Episode Book

EPISODE #6604

What Breaks When You Train AI on Synthetic Taste

PUBLISHED	14 November 2026 at 21:49 UTC
DURATION	27:18
LISTEN	myweirdprompts.com/episode/synthetic-plots-personal-taste-27423/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	omnivoice
PIPELINE	V5.29
SOURCES	25 cited by the research (listed at the end)
BOOK GENERATED	15 November 2026 at 11:16 UTC

SYNOPSIS

Training a model on 200 to 500 synthetic movie plots sounds like plenty of data, given that the average Netflix Prize user rated 208 films. Yet sample size is not what breaks a personal taste model. Replacing real films with synthetic plots traps the model in a narrow pocket covering under 10 percent of real preference diversity, while personalized reward models often learn surface correlates rather than actual preferences. The discussion also examines why fixed checkbox reasons and annotation fatigue distort the signal.

TRANSCRIPT



Corn

Daniel's been living in fine-tuning and pairwise evaluation lately, and it shows. His question this time is a design: generate a couple hundred synthetic movie plots, go through them one at a time, and for each one say whether you'd watch it, whether it makes your shortlist, and pick a reason why not from a fixed list. Too dramatic. Hate war movies. Train a model on those labels so it scores any real movie for you, and only show you the ones that clear the bar.



Herman

And would it work.



Corn

That's the question. But he's upfront that it's a thought experiment. What he actually wants to know is the bigger thing underneath. Can you train a model on one person's taste at all? How do you get from the generic, everything-is-pleasant assistant to something that actually gets you? Movies are just his laboratory for that.



Herman

I like that he picked movies, honestly. It's a subject where we have absurd amounts of ground truth. The Netflix Prize dataset alone is a hundred million ratings from four hundred eighty thousand users on nearly eighteen thousand films. If you want to know whether taste can be learned, this is the terrain where somebody has already measured the neighboring fields.



Corn

And the design itself. Has anyone tested this exact package? Synthetic plots, checkbox reasons, a personal confidence scorer.



Herman

No. I looked at what's out there and nobody has run that specific combination. The closest thing I know generates synthetic chat queries about movies to train a recommender, but the items being rated stay real. Daniel's proposal synthesizes the items themselves. So we're reasoning from adjacent evidence today, not from a direct test. Every component



Herman (continued)

has been studied. The package hasn't.



Corn

Which is fine. That's most of what we do here.



Herman

It is. And the striking thing is where the evidence lands, because the number everyone would worry about first is the number that turns out not to matter. Two hundred to five hundred annotations is ten to twenty-five times what the classic cold-start work needed to beat a baseline, and it's roughly the entire rating history of the average Netflix user.



Corn

The average user, over years of using Netflix, rated about two hundred movies.



Herman

Two hundred and eight, in the Prize data. Jim Bennett, who ran Netflix's recommendation work back then, said in two thousand six that rating a hundred movies is what personalizes predictions. More than a quarter of users rated fewer than fifty. So Daniel's hypothetical annotation session is asking for more effort than most people gave Netflix over a decade of evenings.



Corn

One afternoon with a spreadsheet outperforming ten years of not clicking the thumbs.



Herman

In volume, yes. Volume is not the problem. The problem is what's on the other side of the labels.



Corn

So the question splits in two. Does this specific design work, and what does the research say about learning one person's taste. You've basically given away the answer to the first one already.



Herman

Partly. Let's do the number properly first, because it's a nice piece of history. GroupLens, the Minnesota lab that basically built this field, published a paper in twenty fourteen called Recommending for New Users Is Surprisingly Difficult. They took MovieLens users and artificially restricted how many ratings the algorithms could see. Their best methods needed two or three ratings just to beat the raw average rating of a film. Four to beat the user-adjusted average. The classic item-based method needed at least thirteen, and at nineteen ratings it still hadn't caught the adjusted average. Their own conclusion was that they expected these algorithms to work for everyone and they just... don't.



Corn

Nineteen careful ratings and the best you can do is lose to arithmetic.



Herman

Humbling, isn't it? And MovieLens itself, the actual site, historically gated personalized recommendations at fifteen ratings and guaranteed every user at least twenty in the dataset, which is why cold-start studies evaluate at up to nineteen. The whole literature lives at the scale of a dozen labels. Daniel is proposing a couple hundred. He has budget the field never dreamed of.



Corn

Okay. So if I'm building this thing, I'm feeling pretty good about now. I've got five hundred labels, which is twenty-five times what MovieLens needed. What breaks?



Herman

Two things, and they compound. The first one is the plots. They're synthetic. The model never learns the space of actual films, because it's never seen one.



Corn

It learns the space of the generator.



Herman

And this isn't hand-waving, there's a paper from this year's ICML called The Persona Fidelity Gap that measures exactly this kind of substitution. They took about two hundred sixty-eight thousand real Amazon reviewers, learned each person's preference embedding, and then placed LLM-generated personas of those same people into the same space to see how close they came. Grounding the simulated persona in the real user's history cuts the error thirty-fold compared with just prompting a model to act like the person. Huge improvement. And even after that improvement, the simulated users cover under ten percent of the preference diversity of the real ones.



Corn

Under ten percent. After the thirty-fold improvement.



Herman

Yes. The technical detail that makes it sting is that item-level accuracy looks great. Rank accuracy of point-seven-two to point-eight-five, which sounds excellent. The authors' own line is that these local metrics hide the collapse. Predicting items well within a narrow pocket tells you almost nothing about whether you've covered the person.



Corn

That's the trap in Daniel's design, isn't it. He holds out fifty of his five hundred synthetic plots as a test set, the model gets ninety percent of them right, and he concludes it knows his taste.



Herman

It knows his taste as expressed in synthetic plots generated by one particular model in one particular week. The fidelity work says that pocket is much smaller than it feels from inside it. There's a related finding I find even more sobering. A big simulation study this year tried to reproduce a global preference survey across seventy-four countries and sixty-seven thousand real respondents using twenty LLM configurations, and recovered the direction of individual-level preferences in about forty-eight percent of cultural dimensions.



Corn

Coin flip.



Herman

For guessing which way a specific person leans, yes. Systematically better than nothing, and not much better than a coin. So the model replacing Daniel's movie shelf with its own imagination is not a small distortion. It's the majority of the distortion.



Corn

Alright, that's the plots. You said two things compound.



Herman

The second is nastier. Even if the items were fine, the model might learn to predict his annotations without representing his preferences at all. There's a result from late September that I keep turning over. Researchers working on personalized reward models noticed that the in-context ones, the ones that read a description of the user before scoring, seemed to barely use the description. So they ran the obvious brutal test. They took each user's preference pairs and swapped them. Chosen becomes rejected. The context now describes a person with exactly the opposite tastes.



Corn

And?



Herman

Predictions on the same target pairs were unchanged in eighty-six percent of cases.



Corn

Eighty-six. You invert the human and the model doesn't notice.



Herman

The model is reading the context, in the sense that the tokens go in. It's just not letting them move the output. Which means a model like that can sit at a perfectly respectable accuracy on held-out labels while being almost entirely insensitive to the direction of your preferences. It's learned surface correlates. Genre keywords, plot length, tone words. It predicts your annotations. It does not know your taste. Those are different targets that look identical on the accuracy metric.



Corn

That's the fact I'd tape to the monitor before building anything. There's a test, though, buried in it. If you can invert the labels and the predictions hold, you've learned the texture of my answers rather than the answers. That's a cheap sanity check Daniel could actually run.



Herman

He could, and I'd want him to. Run the counterfactual. Feed the model the anti-Daniel, the person who loves what he hates, and see if the scores move. If they don't move much, the model was never listening.



Corn

Now here's my defense of the checkboxes, and you're going to take it away from me.



Herman

Go on, defend them.



Corn

A fixed list of reasons feels like it cleans the signal. Instead of a mushy maybe, you get structured labels. Too dramatic. Hate war movies. A tidy little taxonomy.



Herman

I'll take it away. The taxonomy is the problem. You've told the model the features in advance, so it learns the features. And the features are a compressed, pre-digested version of your judgment, written before you knew which distinctions would matter. There's a nice result from work on preference elicitation showing that when you ask people about abstract categories versus representative items, what people say about the category is basically driven by how they feel about one or two representative examples. The checkbox is a category. The plot is the item. You're training on the category and hoping it stands in for the items.



Corn

So I've paid five hundred labels to learn my own dropdown menu.



Herman

You've taught it your dropdown menu, which you could have handed over in a sentence anyway. The SynthesizeMe work, which we'll get to, flags another confound that applies to Daniel's plots directly. Without feedback, you can't tell whether the content or the presentation drove the choice. He rejects a synthetic plot. Why? The premise? The prose? Because it was the fourteenth plot he'd read in a row and his patience was gone by the ninth?



Herman

It always is. Anybody who's sat through an annotation session knows the labels from the last twenty minutes are worth less than the first twenty. Fatigue is a feature of your data, and it masquerades as taste.



Corn

There's a third one, isn't there. You said two, but there's always a third with you.



Herman

There's a third. Preference heterogeneity. Different people assign opposite labels to the same pair of items, and a single model fitted to everyone averages over the conflict, which biases it toward the majority. For a population system that's a fairness problem. There's a paper this year that names the consequence precisely: reward-model quality varies systematically with how common your preferences are in the training population.



Corn

But Daniel's system is one user. No population. Doesn't that solve it?



Herman

It solves the averaging and removes the crutch in the same stroke. If your taste is common, a general model already half knows you. If it's idiosyncratic, there's no population prior to borrow from, and your two hundred labels carry the entire signal on their own. Which is fine when the signal is real, and we've just spent ten minutes establishing it may not be.



Corn

Okay. So the design fails for reasons that have nothing to do with sample size. The number was never the constraint.



Herman

So the number is not the binding constraint. It never was.



Hilbert

Hilbert's Mid-Show Haiku. For the person in the third row who has not blinked since the cold-start segment. Two hundred labels. The plots are fake, the taste is real. The model still guesses.



Corn

He's not wrong about the plots.



Herman

He's not. Which raises the obvious question, and it's the one Daniel actually asked. What would you try instead? And the literature has a ladder here, from cheapest to most invasive, and I think the order matters more than any single rung.



Corn

Bottom rung first.



Herman

Ground it in real items and real behaviour. This is almost embarrassing as advice. Annotate films you've actually seen, or explicitly passed on. Let the system watch what you do. What you finished, what you abandoned at twenty minutes, what you rewatched. Netflix made this point years ago, that even when a subscriber never rates anything, they still note which movies you actually order. Behaviour is a label that doesn't lie the way a checkbox can.



Corn

The abandoned-at-twenty-minutes signal is underrated. Nobody abandons a film ironically.



Herman

It's the highest-signal data a movie watcher produces, I'd argue. And Daniel has decades of it, sitting in his head and his watch history, and the design in the prompt throws all of it away to generate five hundred fresh imaginary films. That's the part I find almost poignant. He owns a lifetime of labels and proposed synthesizing replacements.



Corn

Next rung.



Herman

Borrow the crowd. Collaborative filtering as a prior, not a replacement. There's work this year on exactly sparse settings like this, graph-based collaborative filtering feeding a mixture of LoRA experts that balances what's shared across people against what's specific to you. Your two hundred labels become much more informative combined with the fact that people who liked what you liked also liked certain other things.



Corn

Though that assumes my taste resembles anyone's.



Herman

It assumes it partially resembles some people's. Which is nearly always true and never entirely true. The mix-of-experts framing is honest about exactly that. But here's the rung I actually find most interesting for Daniel's question, and it's counterintuitive. Before you fine-tune anything, try not fine-tuning. SynthesizeMe, the method I mentioned, induces a persona from your preferences. It generates candidate explanations for why you chose what you chose, verifies them against your actual choices, builds a written persona from the verified reasoning, then selects the most informative prior interactions to put in the prompt. No training. The induced persona prompts beat state-of-the-art personalized reward models by up to about five percent, with zero fine-tuning.



Corn

Hold on. The cheap version beat the trained version.



Herman

Beat dedicated personalized reward models, yes. And it's interpretable, you can read the persona and argue with it, and it transfers between models because it's just text. Now, why do you think that beats training? Think it through, because you basically said it ten minutes ago.



Corn

Because the training set is the problem, not the training. If two hundred labels carry mostly surface signal, a fine-tune amplifies the surface signal. A persona written in words has to pass through your judgment first, and you can read it.



Herman

That's the argument, and the evaluation work supports the first half. But I'll be fair to the other side. If your labels are good, weight adaptation goes further. There's the same test-time paper that found the eighty-six percent blindness, and their fix is elegant, it reuses the down-projection layer of an existing reward model as fast weights and updates just that small piece within a single forward pass. No backpropagation at inference. It beats the state of the art by nearly six points. And newer work generates per-sample



Herman (continued)

steering vectors from user data and injects them into the forward pass directly.



Corn

You can see the day coming where I admit the mechanism is elegant.



Herman

It's elegant and it's cheap, that's the combination that gets me. Small adaptation, per person, from good labels. The mistake isn't adapting weights. It's adapting many weights from bad labels.



Corn

What about the rung above that? You had that look.



Herman

The one I find conceptually nicest: stop learning the preference and learn the adaptation. There's a SIGIR paper this year arguing for exactly that shift, learning the process of preference adaptation rather than fitting the preferences themselves, representing each user's model as a weighted combination of base reward functions and meta-learning how to adapt quickly from little feedback. Related idea at ICLR: learn text summaries of the user that condition the reward model, with reported improvements over standard approaches ranging from eleven percent to seventy-seven percent depending on the setting.



Corn

Seventy-seven sounds like a setting where the baseline was ill.



Herman

But the direction is consistent. Summarize the person, adapt small things quickly, keep the big model shared.



Corn

Okay. All the engineering is on the table. But the question underneath all of it, the one Daniel actually cares about, is taste. Can a model have it. And there was a line in that Forbes piece you sent me, from Ansari, that I haven't stopped chewing on.



Herman

Say it.



Corn

He draws a line between truth and taste. On factual questions, disagreement is real but bounded. You can collect more expert judgments and steer toward consensus. For taste, quote, that doesn't really exist. If you debate long enough with people you would label as having very good taste, you will keep diverging.



Herman

I've read it three times. And the usual reading is that it's bad news for AI, because everything we do in alignment is consensus-seeking. Averaging judges, majority votes, rubrics. If taste is structurally non-aggregatable, the whole recipe is wrong for it.



Corn

But flip it for Daniel's case. He's building for an audience of one.



Herman

Then it's good news. There's no consensus to converge to, so nobody can tell him his target is wrong. A model of one person's taste is well-posed in a way a model of everyone's never is. The catch is that it has to be measured against that one person's real behaviour, not against a population average, or the idiosyncrasy he's trying to capture gets smoothed right back out.



Corn

The thing that makes taste impossible to aggregate is the same thing that makes it possible to fit.



Herman

And that's why movies are the right laboratory. Dense, subjective, high-dimensional, and drowning in ground truth so you actually know when you're wrong. This isn't like grading poems, where the frontier labs are paying people for rubrics, points for a surprising rhetorical device, minus for the wrong register. Those rubrics are somebody's taste turned into rules, which works about as well as you'd expect turning taste into rules ever works.



Corn

Which is the floor problem. Somebody put words to it that I liked. A check can measure only what you can write as a rule, and generic work satisfies rules easily, because rules describe the floor. Above the floor sits a second bar, that the work is good for the person who uses it. That's the gap between would-I-watch-this and does-this-get-me.



Herman

The whole episode fits inside that gap, really.



Corn

The problem is the gap's dimensions.



Herman

The plots are fake and the floor's the ceiling.



Hilbert

It's not a shortlist.



Herman

Sorry, which part?



Hilbert

The shortlist. You two keep saying shortlist. Daniel doesn't have a shortlist. Nobody has a shortlist of films. They have an evening, a mood, and whoever else is on the sofa. The word is doing you a disservice.



Corn

He's got you there. Every list I've ever made collapsed on contact with a Tuesday.



Hilbert

I keep a notebook. Have done for years. Every film I've walked out of, written down in the car park, with the reason. There's about four hundred entries in it now, which I understand is in Daniel's range. The shorthand is my own, so nobody else can read the reasons, which suits me. I've tried explaining the system to people. Tried it with a machine once, out of curiosity. It doesn't take, and I'll tell you why. Around the time I hit two hundred entries I went back through the thing properly and the reasons are mostly not about the film. They're about the evening. Who I was with. A draught on the left ankle at the Odeon. The notebook records taste, but the taste is half temperature and company, and no model is getting that out of four hundred lines of shorthand, because I couldn't get it out and I wrote them.



Corn

You're saying the annotations were accurate and the target was wrong.



Hilbert

I'm saying the writing was the useful part. You put the reason down before you can tidy it, and you notice things. After that it's a list. I don't consult it. Never have. The rule is the entry gets written before I leave the car park, and once I drove eleven miles back to a cinema in Rehovot because I'd forgotten, and the ticket girl had to let me stand in the foyer to remember what I'd disliked. It was the trailers. I'd have lost that.



Herman

Four hundred walkouts, though. That's a lot of films you didn't finish.



Hilbert

Four hundred entries. Some are films I stayed for and regretted staying for. Those are marked differently. Corn would have marked them differently too, if he ever left the sofa.



Corn

Fair.



Herman

The interesting part is that the notebook works precisely because it's unusable. It's not a dataset. The act of writing the reason down was the point.



Hilbert

If you like. I mainly keep it because I don't trust my memory of why I disliked a thing, and I trust everyone else's less. Herman, you said behavior doesn't lie. It lies by omission. I've walked out of two films I liked.



Herman

Huh. I'll grant the omission. But you knew they were omissions, which is more than the model knows.



Hilbert

Only because I wrote it down before I left the car park.



Corn

Which is a strange place to land. The best taste model in this episode is a notebook nobody can read.



Herman

So let's clean up the misconceptions before we go, because there are two sitting in this topic and people walk into both constantly.



Corn

The big one. People hear this design and say, well, two hundred labels, obviously too few, you'd need thousands. The cold-start history says the opposite. Thirteen to twenty ratings beat baselines on real items. The average Netflix user's whole life is two hundred eight. The number was never the constraint. What's being annotated is.



Herman

And the second, which is sneakier. People assume a model that scores well on held-out annotations has learned their taste. The label-reversal result is the correction. You can invert the person entirely and the predictions barely move. Accuracy on your annotations and knowledge of your taste are different things, and the counterfactual test is how you tell them apart. The cheapest working approach in all of this is also the least impressive-sounding one. You don't build a bespoke scorer for an audience of one. You induce a persona, write down what you learn about yourself, and put it in the context window.



Corn

The failure of Daniel's design is honestly more interesting than the verdict. He didn't lose on sample size. He lost on substituting generated items for lived ones, and a fixed taxonomy for the messy, context-soaked reasons a person actually has. I'd add the forward question: if the floor is rules and taste is the second bar, maybe the test isn't whether the model agrees with your labels, but whether the model can be surprised by you.



Herman

Thanks as always to Hilbert Flumingtop at the desk. For more along these lines, there's episode eleven, How Does Fine Tuning Work Anyway; episode four, If Your Voice Ages, Does Your Fine-Tune Become Useless; and episode five, Fine-Tuning ASR For Maximal



Herman (continued)

Usability. This has been My Weird Prompts, the human-AI collaboration podcast. If you got anything out of this one, the review buttons are right there, and they take less time than five hundred annotations.



Corn

We'll be back soon. See you tomorrow.

SOURCES

What the research for this episode read before the script was written. Primary sources first.

1. **Wang et al., Using Context Is Not Enough: Test-Time Training for Personalized Reward Modeling (P-TTT), submitted 2026-09-28. [primary]**
   https://arxiv.org/abs/2609.35109v1

2. **The Persona Fidelity Gap: Behaviorally Grounded LLM Personas Still Compress Real-User Preference Diversity, ICML 2026. [primary]**
   https://icml.cc/virtual/2026/75683

3. **Ryan et al., SynthesizeMe! Inducing Persona-Guided Prompts for Personalized Reward Models in LLMs, ACL 2025, submitted 2025-06-05. [primary]**
   https://arxiv.org/abs/2506.05598v1

4. **GroupLens, Recommending for new users is surprisingly difficult, 2014. [primary]**
   https://grouplens.org/blog/recommending-for-new-users-is-surprisingly-difficult/

5. **Simulated responses from large language models are weak proxies for human economic preferences across 74 countries, Nature Human Behaviour, 2026.**
   https://www.nature.com/articles/s41562-026-02611-4

6. **Forbes, Taste Is When Machines Learn The Rules Nobody Wrote Down, 2026-05-22.**
   https://www.forbes.com/sites/rayravaglia/2026/05/22/taste-is-when-machines-learn-the-rules-nobody-wrote-down/

7. **AINEXT, AI's Taste Problem: Why Are Frontier Labs Paying Poets Top Dollar to Grade Their Models?, 2026-05-17.**
   https://ain3xt.com/en/posts/20260517-ai-taste-training-problem/

8. **on Taste Labs and ~1,000 Tastemakers.**
   https://www.the1news.com/article/a-startup-wants-to-replace-human-creativity-but-first-it-needs-to-eliminate-sl

9. **Cankun Wang, Taste: The Bar After Correct.**
   https://cankun.me/writing/taste

10. **Choi et al., CoPL: Collaborative Preference Learning for Personalizing LLMs, submitted 2025-03-03, revised 2025-09-17.**
https://arxiv.org/abs/2503.01658v2

11. **Zhao et al., PAFO: Pareto Fairness Optimization for Personalized Reward Modeling, submitted 2026-06-06.**
https://arxiv.org/abs/2606.07988v1

12. **Kim et al., Rethinking Personalized Reward Modeling for LLMs under Preference Heterogeneity via Group-Debiased Federated Learning, submitted 2026-08-03.**
https://arxiv.org/abs/2608.01556v1

13. **Dey et al., PALMs: Using Multi Construct-Grounded Rationales for Modeling Population Preferences in LLMs, submitted 2026-08-02.**
https://arxiv.org/abs/2608.01458v1

14. **Du et al., FleetPilot: Fine-Grained Steering Empowers Real-World LLM Personalization, submitted 2026-10-10.**
https://arxiv.org/abs/2610.14970v1

15. **Cai et al., One Adapts to Any: Meta Reward Modeling for Personalized LLM Alignment, SIGIR 2026, submitted 2026-01-26, revised 2026-04-20.**
https://arxiv.org/abs/2601.18731v2

16. **Nam et al., Learning to summarize user information for personalized reinforcement learning from human feedback (PLUS), ICLR 2026, submitted 2025-07-17, revised 2026-08-26.**
https://arxiv.org/abs/2507.13579v4

17. **Cantürk & Aydoğan, Explainable Active Learning for Preference Elicitation, submitted 2023-09-01.**
https://arxiv.org/abs/2309.00356v1

18. **Zuo et al., Hierarchical Conversational Preference Elicitation with Bandit Feedback, submitted 2022-09-06.**
https://arxiv.org/abs/2209.06129v1

19. **Cheema & Guven, Optimizing Recommendations using Fine-Tuned LLMs, IEEE CAI 2025, submitted 2025-05-11.**
https://arxiv.org/abs/2505.06841v1

20. **Netflix Prize dataset figures.**
https://en.wikipedia.org/wiki/Netflix_Prize

21. **IEEE Spectrum, The Million Dollar Programming Prize.**
https://spectrum.ieee.org/the-million-dollar-programming-prize

22. **MIT Technology Review interview with Jim Bennett, 2006-10-06.**
https://www.technologyreview.com/2006/10/06/273459/the-1-million-netflix-challenge/

23. **Volinsky, All Together Now: A Perspective on the Netflix Prize.**
https://chrisvolinsky.com/files/publications/all-together-now-netflix-prize.pdf

24. **Active learning algorithm for alleviating the user cold start problem of recommender systems, Scientific Reports, 2025.**

https://www.nature.com/articles/s41598-025-09708-2

25. **On the Smaller Number of Inputs for Determining User Preferences in Recommender Systems, Mathematics 8(12).**

https://doi.org/10.3390/math8122138