that we're finding here also were found on stress corrosion cracks. You need adequate data sets. You need representative data, data that captures the procedures and all of the variability of the procedures, the representative noise characteristics in the data, et cetera. So in that sense, I think there is hope that if it works for these types of flaws, with some modifications perhaps. But it should also be useful for SCC as well.

MS. NOVE: Okay. So moving on, so in terms of the findings to date, our analysis showed that machine learning is capable of true positive, low false positive, and false negative outcomes. The convolutional neural network or similar machine learning algorithm may be able to learn key features of flaws and non-flaws using data from simple flaws to saw cuts and generalized weld to other flaw types. Generalization performance may vary depending on the flaw size and location. Shorter and shallower flaws tend to be more difficult to detect using machine learning.

And the difficulty appears to increase the flaws in the vicinity of the weld. Such an effect may also be applicable to the manual analysis of the data and may indicate an inherent challenge in the data itself. The comparison of the machine learning performance with manual analysis is necessary and is part of the ongoing research.

A retraining procedure using a trained network as the starting point when adding more data may be helpful in improving classification performance. This type of transfer learning may be useful in, for instance, improving the performance of a qualified CNN with site specific data. The implications of these findings are that machine learning is likely here to stay for some applications of NDE and nuclear.

However, there needs to be well defined measures for quantifying the performance of machine learning in each application. And the need to ensure that algorithm qualifications account for the variability of flaw sizes and locations likely to be encountered in the field as well as the qualified inspection procedures deployed in the field. In terms of data, data sets used for training shall be diverse and representative of the types of data expected to be encountered during use.

For example, if machine learning is

intended to be applied to dissimilar metal weld specimens, then the training data should contain data from similar specimens and flaw types. In addition, the training data should reflect the expected diversity and flaw types and locations to ensure the machine learning training is adequate. It should be noted that variations in inspection angles and probe frequencies are likely if the inspection procedures have changed.

As a result, machine learning models are unlikely to be sufficiently accurate when applied to data collected using a different procedure to your question earlier. After such an inspection to welds may impact the performance of the algorithm, those such factors may be accounted for by the use of an expanded training data set, especially as machine learning algorithms, will be expected to accommodate nominal weld geometrical variances and associated noise in the B scan images. Our findings point to the need for careful selection of qualification data sets and perhaps defining metrics to ensure the qualified machine learning models are applied only where appropriate.

Specifically, these point to the potential need to determine if the machine learning models are being applied to the right inspection procedures and whether the inspection procedures are using the right set of essential variables. And typically machine learning solutions for data analysis will compute an output given an input. Checks on the validity of the input data will need to be performed separately and that machine learning outputs do not normally include uncertainty bounds.

Estimates of uncertainty may be helpful in determining confidence in the machine learning predictions. In terms of metrics, our work shows the desired performance thresholds are likely going to be dependent on the particular use case with regards to the commonly used metrics of true positive rate, false positive, and false negatives. A high true positive rate may not by itself be sufficient to show adequate performance and other measures may need to be tracked as well.

These other useful measures may be helpful in better understanding the capabilities and limitations of machine learning. It's important to also point out that we are simply identifying which metrics may be useful and not what the desired targets are for each metric. These targets will vary from application to application and for specific qualification requirements.

While some of those targets are currently used for qualifying personnel and procedures, whether these targets should also apply to machine learning applications and NDE is an open question and will likely depend on the type of application, so AI assisted or fully automated. So for instance, we know that a low false positive and false negative rates along with high TPR is desirable in general. But what happens if you use machine learning for screening data and only provide the analyst with a portion of the data to review, something you asked.

In this case, we'll have to determine whether or not we need to have a zero false negative, a low false positive, and 100 percent TPR required in order to use the system. It's noted here other measures such as receiver operating curves and machine learning training curves are also useful to measure performance of the ML algorithms.

MEMBER MARCH-LEUBA: I'll let you finish to --

MS. NOVE: Okay.

MEMBER MARCH-LEUBA: -- make a statement.

MS. NOVE: In terms of best practices, several best practices that exist in the broader ML community were also found to be applicable to NDE. For instance, consistency in how data is obtained and handled prior to applying the ML model is important. Consistency and pre-processing and labeling the training data for supervised ML methods, tuning and selecting parameters that control the learning method, and retraining a trained network with additional data have all been found to improve performance. So in terms of the status of our two-prong program, in the case of the rule-based data analysis, there's a technical letter report that's in the review cycle entitled Evaluation of Commercial Rule-Based Analysis -- Rule-Based Assisted Data Analysis.

And we're getting ready to move on to a confirmatory analysis of the commercial ML system being tested by industry and field trials. This next phase of work is going to focus on upper head examinations since that is the likely -- the primary case that industry is looking to use right now. We're in the process of designing and fabricating mockups.

And our assessment is going to be twofold. We're going to use the pre-trained algorithm that industry has developed. And we're going to have a qualified vendor who does these field inspections, collect data on our mockups.

And we will test the pre-trained algorithm with this data. And then we'll use the data, both the vendor collected data and PNNL collected data to train and test the commercial system as well as the Oak Ridge system. So that's where we're going in terms of looking at the commercial system.

MEMBER BALLINGER: Where's that technical letter report at?

MS. NOVE: The technical letter report, the NRR review was just complete last week. It's gone back to PNNL and they're sending it through their information release process. And it should be available I would say in the next week and a half, two weeks at most. So I will get that to you when that's done.

MEMBER MARCH-LEUBA: And we'll get a copy, right?

VICE CHAIR HALNON: I always saw machine learning as machine teaching itself. But in all this, it looks like machines making a conclusion based on an algorithm. And then somebody is looking at it saying, no, you missed it. We train you to focus it better. Is that more accurate than machine learning teaching itself?

MEMBER MARCH-LEUBA: No, no, no. Machine

learning involves a training data set in which somebody tells them what's good and what's bad. To train all these events and language models, they use thousands and thousands of --

(Simultaneous speaking.)

VICE CHAIR HALNON: Explain to me how it's better. I know she said what's good and what's bad. But do we just tighten the rules or do we --

MEMBER PETTI: Is it more data?

VICE CHAIR HALNON: Or is it more --

MS. NOVE: More data.

MEMBER MARCH-LEUBA: I'm going to make a statement. I think we're doing this all wrong. Just give me your attention. From the safety point of view, these algorithms need to be able to detect what's good.

From the cost and monetary point of view, you should be able to distinguish between what you call geometry, what you call fault. Going through the pipe, a normal pipe that doesn't have any welds, doesn't have any failures, it's easy to indicate what's good. That's a nice pipe.

And the difficulty of these algorithms is distinguishing between geometry of a flaw. What I mean to say here I read somewhere that to be an expert, you have to have 10,000 hours of work in this subject. That's about three years including overtime and 5,000 off on the subject to become an expert. We are going for two hours on this.

But if I was doing this, I would take the data. I will make two algorithms. The first one to the take was good. The area of pipe which is 90 percent of your data which the geometry and there are no flaws. Throw them out with extremely high confidence. You don't even need to review that. And then you're separated with a subset of things that have geometry and flaws and attempt to develop an algorithm and distinguish between geometry and flaw.

MS. NOVE: But see, we're only required to inspect the latter.

MEMBER MARCH-LEUBA: Right. But you've got to define it first. And you're trying to develop an algorithm that does all three at the same time. You have three possible answers, normal, geometry, flaw. I would make an algorithm that says normal or abnormal, geometry, or flaw. And then another subset that defines between geometry and flaw.

I bet it would be a lot simpler and more accurate. And simply you're able to include 90 percent of your data because it's normal. And I put my hand on fire. This is normal.

I tighten up all my weights so that it is normal and only put in 10 percent back to the next -- work on that if you want to improve on your cost. But the safety is covered by the 90 percent which you can put your hand on fire.

MR. RAMUHALLI: So this is Pradeep. And as I mentioned earlier, I think those two threads are basically what we are pulling at the moment. What is presented what has been -- we are further ahead on the classification aspect because that's where a lot of the initial work and the initial effort and the past which your point goes to, in terms of where the state of the art is.

But over the last few months, that second phase that you mentioned, this normal versus abnormal. What capabilities there are for the algorithms and how well are they able to distinguish that. Is it 90 percent? Is it 70 percent? That aspect is something that is an ongoing activity. And hopefully we'll be able to report more results on that.

MEMBER MARCH-LEUBA: To -- just as long as to the fusion guys, it was over 50 years ahead. And suddenly, there was a breakthrough and they were able to do it. I'm not against cheating. It starts with a simple program: normal, abnormal.

And the normal, I can put my hand on fire. It is normal. And then improve on the other one to improve your cost. But the safety concern, the regulatory concern, it is normal. For sure, it's normal. I don't have an 80 percent failure detection there.

MR. RAMUHALLI: And I think that is the question that we are trying to examine is when you do have this normal, abnormal, for example, or even in the case of flaw versus non-flaws, what is that confidence? Is it a 90 percent? And I'm not talking about the probability of detection here.

That is a separate metric that we will have to track. But in terms of what is that confidence in the result, is it 90 percent? Is it 70 percent, 50 percent? Is it a coin toss? How one gets to that point of something that we continue to --

MEMBER MARCH-LEUBA: Did you change -- I mean, you don't know the answer. You're probably asking yourself the wrong question.

MR. RAMUHALLI: Fair point.

(Simultaneous speaking.)

MEMBER MARCH-LEUBA: So since the question to cheat make your program easier. Solve the easy problem. And I only have 10 percent of the data to work with.

MR. RAMUHALLI: Point taken.

MEMBER MARCH-LEUBA: Maybe all these young guys create some specific, PNNL, they're trying to get an LLM type -- if I put more layers, and more -- just to be able to do everything at once. And is there a model that works? Yes, there is. We have to put three unknowns to do it. But --

MR. RAMUHALLI: Responding to the data, data needs for that are going to also blow up as you do that. And we've looked at not just the CNNs. We looked at simpler techniques as well. So your point is well taken.

MEMBER MARCH-LEUBA: And I would concentrate on a high quality, extremely high accuracy. I mean, we're talking 9s accuracy. This is good. This pipe is good.

And then try to identify different types of geometry. This is a V shape weld. It is an L shape weld. And then, next year --

MR. RAMUHALLI: Yeah. And yes, and hopefully we'll have those. We have some initial results that so far seem to be looking good. But again, it's a question of how many 9s of confidence can you get to? That's the question we are looking into.

MEMBER MARCH-LEUBA: But 80 percent deficient quality is not good. I mean, hear this out: autonomous vehicles killed 20 people last month. Yeah, you killed less people than the real drivers, but still, you killed 20 people.

MEMBER PETTI: Carol, just a question on the project. These are three-year --

MS. NOVE: We have a five-year contract.

MEMBER PETTI: Five-year. And where are you?

MS. NOVE: This work started in 2022. So we're about a year and a half into it.

MEMBER BIER: And if the goals is to be able to support industry initiatives or approve industry initiatives or whatever, obviously you're not very far along yet. I'm not asking for a prediction. But where are you in the spectrum from, yeah, give us more time. It'll all be okay to kind of Jose's reaction of, no, this is not looking close enough to be useable. Do you have a sense of that?

MR. CUMBLIDGE: It'll really depend on what industry tries to do with it and how they try to qualify it. So that's the what approach to take.

We're assessing what industry is trying to do. We can't tell them what to do and we can't develop it for them.

They say we're really kind of -- this is confirmatory research on what they are doing. We can't get ahead of them. If we get ahead of them, then we're --

(Simultaneous speaking.)

MEMBER BALLINGER: You're trying to be sure that you're smart enough so that when they do come back --

MR. CUMBLIDGE: We have an informed --

MEMBER BALLINGER: -- you know what you're talking about.

MR. CUMBLIDGE: We want to be well informed and well understand the problem and understand the benefits and pitfalls when it comes to us to make a rulemaking or accepting something in code. We can't suggest, oh, you guys, you should be doing it like this. That's not what we can do.

CHAIR KIRCHNER: Carol and Steve, may I interject? I know you have about a half dozen-plus slides left and we have less than ten minutes budgeted. Could you go through and just highlight any messages you want us to go away with in terms of the material you wanted to present?

MS. NOVE: Yes, I just want to -- on this slide here, we have a lot of things we still need to do with the bottom line being that all of the work we're doing out of the Office of Research is to help NRC determine what's necessary for establishing confidence in the machine learning based data analysis. You should have received a report that was issued just a few weeks ago. This is the Oak Ridge analysis of machine learning to date. And there's also two referee journal articles that are out on this work. And then I'm going to turn it back over to Stephen.

MR. CUMBLIDGE: Okay. I'll go through this slide very quickly. Qualification pathways for automated data analysis. Now if you wanted to go the fully automated, there really isn't -- the rules would have to be written for that.

There are no rules for doing it without a person and in Section 11. So that would require a large lift by industry. We can actually use the current rules for automated data analysis if a person is involved.

And the path forward is -- and this EPRI is you bias towards detections. That is you go through all the qualification data and you make no misses. And then you have a higher tolerance for false calls with the people. And then the qualified analyst would weed out the false calls. And then the acceptance criteria for the algorithm would be biased towards detection. And it's --

MEMBER BALLINGER: I'd change the words and say no bad misses.

MR. CUMBLIDGE: Well, they would have known -- they want to miss nothing in the qualification --

MEMBER BALLINGER: Right.

MR. CUMBLIDGE: -- data set. And also they only include flaws that are expected -- that you would expect someone to find. They didn't put things that are too small to be found.

MS. NOVE: The flaws and the qualification data set start in some cases at 10 percent through all. They don't have tiny flaws. And the other thing is the inspection procedures are expected to provide 100 percent detection. But personnel again are not necessarily expected to do 100 percent detection. They have some allowances for false calls.

MEMBER BALLINGER: The field, you want to avoid bad calls.

MS. NOVE: We do, but they do happen. And missed calls happen.

MR. CUMBLIDGE: People miss calls.

MEMBER BIER: I have another comment regarding that issue of missing things in the qualification set. Presumably, the qualification set is kept separate from the training data. Is that correct?

MR. CUMBLIDGE: Yes.

MEMBER BIER: Okay. Because otherwise, you can have situations like if I'm trying to train something to detect a fire in this room and the example has a trash can fire in that corner, it's going to detect that corner and ignore the rest of the room, so --

MR. CUMBLIDGE: You absolutely have to keep the training data separate from the qualification data. You cannot train on the qualification data. It's like taking an open book test with the answer key. It doesn't work.

So right now the way the rules are written in Section 11, you can use the automated data assisted analysis where you have a person and the ADA. Actually, the rules are there for it right now. And the problem would only cover encoded data. And there are a lot of things that while you can do it, they should probably add some things to Section 11 to facilitate its use and make it easier.

CHAIR KIRCHNER: Encoded data is code for what?

MR. CUMBLIDGE: Recorded data. That's where you have -- instead of, like, a person looking at a screen while they --

(Simultaneous speaking.)

CHAIR KIRCHNER: What QA requirements are put on that data set? Is that NQA-1?

MR. CUMBLIDGE: That would be just in the procedure. It would say check the data and the following things have to be met for the data to be acceptable. But that'd be in the qualified procedure of the data acceptance.

CHAIR KIRCHNER: Okay.

MS. NOVE: So in this picture, this person has -- there's a scanner arm. And that scanner would have an encoder. So it would encode the position of the probe. And that would track to the scan itself.

MEMBER MARCH-LEUBA: I mean, that picture, the darker band is a weld?

MR. CUMBLIDGE: Yes, a weld in the middle.

MEMBER MARCH-LEUBA: So he's counting the weld.

MR. CUMBLIDGE: Yes.

MEMBER PETTI: And I assume you guys sit on some of these Section 11 committees like you sit on the other code --

MR. CUMBLIDGE: Yes.

(Simultaneous speaking.)

MEMBER PETTI: -- committees?

MR. CUMBLIDGE: And I've thrown the gauntlet down and said, guys, technically you can do this. We should probably rewrite the rules a bit. And the big issue is that so you train your algorithm in the lab and then you run it. And it passes qualification. Then you take it in the field, and it makes a large number of false calls because it gets confused by -- so you want to retrain it.

So they get up there and they go, okay, this isn't working because it worked in the lab. But here it's not working. We want to retrain it on site specific data.

Well, if you do that, you have to go back to EPRI or performance demonstration initiative and have a person run through the test again with the current rules. This is not fast and it's not convenient. And they might be busy and getting them on the schedule might be challenging.

So they really should rewrite Appendix 8 to allow for more rapid requalification of the data. So you do this site specific retraining and then have the algorithm look at the qualification data again. If it finds everything and doesn't make too many false calls, you're good.

Basically, the analyst gets what it feeds you at the end. They add things to Section 11 to allow for site specific retraining which by the way would also help prevent a common mode failure. Everything is a site specific retraining and every algorithm is somewhat unique.

Then it wouldn't be a common cause failure across the fleet if one of them misses something. But we think that field friendly retraining has to be part of the rule set. But it has to be one where, again, you find all the flaws in the qualification training data.

Qualification testing data, not -- you can't miss things. So next slide. So as we look at what's going on, right now all the inspectors are qualified. They're very highly skills. They passed Appendix 8 quality performance demonstration testing. They've been doing this for years. They know what they're doing.

Also, we have really good AI experts who are engaged in making everything work. And this combination resulted in very highly optimized procedures. So if you have really good people with really highly optimized AI algorithm procedures, you probably get the best outcome.

You have the experience of the inspector. You have the vigilance of the algorithm. There's something new in the field. The person might be able to catch that.

Now our concerns are -- and we've expressed this to EPRI in the industry in general is that if you have these tools, will the new inspectors not have the same skill level. Or will the current inspectors lose their skills over time? Then you wind up with unskilled inspectors.

Also, these AI experts who are really highly engaged, they might look out for other dragons to slay and other mountains to climb. And then you -- they've left this really complicated process behind that no one really understands how to maintain it because they've moved on. And then you don't have the AI experts.

They have people trying to make procedures using this and the inspectors aren't very good and the AI experts are gone. So you're going to have bad procedures. That doesn't get you at a good place.

So now you're in a place where maybe the current inspections will be done okay because the old AI would be good enough. But the inspectors wouldn't be that good. New degradation, if something new happens, you're probably out of luck. And developing new procedures for new challenges would also -- the expertise wouldn't be there. So you really have to work with industry to make sure we stay on the blue and don't get to the red.

MEMBER BIER: Another example of that kind of thrift is what happened to NASA where it used to be a really cool place to work and all the engineers wanted to work there. And after a while, the cool thing moved on to something else. And they had a lot of degradation of talent and recruiting and whatever. So if this becomes not the cool thing in ten years because it's been done and is boring now --

MR. CUMBLIDGE: Right.

MEMBER BIER: -- it would be problematic.

MR. CUMBLIDGE: We can see things getting started off really well and looking great and nosediving hard in the not terribly distant future if we're not careful.

VICE CHAIR HALNON: So one question I had, and we don't need to go into any detail since we're out of time. But one thing that's not on here is ethics. We rely on the ethics of Level 3s to not pass through things and stuff. Is there any concern with that with the AI aspect?

MR. CUMBLIDGE: Well, the performance demonstration, if someone were to lie about the performance demonstration, it would have to go through the qualification data set and find all the flaws.

VICE CHAIR HALNON: But then an AI programmer could program in --

MR. CUMBLIDGE: Okay. This is not covered malfeasance by people in the process. That's beyond --

(Simultaneous speaking.)

MEMBER PETTI: I'm just saying looking at the equivalent between AI and an inspector, we do rely on the ethics of the inspector to not pass things through that would not otherwise --

MR. CUMBLIDGE: We do rely on -- we rely on the performance demonstration initiative to not give people the answer key before -- the malfeasance and whatnot is beyond the scope of what we're doing here. So again, the last side, we're working on for the future, we've been telling the industry if you want this to be more than just a curiosity or a fun thing, maybe it'll hit the upper head for a bit. They had to build an infrastructure around the use of this to allow for the future of it and to allow it to grow.

They had to essentially make qualifications for AI people. They need something to keep the expertise and really quantify the expertise. Also, they had to make very solid rules for re-qualifying an algorithm after it's modified. So again, the example, find all the flaws in the qualification data without too many false calls.

And they make rules of that and put them in code. And also for the people, one suggestion so the people don't lose their skills is that you can only use one of these automated data assisted algorithms after you pass the performance administration test on your own without it. Like, you have to be an actually skilled person before you can use one of these.

You can't just give the janitor an AI and have them pass which that terrifies the vendors. You wind up, like, wildly unskilled people or the people who wind up being cheap scanners. The vendors are very engaged at preventing the last bullet. They were working within industry to, again, stay on the blue, the top blue level and I'll get to the red level at the bottom.

CHAIR KIRCHNER: Thank you very much, Carol and Steve. Let me turn to Ron. Any final comments, Ron or Dave?

MEMBER BALLINGER: Follow the money.

(Laughter.)

CHAIR KIRCHNER: Okay. Dave?

MEMBER PETTI: I just want to thank everybody for coming. Very informative. Nice change of pace, our day jobs.

CHAIR KIRCHNER: Right. Thank you very much, yes.

MEMBER MARCH-LEUBA: Even Ron said at one meeting a couple years ago, the secret to success is NDE.

(Simultaneous speaking.)

CHAIR KIRCHNER: At this point, we're going to take a break. Before we take a --

(Simultaneous speaking.)

MEMBER BALLINGER: -- whatever he wants and keep him happy.

CHAIR KIRCHNER: Gentleman, excuse me. So at this point, we're going to take a break. I think we're done with the court reporter for the rest of the afternoon. And I think we need you back tomorrow at 10:30.

(Whereupon, the above-entitled matter went off the record at 3:04 p.m.)

SRM-SECY-22-0076 Implementation: Branch Technical Position 7-19, Revision 9

Advisory Committee on Reactor Safeguards Briefing
March 6, 2024

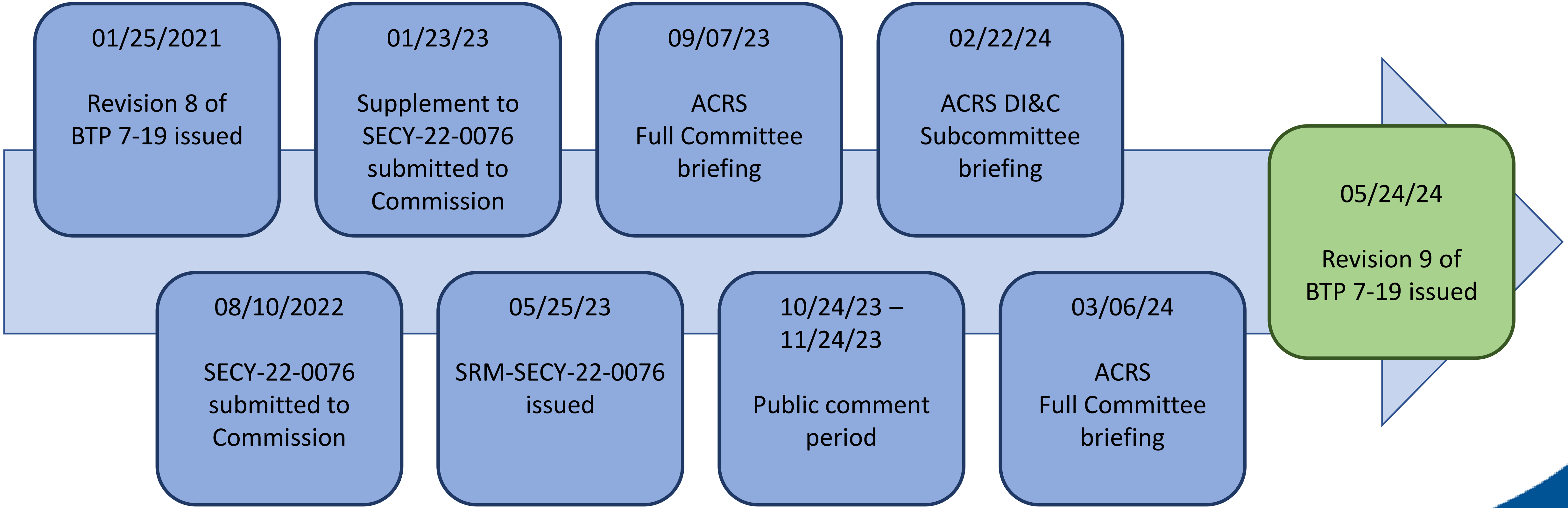
United States Nuclear Regulatory Commission
Protecting People and the Environment

Opening Remarks

Presentation Outline

- Background
  - Timeline
  - SRM-SECY-22-0076 Direction and Staff Response
- Changes from Revision 8 to Revision 9
- Changes since the September 7, 2023, ACRS Briefing
- Key Messages and Next Steps
- Closing Remarks

Recent Activities



SRM-SECY-22-0076

- The Commission approved the staff's recommendation to expand the existing policy for digital I&C CCFs to allow the use of risk-informed approaches to demonstrate the appropriate level of defense-in-depth, subject to the edits provided
- The Commission directed the staff to clarify in the implementing guidance that the new policy is independent of the licensing pathway selected by the reactor licensees and applicants
- The Commission directed the staff to complete the final implementing guidance within a year from the date of the SRM (May 24, 2024)

Staff Response to Meet the SRM for LWRs

- Drafted Rev. 9 to SRP BTP 7-19
  - Allows the staff to review risk-informed applications
  - May result in use of design techniques other than diversity
  - Focused the revisions on implementing the expanded policy
- Staff briefed the ACRS Full Committee on September 7, 2023
- Staff received and dispositioned public comments
- Staff briefed the ACRS DI&C Subcommittee on February 22, 2024

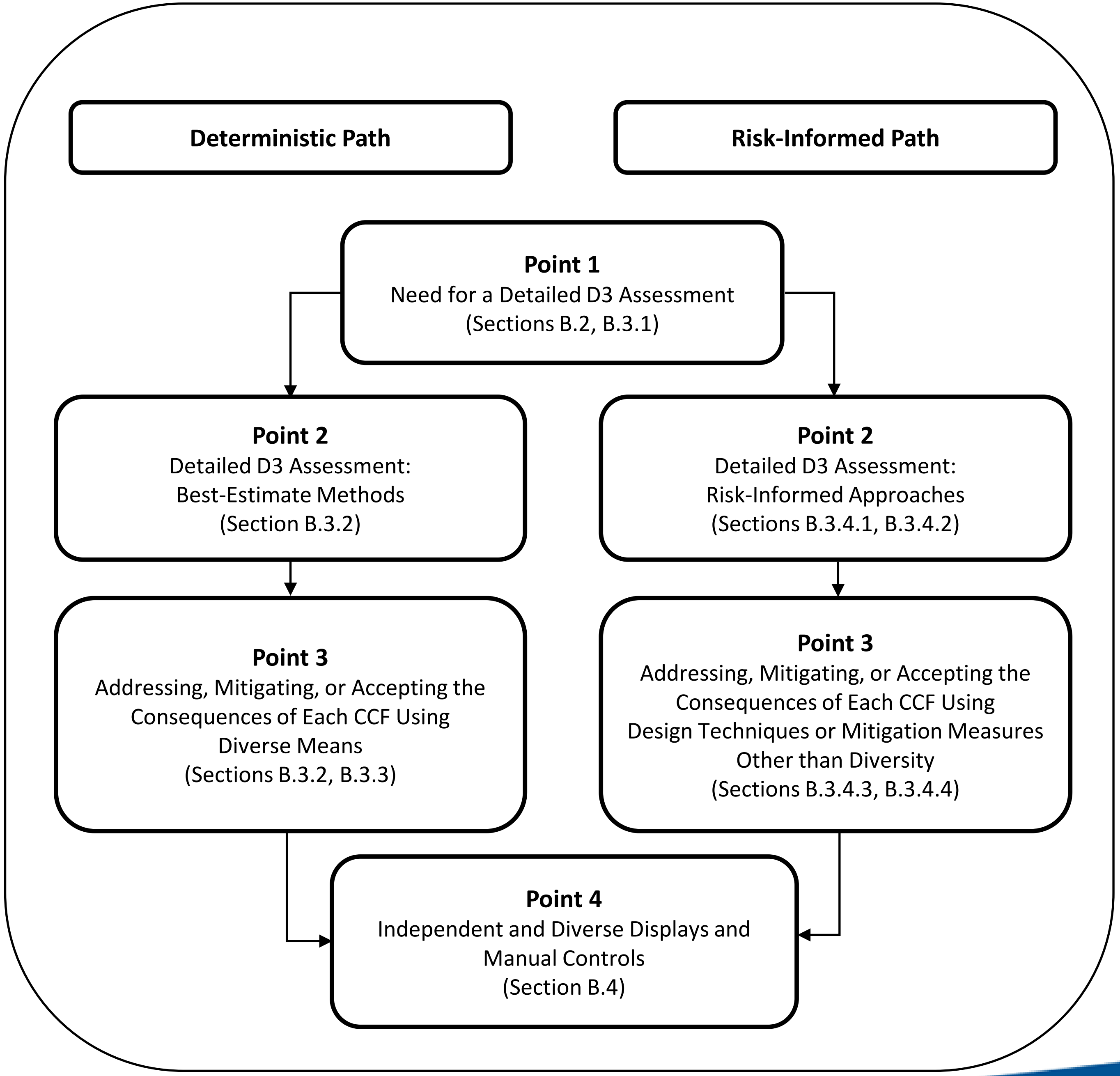
Staff Response to Meet the SRM for Non-LWRs

- SECY-23-0092 provides the following approach for addressing the expanded CCF policy for non-LWRs:
  - The staff is using the guidance in the Design Review Guide (DRG) and RG 1.233
  - RG 1.233 is risk-informed and includes guidance on the adequacy of defense-in-depth
  - The DRG is aligned with RG 1.233; together with the SRM, they provide reasonable guidance for addressing DI&C CCFs
  - The staff is using pre-application engagements to discuss use of the expanded policy with non-LWR applicants to address any questions or concerns
  - The staff will communicate the Commission's policy to stakeholders during advanced reactor I&C public workshops
    - Next workshop is scheduled on March 14, 2024

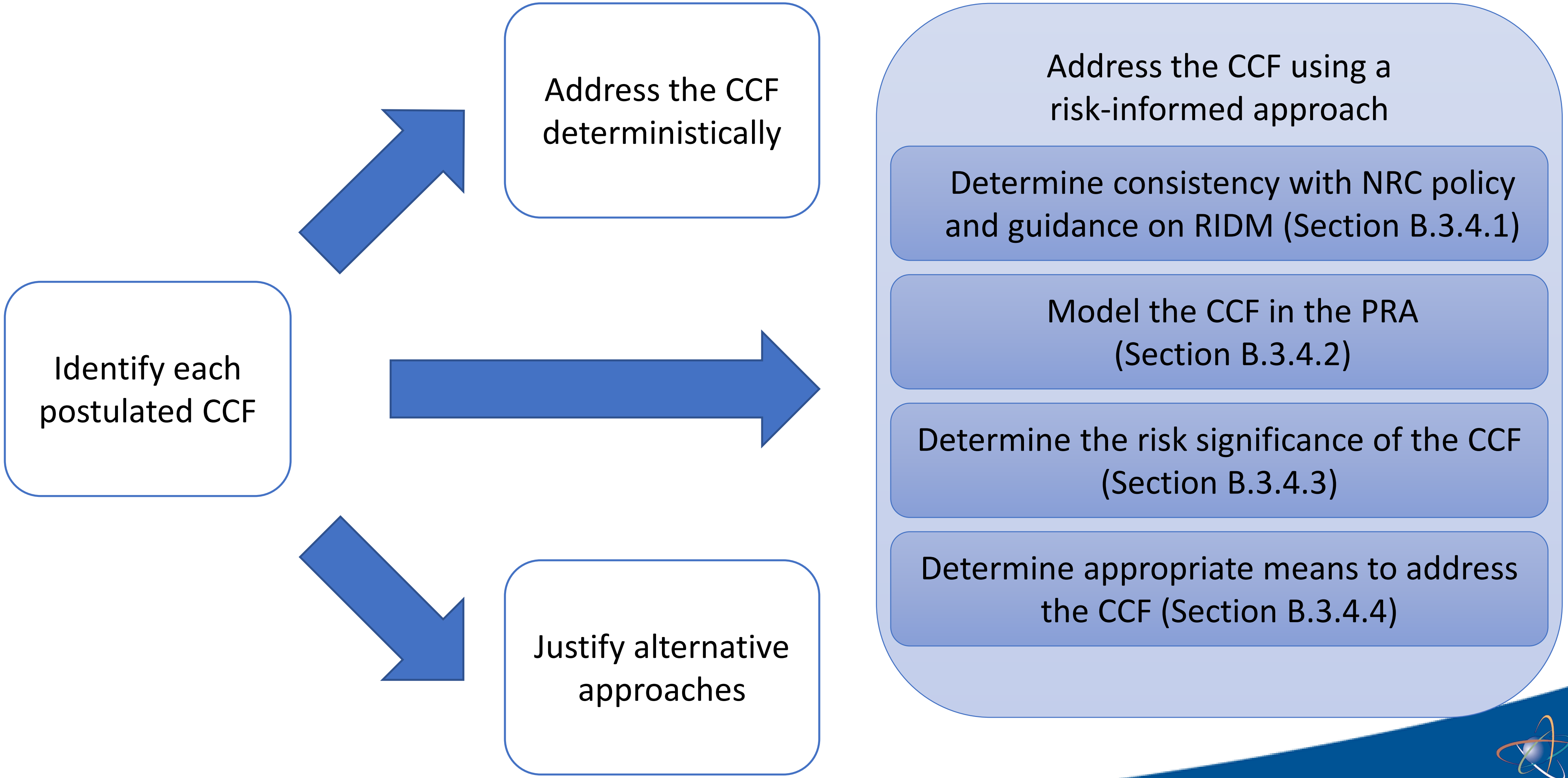
Substantive Changes to BTP 7-19 (Rev. 8 – Rev. 9)

- Revised Section B.1.1 to reflect the updated four points in SRM-SECY-22-0076
- Revised Section B.1.2 for clarification of critical safety functions
- Revised Section B.3.1.3 for evaluation of alternative approaches
- Added Section B.3.4 for evaluation of risk-informed D3 assessment
- Revised Section B.4 for evaluation of different approaches for meeting Point 4
- Added five flowcharts to facilitate the review
- Added language from RG 1.152 to address a prior commitment to ACRS regarding communication independence and control of access

Overview of BTP 7-19, Revision 9



Risk-Informed D3 Assessment Process (Section B.3.4)



Risk-Informed D3 Assessment

Determine Consistency with NRC Policy and Guidance on RIDM

- Review applications that use risk-informed approaches for consistency with established NRC policy and guidance on RIDM

Model the CCF in the PRA

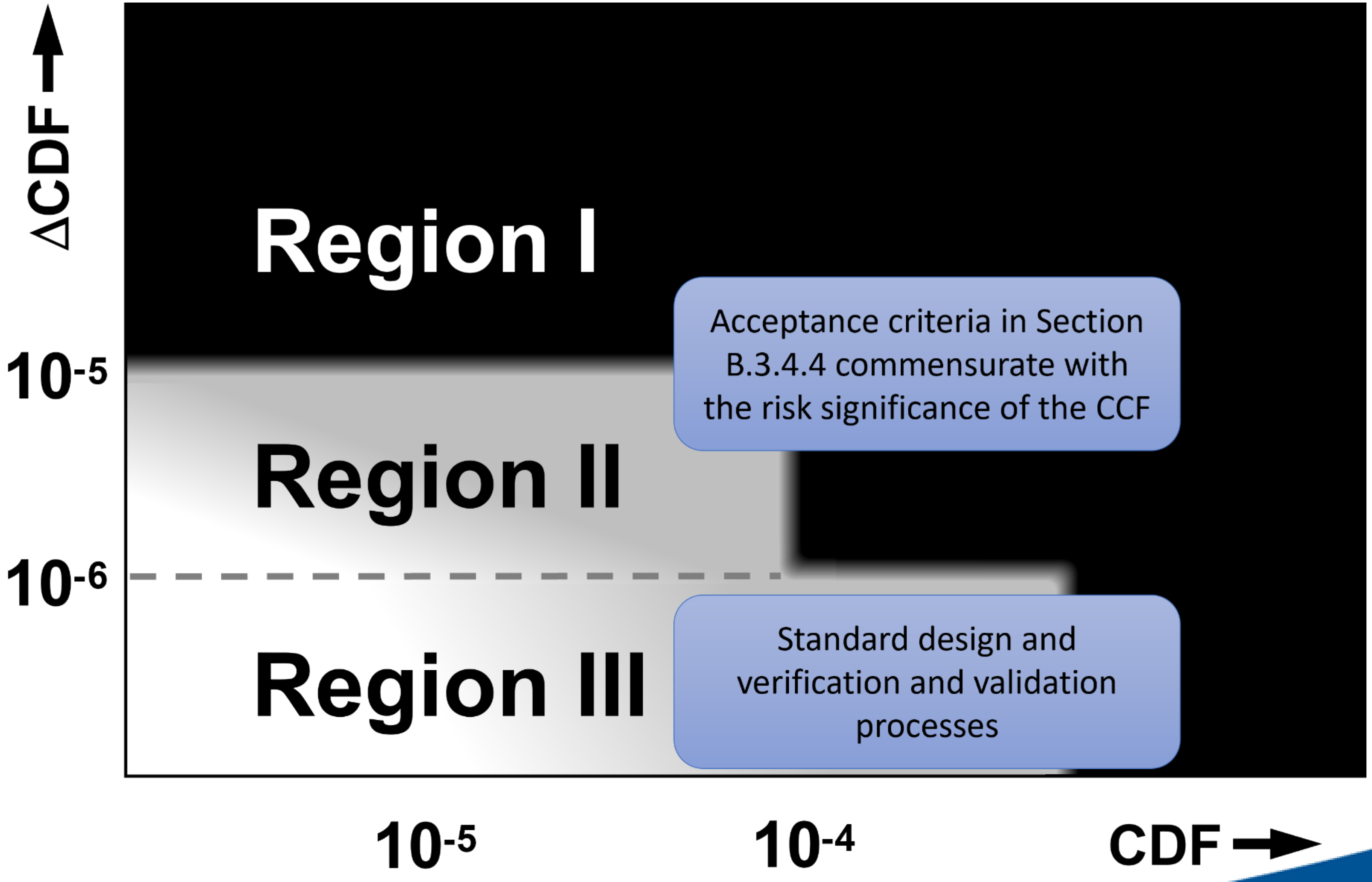
- Determine if the base PRA meets PRA acceptability guidance identified in the application
- Evaluate how the CCF is modeled in the PRA and the justification that the modeling adequately captures the impact of the CCF on the plant

Risk-Informed D3 Assessment

Determine the Risk Significance of the CCF

- The risk significance of a CCF can be determined using a bounding sensitivity analysis or a "conservative" sensitivity analysis
- A bounding sensitivity analysis assumes the CCF occurs
- A "conservative" sensitivity analysis assumes a probability less than 1
  - Provides a technical basis for a conservative probability of the CCF
  - Addresses the impact of this assumption on PRA uncertainty and the key principles of RIDM
- A CCF is not risk significant if the following criteria are met:
  - The increase in CDF is less than 1×10^{-6} per year
  - The increase in LERF is less than 1×10^{-7} per year

Risk-Informed D3 Assessment



Approaches for Meeting Point 4 (Section B.4)

- Per SRM-SECY-93-087 and SRM-SECY-22-0076, the independent and diverse displays and manual controls are not required to be safety grade or hardwired
- Section B.4 provides six acceptance criteria for independent and diverse main control room displays and controls for manual actuation of critical safety functions
- The acceptance criteria calls for displays and controls that are independent and diverse (i.e., unlikely to be subject to the same CCF) from the equipment performing the same functions within the proposed safety-related DI&C systems
- Applications that propose a different approach (i.e., one that does not meet all the acceptance criteria in B.4) provide appropriate justification

Changes to BTP Since September ACRS Briefing

- Clarifications made throughout the BTP to address:
  - Public comments
  - Discussions during the September 7, 2023, ACRS briefing
  - Comments from Member Brown and Member Roberts (attachment to transcript)

- No public comments received involved non-LWRs or the DRG

- No substantive changes made to analysis methodologies or acceptance criteria

Notable Changes to the BTP

- Revised the BTP to consistently use the term "digital I&C system"
- Clarified that the BTP is intended to provide review guidance to the NRC staff for ensuring an application meets the policy and applicable regulations
- Removed the pointers between Sections B.3.1.3 (alternative approaches) and B.3.4.4 (appropriate means to address the CCF)
- Provided "a well-designed watchdog timer" as an example of an alternative approach that may address certain vulnerabilities
- Added a sentence on manual control connections

Key Messages

- BTP 7-19 revised to incorporate SRM-SECY-22-0076
- Changes made after September 2023 ACRS Full Committee briefing in response to public comments and ACRS member feedback
  - Clarifications made throughout the BTP
  - No substantive changes made to analysis methodologies or acceptance criteria

Next Steps

- The staff is planning to issue the final BTP 7-19, Rev. 9 in May 2024

- The staff is planning to brief the DI&C Subcommittee in June 2024

Closing Remarks

Acronyms

ACRS	Advisory Committee on Reactor Safeguards
BTP	Branch Technical Position
CCF	Common Cause Failure
D3	Defense-in-Depth and Diversity
DI&C	Digital Instrumentation and Control
I&C	Instrumentation and Control
NEI	Nuclear Energy Institute
NRC	Nuclear Regulatory Commission
PRA	Probabilistic Risk Assessment
RG	Regulatory Guide
SECY	Commission Paper
SRM	Staff Requirements Memorandum
SRP	Standard Review Plan

References

- Transcript of September 7, 2023, ACRS Full Committee briefing and attachment with comments provided by Member Charles Brown and Member Thomas Roberts (ML23264A865)
- NEI Comments on Draft BTP 7-19, Revision 9, dated November 21, 2023 (ML23326A117)

Artificial Intelligence and Machine Learning in Nondestructive Examination and In-Service Inspection Activities

Carol A. Nove: RES/DE
Stephen Cumblidge: NRR/DNRL

ACRS Full Committee Briefing
March 6, 2024

Outline

- Introduction and Background
- Research Program
  - Evaluation of commercially available automated data analysis
  - Evaluation of machine learning for ultrasonic NDE
- Research Program Outcomes

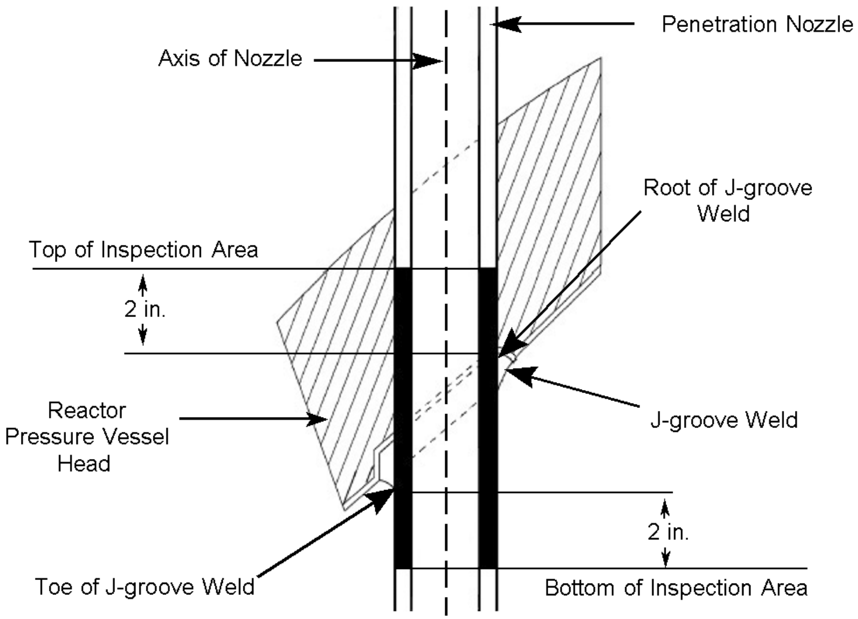
Acronyms

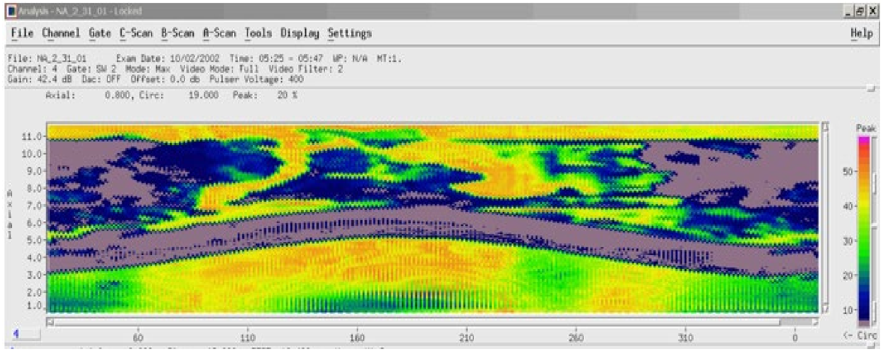
ADA – automated data analysis
ASME Code – American Society of Mechanical Engineers
Boiler and Pressure Vessel Code
CASS – cast austenitic stainless steel
CNN – convolutional neural network
CS- carbon steel
DMW – dissimilar metal weld
DNN – deep neural network
DR – detection rate
EPRI- Electric Power Research Institute
FPR – false positive rate
ISI – inservice inspection
ML – machine learning
NDE – nondestructive examination
ORNL – Oak Ridge National Laboratory
POD – probability of detection
PNNL – Pacific Northwest National Laboratory
ROC – Receiver Operating Curve
RVUH – reactor vessel upper head
TFC – thermal fatigue cracks
TPR – true positive rate
UT – ultrasonic testing (ultrasonics, ultrasonic examination, etc.)
UV – UltraVision
VP – VeriPhase
WSS – wrought stainless steel

Introduction and Background

Nondestructive Examination (NDE) in Nuclear Power Plants

- 10 CFR 50.55(a)(b) incorporates by reference the American Society of Mechanical Engineers Boiler and Pressure Vessel Code (ASME Code), Section III, *Rules for Construction of Nuclear Facility Components*, and Section XI, *Rules for Inservice Inspection of Nuclear Power Plant Components*
- NDE needed for timely detection of service-induced flaws
- Plant aging increases likelihood of service-induced flaws
- Accurate & Reliable NDE increasingly important due to industry trends to reduce:
  - Inspection time during outages
  - Radiation exposure
  - Number of examinations





U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Drivers for Automated Data Analysis (ADA)

- Section XI, Appendix VIII, Performance Demonstration for Ultrasonic Examination Systems, provides requirements for performance demonstration for ultrasonic examination procedures, equipment and personnel to detect and size flaws
- Industry projecting potential shortage in NDE technicians with proper skillsets to conduct NDE to meet future fleet needs (ML24026A087)
- Some UT inspections such as upper head exams yield large quantities of data that must be reviewed by multiple qualified inspectors during the outage period. (EPRI 3002023718)
  - High level of focus required for long periods of time
  - Human factors related to fatigue and momentary loss of focus can challenge reliability of results

ADA Is Coming

- Widely available, open-source ML tools have enabled the development and application of ML algorithms for many uses
- These tools are becoming more powerful and easier to use over time
- The nuclear industry is funding work to use these tools for automated data analysis algorithms to analyze NDE data

ONNX

GLUON

PyTorch

TensorFlow

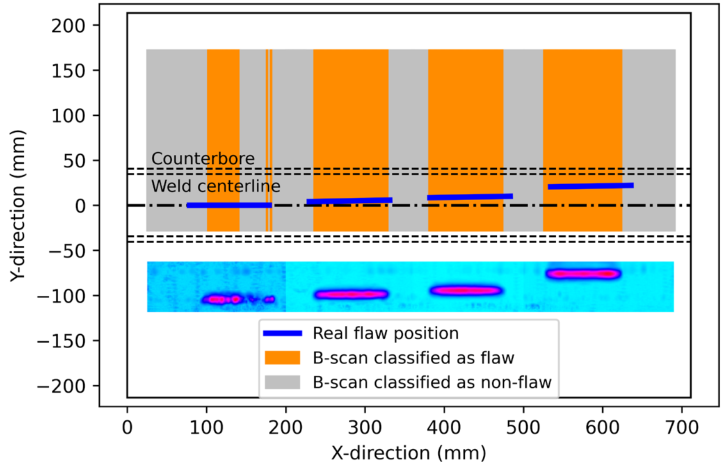
scikit learn

K

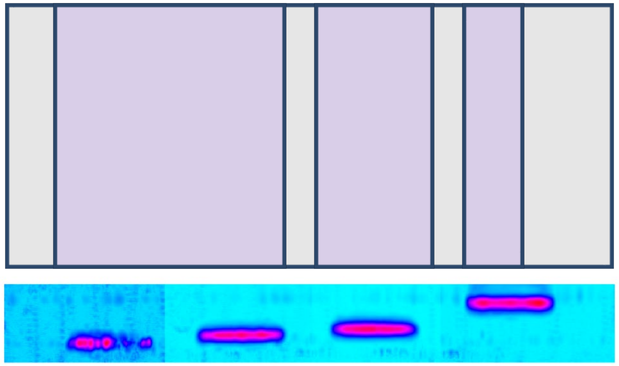
PyTensor

ADA/ML Use Cases for Ultrasonic NDE

- Near term
  - Analysis of encoded (recorded) data
  - Screening: Identify regions that are indication-free
  - Classification: Identify regions that contain flaws
  - Quality Control for NDE Examinations
- Longer term
  - Data compression
  - Generate NDE reports
  - Real Time data analysis of unencoded data
  - Synthetic data generation for training
  - …



Flaw Classification



Flaw Screening (Hypothetical Example)

Two Ways of using ADA

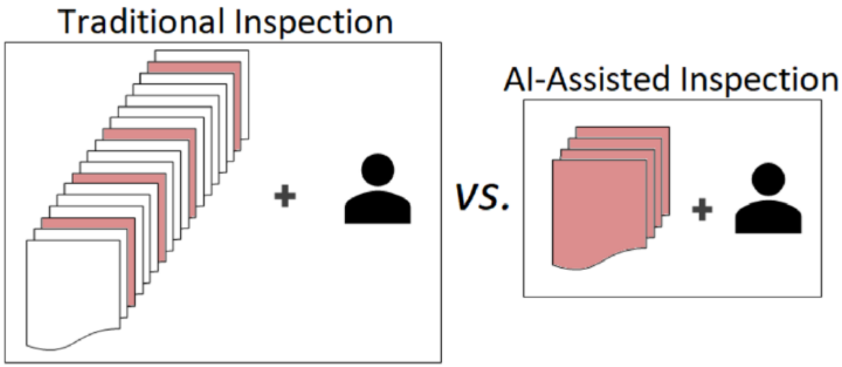
- ADA-Assisted Examination

  A fully-qualified inspector uses hints or highlighted areas to analyze the data, but the qualified individual makes the final calls

- Fully-Automated Examination

  The ADA algorithm makes the calls without human input

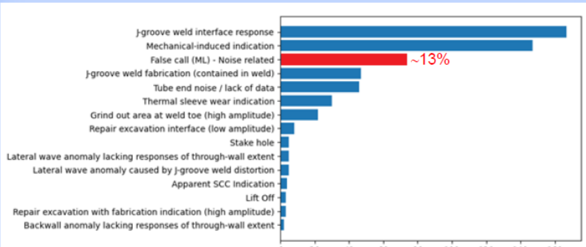
Automated Data Analysis -Assisted Procedures



Amount of Data Requiring Review

Pre-AI		Post-AI	
4.4	miles	463	feet
7.0	km	141	meters

All values are approximate



EPRI

- One suggested approach by EPRI is for an ADA algorithm to flag areas with flaws, and the algorithm must find all flaws in the qualification set
- The algorithm can produce more false calls than allowed in the given supplement
- It will be up to the inspector to determine which of the areas flagged by the algorithm contain flaws, and ultimately the inspector is responsible for the results

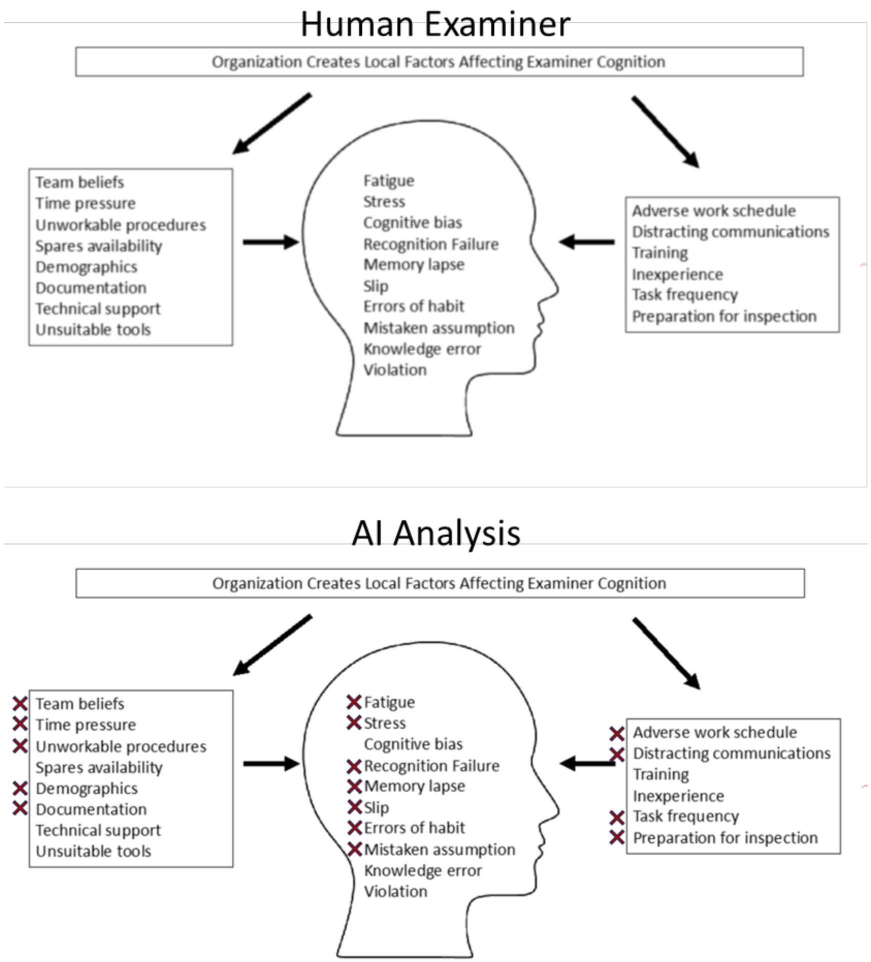
Graphics from EPRI 3002023718

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Automated Data Analysis – Possible Benefits

ADA has the potential to improve detection of flaws and improve the human factors of an examination.

- In-service flaws are rare in the nuclear industry. Computers can maintain vigilance in cases where humans struggle.
- Humans and computers make different types of mistakes, and a qualified analyst paired with an analysis run by ML gives the best of both worlds.
- Reduced dose to inspectors if ML used to support manual UT examinations.



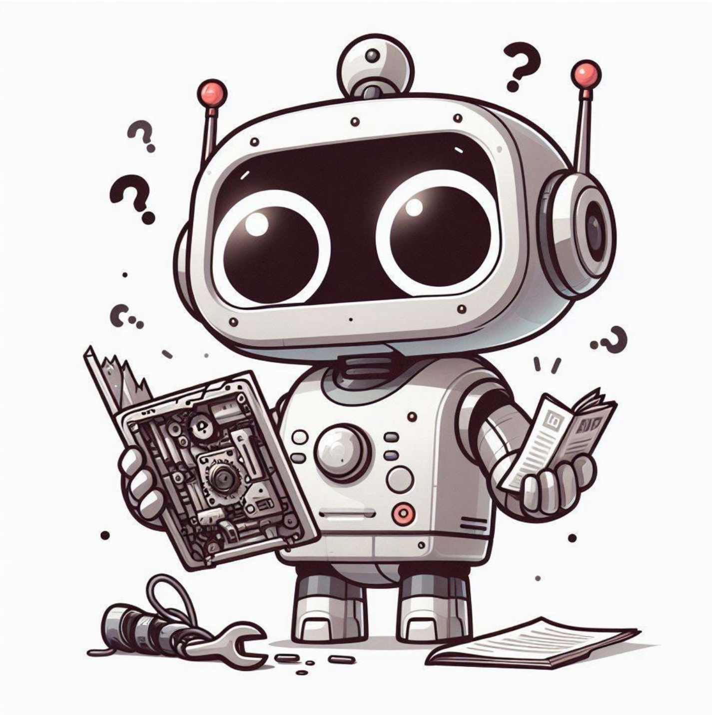
Graphic adapted from NUREG/CR-7295

Automated Data Analysis – Possible Hazards

- ADA has the potential to introduce common-cause failures of inspections across the fleet
- Licensees may not understand the capabilities and limitations of ADA, which could lead to improper use of ADA
- ADA assistance may allow people to pass Appendix VIII qualification testing without the skills to recognize unknown degradation in the field
- ML algorithms can be challenging to train and retrain, possibly making the ML algorithms unreliable
- ML algorithms require a new class of experts to support UT examinations

Automated Data Analysis – Expect the Unexpected

- As plants age and new reactors are designed, it is almost certain that new degradation mechanisms will emerge, and flaws will appear in unexpected places
- ADA methods can be very good at handling known problems but may not work on new forms of degradation



Research Program

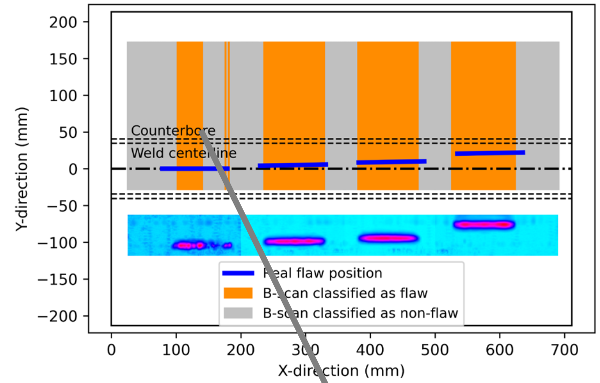
Research Program on Automated Data Analysis

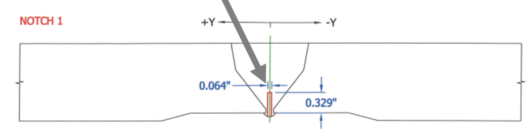
User Need Request for Evaluating the Reliability of Nondestructive Examinations, (NRR-2022-007), Task 4, Automated Data Analysis, requests that RES provide a technical basis describing current capabilities of machine learning and automated data analysis for nondestructive examination (NDE).

RES activity to address UNR request:

- Evaluating machine learning (ML) for Ultrasonic Examinations (UT) - Oak Ridge National Laboratory (ORNL)

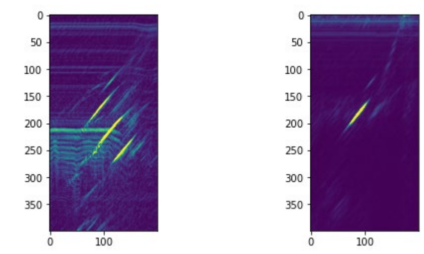
- Evaluate commercially available automated data analysis platforms including rule-based and ML-based systems - Pacific Northwest National Laboratory (PNNL)





Flaw inside the weldment

Longitudinal 45˚ Shear 45˚



Variation in Data (probe/mode)

U.S.NRC United States Nuclear Regulatory Commission *Protecting People and the Environment*

Automated Data Analysis – Types of Algorithms

Rule-based

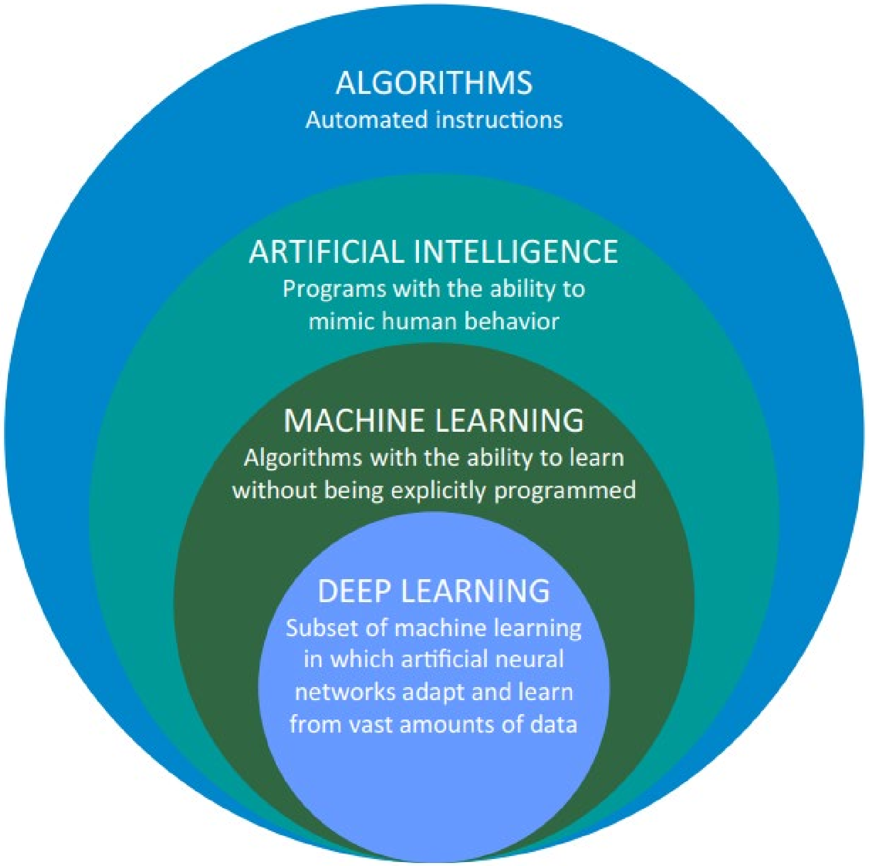
- Decisions made based off explicit rules
- Easy to determine why specific decisions are made

Learning-based

- Decisions based off training data
- Difficult to determine why specific decisions are made

Analysis

- Assisted – ADA provides analyst with flagged dataset
- Automated – No analyst



Vrana and Singh, 2021, https://doi.org/10.1007/s10921-020-00735-9

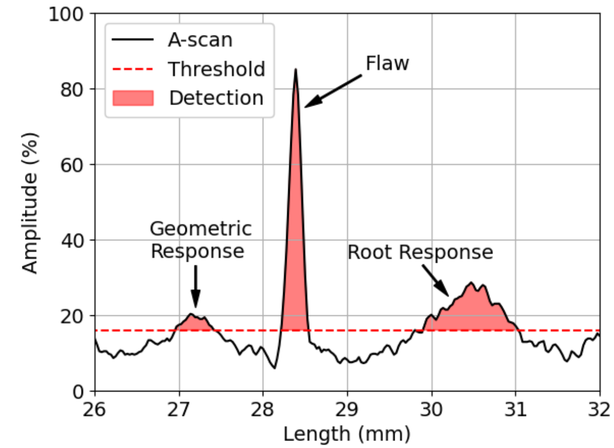
Evaluation of ADA for UT

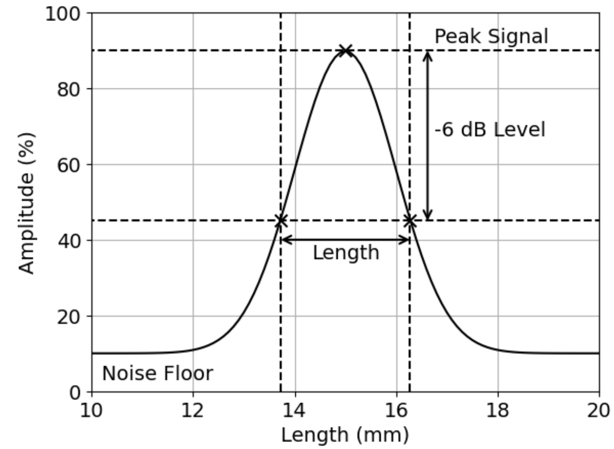
- Objectives
  - Assess current capabilities of ADA for improving NDE reliability
  - Provide technical basis to support regulatory decisions and Code actions related to ADA for NDE
- Expected outcomes
  - Identify capabilities and limitations of ADA for UT NDE applications
  - Identify factors influencing ADA performance and their impact on NDE reliability
  - Recommend verification and validation approaches and methods for qualifying ML (and ADA, as appropriate) for nuclear power NDE
  - Identify gaps in existing Codes and Standards relative to ADA for UT NDE

Assessment of Rule-Based ADA

Takeaways from Literature Review

- Almost all recent publications are dealing with learning-based analysis
- Rule-based ADA is usually used for flaw detection and signal processing
  - An amplitude threshold can be used to identify flaw signals above the noise floor
  - Signal processing can help improve signal to noise ratio
- Rule-based ADA can achieve high detection rates but also high false call rates
  - Not able to consistently distinguish between geometric and flaw responses





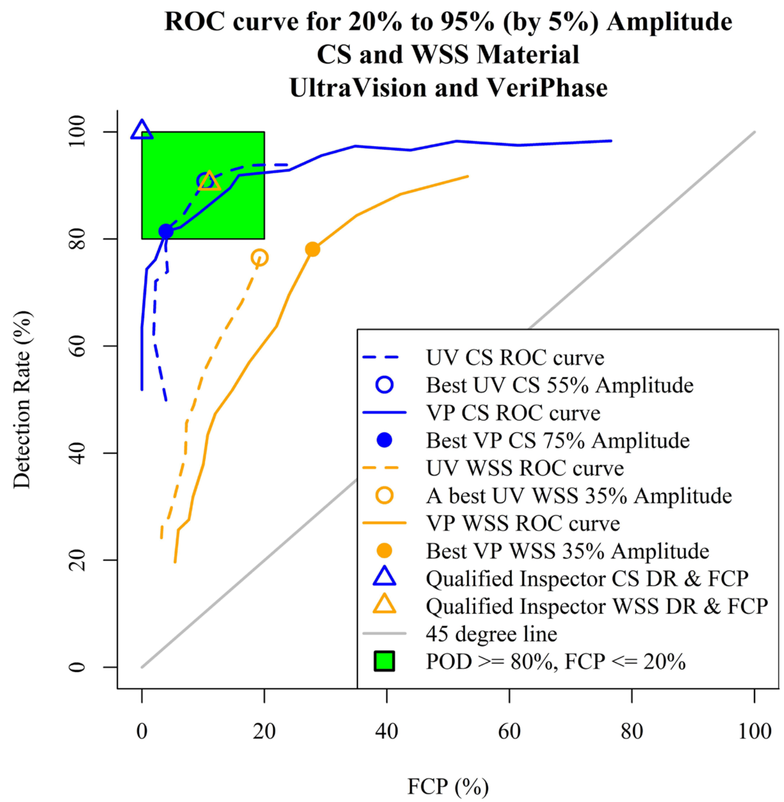
U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Assessment of Rule-Based ADA

Empirical Evaluation of Commercial ADA Systems

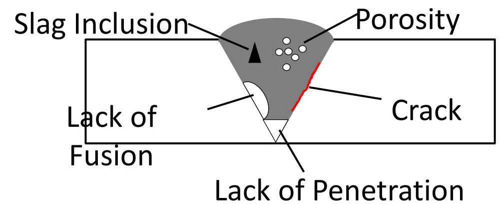
- Data analysis with two different commercial ADA software packages compared to analysis by qualified Level III UT analyst
- Statistical analysis of results using established methodologies

- Rule-based ADA is likely not fit for nuclear pipe inspections on its own
- Rule-based ADA could potentially be used alongside learning-based methods depending on the use-case

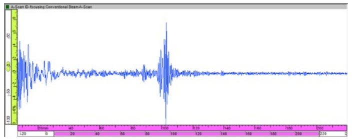


Assessment of Machine Learning (ML) Algorithms

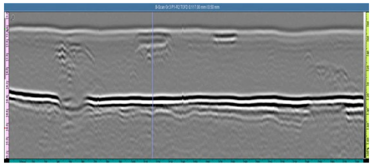
- Limited to ultrasonic NDE classification problems with data from weld inspections
  - Materials: Steel (*austenitic stainless steel, DMW, etc.*)
  - Flaw types: saw cuts, *EDM notches, thermal fatigue, stress corrosion cracking, weld fabrication flaws*
  - Inspection procedure assumed to be appropriate for weld inspections



https://www.zetec.com/blog/destructive-and-nondestructive-testing-of-welds-how-ndt-ensures-quality/

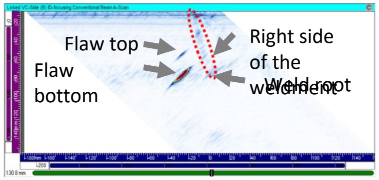


Example A-Scan



Example TOFD Scan

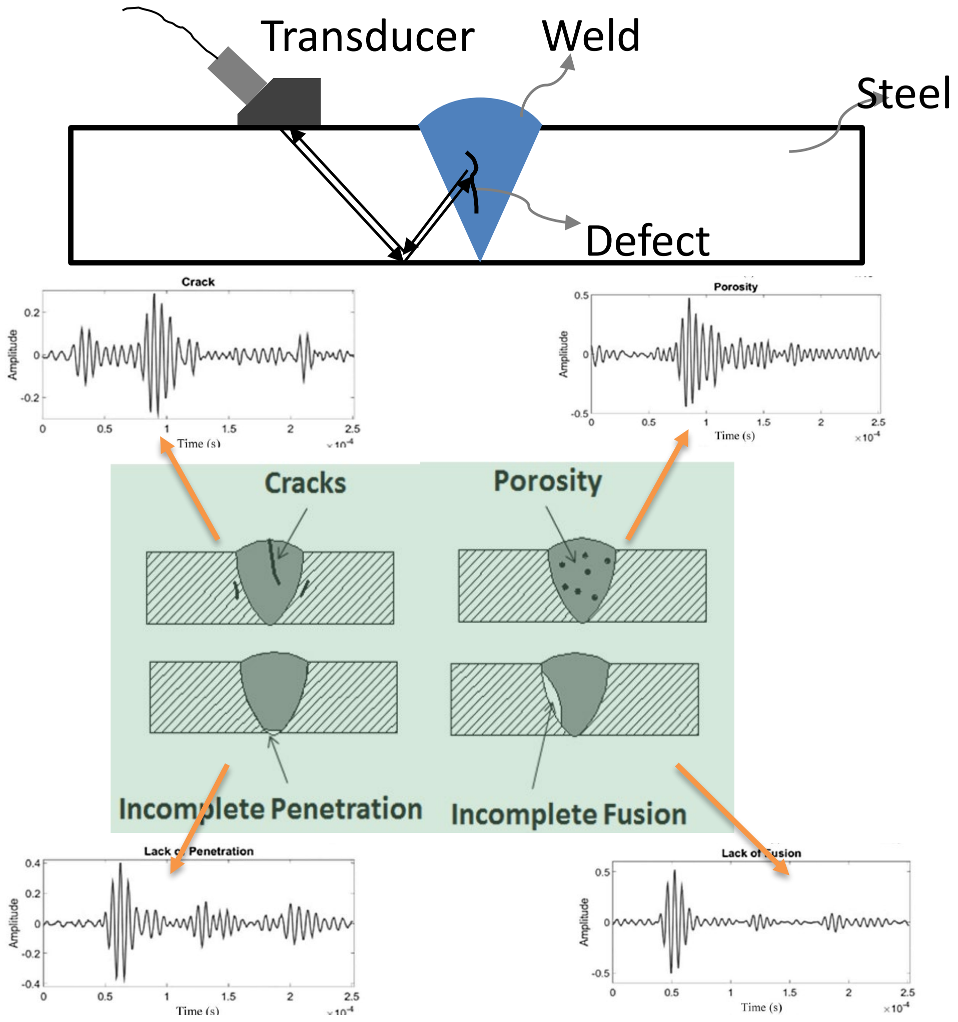
https://www.olympus-ims.com/en/applications/introduction-to-time-of-flight-diffraction-for-weld-inspection/



Example B-Scan

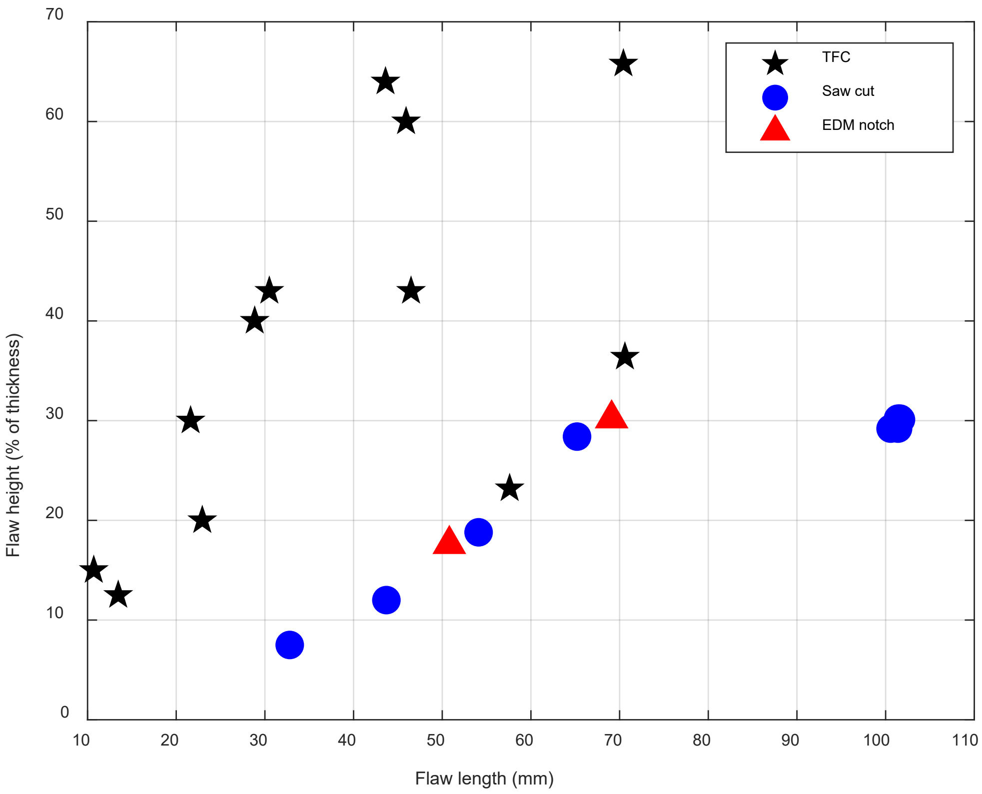
Empirical ML Research Objectives

- Determine capabilities and limitations of ML for NDE
- Identify factors influencing applicability to other inspections (CASS, DMW, RVUH, etc.)
- Assess effects of data augmentation, including using simulated data
- Establish methods to quantify confidence in ML results
- Assess capabilities for flaw size quantification from UT data



Generic Workflow for Assessment of ML for UT NDE

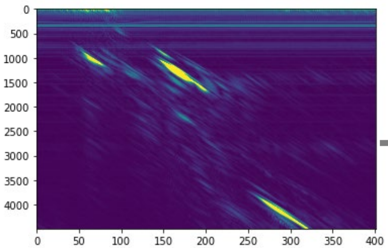
1. Collect ultrasonic NDE data from a variety of materials with multiple probe designs, frequencies and wave modes
2. Pre-process the data to remove noise and outliers
3. Train a machine learning algorithm on the preprocessed data
4. Use the trained algorithm to analyze new ultrasonic data
5. Assess the results using multiple metrics



Flaw size distribution for four stainless steel and two DMW specimens.

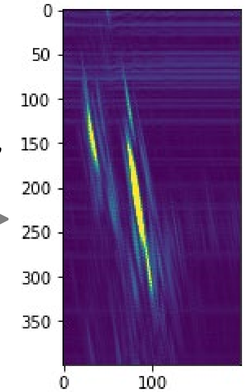
Overview of Empirical Assessment

Original B-scan



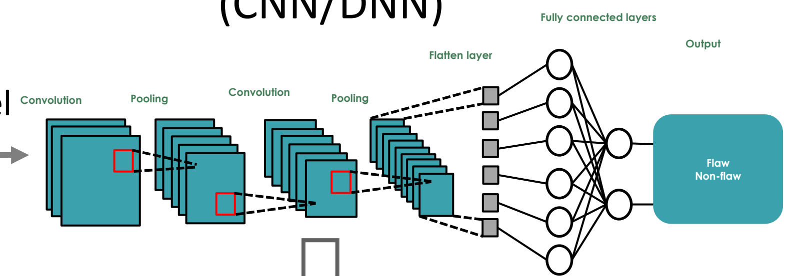
Preprocessing
(Crop, normalization, downsampling, flip)

B-scan input



Input to ML model

ML modeling
(CNN/DNN)

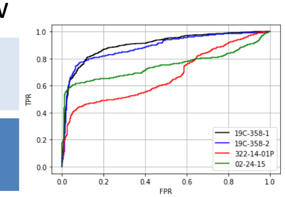


Accuracy=0.88
TPR=0.83
FPR=0.05

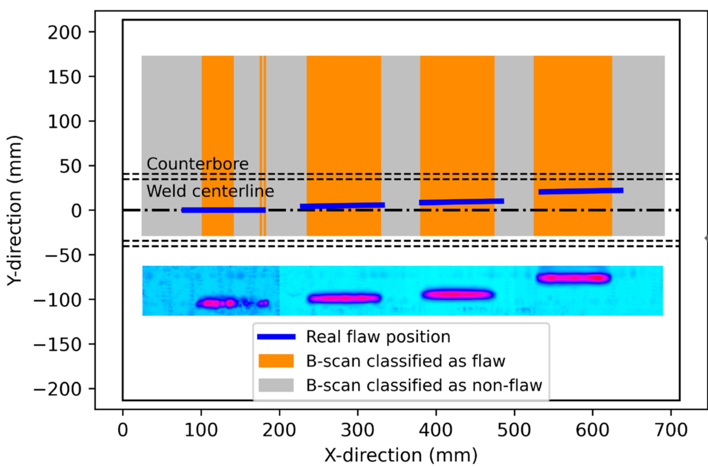
Accuracy, true/false positive rate (TPR, FPR)

	Flaw	Non-flaw
Predicted as Flaw	339	14
Predicted as non-flaw	67	274

Confusion matrix



ROC curves



Visualize Results

Examples of Results

Training with →

A	**Plate**	**4 saw cuts**
B	Plate	4 saw cuts
C	Pipe	3 TFC
D	Pipe	4 saw cuts 3 TFC

Specimen B

	Flaw	**Non-Flaw**
Predicted as **Flaw**	**339**	**14**
Predicted as **Non-flaw**	**67**	**274**

Accuracy=0.88
True positive rate (TPR)=0.83
False positive rate (FPR)=0.05

Specimen C

	Flaw	Non-Flaw
Predicted as Flaw	**40**	**27**
Predicted as Non-flaw	**88**	**325**

Accuracy=0.76
TPR=0.31
FPR=0.08

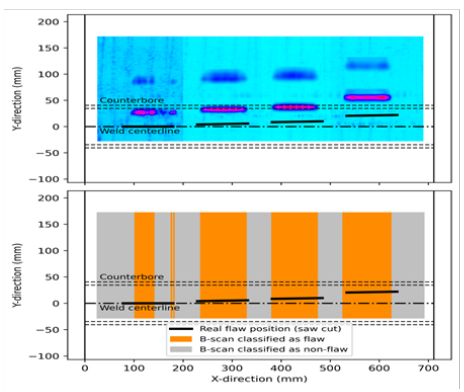
Specimen D

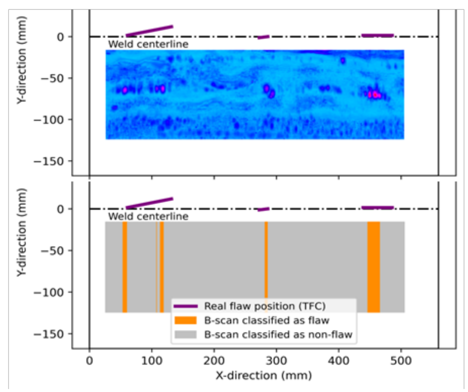
	Flaw	Non-Flaw
Predicted as Flaw	**243**	**58**
Predicted as Non-flaw	**29**	**352**

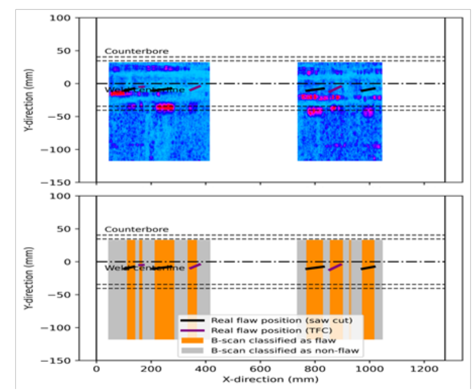
Accuracy=0.88
TPR=0.89
FPR=0.14

UV (top view)

Scan area







Low true positive rate on flaws close to weld centerline and on smaller TFC flaws

Transfer Learning Example

Test results using the retrained model

Specimen A (4 saw cuts)

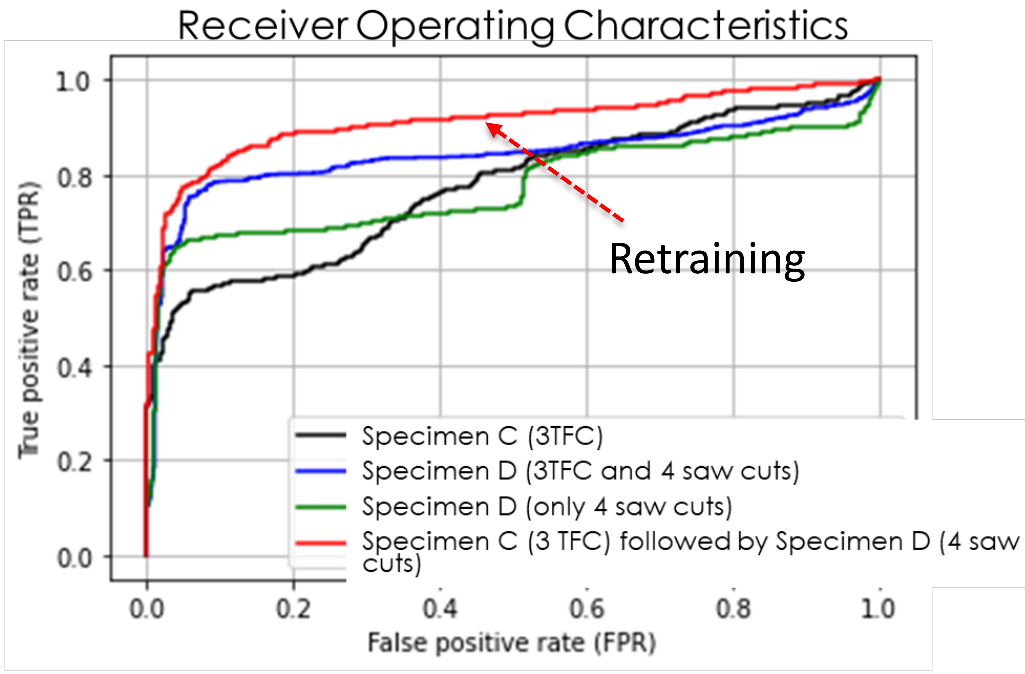
	Flaw	Non-Flaw
Predicted as Flaw	**362**	**26**
Predicted as Non-flaw	**44**	**262**

Accuracy=0.90
TPR=0.88
FPR=0.09

Specimen B (4 saw cuts)

	Flaw	Non-Flaw
Predicted as Flaw	**338**	**65**
Predicted as Non-flaw	**68**	**223**

Accuracy=0.81
TPR=0.83
FPR=0.22

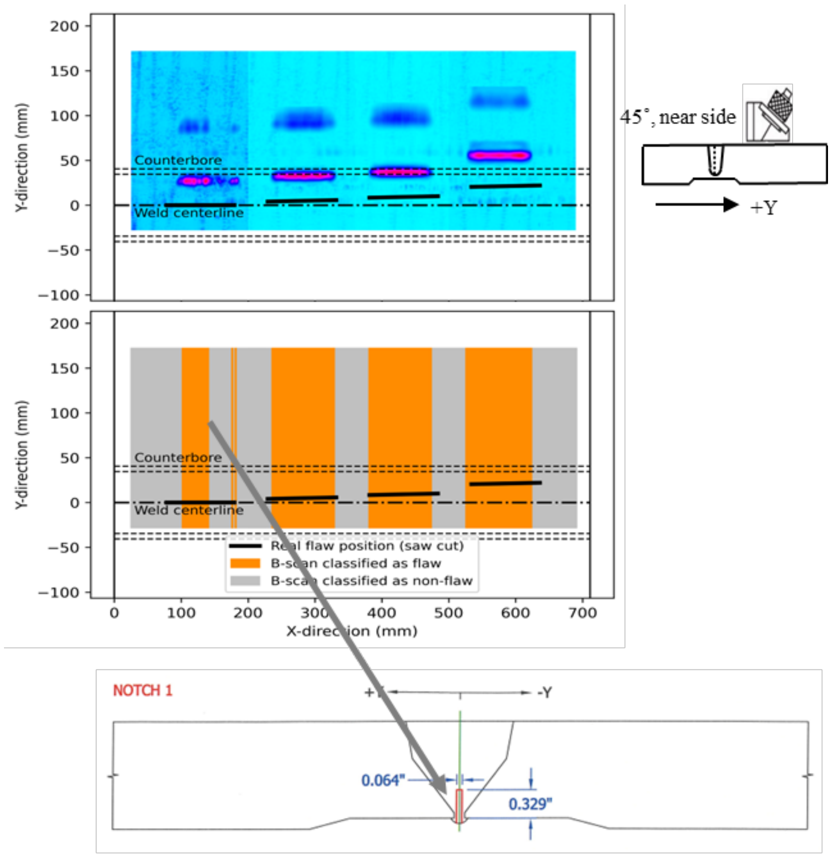


ROC curves on specimens A and B from different models

Retraining and incorporating transfer learning methods may help to improve the performance when the model encounters new data.

Findings to Date: ML

- Capable of high TP, low FP and FN
- May be able to learn key signatures using data from simple flaws (e.g. saw cuts) and generalize well to other flaw types (e.g., TFC)
  - Generalization capability may vary with flaw size and location
- Transfer learning techniques may be useful for improving accuracy with new data sets
- Model type (for instance, NN vs DNN) may not *significantly* change results

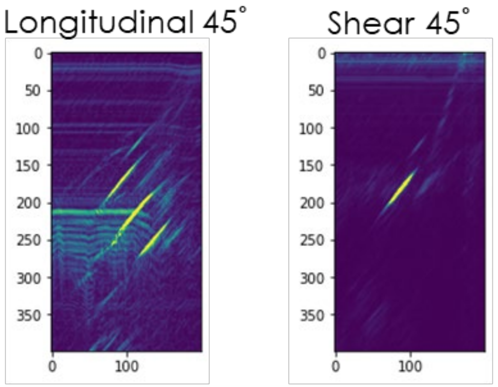


Flaw inside the weldment
(TPR=0.54)

ML , *if used with care*, can be used for NDE data classification

Findings to Date: Data

- Training data should be representative of the types of data expected during testing
  - Expanded training data sets may allow ML to accommodate nominal weld geometrical variances and associated noise

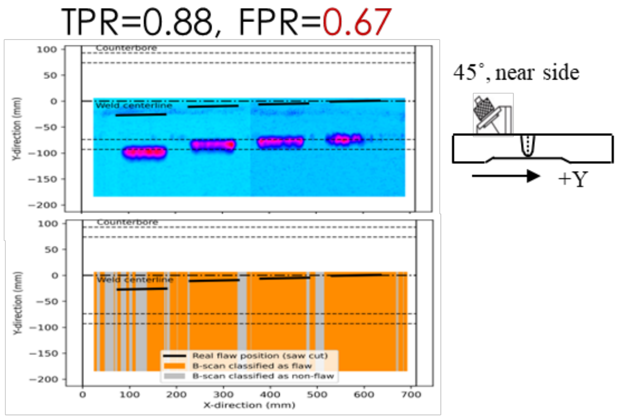


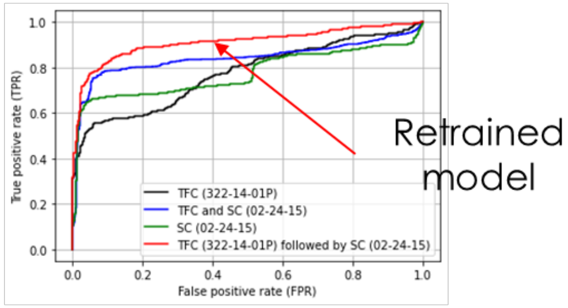
Variation in Data (probe/mode)

- High accuracy possible if test data is “in distribution” relative to training data
  - Consistency across training and test data sources important for high classification accuracy

Findings to Date: Metrics

- Desired performance thresholds likely dependent on use case
- Commonly used metrics: TPR, FP, FN
  - Low FP and FN rates, high TPR desirable
  - Zero FN, low FP, high (100%) TPR for screening?
- Other useful measures
  - Receiver operating characteristic (ROC) curves – TPR vs FPR
  - ML training curves – can indicate overfitting and potential poor classification accuracy if deployed



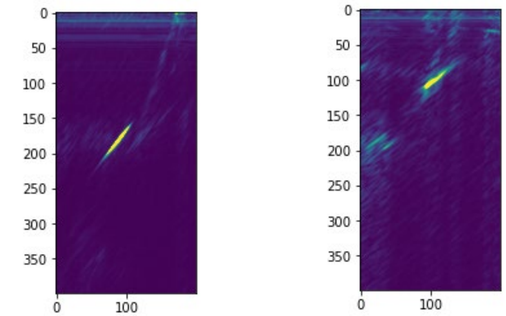


U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

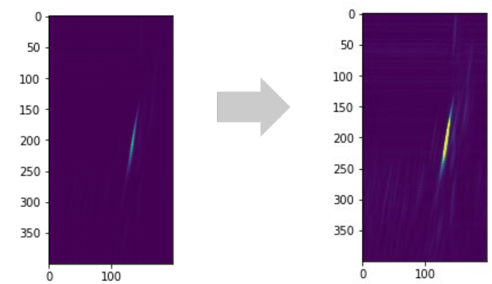
Findings to Date: Best Practices

- Consistency in preprocessing procedures (crop, normalization, down-sampling, etc.)
- Review and correct, if necessary, output labels
- Tuning and selecting parameters that control the learning method
- Retraining a trained network with additional data to improve performance and tune ML to site-specific data

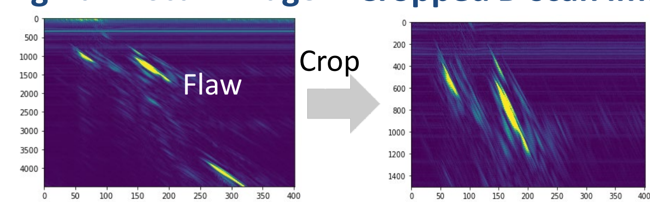
B-scan of saw cut   B-scan of TFC



Normalized to 40 dB



Original B-scan Image   Cropped B-scan image



Status of RES Program – Assessment of Commercially-available Algorithms/Systems

- Technical Letter Report entitled "Evaluation of Commercial Rule-Based Assisted Data Analysis" in the RES/NRR review cycle
- Confirmatory analysis of the commercial ML system being tested by industry in field trials has recently begun
  - Focus on upper head examinations
  - Mockups being designed and fabricated
  - Assessment will include:
    - Pre-trained algorithm tested with vendor collected UT data on NRC-owned mockups
    - Training and testing with PNNL/ORNL data with comparison of results to ORNL ML algorithm results

Status of RES Program – ML for UT NDE Ongoing Research

- Impact of ML on POD, and comparison of ML results with manual analysis performed by a qualified analyst (including comparison of ML performance against Appendix VIII requirements)
- AI-Assisted vs Fully-Automated analysis: Detection and sizing of degradation that the ML system has not been trained on, validation/qualification requirements, and essential variables
- Qualification of ML
  - Training, test, validation data requirements, and benchmark data sets
  - Acceptable performance thresholds and requalification processes
- Methods for establishing confidence in ML results
  - Verification and validation of data and methods
  - Uncertainty quantification, ML interpretability, and related criteria (if any) for qualification

Status of RES Program – ML for UT NDE

- Technical letter report entitled “An Assessment of Machine Learning Applied to Ultrasonic Nondestructive Evaluation” (ORNL/SPR-2023/3245) published February 2024 (ML24046A150)

- Other publications
  - H. Sun, R. Jacob, and P. Ramuhalli, “Classification of Ultrasonic B-Scan Images from Welding Defects Using a Convolutional Neural Network,” *Proc. 13th NPIC&HMIT 2023,* Pages 272 - 281 . ISBN 978-0-89448-791-0 (ML23241A961)
  - H. Sun, P. Ramuhalli, and R. Jacob, “Machine Learning for Ultrasonic Nondestructive Examination of Welding Defects: A Systematic Review,” *Ultrasonics,* Vol. 127 Issue 1, Jan 2023, Pages 106854 (ML22284A071)

Research Program Outcome – providing the technical basis to answer…

Potential Qualification Pathways for ADA (including ML)

ADA for classification (flaw detection)

- Can adopt approach similar to existing Section XI, Appendix VIII for performance demonstration
- Assumed standard for performance:
  - Greater than or equal to current practice (i.e. human performance)
  - Could adopt similar acceptance criteria for performance demonstration

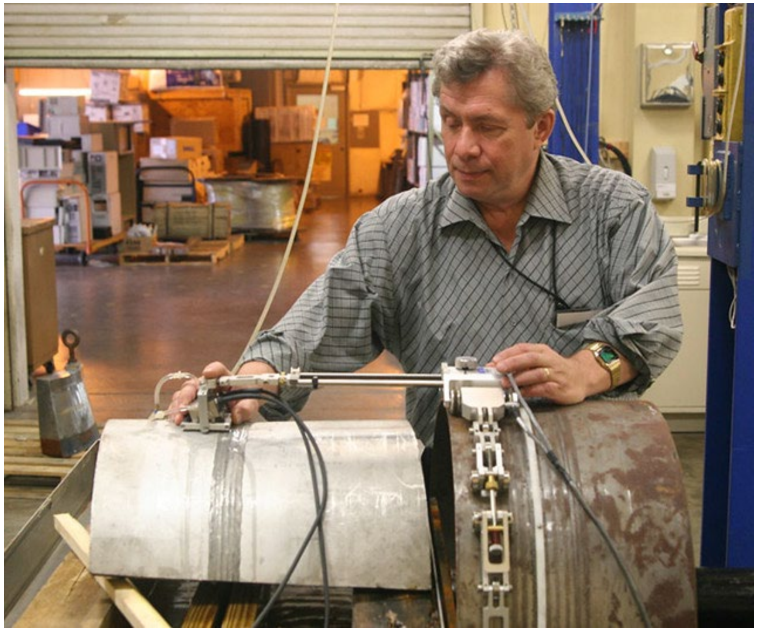
ADA for screening (excluding unflawed regions from evaluation)

- Can adopt approach similar to existing Section XI, Appendix VIII for performance demonstration
- Biased toward calling "detections"
  - Goal is to have **no** "misses"
  - Tolerance for high false call rate
  - Qualified UT analyst responsible for all calls
- Acceptance criteria should reflect the bias toward detection
- Do training/qualification specimens need to incorporate non-flaw features intended to generate a "detection" response with the algorithm?

If ML-based ADA has the potential to be better than current practice, then should ADA be held to a higher performance standard?

Initial Qualification Requirements for ADA-Assisted Examinations

- A UT procedure that uses ADA-assistance can currently be qualified using Appendix VIII as the user of the procedure is a UT Level II
- How should the qualification requirements specified in Section XI, Appendix VIII be updated?
  - Currently only covers encoded data
  - There are many complexities associated with training ML algorithms not captured in current rules



Implications Related to Retraining ADA Algorithms

- If an ML algorithm is retrained, the algorithm has been altered and is a change of an essential variable in the procedure
- In ASME Code Section XI Appendix VIII, a procedure must be requalified via a successful personnel qualification if an essential variable is changed
- The NRC understands the potential benefits of changing the ASME Code to allow for field-friendly implantation of ML (e.g. requalifying a retrained ML algorithm on-site)

Paths to the Future for ADA

Near Future on Current Trajectory

Skilled Inspectors + Engaged AI Experts + Optimized Procedures = Best Outcome

Current Inspections Improved

New procedures can be developed

New degradation can be found

New inspectors become overly-dependent on AI tools

AI experts move on to new tasks

Lack of AI experts prevents effective development and retraining of algorithms

Unskilled Inspectors + No AI Experts + Unoptimized Procedures = Worst Outcome

Current Inspections may be OK

New degradation may not be found

New Procedures may be challenging

Care must be taken to prevent this outcome.

Avoiding Future Problems

- Industry needs to build the infrastructure to allow for the effective use of ADA
- Create rules for requalifying an algorithm after modification that does not require a person to pass a personnel test
  - e.g. Finds all flaws in qualification data without too many additional false calls
- Requirements for personnel to use ADA-assisted procedures to assure that they have appropriate skills
  - e.g. Pass an Appendix VIII tests for the same Supplement without ADA assistance