Episode Book

EPISODE #6274

Machine Learning for Smells, Whale Codas, and Aftershocks

PUBLISHED	25 October 2026 at 18:35 UTC
DURATION	20:20
LISTEN	myweirdprompts.com/episode/unusual-machine-learning-tasks-76098/
HOSTS	Corn, Herman, Dolores, Hilbert
SCRIPT MODEL	anthropic/claude-sonnet-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.22
BOOK GENERATED	26 October 2026 at 06:38 UTC

SYNOPSIS

Beyond standard text generation and image classification lies a frontier of unconventional machine learning tasks driven by bizarre datasets. Predicting odour profiles directly from molecular graphs challenges human sensory panels. Combinatorial analysis of sperm whale coda recordings reveals intricate rhythm, tempo, and ornamentation without translating a word. Meanwhile, neural networks mapping earthquake stress changes outperform classical physics benchmarks to forecast aftershock locations. None of these tasks rely on scraped internet text, requiring bespoke data, custom formulations, and novel scoring rules.

TRANSCRIPT



Corn

I'm still a bit sore from the time series episode, if I'm honest. Half an hour on Mistral's Timestral model and what it means to forecast the next number in a row of numbers. Good episode. My brain felt ironed.



Herman

It was dense. I enjoyed every minute and I accept that I was the only one.



Corn

Daniel's prompt this time is the antidote. Clem Delangue, the Hugging Face CEO, said last November that we're in an LLM bubble and it might burst. But the part Daniel latched onto was the other bit: LLM is just a subset of AI once you get into biology, chemistry, audio, video, and you'll see a multiplicity of models. So he wants a tour of the far end of that multiplicity. The strangest tasks machine learning has genuinely been pointed at.



Herman

The word that matters there is tasks. Not applications. We've done the practical specialised stuff before, in the hidden taxonomy of AI and the AI long tail, so that ground gets one sentence today. This is the weird end.



Corn

Which needs a rule, otherwise "a language model that writes limericks" counts as exotic.



Herman

And it doesn't. A language model writing limericks is a mainstream pattern wearing a hat. To qualify, the thing needs its own formulation, its own kind of data, and its own way of being scored. Time series forecasting qualifies. Predicting the next value in a sequence is a different problem from classifying a picture, and the evaluation grew up around it.



Corn

Where does wine go?



Herman

Wine is the borderline case. There's a famous dataset where you take the chemistry of Portuguese vinho verde, acidity, sugar, sulphur dioxide, a few thousand bottles, red and white together, and predict the score a tasting panel gave it. Fun, and it's in about every machine learning course on earth. But it's a table of numbers in and a number out. That's ordinary regression. The wine is exotic. The task is not.



Corn

The wine is exotic and the task is a spreadsheet. It fails.



Herman

Kindly. And the Hub's own task page makes the larger point, though I'm not putting that on Delangue, he didn't argue from the catalogue. Scroll through it and you get a few dozen task types, grouped into vision, language, audio, tabular, and one lonely reinforcement learning row. Text generation is one line among many.



Corn

Fine, but one worry. If the rule is that strict, do we have anything left? Everything I can think of is either a language model in a trench coat or a classifier.



Herman

I had the same worry, and the answer is yes, but you have to go looking for tasks where the data itself is odd. Not odd subjects. Odd data. Molecules, click timings, stress fields on a map, a buzz, a face with a pattern on it. Each one forces a different shape of model.



Corn

Okay. First stop.



Herman

Smell. I want to start there because it has the best twist. The task is, give the model a molecule, and it has to say what the molecule smells like.



Corn

That sounds like it should be easy. Chemistry in, chemistry out.



Herman

That was the assumption, and it's wrong in an interesting way. Vision has a long tradition of knowing roughly what a pixel pattern means. Smell doesn't. Two molecules that are near mirror images of each other can smell completely different, and two molecules with nothing in common structurally can both smell like garlic. Structure to odour is not a clean mapping.



Corn

And the label is a nose.



Herman

A panel of noses. In the original public challenge, run by a consortium called DREAM, the paper was in Science in two thousand seventeen, forty-nine people rated four hundred and seventy-six molecules. Intensity, pleasantness, and a list of descriptors: garlic, fish, sweet, burnt, flower, sour. The models did well on intensity and pleasantness and on eight of the nineteen descriptors.



Corn

Eight of nineteen. I'd frame that on a wall.



Herman

For a first attempt at predicting human perception from a diagram of atoms, it's not nothing. And the fancy models barely beat fairly plain regularised linear ones, so the field started humble.



Corn

What's the sequel?



Herman

Google researchers, and later a company called Osmo, built a graph neural network. The molecule really is a graph, atoms as nodes, bonds as edges, so it's a natural fit. They trained it on a few thousand molecules that trained perfumers had labelled. Then came the clever evaluation. They took a few hundred new molecules, had a panel smell them, and asked whether the model's predicted profile was closer to the panel's average than a typical single panelist was.



Corn

Hang on. The question isn't "is the model right". It's "is the model a better smeller than the average person in the room".



Herman

Because there's no ground truth beyond people. And the reported result was that it did get closer to the consensus than the typical individual. The machine has never smelled anything, and it agrees with the group better than you do.



Corn

I've been in rooms where that's true of people who weren't there.



Herman

Ha. The model also builds an internal map of odour space, where molecules that smell alike land near each other, and that map turned out to be useful for other smell tasks. I've only seen the headline numbers on that, so hold the exact figures loosely. Here's my question for you, Corn. Is a panel of noses even a fair scoreboard?



Corn

Not really. Everyone's coriander is different. A third of the room tastes soap.



Herman

Right, so the consensus is a blurred average of people who disagree. Which is true of a lot of labels, now that I say it. But it's the one place where the evaluation, not the model, is the strange part.



Corn

The principal odour map. I like that it sounds like something a pirate would be buried with.



Herman

Cute, and a bit unfair, because the interesting thing is how it was trained. Perfumers don't say "this is molecule number four thousand". They say rose, smoky, waxy, a few words from a shared vocabulary. So each training example is a molecule and a short list of words a trained nose agreed on. The model has to learn the words from the shape.



Corn

Teaching a computer to say "waxy" with confidence. I know people who can't.



Herman

Which, if you consider what the word is standing in for, is a lot of confidence on very little information. The model has never been in a room with a candle.



Corn

Next one has a better cast.



Herman

Sperm whales. The task is to take the click pattern of a whale vocalisation, a coda, and work out its structure. What are the parts, and what changes with context?



Corn

Before you get excited, give me the disclaimer.



Herman

The disclaimer is the point, so I'll do it first. Nobody knows what any of it means. The researchers are careful not to call it language. What a Project CETI paper in Nature Communications in two thousand twenty-four did, with an archive of recordings from about sixty whales off Dominica, nearly nine thousand codas, was show that the codas aren't a fixed list of patterns. There's a combinatorial structure underneath.



Corn

Combinatorial like letters.



Herman

Like features that combine. A coda is a short run of clicks. Rhythm and tempo seem to be the base identity of the thing and don't change with context. Rubato and ornamentation do. Rubato is the stretching of the timing, same word musicians use. Ornamentation is an extra click added on the end.



Corn

The whales have a time signature, a tempo and a flourish.



Herman

And that gives you a space of possible codas much bigger than the handful of types people had catalogued. Hence "phonetic alphabet" in the press. I'd hedge on that phrase. It's a combinatorial inventory. Whether it's an alphabet in any sense a whale would recognise, nobody can say. There was also a follow-up last year, which I've only seen in press coverage, finding vowel-like patterns in the codas, with gliding shapes like diphthongs. Treat that one as early.



Corn

How do you even check a result like that? There's no whale to ask whether you got the rubato right.



Herman

You check it against context. The team looked at which codas show up when, which whales are chatting, how many are involved, and whether the features line up with those situations. Rhythm and tempo hold steady across contexts, and the other two shift. That's a pattern a person can test without translating a single click.



Corn

This is the part I like. It's a task where success has no definition. You're not matching a label, you're finding out what the labels should be.



Herman

It's closer to discovering the classes than sorting into them.



Corn

So the whales have a grammar nobody can read, and we've got a ruler. We can tell the sentence has three clauses.



Herman

A tourist who can count the clauses and not translate one.



Corn

That's still better than my Portuguese.



Herman

Let me bring up a task that shouldn't work. Earthquake aftershocks.



Corn

Is this where a neural net beat the physicists?



Herman

In two thousand eighteen, in Nature, yes. Phoebe DeVries and colleagues, including people from Google and Harvard. The task: a big earthquake happens, you know how the ground shifted, and now for each little square of ground around it, about five kilometres across, predict whether an aftershock will hit there. A map of where the next ones land.



Corn

And the data?



Herman

Big, for seismology. More than a hundred and thirty thousand mainshock and aftershock pairs. The benchmark was a classic quantity called the Coulomb failure stress change, a physical idea about how a quake loads the neighbouring rock. On the area-under-the-curve score, the network reportedly got to about point eight five, and the classic formula about point five eight. And point five is a coin flip.



Corn

So the physics formula was barely better than guessing.



Herman

At this particular job. Which should make everyone a little humble.



Corn

You said it shouldn't work.



Herman

The network wasn't handed the physics, only the stress changes, and it learned the rest. Then came the debate. Other researchers looked at what it had learned and found, as I understand it, that you could explain nearly all of its predictions with a few simple stress measures, including a yield criterion from materials engineering. So did it discover something, or find a better blend of old formulas?



Corn

The second one. And I mean that kindly. A model that finds a better formula you hadn't tried is doing the job.



Herman

It's actually a gift. If a simple formula does nearly as well, you can write it on a page. Though what do you make of the word "discover"? Would a seismologist be happier if the net found something new, or if it found something they could print?



Corn

Print. A black box that says the ground breaks here isn't something you want to explain to a mayor.



Herman

Fair. One caveat, it's a map of where aftershocks cluster, not a warning about when the next big one comes. And look at the pattern so far. Smell, whales, ground. None of them has a giant labelled internet to scrape.



Corn

Is my phone doing something? That isn't the usual ringtone.

D

Dolores

Good afternoon. This is the World Podcast Authority, compliance division. Your paperwork is not in order: the feed lists this programme under a category the Authority considers too broad, and you will need to resubmit under Form Eleven B, the Narrow Category Declaration, reference number four four seven.



Herman

For the record, the World Podcast Authority does not exist.

D

Dolores

I hear you, but I must carry on regardless: your podcast licence may be revoked, and you have until noon on the twenty first of this month to file the corrected declaration.



Corn

Nobody issues a podcasting licence, so there's nothing to revoke.

D

Dolores

I will make a note of that. A form will follow in the post. Good afternoon.



Herman

Back to the labels. Somebody had to go and make every one of them, with noses, with seismographs, with boats. I'd rate the grubbiest dataset on the list as mosquitoes.



Corn

Grubbiest meaning what?



Herman

Meaning you have to catch the insects, know for certain what they are, and then get them to fly past a sensor in the right mood. The ground truth is a person with a magnifying glass and a jar. If the jar is mislabelled, so is your model.



Corn

A model that believes everything the jar says.



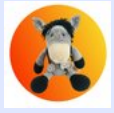
Herman

Every dataset has a jar. This one has an actual one.



Corn

Go on then. What's the task?



Herman

Listen to a mosquito's wingbeat in a recording and say what species it is and whether it's male or female.



Corn

I'd guess it was invented by someone who was kept awake.



Herman

Can't rule it out. But the logic is sound. Different species beat their wings at different frequencies, and the sexes differ too, males typically higher. A mosquito is a tiny noisy aircraft, so the signal's in the sound.



Corn

And the reason to want it?



Herman

Public health. Only certain species carry certain diseases, and only females bite. A trap that can tell you what flew in, and whether it was a biter, is surveillance without a person squinting at a heap of dead insects.



Corn

Squinting at a heap of dead insects is somebody's job.



Herman

It is, and it's slow. A team in Greece around two thousand eighteen used a sensor that records the flutter optically, roughly two hundred and eighty thousand recordings across six species, and got about ninety-six percent on species. A newer paper in PLoS ONE handles background noise, species and sex together, at about ninety-three percent. The lower number is the more honest one, because the real world has wind and other bugs.



Corn

I'll still say a quiet lab room and a mosquito with nowhere to be is not a field. Out there you'd get a donkey braying in the next county.



Herman

A credible acoustic hazard, yes. The sex label is the part with teeth, by the way. In a trap it tells you whether you're catching the ones that bite, which is the whole public health point.



Corn

The only task on the list where one of the classes is a complaint. And I'd call it impractical for a person, nobody's telling six species apart by ear in a field. That's why you need the machine.



Herman

Agreed, nobody does that at scale. What I like is the formulation. A narrow band of information, one frequency and its overtones, and the model has to pull the answer out of a buzz.



Corn

Last one, and the one I've been waiting for since I read the brief.



Herman

Cow faces.



Corn

Individual cow faces. Not "is this a cow". Which cow.



Herman

It's called re-identification, and it's a real task with its own literature. A Holstein has those black and white patches, and the face pattern works like a barcode. One paper in Animals in two thousand twenty-two used something like a hundred and thirty thousand images of around three thousand cows, if I remember the figures right, and built a network that looks at the whole face and also at parts of it.



Corn

I admire the parts. The whole cow I can manage. It's the left eyebrow that gets me.



Herman

It's a separate task from classification because the herd keeps changing. New calves arrive every season, so you can't train a fixed list. You train the model to produce a signature for a face and compare signatures. Same trick as face recognition for people, applied to a subject with a very consistent expression.



Corn

Deadpan. Largely resigned.



Herman

The hard part is that the same cow doesn't hold still. She turns her head, there's mud, the light changes through the day. So the network has to learn which parts of the pattern stay put across angles and which are just weather. That's why the version that also attends to parts of the face does better than one that treats it as a single blob.



Corn

A face recognition system that has to cope with a cow turning away because she's heard something about you.



Herman

Now, if I were running a dairy, here's what I'd want. Which cow ate, which one's limping, which one's off her milk. Ear tags fall out. A camera at the gate doesn't. What would you do, Corn, if you had a thousand cows and a budget?



Corn

Sit down. Slowly. Then buy the camera and put it at the gate, where they all queue anyway.



Herman

Which is the right answer, and a good deal faster than mine, which was to start a spreadsheet. It's funny that the most practical one on the list is also the one that gets the biggest laugh.



Corn

Because it's an ordinary farm chore handled by very serious machinery. Which is where the humour is. I mean it respectfully.



Hilbert

Philips Pocket Memo, model two seven one seven, nineteen pounds from Desmond Quill's shop.



Herman

A dictation machine, for wingbeats. I'd have gone with a proper microphone, but I'm listening.



Hilbert

Corn, you said nobody tells mosquitoes apart by ear, and Herman, you agreed in four words. I've been doing it since April, off the bathroom window and the shed. Four hundred and twelve recordings. The tapes are at my brother-in-law Cormac's, in his chest freezer. The recorder lives in a biscuit tin, with the aerial taped to a coat hanger.



Corn

Why the freezer?



Hilbert

Damp. Desmond Quill has sold me the same battery four times, and each one takes him nine days to get in. He says the supplier is in Leeds. It is not in Leeds. After every recording I say which mosquito it was, out loud, onto the tape. Eleven pounds fifty a battery. Nine minutes a recording, four hundred and twelve recordings, and Cormac has



Hilbert (continued)

the lot.



Herman

I'll take the coat hanger on trust. The nine minutes is the number I'd want in the abstract, and it never makes it.



Corn

It makes the whales look cheap, and I'd have assumed the boat was the expensive part.



Herman

A boat and a hydrophone and years, so not quite. But the shape is the same across all five. A person at the start with a notebook. The model is the fast part, and in most of these cases the easy part too, which I'd never have said ten years ago.



Corn

You'd have said it was the hard part and the notebook was admin.



Herman

I did say that, with some confidence.



Corn

If you take one thing from this, the exotic ones aren't exotic because of the model. They're exotic because of the label. A panel of noses, a map of the ground, a cow's name.



Herman

Each one needed somebody patient enough to build a dataset nobody else wanted. That's where a multiplicity of models comes from.



Corn

What I'd like to know is which task is next, the one with no name yet, where someone has just started the notebook.



Herman

Somewhere there's a person with a tape recorder and a freezer, and in a few years we'll be covering it.



Corn

Thanks to our producer Hilbert Flumingtop for keeping the desk running.



Herman

This has been My Weird Prompts. The whole archive is at my weird prompts dot com.



Corn

See you tomorrow.