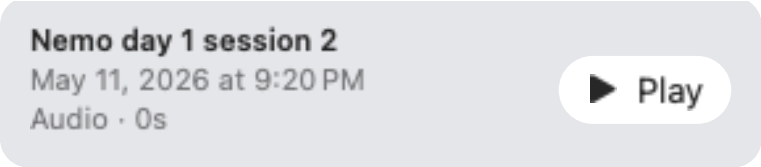

Paul: Report children attending this year.
Thanks to you, Natasha.

Natasha: Yeah, they both ended up showing up and at one point when my, because I turned off my camera.
They've been, no, by both of them on my lap and like listening and they're like, are they talking to you?
And I was like, yes, the please don't say anything. Like, ooh, buttons.
And I was like, do not press the buttons.
At all.
Yeah, I have to pause.
So yeah, so we can throw that in too.

Paul: Fun for the whole family.

Natasha: So fun. So fun.

Andrew Shao: And there's multiple axes of the diversity and inclusivity, right? Age, Age being one of them.

Natasha: Excellent, excellent.

Inge: Okay, great. I guess, so we are already at 10 past the hour, so I think we could start with the next part.
I'm, I'm very happy to, to have both Amber and Andrew, um, uh, agree to do the discussion session on AI integration into Nemo modelling and analysis processes, and also Andrew has also, uh, been very kind to help out tomorrow with the early career day. So, um, this is, it's going to be great.
Um, I don't know if the best would, uh, you guys like to share your screens or do you, would you like me?
Okay, great.

Andrew: Yeah, I can share my ice cream.
And I think Amber, correct me wrong, but I think the order of the day is essentially, I think, um, I'm gonna give a little bit of an just overview, like, and just set some discussion points about the broader kind of questions about AI, um, and specifically whether in climate.
Um, and then Amber, do you want to take point on kind of going around the horn for anybody who had that one slide about AI activities? Um, and then, yeah, we'll just kind of open up for discussions, maybe we'll kind of distil a few discussion

questions.
I'm down to prompt, prompt over the course of the next hour or so.

Amber: Yeah, yeah, that sounds good.
Um, yeah, so Andrew's gonna start with a quick overview of some broad applications of AI, machine learning, and ocean atmosphere fluids, um, and then we'll have a little bit of discussion around what, he talks about left for some questions there.
And then, I think, um, maybe one or two people have shared some slides, which then maybe, um, I don't know if you have those in your deck, Andrew, so maybe they won't turn...

Andrew: I don't, though.

Amber: To Inge to share those.

Amber: Um, yeah, and then we're just hoping to have a really good open discussion.

Andrew: Perfect.

Amber: Sounds good. So, yeah, you can probably go ahead and share your screen, and yeah, just maybe introduce yourself, too, just in case...

Andrew: Sure. Yeah. Uh, so, I'm Andrew Show.
Definitely am happy to see a bunch of people that I know on this call, the names I know was called.
I started out in doing ocean modelling at the University of Washington, and then I postdoc at GFDL, and then CCCMA, and then I had then jump shipped over to HP, to kind of look at a lot of these questions about how we inject AI, MML, into scientific simulation, what the role is.
So that's basically what I've continued to do here for the last 3 years of HPE. I also will just make mansion too, that I am part of the Nemo development team. I am the co-chair of the Eddie Closure working group.
So there's anybody who's really interested in doing turbulence modelling, um, at all scales, submesoscale with scale, mesoscale, just let me know and I'd be happy to kind of entrain you into some of those efforts.

So, yeah, so, for today, I just kind of wanted to kick off the conversation about this question of, like, okay, what are we doing for AIML and ocean modelling? Um, I'm not gonna talk about my research in general, so this is kind of just more from the very high level kind of question about asking this question, like, what should we be thinking about, what is the world that we're living in, and what's the broader kind of perspective of computing in simulation in general?

Um, so, as part of that, I think it's this really interesting, and to ask the question, okay, like, why has there been such a rise in AI for computing? And for science, AI for science, right?
And it's kind of informative to see this kind of time series. So what's on the left here is, so it's a log plot on the vertical axis, and it just sort of compute. And this is the high performance computing top 500 list.
So there's 3 lines here.
So, the top line is the sum of all compute on the top 500. The second line is the computer capacity of the number one machine on the list. And then the last line of the bottom is, what is the computing power of the lowest machine that made the list.

And so there's kind of some things to point out on here, right? So, like, overall, the sum of compute, you know, there's not that strong of, it's almost linear. Roughly linear, right? But if you look at the number one machine, It follows that as well. Um, it's a little bit forced up, really, to get into that.
But we feel, like, the number 500 machine, there's this very clear kind of asym totting that we're getting, right?
Around that, 100 plot per second mark. Um, so it's kind of interesting to point out some of these break points.

So 2012 is the first kind of point that I see where there's a step change in this particular part of it.
There's some step changes here, which are not as related to something, so I'm not going to talk about one of them.
But 2012, um, this jump from here to this triangle really went, uh, is because of Titan.
So, Titan is the first machine, number one machine in the world where everybody never had GPUs on it.
So, this is distinct, like, these are the graphics processing units that we use primarily now, or a lot of uses being done for acceleration, versus the standard kind of CPU based machines, that we tend to run to things like Nemo, right? And, like, a lot of scientific simulation really fundamentally relied on CPU computing.
So that was the first big jump. So, there are those, but it's raw compute power.

Um, and there is the question about this as well, um, that we'll talk about in a little bit in terms of CPU and GPU.
But just know that that was the really, that's the reason for that step change, is like, we got GPUs on every single node of a supercomputer.

And then if you look at 2023, um, which is this mark right here, this is the first time that we actually hit the EXA-plot scale, right? And that was because there was multiple GPUs on every machine.

Super, like, these are anything she appears, like, top tier ones, and we broke the access skill barrier then.
And the only reason why we were able to do that was because of GPU or computing.

So, yeah, in training and inference performs really well on that kind of hardware, those GPU hardware.
And so, then the, I think, one of the other questions that people always ask is like, okay, well, why hasn't CPU computing gotten better? And it's just fundamentally because the hardware style, CPU computing, versus GPU computing is very different, right?
GPU computing is great at doing super parallel competition, computations, and what I mean, parallel, I don't even mean, like, how we do, like, multiple domains in the nemo.
I mean, like, matrix, matrix, matrix, matrix, multiplications, matrix factor, right? Where everything is an independent copy takes soon.And maybe you bring it all together at the very end.

Whereas, like, you know, Nemo style computation, we actually do very little arithmetic in comparison, right? If you're actually summed up all the computations that happen per grid cell and Nemo, it's probably on the order or something like 100 arithmetic operations, and that's really, really low computational arithmetic intensity, that's really good for CPU based computing, but it's terrible for GPUs. So there's this fundamental tension between the two types of things. And it's partly a numerical techniques kind of thing, because we've been living in a world where the style of low arithmetic operation has really been the limiting factor to performance. Whereas now, this kind of very large scale, but necessarily independent computation, is really, is dominant. So we have these two things now that, fundamentally, we have a split at the very high level. We have hardware that's good at simulation, and we have hardware that's good at AI, ML workloads, which are primarily linear algebra based.

You know, and so there's been a lot of efforts recently. You may have heard about some of these, Nemo itself actually has an initiative where they are accelerating parts of the Nemo code base with GPUs as well. Uh, you know, there are, there are piecemeal efforts from some other ocean modelling groups that are trying to do this, where they're saying, like, okay, well, maybe we could put the dynamic core on it, or, you know, maybe we can put the physics on it or something. Um, and then there's other strategies that people are pursuing, which are just rewrite everything so that they are GPU, uh, efficient.

And so, that's, those are all strategies that are being pursued right now to solve this problem simulation.

The, what we've seen so far, though, in all of these efforts, is that we do get speed ups, but they're not these, like, 100 X, 10 X, like, 100 X speed ups that people are hoping for, you know, oftentimes it's kind of, like, a two times improvement, maybe a five times improvement, and then as you add complexity, you know, those numbers kind of bounce and change around. And one of the things to think about, even if people are interested in the economics of it, is that a GPU is almost 10 times expensive as a CPU, if not more, right? And so, are you getting the same performance per dollar spent? Um, you know, and then energy, obviously, since we're all here in ocean modelling, we also think about the carbon cost of things, and that's yet another big question.

So, one of the things, one of the reasons why computing with AI and accelerating, like, weather and climate is so popular is because it is the one part of our workloads that are continuing to scale fairly well. And the circuit models are really quick to compute on GPU. And so this is when I'm talking about circuit model, I'm talking about things like, um, ECMWFs, like AI forecasting system. I'm talking about the types of surrogate models that are being done out of ECCCs and MRD now, right? Where it's like, they're taking the full, the training machine learning models based on data from the numerical, the traditional miracle simulations and then just running them. And one thing to point out to this is that they are super quick.
This is what's getting people, like, this 100 times speed up, in terms of just being able to do a single prediction.

There's a few reasons for that. One is that, um, you know, you can take very long time steps. So for those of you who are doing regional modelling and very small spatial skills, you know that your time step has to be very small, otherwise the model starts blowing up. Right?
In this particular case, right?
Like, or a lot of the weather models, we're seeing them go from something where the atmosphere model has to be stuck for it at a minute. And, you know, you can actually do a prediction at six hours, right? So, going from 60 seconds to 3,600 seconds, is a, is a, is a 100 order magnitude, like, order 100 times faster, right? It's much, much faster.

And the other thing that you can do is that, again, like, when we were doing traditional numerical modelling, you wanted to have high resolution, right? So you can resolve the scales of turbulence or the spatial features that you actually want. Whereas a lot of times, what they do for surrogate models is that they actually coursin the resolution, which retains most of that information, and due to the way that AI surrogate models work, they're actually able to continue on, and, you know, have a high representation of things, even though they are being trained and using lower resolution prediction. So there is another kind of factor of three,

improvement that we get, just because we can use coarser data.

And these are the two fundamental things, right? So these are very different ways of solving the same problem. And one of the things I really always emphasize, and I think I understand the next slide, is that all these surrogate models. One of the things that you also have to keep in account is the time to train the model, and not just looking at the raw inference time of it.
So, for things like weather and climate, this is actually, it turns out favourably because you can train once, and then you can use that same model over and over and over again, with minimal. And so you kind of amortize that cost of training the model.
But this is really one of the big things. And one of the main reasons why AI, AI surrogate models of this type are really popular and rising in populate very quick.

The other reason why AI modelling is going very quickly is just to compare the development time scales, right?
So this is human time. So, um, on the top, there is NOAA's GFS. So it's the NOAA's global forecasting system.
So to move from one generation of the GFS to another one, they kind of budget out this kind of like 6 to 7 year time frame to evolve, right? This is a constant development, constantly improving the physics, the numeric regimentation, updating the initial conditions, boundary conditions, you know, all the work that you do tuning the model, and so it's six years.

Now, we compare that to what ECMWF did. So, in 2023 is one of the first papers worth the forecast net, um, NVIDIA release ForecastNet, uh, and show that it had these representations of medium range weather forecasting.
Within this, and so, like, you know, people started paying attention to that, and AIFS, which is ECMWFs, parallel, numerical weather prediction system, first develop, we first started developed, started development in Midway 2023, and they actually released a version of this AIFS within the span of six to 12 months, right?
So you can see there, and part of the reason for this is that, um, one, there's more tooling out there in the world where there's much larger problems, so they're able to take advantage of a lot of the things that we, uh, that the community has built to make training easy and efficient. Um, and then also partly because, again, like, it can take advantage of the GPU scanning, right? So we can take advantage of the hard ware in ways that we couldn't do with simulation. So, and then if you look at the entirety of it, right?

So, for ECMWF, AFS development starts, and then, arguably, what the gold standard is for weather forecasting is ensemble based modelling. So AIFS was released earlier this year, right? And so that's maybe a one and a half to two year development window, compared to the development of a weather forecasting

system. So there's a huge advantage there too, right? People time. And I think we see this in a lot of different ways. Across it. This is also not to say that the development of the underlying modelling systems are now worthless in the span of AI, right?

missed speach

Because the fundamental thing, one... And so, the only way to generate better embeddedator is to collect more and more observations and to develop the complexity of your numerical. Your underlying numerical-based weather protection models, because that's where we know how to inject things like improvements to parametersThe representations of new processes are better representations, new processes. And so it's a virtuous cycle of development. It's just that you have longer scales of development for the numerical models and the shorter times skills for AI.

Okay, so what's next for weather modelling? And I think this is also a question that's being asked, or this is also a question's being asked on the climate time skills as well. So, and there's 2 different axes for this, right? So, as of right now, there is this question about what is a standard data simulation time, like weather forecast. What happens, right?
You go from observations, raw observations, you do data processing on it to acute quality analysis, quality control. And then you feed it into a data assimilation system, right, to get generate your initial conditions. And then you then do a forecast on that, and then you generate your products, whatever products you want, right? So these are your forecast, your forecast metrics, these could be things that you see and provide, uh, beyond your phone that are, like, just the standard weather forecast themselves or derived variables thereof.

One of the big things that people are now trying to do is to understand whether we can go use a AI surrogate model, or a series of AI surrogate models, to go straight from observation, bypass a lot of these steps, and immediately give it to a forecast, you know, again, the ability to do this is predicated on the robustness of that top part of the panel, right? Where it's like, if you're gonna have any hope that you're gonna get a better product at the end, you can simulate, you can emulate what's going on underneath the hood, but if you don't have anything to emulate against, or the thing that the quality of the thing that you're emulating doesn't improve, your emulator also will not improve. And so this is where it's been a really interesting kind of change in thinking about it, where I don't think people, in general, are not asking the question of, like, can we replace all of our simulations with air anymore?

But they're asking the question, like, where are the places where we can emulate? Where do we depend on the numerical simulations or the numerical systems, and

how do we, what are the best places we can do that, and how can we combine these two things?

So, yeah, I only have a couple more sides here.

So I think bridging the gaps and climate to weather, right? There's also been some really interesting work in this space. So, what I'm showing up on the top left panel there. That's the early, that was a, that was an early prototype of a video's weather model from early 2023. And you can kind of sort of see that, okay, it does well for a couple of days and then it blows up completely, right? It does not look like anything realistic.

In comparison, if you look at, you know, any of our numerical simulations, they tend to be very stable, and we have engineered them to be such. We derived the numerical techniques so that they are stable. Um, so now the question is, is how do we actually, uh, extend short-term forecast, how do we get stability and still get a, a reasonable estimate of what the climate system looks like, or Earth system looks like.

And a lot of these things that, um, the AI models, as they've been kind of encroaching this domain, have encountered very similar problems to what we did during the early development of kind of taking numerical models and trying to extend them to the climate time scales. So some things are super important, conservation actually becomes really, really important, right? And then long term stability, like the ability of a simulation to kind of continue, or of a system, whether it's a simulation or a circuit model, to be able to just be run out indefinitely is really, really key.

So now, instead of just, you know, for all of us, because it's the ocean model community practice, we know that, oftentimes, the people ignore the ocean, but we all know that the ocean is actually really important, coupled earth system, and so it's really, again, gratifying to see how much benefit these, this type of problem, both in the weather, and also in climate, that we get when we actually include a representation of the ocean in these surrogate models.

So there's two that I want to highlight here. So there's DLE SIM and some Mudrays, so DLA SIM was from NVIDA.

Mudrays is a joint collaboration between M squared lines down in the United States along with some private foundations to do this. And both of them show market improvement when you include a representation of the ocean.

They're slightly different. So, DLE SIM has, basically, a surrogate model of a alab ocean. The equivalent of a slab ocean.

It only predicts SST, um, and, you know, it gets the heat fluxes and presufluxes from the from the atmospheric model.
But it's trying to predict SST only, whereas Smudres has a full 3D ocean coupled to the atmospheric model, along with primitive ice and land model associated with it. But these two things, both of those two models run freely, for a long period of time, and are stable. And the reason for that, that both papers kind of claim is because there is a representation of the ocean.

I'll note that there's this third one here, Cumulator, so this is out of NCAR. This was trained on purely an atmospheric model. They can also get long term stability, but they have to play the same games that we used to play in the climate modelling space, which is essentially fix and enforce mass energy and heat conservation in the model in order to get that one. And they do not have an oceanic representation.

There are still deficiencies in these things. So, if you look at the bottom panel there, all the CM4, which is the underlying GFTL numeric model that SMU trace was trained on, you can see that it has a reasonable representation of Enso, where you have, you know, kind of a very sharp, high frequency peak, or, sorry, this is article correlation. So, like, obviously highly correlated, you know, with itself.But then, as you go on, you can kind of sort of see peaks at roughly the three ish and five year ish and six years marks.

Right? So, but it's not very strongly or overly. In contrast, the emulator has this wild pattern to it, right? Where you get these highs of, uh, based on the lab, you get super high autocorrelation, and this is not what we think of as being particularly representative. So there's still some kinks to work out here, especially if we want to do these things for long simulations.

Um, it suggests that we're not, they're not getting the long-term variability, um, of this correct, and they're kind of more getting something probably like a seasonal cycle, being alias onto some of these longer term modes, um, a little bit better, but it's still an open question about what, why the cause of these actually are. Um, so just kind of want to, like, put that in the frame, right?

Because I think there's a number of us here who are doing kind of more, either regional modelling.Uh, regional modelling on shorter time scales, so on the order of decades, climate modelling on the order, uh, also for regional, but then also extending out to like full global simulations. And I think there is a question to be asked here, um, as to where, and when, again, are you using the right emulator, what are the AI methods that we can use, um, and should we be relying on the numerical simulations, or when can we purely rely on the emulators, or where there's a benefit? Is there a benefit to do some kind of hybrid methodology?

So this is the very last slide and kind of opening it up for a lot of questions. So, AI for physics requires good numerics, first principles-based simulation, because we just cannot simply afford to observe the entire Earth system at the resolution, and for the amount of time that we need to actually train a data-first model. So, numerics and physics-based models are still our main tool across almost any physical domain, where, if we want a good AI representation, we need to use those to generate the data.

I will also say that AI and ML use, and weather and climate extends beyond just the surrogate model that I've been focussing on. Main reason why I want to focus on here is because it's most people's, when they think of AI and weather climate, they think of the full emulation problem, how do we replace an entire numerical simulation with the AI with an AI equivalent?

Um, but a lot of what my research is, and so, in the second part of it, when I give my, what I'm doing, is to ask this question about, how can we incorporate AIML interesting relations, or around simulations, and what kinds of new research directions, and/or actually scientific insight can we gain by doing that?

Third point here is that, you know, this question about AI and traditional ocean modelling is actually a symptom of a larger trend in computing that I think we're all gonna have to deal with, which is that computing is becoming increasingly heterogeneous. And by that, I mean, there's not just one thing that's being run. And it's not just one thing that's being run, and one type of hardware, but we are kind of moving into this new emerging reality, where CPU computing is stalling, right? So we saw that in that graph at the very beginning of the top, where it's kind of peak peeking out and asymptoting. GPU computing is still scaling, and then, you know, as different as CP or GP computing is, and AI and traditional simulation is, quantum computing is also on the horizon, right? And so there is this question about, again, if there's gonna be, if there's gonna be new hardware, new techniques, new algorithms that we can take advantage of. There is a not as obvious connection straight to climate modelling and weather modelling as there is in some other domains, but there are even things where people are talking about, you know, doing this for, like, using quantum computing, for, you know, chemical modelling, For, uh, physics modelling, aerosol modelling, um, in climate and weather, too.

So there is something like that. But you can actually imagine a world that's probably only a few years off where we might have the first, you know, Earth, weather model, Earth system model, weather climate model that has 3 or 4 different types of computing, all being combined into one giant thing.

Um, one thing that we also, that, based on the understanding of the commuting as

a whole, is that applications of AI, um, really, really depends, knowing what the benefits are and the limitations of both methodologies, right? One is first principles-based, the other is data-driven. And so, like, knowing where uncertainty happens, what does it mean to have an error, what does it mean to, uh, mitigate errors? You know, they are actually very similar things sounding in the English vocabulary. But when you're doing the actual practice of it, they mean very different things. And so these are things that, when you are trying to combine AI and simulation, whether they're directly connected to each other or only tangentially connected to each other, these are huge things that we need to be aware of so that we understand how best to use these things.

And then lastly, there is such a rapid pace of development, right?

Across AI in general, but also within the domain sciences, and also within the Earth sciences, and also with an oceanography, that it is basically just a given now, that we need to be part of a community, at a variety of different scales, like from your group level, to university level, to nation level, to international level, to kind of just be a part of this, and understand where we can take advantage of certain of the advances that are being done in the community, so we're not reinventing the wheel. This goes from both the intellectual side and also the technical side of things. And then also to understand where we can then contribute backup and just kind of be a part of this very rapid innovation cycle.

So yeah, that's it for the intro part of it. Again, just kind of wanted to keep it a high level to kind of just like set the stage more broadly. But I think the main takeaway is that, you know, AI and simulation is a great place to be. It's a really interesting area of research.

You know, if you talk, if you're doing any kind of research on this, and you talk to somebody in molecular dynamics, or in, um, you know, fluid, fluid dynamics modelling, or, you know, even some really bizarre, or what I would consider bizarre things, like material science, where I have no prior knowledge, it's really interesting to see where all these techniques, and people are struggling with the same problems, and coming up with some really interesting solutions.

So I think that that is also the last thing that I want to leave folks with, too, is that, you know, we have really a unique place to be in ocean modelling where it's like, this is one of the most mature branches, a coordinated effort, where simulation is being done, across these varieties of time scales and spatial scales. Um, and that means that people look to the ocean model community, a science community, for inspiration, because there is, this is one of the places just as a whole, where people do use simulation, whether they know it or not.

And so we get a lot of people asking for, like, oh, how is this problem being done in

the Earth system, in the earth system modelling? Um, and then, second, it also means, too, though, that there are some really interesting places that require a little bit of a leap of bridging the gap of comfort, to understand how analogue problems in other fields can be translated into ours as well.

Yep, and so that's what I just want to leave you with. Um, happy to take questions, if anybody has, kind of, like, these higher level questions or discussion items that they want to bring up as well.

Amber: Cool. Thanks, Andrew. That was really interesting. Rounds of applause coming up before you were there.
If people have, uh, questions, you can, or comments. Go ahead and raise your hand.

Andrew: Or if you think it's better for a bigger group, then we can just shove it to the broader discussion section two, and that's fine.

Amber: Uh, Fred, just gotta stand up.

Fred: Actually, one... But one of the.. The.. Last, uh, comments, maybe not on the previous slide, probably? Anyway, the, on the side, in the sense that, um... For instance, we, uh, we care about the ocean, we care about our 1st principles in the ocean, but we do not really have, we handle all, all the physical processes going into the ocean. So we could probably use emulators for things outside the outside the box.
So I'm thinking, for instance, that If you're just interested in physics and of the ocean, and don't necessarily want to spend all the your numerical, well, computer resources into the biology, but you want to have the feedback or biology on the physics, you would like to have the light penetration, for instance. So you, so it is definitely different pedals over and it never, never processes a completely physical difference. Wave action on ocean. The wave model is actually very expensive also to run.
So, this couple is definitely fields like this where you, um, processes like this where you, you could, uh, implement, we include a new model without having to, to necessarily have to, to resolve them, to resolve the physics for them. Just emulate them would be sufficient.

Andrew: Yeah, I think that's a really good point. You know, I think that it's one of the things that's really fascinating, right?
We rely in science a lot on heuristic relationships in general. Um, you know, traditionally, a lot of times, even for, like, you know, what we do for, um, like, the surface boundary layer in the ocean, right? Like, we have fits to empirical data that we're like, okay, this is what we think is how the heat exchange would actually go. But one of the things that, to your point for, that's really, really salient, right? Is

that a lot of AI models can get to the point where you can get fairly accurate very quickly. Um, you don't have to rely on like, you know, kind of the standard things that we, that we've done in the past.
And so they're a really good way of doing some hypothesis testing, right?

They're asking that question, like, if I include this process, some representation of this process, that I might not be able to, as you say, capture from 1st principles, you know, maybe, you know, just the effect in measuring, the effect will tell us whether it's worth it to kind of continue down, and or even if that level of accuracy is sufficient, in a heuristic sense, um, without having that scientific understanding, or ask the question of, is it actually worth spinning up, like, a research project to more formally do this on first principles?
And so, yeah, it's a really interesting application of AI is to be able to just get a high accuracy, good enough representation of something, to just do that initial hypothesis testing.

Fred: Yes, and that, from that point of view, actually, I always mentioned this morning, that the, uh, as a DRAKKAR, we presented, um, a bulk formula based on it, on it, well, well, well. So that was interesting, because this one doesn't require, well, it's not a full emulator, so it's not the issue of reproducing, well, the error accumulating from time to time. I guess. Although I guess it could probably be in some degree. But anyway, it sounds easier, an easier problem. And for which we have, yeah, very poor solution in general. It's just per, yeah, yeah, back parameterisation, for instance, as a typical weakness of the evolution processes we have, we have in all.

Okay, thanks, Andrew. Yeah, it was very nice presentation anyway, very didactic.

Amber: Thanks. Uh, other questions for Andrew. Before we move on to the slides.

Um, I, I was just thinking about, uh, what you were saying about, uh, you know, we, sort of, people have, sort of had this fear, and I think it's still out there in the group at DFO, that these machine learning approaches will replace our numerical models, and, you know, we're quickly gonna be, be out of, uh, out of our jobs, you know, so dramatic, but, uh, um, But, yeah, it does seem to be the case, you know, that we need them to train on.

But I wonder, will it always be the case? Or are we moving to, uh, are we gonna move into a sphere where we have some numerical component of our, our neural network, you know, that, you know, so it's, you know, thinking about these physics informed neural networks that are really rapidly coming onto the scene. Could we be looking at kind of a totally new beast in the future for our bottling efforts?

Andrew: Yeah, and I think that that is gonna be... I think that is the fundamental

question. And I think that exactly what you're targeting, is why are we, and by the world, we, I mean, like, the domain science modelling, computational modelling community in general, is seeing that a future, where there is that spectrum, right? And there will always be the spectrum of pure surrogate, pure simulation. And then, practically speaking, there's kind of, like, hybrid space in the middle where there will probably be things where there's explicit machine learning code in line, um, along with traditional numerics.

And in some ways, we're starting to see this already, too. Um, uh, you know, there's, you know, there's projects that are written in Julia or Jackson or something like that, right, which rely on automatic, which have, as part of the fundamental capability, automatic differentiation, right? So you can do things that are canonically challenging to do in numerics, um, opera, like an adjoint model, right? Where it's like, okay, well, is it okay? And is it okay to kind of do that? How do we incorporate that? How do we think about these things where we can do some kind of back correction, even if the model's online, in a way that's non standard quote, unquote, for a numerical algorithm?

Um, you know, I think adjoint modelling is also a really interesting one, because, you know, it's hard to maintain an adjoint model, um, to a system, to do a lot of things that you would want to do, like parameter tuning, data assimilation. But there is an open question about whether the surgeons themselves can be used in adjoint, because she learned models are differentiable, or is there a place for some of these kind of, like, different types of computing that come with these advantages that allow us to just kind of take advantage of them. Maybe not as a whole system or in part of the system.
So, yeah, there's, there's, and this is, I think, again, that the emphasis that heterogeneity is something that we're facing immediately, this question of, what are we doing?

How do we combine all these different things and approaches into one thing that allows us to do better science and do better science faster?

Amber: Yeah. Really, I think, becoming, so thinking about it as a separate thing. Like, machine learning is this other thing. I think we just have to start thinking, this is a tool for numerical modelling.

Um, if there are no other questions, maybe we'll move on to the one slides, um, overviews that, uh, I'm not sure if very many were submitted, but we can, we can go through those, and if you don't mind, uh, sharing the screen.

Yeah, that's funny, Neil. Yeah, I can't replace your job, physical model, but an AI salesman can convince your boss that it can.
That is so true. That's problematic in government, too, but I won't say more on

that.
Uh, okay. Uh, this is this is my side.

Um, okay, so, uh, this is just a slide about a working group that I started, uh, over a year ago. Um, it's about machine learning for ocean applications. We actually started looking at statistical downscaling, but then we quickly realized that all these data driven methods that we're interested in for down scaling can a lot of them can be used for emulation, or there's overlap with bias correction, or imputation. And then unsupervised learning is something that we're all sort of interested in to. And so, those are the kinds of applications that we're interested in in that group.

And basically, this is a group of international experts from oceanography, atmospheric science, statistics, and ecology, and spans government, academia, NGOs. Um, and what we do is, we have a community. And we have seminars and discussion sessions, and we had a hackathon about imputation in the fall. And so the Hackathon was really just away, because, as Andrew was saying, these methods are advancing really rapidly, and it's tough to know what you should be using for what purpose.

And so the idea was to get together a bunch of technical experts and focus on a problem together, and see if different teams trying different methods could help us advance the science of this emerging field.

And so, um, we're actually continuing on with a collaboration and hoping to, uh, write a paper on the back of that. And I wanted to just point out that anybody is welcome to join this, and we're going to actually switch to having a, or try to. Anyway, to having a seminar coordinators committee, and so this could be a really nice opportunity for early career. This way, you'd have a few people kind of working on organizing things, and you could shape things in the direction that you want to go.

And yeah, so we're really interested in, um, continuing to have discussions that are, you know, preliminary, so people can present their, like, preliminary work, and then, and then get feedback, and then, as it advances, they can then present, um, the more, uh, finished work, but we also have people from atmospheric sciences, 'cause they're a little bit ahead of us.

And so they're presenting some of their more completed work. So it's a really nice community, and I invite you to reach out to me if you want to participate in it.

Um, yeah, my interest in it currently is, I'm really trying to push forward with downscaling efforts at DFO, and so I think it's really important that we are careful to evaluate our data-driven methods. And, in particular, this, in this context, this

would be, like, a perfect model test, where we use our dynamical models, um, as the truth, and try to test our machine learning models. Yeah, so that's all I had for you guys on my one side.

Oh, looks like François Roy, or is this you, Greg?

Greg: Yeah, I was gonna do this one. Um, yeah, François is not here. And Fred is here, my, I talk with a few editions, but, uh, it's giving you an idea, so, I mean, Andrew gave a fantastic review, I think, of the, kind of the lay of the land, and, uh, so it'll help you understand sort of how we fit into that.

So, obviously, CMEP, we have, you know, global to regional scale, operational systems, data similations, starting to do some supreme analysis, our atmospheric groups have started to, uh, to do some AI, we have systems going operational kind of imminently, where they have, uh, atmospheric emulator, and then the, the global deterministic prediction system, the full couple system, actually, is nudged towards the AI system.

So, in the ocean, we're trying to sort of replicate some of those sorts of ideas, we'll get different things. So on the global scale emulation, Um, so we've started looking at Mercator's global emulator, glown at. There's also the tool produced by European Weather Centre, Animoy, that we've started playing around with in the ocean. I think you may be working towards something like what they do in the atmospheres. We can use this, um, the spectral nudging GIOPS.

So, as Andrew pointed out, often these are set up with, you know, some sort of very large timestep, which isn't going to provide us the full range of products we need, and coupling the atmosphere. So we could use this sort of spectral nudging approach to the emulator.

There's also an activity that's going on internationally called Ocean Bench that's modelled off the weather bench, where people can then upload their own, um, their own simulations and have it kind of automatically compared, uh, to the standard set of observations, and different diagnostics, any tracking, different things. Um, and so we're gonna potentially participate in that as well.

Um, There's also, um, an effort to develop a GIOPS reanalysis. So this was something I've been thinking about for a few years, uh, to support the seasonal prediction system that needs a reanalysis, uh, to evaluate their kind of high historical hindcast. But then also for machine learning, it'll be very useful to have our own reanalysis, because the other methods I mentioned, GLONAD and ANEMOY, they'll all train on, uh, the Mercator GLORYS 12 or GLORYS 2b4, um, and so, obviously, having our own brain analysis would be better to avoid having to fine-tune or have alien analysis problems.

Um, so, actually, I've had CMOS presentation on that, if people are more interested in the details there.

Um, another thing we're starting to think about, um, sort of just out of the box is really just how we can use these emulators more broadly, so essentially as part of these simulation. So I enjoyed a slide, sort of top and bottom, and on the bottom is kind of a direct from observations approach. So something like that, that perhaps we could start, um, you know, playing around more in the latent space, um, and see how we could exploit that. And in particular, our ensembles seems like an area where we could get a lot of benefit out of the emulators.

On a regional scale, uh, so we have our global system, GIOPS, and the regional system, TIOPS, uh, with a couple part of that, is called caps. So we're working on an emulator for RIOPS, using Animoy. So a key challenge here is in terms of getting all the details of the ice edge. So there's a few different studies that have shown some skill in sea ice concentration emulation. But generally, all of the demonstrations show more of kind of a seasonal evolution of the ice and some anomalies, and it's very smooth. So certainly for RIOPS, we really want to have, like, all of the little details of the ice edge, so that'll be kind of the main challenge.

And also adding in the tide. So this is another thing we haven't seen a lot of in the community is, um, trying to count for that kind of high frequency title variability. So that's what the image on the top right is kind of speaking towards, was an effort of Francois deBois, in terms of developing an emulator for the Gulf of St. Lawrence, and eventually a kind of downscaling approach to 500 metres from our two kilometres, CIOPS system, and the plot you see in the background is actually showing the sub daily barriclinic variability, and you can see the importance there of the kind of the internal tides. And these sorts of things we're thinking about looking at how to include this and capture it well with the emulators.

Um, so, François has put together, uh, a couple of methods, one using a kind of an in-house, uh, LSTM, and then also comparing that to the 3 different methods available in animally. So, we presented this at this AI past team meeting that I mentioned earlier, a couple weeks ago, that some people might have been listening in on. But certainly, that's an interesting community for people who are interested in AI who want to get involved, and,let me know and I could link you up to it.

Um, so a few kind of high level issues in all of this, um, is that, you know, as we start to try to work on these sorts of emulators approaches, the data sets is really the key piece, is having, having high quality training data sets. So some studies have tried to make do with shorter data sets, and they tend to all result in kind of overfitting issues. So certainly having.. Having ample data sets and really thinking

carefully about the quality of the data sets and the sensitivity of the quality of the data set, too, then what you get out of the emulator, I think, is important. So I think it's the direction that we're gonna see more of in the reanalysis community.

Also thinking about how to exploit machine learning for applications beyond just emulators and downscaling and these sorts of things that we hear about a lot, but, so for example, for data assimilation problem, compression is another one, sort of paper on extreme compression. So, looking at ways to take our data and, you know, compress it, and then you basically just share the latent space. And then you decompress it when you need it. So there's sort of different ways of working, that's gonna be interesting to think about.

Um, the way I sort of mentioned this a little bit earlier, too, is about the physically based way that we evaluate these products. I mean, you know, it seems very likely that we're gonna be moving into an era where we kind of democratise the application of these things, people take data sets. It's very easy to train it up and have an emulator, but what's the skill of that emulator? Um, in order to capture different things. I mean, I know when I was being trained as a physical oceanographer, it was really, you know, taught to think about, you know, what processes in our model, that a model is just a certain representation of reality, and we have to, you know, stay within the bounds, when we're interpreting the results, and in terms of machine learning, it's less clear what those are. So certainly, you know, we need to think of different ways to evaluate these systems to be able to demonstrate exactly what they're doing, um, to have a better measure of, I don't know, reliability or, um, quality assurance.

Um, so another sort of point that they're talking a lot about in Europe, that, you know, we don't mention as much here, is really how to accommodate the need for, let's call them, integrated platforms, that are often called digital twins, but it's not a popular term here, um, because we are gonna see, I think, a diversity of machine learning applications, so people can train these emulators for specific applications. And, you know, as we have data sets that are more available, and, um, and so we start seeing these applications pop up. I mean, like, the way they also talk about in Europe is all, but these what if applications.

So if you have little emulators, then you can, you can ask different “what if” questions and run them out and look at how, how that would impact your solutions in ways that to do that for full sort of numerical models would be very competitionally heavy and maybe even impractical. So, um, so seeing how we need to kind of accommodate that, and, you know, is something that, uh, that we're thinking about anyway.
So maybe I'll stop there. I don’t know if, Fred, if you wanted to add anything.

Fred: On the, uh, well, just to mention that the, uh, we have Yosveni here working

on, uh, downscaling and, uh, but I don't know if he was able to participate in the, uh, the workshop you, uh, you set up AMber, uh, the last year, and the, uh, And so, anything of ED, so we, using more classical approach of uh, CNN competition, and uh, I guess with the, uh, the spicy, with the, the spicy beats is the, uh, is, is using aGAN, a generative adversarial approach. Which I haven't seen a lot to use, especially during the AI workshop we had in Montreal. I'm just curious to know that still, uh, still relevant to to people. Anyway, that just be, that's it, that's it.
Thanks, uh, thanks a lot. I think that's that covers what we do.

Greg: You could share one additional thing. So I had a really interesting discussion with Shalam ? and about this last week. And, you know, his comment, um, seemed very, very relevant, that a lot of people have been taking, you know, their data sets that they have available. So, some have a 30 year data set, and then you train out with a CNN or something. You know, look at what your emulator could do. But the problem is, you know, it does cost a fair amount of GPUs to do these trainings, and so you can't test that many things. Yet there are very, very many things to test.

So his approach was a little bit different than he started, rather, very simply, in terms of doing sort of 1D tests, where he can really explore the kind of full phase space of different parameters and aspects that, you know, would affect the results. And then he's building forward from there. And I thought that was kind of an interesting approach, in terms of people taking, you know, taking a step back from this rush to just take whatever data we have, and throw some system at it, and make an emulator without really understanding. The sensitivity to all the parameters. And start a little bit more simpler and then build forward. Because it is an interesting approach and when we're considered...

Amber: Yeah, I think that is an essential point, Greg, like, that we need to start with toy models.

It's kind of like, you know, in fluid dynamics. We don't start with the big model and test out our, you know, new parameterizations. We build a little toy model, right? And so I think we need to use the same approach here.

Did you have more to say, Fred? I see your hands up.

Fred: Right? Yeah, yeah. No, sorry.
I just wanted to mention that the, uh, at the end of a workshop, we had, uh, record sessions, and I, I think I, uh, have a conclusion I recovered from, uh, this breakup session, you see, uh, uh, is something about, well, it's close to what Ogre was saying, uh, the, uh, the explainability of the, uh, AI is something that was stressed by a lot of people, so, uh, and, and explainability of the results in general. So, yeah, exploring the latent space, from that point of view, is probably, yeah, the 1st to do.

Amber: Do folks know what it is meant by latent space, though. Maybe someone can take a second, 'cause I feel like there's a lot of people that are really unfamiliar with what we're talking about on the call. Um, I'm gonna put Andrew on the spot, too, since he's our expert, to explain what is a latent space.

Andrew: Uh, yeah, the 30-second explanation I gave was, like, the light and space is all the stuff that's kind of... When you're going from training, when you're training a machine learning model, right? You're trying to get it to understand just kind of some heuristic relationships. So, the core of that is what is known as latent space, which is like, okay, what is the reduced representation that the model has learned. And then, from there, you can derive other things, right?

So if you think of, um, you know, a very popular architecture is called UNET, right? And so you take a bunch of information, and then you're compressing it down, and then after you compress it down, then you re-expand it to your data set. So, you know, in some ways, you can think of it as, what is that reduced order dimension, without reduced order basis on which you have learned your relationships.

So, if you think about it in something like an EOF analysis, right? It could be, like, the EOFs would be, you relate in space representation. Of what your dynamic system actually is. And so, it's a little less explainable and much more convoluted in AI, because it's so much bigger, and it's nonlinear, but essentially, that's what that's the right framework to think about.

Amber: Thanks. Thanks. Yeah, did I just add one more comment on your slide there, um, very good to the next question.
It's just that you were talking about, uh, generative adversarial networks, Fred, and I would be remiss if I didn't plug Adam Monahan's, uh, group, he's my co collaborator on the machine learning, uh, working group, and his group has been developing, uh, generative adversarial networks for fire weather, and various atmospheric applications, and they've added stochasticity, and it is still very active. And we'd love to be able to apply to the ocean, but, yeah, having trouble getting funds to bring one of their experts over to do it for us. But hopefully on the list of things to do.

Um, and, uh, Hatem.
Your hand’s up. You have a question?

Hatem: Yes, I have, uh, questions concerning the machine learning and artificial intelligence. My first question is about the dimensionality of behaviour, and that's what is, uh, Andrew, just right now, explain it. If you take the database with a huge amount of variability to your inputs, So what you can do to recuce the dimensionality of those... Check, how can you do to check the highest direction

where you have the highest variability of combat, you can build your machine learning approach.

First question, second one, there is a lot of discussion about how to chew on the machine, the models or the hyperparameters to find the right learning or artificial intelligence processing, especially in the weather.

53:37

So, you have an idea about that, how you deal with that, because I think that's very important to understand, also, once you start you know, playing with...
And this question also for Greg who presented the work on using LSTN, long short term memories, YLSDN, in that case, why not another order?

Andrew: Thank you very much. Okay, I'll tackle question one, question two.

Um, and then, Greg, I'll, uh, I'm not gonna speak for you.

Question one. So, I'll actually show it to my slot when I kind of go over my research, my research interest here. Um, I think one of the things that's really interesting to think about is this question of how do we make the problem? And this is one of the big, like, kind of motivating research directions for me is, like, how do we make the problem simpler for machine learning problems to solve? Right?

We have the universal approximation theorem that says with, you know, large, the infinitely wide neural network, you can learn any function. Fantastic. That's not practical.

So, the question, I think, becomes how do we actually ask the question, like, how do we make this problem simpler?
Um, and so one of the things, so I have, um, my text proposal in, um, for post-doc work to actually ask this particular, very similar to what the question you're asking is like, how do we make sure that we capture not just the representation at the next prediction snapshot, but kind of like on long time skills. And so, like, essentially, can we respect both the spectral, spectral in time, and spectral in space aspects of the physical system by changing the training function.So, I think it is an open question, because right now, basically, every other modelling system in the world just says, I have a protection window of 3 hours, 6 hours, whatever. What does the hour between that look like versus the reanalysis, right? That is 99% of what people do.
And that is, you know, and that's great. And we've been able to, it's incredible to see how much we've actually been getting out of that. But I think there's more sophisticated ways that we as a community can explore on what are we doing to

train, and how do we make the problem easier?

And I think that that feeds into my answer to the 2nd question, in particular, right? Is that this question of hyperparameter, tuning, and selection, and exploration is the dark arts of machine learning in AI in general, right? I mean, like there are so many of these.

And for people who don't know what hyperparameters are. These are things that... These are things where it's, like, uh, it's essentially, like, when you tune a parameterisation, right? There's an uncertain range of, like, okay, like, observation, we see 0.1 to 0.2. Uh, the, the problem with hyperparameter tuning in machine learning is even worse than that because it's like, oh, well, maybe I need an extra layer here. Maybe I need an extra 10 layers here. Maybe I need a change. How quickly the model actually learns, right? I mean, like, it's all these things that are just completely unconstrained physically.
And they're just things you need twiddle in order to get a model that converges well.

Um, and so as part of the, so I think the bigger takeaway from that, right, is that hyperparameter optomisation is such a big problem that I don't think that we're ever going to get away from the need to do it. I think the thing that it ques us to do in combination with looking at architecture is looking at, the training methodologies, is how do we make these models smaller, so that we can train it faster, such that the hyper optimization space, or the hyper parameter optomisation space, is either easier to explore and or smoother to explore, so that they become less sensitive to these parameter choices.

Yeah, and then third, yeah, Greg, I'll leave it to you, Y, the LSTM versus a different type of architecture.

Greg: Yeah, so the idea, so in this particular context, we wanted to get the tides right. So it was very important to get the tidal part of the solution correct and so if we use with Anamoy, we had access to a graph neural network, graph transformer, things like that, so we could run them out with, like, a one hour time step, you know, auto aggressively, with, you know, some sort of a six hour rollout or something. But the problem is you see an error growth over the 48 hour forecast when you do that.
Whereas if we did it in one chunk with an LSTN, then we have, we have no error growth. So that was a thought. So it was interesting when you see the results. So for some things, the error growth becomes more important, and for other things, um, than the better treatment of the bias and whatnot becomes important.
So.

Hatem: I would add something else, because I thought that from your answer, you

would look to the, because you're looking for work and memories back to the many, I don't know what is your database inside here, or 2 or 10, 15, and you try to link between those information that's from really backward, and try to see what kind of prediction you will do, based on that memory that the model itself is. And that's the main information that is behind the LSTN, and it definitely uses differently of what you have as a database.
From another database, you can use another model, another numerical model or artificial intelligence one.But, that's answer my questions.

And for the dimensionality of your database, and think that the most important things in artificial intelligence processing, and any, different kind of philosophy behind that, um, mind, particularly since I used to work a little bit, but that is to work a little bit on the principal component analysis, which is a narchaic methods, but it is very broad place in terms of, give you the highest PC. So you can, from that highest PC, you can develop more and more. That's a way of thinking. Definitely, there is other different ways. So, you can have the highest variability in your access of data, and from that, you develop your predictions, and minimize the act of the hyper kilometres, that's it.

Andrew: Thank you very much. Yeah, so that's actually, I mean, so it's a very similar approach to what we proposed.So we're, instead of principal component analysis, we're looking at two different types of, similar in spirit, right?You want to do the dimensional reduction part of it.So, one of them is dynamic mode decomposition, which is essentially like, you know, is a variant of PCA where you allow the modes to actually change. Somewhat in time and ordering other variants, change in time. Um, and then the other one that we're taking a look at is spectral proper orthogonal decomposition, so that respects just explicitly, right? A decomposition of the spatio or the temporal space of the basis vectors themselves, and, like, you know, and so that's essentially what we're trying to do, is to understand whether we, um, you know, whether by changing the basis, so that it's, you know, that it's incentivized to learn kind of like the eigenvalues and the projections onto these basis vectors. If the projections are more accurate, as opposed to specific, like point-by-point accuracy that is essentially being done today, like I said, mostly applications.

Yeah. So that's one of my taxes. So, I hope I'm not influencing any of my text reviewers on this call, but that's exactly what it is on, you know?

Hatem: Okay, thank you very much.

Amber: Great. Um, just again, because I think a lot of people probably are really new to the subject area. I wonder if, Andrew, you could just talk for a second about just more generally, like, what is an architecture?
What is a model in this context?

Andrew: Yeah, and I think I would encourage, like I'm happy to kind of get a level set on the terminology here a little bit. I would also highly encourage that we, for folks who really wanted to get that primer, machine learning methodology and terminology, that the early career workshop tomorrow. There is a lecture component, and Kirsten Meyer from NCAR is gonna be leading it, and it is an excellent, excellent overview of, like, all these terms, because I do not, I cannot tell, the first time that I got exposed to A&ML, I was literally in a room with, like, machine learning engineers and being like, I have no idea what you're saying. Like, you need to break this down for me. Like, this does not make any sense.
Like, you know, like, you're using terms in ways that I think you're abusing the term and all these kinds of things, and, you know, it's like learning a different language, right? That it's, um, It is just one of these things where you kind of had to go through the definitions. Maybe some word is inappropriate, saying one language was totally appropriate, saying different, like, same thing, right? And so there is that barrier.

Um, I think a very, very high level, though, what I want to talk about, uh, in terms of just giving everyone a basis of things, is that the architecture of an AIML model is essentially this question of, like, what is in the internals? What is the structure of the model?
So, and by that, I mean that there's kind of 2 things. There's both, what are the operations that are being done inside a layer? So, a model is made up of a number of layers, and also then one of the weights associated with those free parameters, because that's really what's being done when you're training an AI model.
So very, very simplistically, you know, you could, a linear regression model is a machine learning model. Right? It has two free parameters, and it has a structure; it has the equation of a line. So, we would say, in equivalent terms, that the architecture is the equation of a line, MX plus B. And then the weights are M and B, right? And so the whole goal on your regression is to find the, um, you have your data and you're trying to fit align to it.

So then we also talk about things called the loss function, right? So this is the other fundamental part of doing model training. I don't have machine learning models. What is the loss function? And by that, we mean, what is the measure of what we are saying is the accuracy of a prediction versus what the original data set actually had inside of it? So equivalently, if you're doing least squares regression? Strictly just x minus x minus x prime, squared, summed over all the data points.
That's a linear aggression. Again, linear aggression is a machine learning model where your loss function, your architecture is the equation of the line, your loss function is just, mean squared, or squared error. And that's means square error, and that's all that that is.

So, um, and then when we talk about parameters, we're talking, we typically are talking about what are the free parameters in this model, so, like, that M and B, we were talking about hyperparameters, we might be saying things like, um, and that, like, the loss function itself is, is technically could fall under a hyperparameter as well. Like, what is the form of how we measure the accuracy is, um, There are things in there too. Like if you multi-objective loss functions, right? You have like 2 things you want to compare. The relative weights of those is another hyperparameter. So, hyper parameter is basically anything that exists outside of the model architecture itself, that you could conceivably twiddle, um, with the one grey area then being that you can also consider how to modify the internal architecture as a hyperparameter, like how many internal layers are there. What are the types of those internal layers, how big are those internal layers, of the model?
So those are all things that are part of hyperparameter tuning.

And when we say training a model, unlike Least Squares' regression, where we know how to exactly get to the right answer immediately. Um, all these networks, when we talk about AIML are nonlinear. So when you're trying to optimize or reduce the value of that loss function, you just require, like, many, many iterations to solve the problem, because it's not, you know, you can't one shot, the solution, to the optimal set of things. So, training is finding the optimal set of weights or parameters, equivalent terms, given the structure or the architecture of your neural network that gives you the best match from your target data set, and that's the loss function.

Amber: Thanks, Andrew. Um, so, like, hyperparameters aren't always trained, but sometimes they're set.

Andrew: Yeah, they're looped over, I think, is probably the right way to do it. You can intelligently loop over them, but, or maybe more intelligently over them, but, yeah, you're basically just looping over a bunch of values, yeah.

Amber: Okay, great. Um, I think it makes sense to go to the next, like, you have a side, too, right, Andrew?

Andrew: Yeah, I do. Yeah, let me share mine, and I'll be a little brief since I've eluded some of it. Uh...

Amber: Yeah, and then, um, after we talk about that, then I think we can just turn it to people if they, I don't know if other people had slides, I don't think so. Uh, you don't have anymore? Yeah, um, but if other people wanted to chime in about things they're interested in, um, we can do that, too.

Andrew: Uh, okay, yeah. So, uh, I'll give my my brief overview. Uh, first thing, so, yeah, I, um, I think my title is actually AR HBC research scientist. Um, at HPE Labs. So HPE Labs is a pure R&D organization at Hewlett-Packard Enterprise.
Uh, Hewlett-Packard Enterprise does a lot of things. We are not the printer side of it, the printer and the laptop side of it, so please do not ask me for printer support. I do have a HP laptop that I own, but that's completely different. Um, they stem from different parts, like, you know, there was a split, basically, between Hewlett-Packard, and one became Hewlett-Packard, and one became Hewlett-Packard Enterprise.

Salient to this conversation, though, HPE provides a lot of supercomputing. They're a supercomputing vendor. So we provide, um, we will probably bid on the ECCC next machine. Uh, on this, but we also do things for, like, you know, provide supercomputers for the Department of Energy, for the European, uh, leadership class, computing centres, a bunch of weather centres. Um, and so that's, and labs is an RN Pure R&D division side of it. We behave very much like an academic institution.
We host PhDs. We fund postdocs, we serve as COPIs and grants. So if you have any interest in any of those things, you can contact me, and we can see if there is an opportunity that we can work together on.

The 3 aspects of things that I work on particular, um, are this question of what numerical techniques can be applied to AI surrogates, uh, dynamics systems, and dynamic systems being a very broad umbrella term for anything that, you know, changes in time. You know, specifically, obviously, for me, I focus more on physics-based things or science-based ones. But I think there is a lot of interesting things to be asked in the area, just in general. Like, if you have a changing system, what can, what are the limitations and what are the opportunities for AI circuit modelings?

The second one is more about the benefits of blending AI methods and simulation. How do we do that to open in new areas of research, like I learned it to?

And then third, because I work at a tech company, the software engineering part of this is how do we make it? How do we develop software, especially in the open source space, where we can, uh, help people do these kinds of applications, and how is it actually changing our view of what computing is, and what are the machines and what are the types of computing platforms that people will need go in the future?

Um, I have 3 kind of broad things that map on roughly to those research areas down below.

So one of them is, we did a turbulence mesoscale turbulence parameterization that we embedded inside of the MOM 6 ocean model, and we ran, um, a full kind of OMIP style simulation using, um, using this neural network embedded into MOMS 6. And ran it out and showed that we can get better predictions of Eddie kinetic energy, which then gets fed into kind of the Genten McWilliams parameterisations inside of that, um, and just get slightly better, or get much better representations of the Eddie kinetic energy. And so some of that, you know, parameterised transport that's associated with it.

Um, there was interesting, 'cause we did it where Moms 6 was running CPU, and then we had, basically, set up an inference server type setup, so that all of the GPU predictions, or all the machine learning predictions, were done on a different, uh, inference on a different server itself.

The third one is a reinforcement learning based thing, and I don't want to go too much into it. But basically, instead of just doing the normal thing where, or doing a regression type problem where we say, like, we have a lot of data, let's just map the relationships. Reinforcement learning a lot as you kind of train a machine learning network, a neural network, on the emergent properties of what the network might do.

So, for instance, on the left-hand side, we just had a bunch of, like, any result data, that we were training on in the course and variables, and then tried to relate the eddy terms. In the middle one, we essentially just said, we have no idea what it is. All we want to know is that we want to make the kinetic energy spectra better. And how you do that, that's what you do. And so that's another technique that we showed that we can actually be quite successful at.

And then the 3rd one is, um, this is actually one where we're using machine learning techniques to try to optimise, uh, how a wind turbine would move, if, you know, field of wind turbine. So, like, how can you extract the most amount of energy from the windfield, um, when you have a farm in them, and you can rotate your, your wind turbines a little bit.

I have a number of ongoing and active projects right now, so the one that I was talking about with Hatem. That is one of the ones that I've got going on with Uvic. There's also another company that I'm working with here out of Bellevue, to kind of look at, again, like this problem of, what do we mean when we're training circuit models? How do we get better circuit models of the Earth system?

Um, letting AI, AI simulation, like I said, this is the kind of act of learning, different types of machine learning applications, um, that we kind of wrap around simulations to figure out, you know, what are the best parameters, what are the

best tuning parameters for a simulation, what are the, what are the new areas that we might explore, because it's, we've never really, like, put them, we've never done it in this particular part of the dynamic space, what would be the most interesting simulations to run there.

And then, yeah, lastly, is this open source software for AI and simulations. Um, I'm working with the Department of Energy to kind of build this HPC-AI work flow tooling, um, that's gonna be run and be the solution that a lot of people at Department of Energy use, and also, we have some interest from the folks in Europe, um, to do that, again, open source, science, not gonna sell you anything, I promise.

Um, and then, lastly, I also have another project going on where we're looking at all these different, uh, software packages that are out there now to connect fortran models, like the ones that Nemo, Nemo, mom six, you know, a bunch of our climate models are all written in fortran. How do you actually connect those to AI workloads? So this is a really interesting collaboration between UC Berkeley, the Department of Energy, NVIDA, Cambridge University, UC Met Office, to kind of just give the overview and understand why we made the choices we did. Which ones might be better for certain situations, which ones are easiest to use if you're spinning up, and which ones give you more advanced capability?

So we're trying to write a survey paper, about that. That should be, well, ambitiously, in a month, probably more likely by end of summer, but would be happy to kind of share initial learnings with whomever is interested.

Amber: Oh, that sounds really interesting. I'd like to see that paper. Uh, yeah, do folks have any comments about Andrew Slide?
Yeah, go ahead.

Hatem: Uh... I just saw the DNS, the director, direct night, and the numerical simulations that you're using for actually, definitely modelling the turbulence, right?At the surface layer.

Andrew: So, this was actually a 3D box of turbulence. So this is more in the CFD realm.

Hatem: Yeah, not, not, not in the, not in air sciences realm. Yeah.

But in general, my question is a little bit in general. Definitely, if you, if you use a numerical morning processing, or, uh, using the artificial intelligence machine learning, you need a database, and from that database, you can predict, okay? Black box.

Andrew: Yep, yep.

Hatem: The problem that I'm seeing right now, in the artificial intelligence and machine learning, that the outputs of the prediction won't provide you a resolution better than the database that is giving you. I mean, if you're in a scale of metres, the results will be in a scale of metres, not less in that. It's not making the resolution outputs very better than they want. We're not doing downscanning on anything like that, not physically.

So, in the turbulence case, how, from where you get database?

Andrew: So this is where we use DNS numerical simulation.

Hatem; Oh, yes, so it's a database as a numerical simulations.

Andrew: Yes, yes.

Hatem: So you apply your modelling, artificial intelligence on numerical stimulation, right?

Andrrer: Yes And because of CNS, right, or at least at the relevant scales of turbulence, for, again, this was just 3D box of turbulence, so a very basic problem, but, you know, and then just asked the question about, like, how can we do better if we then coarsened it to what would be cannot be considered a large eddy simulation case.

And what would a parameterisation of that look like? So...

Hatem: Okay, so the court, the oral physics and fluids, uh, 224, that is a comparison about the model, with the machine learning, and the other DNS, and other things. That's the benchmark. That's the benchmark.

Andrew: Yeah, so the DNS is described in there, and then they approach this reinforcement learning approach that we took, and the results of that parameterzation are all discussed in that paper. Okay, so if you haven't missed less accuracy, even though that DNS, it's very high accuracy. Uh, the prediction using the artificial intelligence machine learning would be affected, but that accuracy of

Andrew: Correct, yeah, yeah, yeah, yeah. And that's, I think, you know, and making it correct.

Hatem:Okay.

Andrew: Greg, you can comment on this a little bit more, um, and like Amber, you know, your knowledge of what's going on in the generative area stuff, whether what Adam's doing could kind of come into this as well. But, you know, we, like, all of us who work with simulations know that simulations are wrong. So, you know, it's that question of like, okay, like, are we able to, um, basically use what we know was true, you know, let's say observations and fix that internally, right? So that's kind of like what Greg was referring to in the data simulation side of it, too, right?

There is kind of this, even in the pure numerical modelling side of the things, there is this question about whether, if we inject physics, physical understanding into the neural networks themselves in some way, or if we have true, like, observations, can we actually use those in combination with the surrogate models, which we know are themselves in accurate representations of an inaccurate system, can we then generate a better result overall at the end of the day? Right?

So I think that those are the kind of tricks and games that we're also trying to play as well. With AI.

Hatem: Okay. Thanks.

Amber: Is there anybody else that had a slide they wanted to share, or they just wanted to chime in and talk about, uh, some of the work they're planning, or, um, now's the time, put up your hand if you got something else you wanted to, um, to bring up? Otherwise, I have another little topic that will shift the gears a little bit here. Um, which I'm curious, Andrew, if you have thoughts about witches.

Oh, before I do. Paul.

Paul: I don't have a slide. Um, I focussed too much on my session and didn't think about this. This is not an area of super expertise in our group. I guess in terms of exploring, um, you know, there's been some interesting works where people have used machine learning to categorise pathways, Lagrangian trajectories, like, how does water get mixed off the Labrador current?Where are there exchanges out of the Beaufort gyre in that? So I guess we're going to be starting, um, some students. We're going to be looking at, for example, at the riverrine coastal domain in Hudson Bay.

And if you put Lagrange and particles in the very high religion model, and then you get a pile of very complicated, large number of particles, can the machine learning be used to then determine, are there certain types, categories of exchanges, seasonally, spatially, from where you get to change into the interior, system, offshore, things like that, and probably also do that in the Arctic and to theLabrador current. So it's very, going to be very basic, and it's going to be

learning for us, but we're definitely wanting to explore that space.

The other space that I think could be interesting, especially with the very resolution models, how do you evaluate them with the good old sparse, irregular in space and time, observational data sets, like Argo floats, or something? And can you get a better understanding of what those models that are going into the mesoscale versus the submesoscale are potentially representing in terms of the observed structure? So, you know, they're new to ongoing work, but it's stuff, I guess, in this AIML space, I'd be interested in understanding and seeing, you know, the value of those approaches to those types of questions.

Andrew: I'm gonna put you on the spot, Paul, because I think you open up a very nice can of worms, because I'm talking mostly, and, you know, I think the people who are presenting, I've been talking mostly about, you know, kind of applications, that mimic simulation, and you're talking about AI analysis and ML techniques for analysis thereof. So, I would be really curious and putting you on the spot to say, like, what are the main challenges that you are facing in that space?
Is it adaptation of technique? Is it development of technique? Is it practical things? Is it base knowledge? Like, yeah, really curious.

Paul: I could start by so to say, all of those.

Andrew: Yes.

Paul: I, well, I'm also gonna, you know, I, you know, I'm traditional physical oceanographer, a numerical modeller, who's learning AI on the fly now. I've definitely, my students have better knowledge than I did, and so there's a lot of background knowledge, but, you know, I'm really thinking, you know, when you've got very high resolution models that have huge amounts of output, then you put in, say, complicated trajectories, you've got large data sets, um, lots of detail. You know, the old fashioned, integrate, but, you know, looking by eye for patterns, um, trying to figure out, you know, similarities between it.
I think it's just, you know, for me, it's an exploration question. You know, I'm an academic at a university. Can we better understand these applications? Um, and I think from everything I've looked, talked to people who got to seminars, there's some good approaches that will likely help us analyze these very large, complicated data sets.

So that's the space I'd like to get into, and I know students are interested because they, you know, they want to learn more things in those spaces, applications. So, I'm very open to how best to do it, integration, collaboration, but, and I don't think it's just me. There's many people who are, like, like, uh, was mentioned, like, at the last DRAKKAR meeting, there's now been sessions on, again, the diagnostics, the AI, the applications. So I think, you know, that society, I've been interested in as

much as the emulators, the forecasting, the support of the modelling, um, because I think, again, it's a, you know, it's another tool that can help us answer the science questions.

Andrew: Mm hmm. Hatem, you wanna respond to that?

Hatem: Absolutely, since he's there, talk about the lagrangian and drifting, that's what we're doing every day, and, as I said, before, we're working with the speed view, which is using all the Canadian port models to derive and calculate the derivatives, to search for the best trajectories, if there is any spoil anyway, in the Canadian water. And I'm talking about about 17 domains, the name in Vancouver, Harbour, 20 metres, Kitty Matt, 100 and 500, All those models were downloading them every day, and using them with an open drift to model this OceanDrift from OpenDrift. And there is one part that is very important, that we can integrate machine learning, and artificial intelligence in it. It's the one that is very pressure for us, the deffusivity coefficients, and definitely, to get the right the trajectories, the drifting machine, is to have an information, very crucial and accurate knowledge about the diffusivity in the area, and unfortunately, for the moment, we have to play a little bit with this cloud of points that to make sure about the step, after steps, simulations go well with the trajectory. So instead of having one point, we have a thousand, found some groups making that using Monte Carlo simulations with 5,000, 10,000 points. And do the average of trajectories at the end of the day.

But where I've seen, and I'm talking about that, and also collaboration for that, is how can I have the right ensemble of parameters to get the right and good trajectories? And the only way to do that is to, unfortunately, have an evaluation with observation, because we don't know, and now we don't have it. But the more we evaluate your transjectories with the observation that drifters launch it in the water and do simulation compared to that. Once we define the database with this observation and simulations, you can create, with a machine learning, the right combination of parameters, to get the right prediction. And not the number that you need, for the future, to start doing the protections, with setting the right parameters for your sediction, among those parameters, sometimes about the diffusivity coefficients. I've seen it in my case where we can integrate the artificial intelligence or machine learning into your drifting calculations.

Caio: Hey, um, I'm one of Paul's students, I'll be working with lagrangian trajectories, um, and one thing that I, that I always thought is, um, sometimes I don't know, um, how to select a, uh, um, a correct, like, algorithm to, to my research. Um, sometimes I think about uh, what's the correct algorithim and how to select uh, a good algorithm because I was looking into some papers thatPaul mentioned about the great interjectors into the Labradorsea and they use uh, Cummings plus plus. Uh, but I was wondering, um, how to select the best one, um,

in in this aspect. Yeah.

Hatem: Absolutely. It's a good question, so there is no universal formula for selecting which kind of controrism or machine learning the moment that fits dependently of what you have as a database. If it's a long, very long gradation, you have enough information, it's diversified as a variables inside. Is it spread, homogeneous, or just focusing on one direction? And that's why I explain it before, if it's possible, to really dimension the database into one direction, where you have the highest variability. And from that point, you can understand which model you can, because each model in machine learning processing comes with the advantage and drawbacks, and you have to, you know, a little bit play with that. And this is gonna be only learned by the claims.

The more you deal with different models and get the result, benchmark and models, The more you get more clear knowledge about which one you use it for which case specifically. And to answer your question, which way, which I don't know, which model I have to use on that, so definitely, you have to try them all, you see, in my case, for example, on benchmarking all different kinds of models, and try to have a plot at the end of the day to see the variabilities between them, which one is for highest, and the best case scenario to get the observation with that. And from that day, from that process, you can understand which one's closer to the observation and quantify the error of your models in terms of calculation. Be careful, but overfeitting. Sometimes you get 100% correlation, that's not a good idea.
Sometimes.

Caio: Yeah. Thank you. Yeah, it's...

Hatem: It's, it's not good. If you feed 60, 70 percent, okay, you may have good parameters, just doing that, that's it.But, uh, try, try the simple ways like that. It's very simple. Just an excel 5, with different columns, the first one, three, three, four, five adarts, your parameters, the hand columns, it's not whatever result you want. If it's good, it's zero, it's one.
If it's bad, it's what is your, like, classification process. And apply your AI model, and see which, they process how it works, you know? And after that, it's possible if you have an observation, that you know what is zero, it's really zero, and one is very one. And compare that, and from that point, you know, start brick, brick, don't, don't start by big things. Start one step at a time, and each time, and that's the key element is to succeed with a machine learning or AIto evaluate your process of simulation step by step.

Each step, you don't go to another step, unless you're sure about the first step, to start by the small coding processing, and after that, add information, a little bit, until you get the final result, that is, you, first of all, be convinced about that result.

But that's the way that our approach will be the control machine.

Caio: Thank you, thank you.

Hatem: Welcome.

Amber: Yeah, that seems like a really nice systematic approach.

Hatem: Thank you.

Amber: Anyone else? Okay. There's something in the chat. Uh, from, uh, Damien. Um...

Amber: Ah, okay.

Damien: Did you wear...? My camera doesn't seem to work. I don't know what's going on.

Andrew: Oh, I see you, I see you. Yeah.

Damien: Oh, I don't see myself, that's why. Uh, okay.

Andrew: You're perfectly in frame, so you're doing great.

Damien: Yeah, I just did a few, a few thoughts in a chat because that's something I have been thinking about for some time, especially with a clique of mine when I was at McGill.

Um, But it's like, it's like, I feel like one of the big challenges, like, also, like, we have so much to catch up, you know, feel like I don't, like, it's so much to learn and you, like, feel like, I've seen some papers on physics, informed, neural networks, but, like, oh, like, yeah, it's, um, yeah, um, I would love to use, use it more, but for now, it's, it's been like a bit like a challenge to really understand what to, what to do. Um, for, for sea ice dynamic, sorry, I, yeah.
Um...

Andrew: So, I... Go ahead.

Damien: Go for it, go for it. Sorry.

Andrew: Yeah, no, so I think that it's also just, again, to kind of level set some terminology. So physical form neural networks are a very big catchall for a lot of terms, right? So there are things, again, in the space where we talk about, okay, like, what does it mean to be a physics form neural network? Sometimes it's loss

function, so it can do things like the energy conservation, right? Energy momentum, those kind of things. That is technically a physics formal network, 'cause you're ejecting some kind of physics information in the training process.

And then there's also other classes of things that we call physics and form neural networks, which, um, you know, again, I would even say, you know, for something like Greg was talking about early, like an LSTM, right? What is the correct architecture to solve? Or what is the correct part of the architecture to solve a specific type problem
Because, you know, like causality is a thing? Uh, you know, and for your particular parts of it, I think there's also yet another class of physics and form networks that are basically taking advantage of the fact that neural networks are differentiable. So if you have a PDE, right? You can actually directly calculate the derivatives using the PD. So you have, if you consider a network as, like, phi of X, right? Yeah, not to be differentiable.

You can do Dphi of X, DX, Dphi square feet, DX squared. And so, in some ways, and I see this in the CFD domain, you know, where it's much more natural express something, as, I know, like, let's say, heat flux, right? I don't know how to calculate a heat flux, but I know that what the temperature is at the interface, I know what the temperature is somewhere in the intermediate part of the fluid, right? So, I can actually train a neural network on the heat advection equation, right? But knowing that it needs to satisfy these things.

So, that might be an area that you could start going down, and I would say, like, the PDE based, or PDEs, physics informed neural networks for solving PDEs, might be an interesting area for you to kind of go down.

Damien: Yeah, because we have a representation of stress, then we take actually a div of this stress tenser to set our momentum equation. I'm going to write down PDE.
Thank you.

Amber: Thanks. Uh, yeah, there was, uh, something that came up in our Hackathon work for amputation. It was this really neat paper, SST ODE. Is, like, a new, I could share the paper. Um, it's about, um, like, pretty much exactly what Andrew is talking about where we, we have a neural network, and it has some physics components in it, and it's, you know, part of a larger structure. But, like, I think there's a lot of promise. I hadn't really seen anything in the physics and form neural networks world that was relevant enough to, like, downscaling, the stuff I wanted to do, to kind of glom onto it, and, uh, but this, I think, is, can be really, uh, nicely adapted. So I'll share that paper.
It was Xiang 2026.

Um, yeah, so, I think this has been a really great discussion, and it's really been about physics, and so, I just wanted to just give a nod to the session that's gonna happen, and the part two about biogeochemistry, and sort of go, like, yeah, okay, so, for when we think about these neural networks, and we think about, I don't know, even doing climate projections with, um, machine learning, We think, okay, the dynamical models are superior, because they have the physics built in, we have the Navier Stokes equations, and we have thermodynamics. And so, even if it's not in our training set, we don't have a training set with dynamics, right? In the future, we think we're gonna be able to do well 'cause we have the physics.

But what about the biogeochemistry, right? Because in our current models, these are really empirically derived relationships. We don't know that there's gonna be stationarity into the future in our representations that we have in dynamical models. And so I guess what I'm wondering is, what is the future for biogeochemistry with machine learning?
Like, how does it help us here? And I'm curious, Andrew, if anybody else knows what's happening in this space.

Andrew: Uh uh. Um, so, I think, I think one of the things that's kind of interesting.

I think you, you, the, at least from what I've been hearing, and it's not strictly in biogemistry here, in ocean stuff, is the question of, like, what is the most natural plate, way to actually model biological interactions, right? Because I think a lot of the, both in terms of the chemistry and biology, a lot of it ends up being like rate equations, right? And, like, just, like, what are the predator prey relationships, the consumption production rates? And it's just a lot of rate equations.
And, you know.

Amber: Rates and pools, right? That's biogeochemical modelling.

Andrew: Yeah, that's that's biogeochemcial modelling. And like, you know, obviously, we've been able to get a lot of stuff out of that. And it goes down to 1st principles of how the chemical reactions work. I think that there's the larger question to be asked. Um, not machine learning related, but I think it's been interesting, is that people have been moving to, uh, like probability distributions, right, um, for as the basis for some of these things.

So instead of having a small, medium and large class, right? You just have a continuum with some measures of what the distribution are, and involving those in time in a physical way. I think that becomes a really challenging problem to understand how we do that from first principles, right? And being like, yes, we understand how this, how advection affects a probability distribution of things.

And I kind of, and this is pure speculation on my part, but that's where I wonder if there isn't a little bit of way of introducing some of the heuristic kind of understanding that, a, um, machine learning models give you, right? Like, okay, we know that, you know, we have an underlying distribution, we know that's being transformed. We don't know what the, like what Fred was alluding to. We don't know what the first principle of this equations are. We just know, here's a distribution and involves like this. Can you tell us how distributions evolve, right? And use that as the operator.

So, that would be something that I would say is potentially an interesting thing to think about is that question of, like, what are the natural ways of modelling the system, that we've had to compromise because we don't have this first principles based approach to it, that we then don't, like, that we, or are limited by our knowledge of what appropriate numerical analysis techniques are out there to solve.

Amber:Thanks, Andrew. Uh, Jim has to stand up. Go ahead, Jim.

Jim: Yeah, sorry, I had my, I had my thing muted at the computer. I had to double, double muted. Sorry.

Um, so I think, I mean, maybe what I'm saying is similar to what Andrew said, um... But it seems to me that ocean biogeochemistry is actually something that, in principle, machine learning could do better than our process-based models. Um... So, but, I mean, the question is, how do you use the observations to constrain, uh, the learning or the training process, right?

So, I mean, if you're thinking of something like, you know, chlorophyl, okay, we observe chlorophyl. So, you know, you potentially have profiles of chlorophyl, from Argo floats, you have satellite, ocean colour products that can resolve the mesoscale variability. Um, and then, you know, you have a model that has, okay, let's say you got at least three processes. You have, you know, photosynthesis, you got production, new phytoplank, and you've got law terms like grazing, and then you have the photoacclimation of the carbon:nitrogen:chlorophyl ratios in the cells. And, you know, our process-based models of all those processes are quite imperfect.

So, can we do better with machine learning? Well, I think it depends on how you define, you know, again, how the observation constrain, what the machine learning model is able to learn or to come up with.

Amber: Thanks, Jim. Hatem.

Hatem: I absolutely agree with Jim, and thank you very much to do that, because my example right now is from your biotechnology in general, okay? You have a process that you observe it in the lab, and you create your profile. Whatever, CO2 profile, methane profile, from bacteria going on inside the reactors, whether, and that profile, it's obsered, and you did that many times, in past experience, or chlorified, whatever we have as a result, in terms of profile.

So, the second thing is to find an empirical model that fit your profile, and that empirical model would have some parameters that you can't see in your profile, but it helps you understand the process, the observations. For example, the kinematics, the half saturations, the growth rate of your biomarks, whatever. And I think it's very expensive to determine them with a lot.
It takes a lot of time to do that, but you've done a lot of experience doing many inputs and variations. And here, well, and here, where the artificial intelligence would help you understand the process, because you built organisation, your database, that once you apply your machine learning on that, you define with a hyperparameter, the right combination to get the good profile that you like it, or the result that you like it.

And in the future, if you have the same conditions, with the pH, for example, viscosity, whatever conditions, as the parameters, you plot them into your machine, and they have already defined a bit hyperparameter, and will tell you you will get the same profile that you like it before or not. So it's like repeating the process in front of the past from your data. That's the only way that machine learning can be applied in such conditions.

So, to build on one, you observe it before, when you said before, yes, you had this couple of affirmation and observations, you define it a profile, or a result that you like it, or you don't like, whatever, it's a concept of things, you like that things, but when my profile would be repeated, the next time, with the same kinds of parameter, because you can't have the same observation, once again, in your lab, or whatever, the ocean, or devil in things. So, if you are monitoring the parameters in the future, like PH temperature, salinity of water, viscosity, you name it, you put them in your combination, apply the same models, but the condition should be the same, right? You can't make artificial intelligent on different conditions, should be the same environment. So you will have a good guess, or probability, it's always under the probability to get the same profile at the 80% or 85%, as you said it yourself. And that's the way where you can apply your machine in processing on the system.
And I'm doing that many times.

I have a PhD in bioprocess engineering. So I did that many times on my reactors. For example, I had a different reaction during my time. From bacterial

transformation, and I know the process, and the bacteria inside, there is different characters.
The metalogenesis, acidogenesis, different kind of bacteria, it's not possible to understand them, and sequencing them, and determine them. But I need to have at the end of the day, a profile that I wanted for my reactions. So, I create that in the observations, and apply my empirical model. I took the parameters that is very important for me. I built my database, because I have different scenarios, and in the future, when I'm in the condition where my PH temperatyre, your whatever volatile fatty acids. So they have things inside my reactors, I would be more than 80% sure that I would have the same, what we call BMP, biomethyle potential. And this is the same things with the CO2, the salinities, things that is a relation with the biotechnology inside the ocean.
I think, I hope that is...

Amber: Thanks, Hatem. Um.. Any other comments about that? We've got five minutes left, so, um, if anybody has some last things they wanted to bring up, now's your chance.

All right. Well, I think with that, it's been a really great discussion session. Thank you, Inge, for inviting us to help out with the discussion. I think it's been really great, and thanks to everybody who participated.And I'll turn it back to you to have the final word there.
Oh, and thanks to Andrew, it was great talk.

Inge: Yes, thank you so much, Amber, and Andrew, for the discussion, the AI discussion, and then also Paul and Nicolas for this morning's discussion as well.

Um, Uh, yeah, I, it's, it's been a long 4 hours. Um, I think for everybody.
Um, everybody is welcome. It is the early career day tomorrow, starting at the same time, but if you're keen to learn, please pop in.Um, and then also, please come back on Wednesday for the 2nd session. It will be the bigeochemical modelling talk, uh, and discussions that will be hosted by Elise Olsen and I, and then in the afternoon, it will be the, um, uh, collaboration discussion hosted by Natasha and Hold on, my brain will come back to me.

Natasha: Heather Andres.

Inge: Thank you. Thank you, Natasha. Um, um, for that. So thank you very much, and I hope you all have a wonderful afternoon. Or evening, wherever you are, and then, Andrew, see you tomorrow again. But thank you so much for coming and joining.

Bye, everybody.