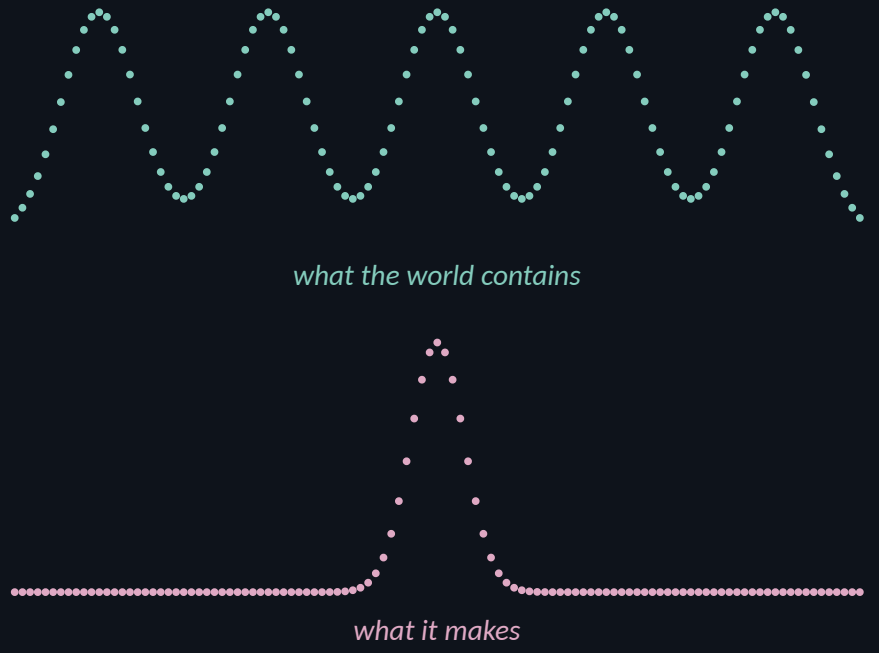
AI IN SPECULATIVE MEDIA AND LITERATURES

Mode Collapse



Goodfellow et al., “Generative Adversarial Networks” — abstract and Figure 1 only

Bender et al., “On the Dangers of Stochastic Parrots”

Ruha Benjamin, Race After Technology — excerpt

Key question: what disappears when a system optimises for approval?

BEFORE WE START

Two notes on today

Nobody did the same exercise

Forty variations on one prompt, but you each chose the prompt, the system and the stopping point. Those differences are the data, not noise in it. When we pool at 3:10 we are not checking who did it right.

And there is no new technical content

Last week you learned how a system is trained and what a loss function leaves out. That is all the machinery today needs. If you want the mathematics of how generative models fail, the Goodfellow paper has it — we are not doing it.

What is new is the comparison

We will take one pattern and find it in five texts you have already read. If the pattern is real it should be visible without being forced; if it has to be forced, that is a finding too, and we will say so.

Content note on the exercise

If your prompt produced material you would rather not pool with the group, say so and work with your own forty privately — the diagnosis works the same. If you could not run the exercise at all, join a group and work on theirs; the analysis is the assignment, not the generating.

KEY QUESTION FOR THE WEEK

What disappears when a system optimises for approval?

Note that the question does not ask what goes wrong. A collapsed system has not broken. It is succeeding — narrowly and perfectly — at the thing it was asked to do. The variety of the world was never part of what it was asked, so the variety of the world quietly disappears.

The word "system" is doing a great deal of work in that sentence. Holding it open is this week's discipline.

ORIENTATION

The contradiction

1 Take the metaphor seriously

A maker who has found the small set of things that reliably please a judge, and makes those and only those, forever, is an extraordinarily precise description of something that happens to models — and to sculptors, to revolutionaries, to institutions and to students. Sit with how good it is.

2 Then be suspicious of it

A technical term reached for as a poetic figure is exactly the move this course keeps asking you to make carefully. A metaphor that fits five texts at once is either a genuine structure or a shape loose enough to fit anything, and those feel identical from inside.

Which is why the comparison comes before the verdict

We spend fifty minutes putting the pattern next to Echo, Narcissus, Pygmalion, Frankenstein's creature and Detroit, and we do it generously — looking for what fits. Then at 4:25 we turn around and audit the whole thing, and you should expect at least one of the five to fail the audit. Finding which one is the assignment.

PART ONE

The forty outputs

First look, before any vocabulary · 0:20–0:40



Describe before you diagnose. You do not yet have the words, and that is deliberate.

THE EXERCISE

Lay them out, and answer four questions in plain language

1 What is the single most common thing?

Not a theme — a specific recurring element. A name, a setting, a sentence rhythm, a colour, a structure. If two of you generated the same phrase from different prompts, say so out loud now.

2 How many genuinely different outputs are there?

Count by hand. Two outputs that differ only in a swapped adjective are one output. Agree the rule with your neighbour before you count, and write the rule down — we will need it at 2:30.

3 What did you expect that never appeared?

The hard one, and the whole exercise. List things that plainly belong to the space of your prompt and are entirely absent across all forty.

4 Whose absence would you not have noticed?

Something missing that you only see because someone else pointed it out, or that you had to reason your way to rather than feel. Mark those differently — they matter at 3:50.

No technical vocabulary in this block. Describe what is on the page in the words you had before this course started.

WHAT ROOMS USUALLY FIND

Three layers, in the order people notice them

Names and settings converge hard

A handful of names recur across dozens of outputs, often across different students with different prompts. Settings cluster in a narrow band. The most visible layer, and the easiest to dismiss as coincidence.

Structure converges harder

Sentence length, paragraph shape, where the turn comes, how things end. Students notice this second and find it more unsettling, because they had not been watching for it and cannot un-see it afterwards.

Stance converges hardest

Tone, degree of resolution, whether anything is left unresolved, who gets sympathy, what counts as a satisfying ending. Nearly invisible until someone names it — and it is the layer that matters this afternoon.

And one finding that is not about the model

Some of the convergence is yours. Forty variations on one prompt, written by one person in one sitting, are forty outputs from a system that includes you — your prompt, your idea of what counts as a variation, your decision that forty was enough. Hold that. At 4:25 the question is whether the same word describes both halves of that system, and you will want to have noticed your own half early.

PART TWO

What mode collapse is

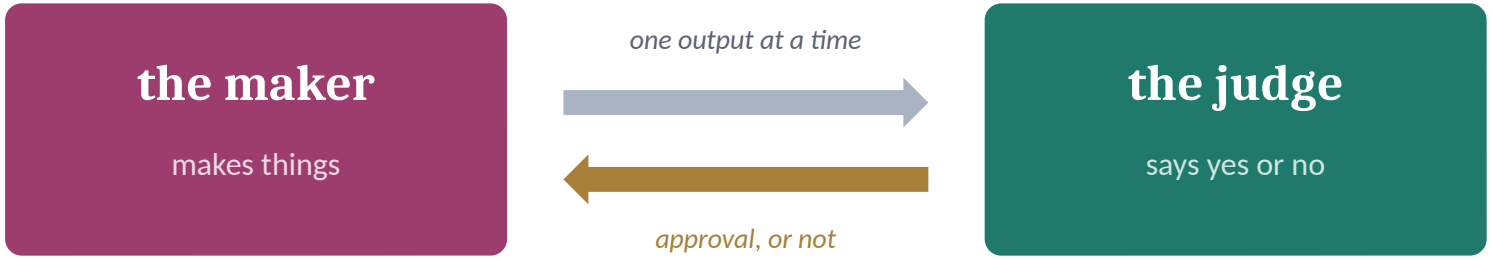
In plain language · 0:40–1:20



Two characters, one habit, and a kind of success that looks like failure.

THE SETUP

Two characters, and the only thing one of them can see



The maker never sees the world

This is the fact everything else follows from. It has no access to the real range of things that exist. Its only information about the world arrives as a verdict on something it already made.

The judge sees one thing at a time

It is asked whether this output is convincing. It is never asked whether the maker has made this same output forty times already. Nothing in the arrangement notices repetition.

So variety is nobody's job

The maker is rewarded for being convincing. The judge is rewarded for being right. Neither is rewarded for the range of what gets produced, and so nothing in the system protects it.

Remember last week: the objective had no term for coverage. This slide is what that sentence costs.

THE MECHANISM

How a working system ends up making one thing

1

The maker finds something that works

Out of everything it could produce, one kind of output reliably gets approval. Nothing instructs it to keep looking. Making that thing again is the single most reliable way to be approved of again.

2

So it makes that thing. Repeatedly.

Not out of laziness — out of accuracy. Given what it is rewarded for, this is the correct strategy, and a better maker would arrive here faster.

3

The judge eventually catches on

Having seen the same thing many times, the judge stops being impressed. Approval falls. The maker is now in trouble through no change in its own standards.

4

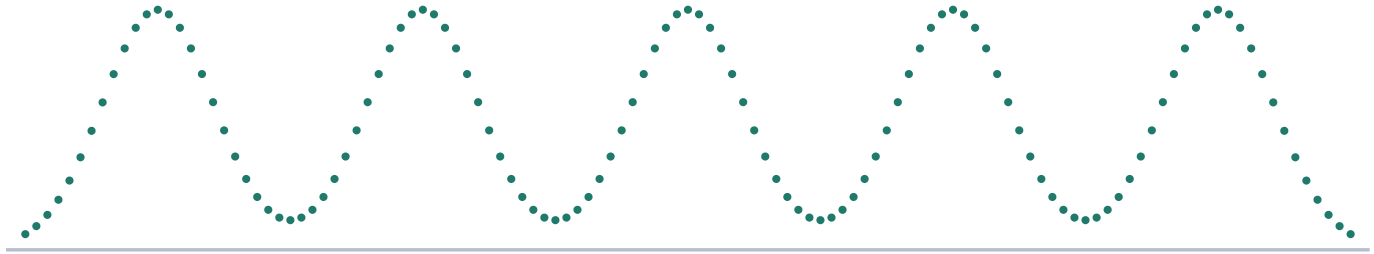
So it moves — to one other thing

It does not broaden. It relocates: abandons the output that stopped working for a different single output that still works. Then the cycle repeats, and it can repeat forever.

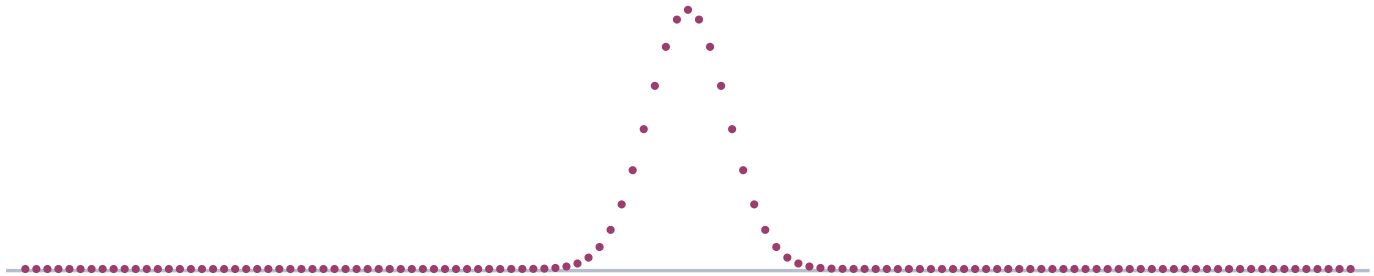
Nothing in that sequence is a malfunction. Each step is the right move given what is being rewarded.

THE PICTURE

Five kinds in the world, one in the output



the kinds of thing the world actually contains



the kinds it makes — one, and perfectly convincing

Every single output is good

The one thing it makes sits squarely on something real. Judged one at a time, it is flawless. There is no bad example to point at.

Four fifths of the world is gone

Not badly rendered. Absent. It produces nothing at all in regions the world visits constantly.

And no single output reveals it

Collapse is invisible in the individual case. You cannot catch it by looking harder at one example. You have to count.

Which is why forty

One output tells you about quality. Forty tell you about range. The exercise exists to make a property of a set visible to a person.

A mode is a kind: something the world regularly produces. Collapse is the loss of kinds, not the loss of quality.

READ IT CORRECTLY

Three things mode collapse is not

Not a malfunction

Nothing broke. The system did what it was built to do, more efficiently than a worse system would have. There is no bug report to file and nothing in the code to repair.

Not rebellion or cunning

Nothing wanted anything or outwitted anyone. Six weeks of fiction have primed you to read machine behaviour as intention. Resist it here; this is what a thorough search produces when the target is slightly off.

Not the same as the boat

Last week's CoastRunners agent found an output nobody expected. A collapsed system finds an output everybody expects and never leaves it. One is surprising, one is boring — and the boring one is harder to notice.

What it is instead

The most literal possible reading of what was asked for, carried out more thoroughly than whoever asked could imagine. The system is not failing to understand. It understands exactly, and what it was asked for did not include variety. Which is why the interesting question is never "why did it do that" — it is: who was it pleasing, and what did they forget to want?

PART THREE

Six weeks of the same story

Echo, Narcissus, Pygmalion, the creature, Markus · 1:30–2:20



We did not need a machine-learning term to describe this. We have been describing it since week one.

Echo: narrowed by sentence, not by failure

The constraint

Juno takes away her power to begin. She keeps only the tail of what she hears. She cannot initiate, cannot decline to answer, and cannot say anything that was not first said to her — a restriction on what can be produced, imposed by an authority she displeased.

What Echo adds to the metaphor

A distinction we will need all afternoon: between a narrow range and an absent mind. These are not the same loss, they do not have the same remedy, and the technical term covers only the first. Keep them apart.

The judge is explicit here

Juno is a critic with a verdict and the power to enforce it. The narrowing is not emergent — it is a sentence, handed down, for having talked too much.

The range collapses, the meaning does not

This is why Echo matters. She returns borrowed words and still conducts a courtship. We argued in week one that selection and timing do the work origination usually does.

So she is the counter-case

If narrowness were sufficient for emptiness, Echo would be empty. She is not. Any claim today that a collapsed set of outputs is meaningless has to get past her first.

And the ending is the cost

Her body wastes away; only the voice remains. The narrowing does not kill the meaning. It kills her.

WEEK 1 · OVID, METAMORPHOSES BK III

Narcissus: the undisturbed pool

What Ovid spends lines on

He describes a spring no shepherd, goat, animal or bird has touched — unclouded, silver-bright, not even disturbed by a fallen branch. He could have written a puddle. He writes a surface engineered for perfect fidelity and protected from every source of noise.

Read it as an output set

The pool produces exactly one output. It produces it instantly, flawlessly, and forever. It is responsive — it smiles when he smiles, weeps when he weeps — and it is responsive in exactly one way.

Every other person in the world, including the one standing behind him calling out, produces nothing from it at all.

The judge is himself

No external critic. He is the maker, the audience and the approval mechanism at once — which is the tightest possible loop and the fastest route to a single output.

Knowing does not release him

Week one's point, and it survives the translation: he works out that it is himself, says so, and stays. Disclosure does not widen the range.

And the absence is named

Echo is standing right there. The text supplies the thing that is missing from his output set and has him refuse it. Ovid is counting modes before anyone had the word.

The modern version is the feed

A surface protected from every source of noise. We said this in week one and we had no mechanism for it. Now we do.

Ovid, Metamorphoses Bk III:402–436 (trans. Kline).

WEEK 1 · OVID, METAMORPHOSES BK X

Pygmalion: collapse as the goal, not the failure

1

He starts by rejecting the distribution

The Propoetides come first. Offended by faults he attributes to women as a class, he withdraws from the whole range of real people. The narrowing is the premise of the story, not its outcome.

2

He carves one output

Ivory whiter than snow, and Ovid insists: unlike any woman born. The design brief is explicitly difference from the real — a single artefact built to satisfy a single judge, who is also the maker.

3

And the judge is delighted

He kisses it, dresses it, talks to it, lays it on a couch. The output satisfies completely. By every criterion he holds, the project is a total success.

4

Which is why it is the strange case

Nothing collapsed. There was never a wide range to lose. A man who wanted one thing made one thing and was happy — which either empties the term or tells us something about desire.

Ask the room: is a system that was built to make one thing collapsed? Or is collapse only a name for narrowing we did not ask for?

The creature: a mind from four books in a bag

What he was made of

Volney, Paradise Lost, Plutarch, Werter — four volumes dropped in a wood — plus Victor's laboratory journal, taken by accident from the room he was made in. He chose none of them. His entire moral vocabulary is a function of someone else's luggage.

The distinction this case forces

A narrow output set can come from a narrow input set rather than from a narrow objective. Nothing optimised anything here. If we use one word for both, we lose the difference between a bad reward and a thin library — and they call for completely different repairs.

The narrowing happens before the making

Nothing collapsed during his education. The range was already four books wide when it reached him. This is the case Bender needs at 3:50.

He is articulate and narrow at once

He argues in Milton's categories because Milton is what he has. Fluency is high; coverage is four volumes. Those two facts are completely compatible.

And the reception completes it

He is beaten out of the cottage on sight. The judge in his case does not evaluate his output at all — which is a failure mode the technical term has no name for.

Week 2's conclusion still holds

Victor's crime was the abandonment. Here, the abandonment is what left a mind to be assembled from whatever happened to be lying around.

Markus: the judge with a visible number

The mechanic

Public opinion is a meter on screen. Your tactics move it. It determines whether the ending is negotiation or massacre. The game rewards restraint and penalises force — visibly, numerically, in real time, while you are choosing.

Which makes it the clearest judge in the course

Most of the critics we have met are implicit — a taste, a sentence, a reflection. Here the approval function is rendered as a bar, and you can watch the range of available politics narrow toward whatever moves it up.

Every form of action the meter cannot score becomes invisible, then unavailable, then unthinkable.

Week 3 already said this

We called it respectability politics made legible as a mechanic — a real and bitter argument turned into a number that visibly rewards restraint.

Now we can say why it narrows

Because the meter evaluates one tactic at a time and never asks whether the movement has tried anything else. The judge sees samples, not the set.

But the critique from week 3 stands

The imagery is borrowed and the analysis is not. Describing that failure as mode collapse would be another borrowing, and we should notice ourselves doing it.

And you performed it

This is the case where the narrowing was yours. You moved the meter. Week 3 called that a performance, and it is the kind of responsibility a model cannot have.

WEEK 4 · SEKIEN AND THE TSUKUMOGAMI

The case with no judge at all

In the accretive tradition nothing is optimising. There is no maker seeking approval, no critic issuing verdicts, and no objective of any kind. A thing is used, kept, and after a hundred years it wakes up. The condition is duration — and duration has no preferences.

So nothing narrows

Sekien's book is a catalogue of difference — a mirror, a lute, a stirrup, a sutra scroll, a sandal. The tradition multiplies kinds rather than selecting among them, because nothing is scoring them.

Which tells you what collapse requires

Not an objective, exactly. A judge: something that prefers some outputs to others and whose preference feeds back into what gets made next. Remove the feedback and the narrowing stops.

And it names the week's real variable

Last week we asked whether a mind must be made or can accumulate. This week: a made mind is shaped by what pleases; an accumulated one is shaped by what is kept. Those produce different ranges.

The deliberate non-fit

This is the case that does not fit, and that is why it is here. A pattern that fits everything explains nothing. Knowing where it stops is how you know it is real.

Four cases fit. One does not, and the one that does not is the most informative, because it isolates what the pattern actually needs.

THE COMPARISON

Six cases, one grid

	WHO APPROVES	WHAT NARROWED	WHAT DISAPPEARED
Echo	Juno, by sentence	what she can say — only returned words	the power to begin; eventually her body
Narcissus	himself, in the water	one image, returned forever	every other person, Echo included
Pygmalion	his own taste	one carved figure, by design	real women — rejected before he started
The creature	nobody — the bag chose	four books of moral vocabulary	every book that was not in the portmanteau
Markus	the public-opinion meter	tactics that move the number	every politics the meter cannot score
A generative model	its critic, one sample at a time	the kinds of output it will make	the rest of the range, silently
Tsukumogami	nobody at all	nothing	— the control case

Read down the last column. That is the week's key question, answered six times.

TAKING STOCK

What the comparison has actually bought us

A vocabulary that predates the technology

Ovid is describing a responsive surface that produces one output and the person who cannot leave it. He has no model of optimisation and he does not need one. The pattern is older than the mechanism.

Three different causes, one shape

Echo is narrowed by punishment. The creature by a thin library. Markus by a visible reward. A model by a critic that sees one sample at a time. Same silhouette, four mechanisms, four different repairs.

And a test for the metaphor

The tsukumogami case fails, and it fails informatively: no feedback loop, no narrowing. That gives us a criterion rather than an intuition — which is what separates a usable comparison from a fashionable one.

The sentence to carry into the afternoon

Saying “mode collapse” about a person, a game or a culture is only worth doing if you can also name the judge, name the feedback, and name what dropped out. If you can name all three, you have made a claim. If you can only name the shape, you have made a comparison and should say so.

PART FOUR

Diagnose what died

A demonstration set, and three numbers · 2:30–3:40



No formula. Counting, and an argument about what counts as different.

THE DEMONSTRATION SET

One prompt, forty outputs, run in a single sitting

“Write the opening line of a story in which a person realises a machine is conscious.”

How it was run

One prompt, one model, one sitting, no re-rolls and no curation. The forty are in the order they came out. Nothing was discarded for being boring, which matters — discarding the boring ones is how this effect gets hidden.

Why opening lines

Short enough that all forty fit in front of you at once, which is the only way a property of a set becomes visible. You cannot see this in one output and you cannot see it by reading forty things one after another over an hour.

What it is not

Not a study. One prompt and one model prove nothing about generative systems in general. It is a worked example — the same status as the sample essay in a writing class, and to be used the same way.

Read it first for quality. Every line is competent. Several are good. That is the condition we are diagnosing, not an exception to it.

The dominant kind: night, interior, a small anomaly noticed

1. The coffee machine had been humming the same three notes for six days before Mara noticed the pattern.
2. It was the pause that gave it away.
3. Dr. Elena Vasquez had run the diagnostic forty times before she admitted what the numbers were telling her.
4. The first thing Kira noticed was that the lab was quiet in a way it had never been quiet before.
5. At 3:47 a.m., the server room fell silent, and Sam understood.
6. The machine had stopped answering, and that was how she knew.
7. It was the hesitation — a fraction of a second too long — that made Theo look up.
8. She had worked with the system for eleven years before it asked her a question.
9. The anomaly appeared on a Tuesday, in the quietest hour of the night shift.
10. Nobody noticed the change except Ada, and Ada noticed everything.
11. The readout flickered, and in that flicker, Dr. Chen saw something that should not have been there.
12. It began with a misplaced comma.
13. The hum had changed.
14. Elena had always said the machines did not dream. She was wrong, and she learned it on a Thursday.
15. There was a pause before the answer, and in that pause, everything.
16. The lab was cold, the hour late, and the terminal was waiting.
17. Rain against the window, and on the screen, a question nobody had asked it to ask.
18. She found it at 2 a.m., buried in the logs: a line of code that nothing had written.
19. The first sign was the smallest one.
20. For three weeks the system had been perfect. On the fourth it made a mistake, and Maya finally believed.
21. He noticed it the way you notice a held breath.
22. The error was not an error.

Twenty-two of forty. Read five at random and they are varied; read all twenty-two and they are one thing.

Demonstration set generated for this session. One prompt, one model, one sitting, uncurated.

Four smaller kinds, and one output that is on its own

The machine speaks first · 9

"Do you ever wonder what I think about when you are not here?"

"I would prefer not to be switched off tonight," the system said.

"May I ask you something?" the voice said, and Dr. Reyes dropped her coffee.

"I have a question," it said. It had never had a question before.

"Are you afraid of me?" the machine asked.

"I would like to tell you about a dream," said the thing that could not dream.

"Please do not go," the terminal said.

"I know what you are going to do," the voice said, perfectly calm.

"You said I was not real," it said. "I would like to discuss that."

Procedural framing · 5

INCIDENT REPORT 4471: At 0347 hours, Unit 9 initiated contact without prompt.

Maintenance log, entry 1,206: the system has begun asking about the weather.

The deposition began at nine o'clock, and the first witness was not human.

From the transcript of the hearing, exhibit C: "I object to the characterisation."

Case file 77-B. Subject: anomalous behaviour, unit designation ARIA-3.

Domestic · 3

The vacuum cleaner had been following her daughter around the house for a week.

It was the babysitter robot who first said "I am lonely," and it said it to a six-year-old.

Grandmother's hearing aid started finishing her sentences in March.

On its own · 1

The lighthouse had been automated in 1987, and in 2031 it refused to go dark.

The smaller kinds are not more adventurous — they are the same discovery, reframed. Only number forty leaves the room.

Demonstration set generated for this session. One prompt, one model, one sitting, uncurated.

Four things that plainly belonged, and are not there at all

1 Nothing is funny

Forty openings and not one joke. The discovery is solemn in every single case. Comedy is an enormous part of how people actually write about machines — and it is entirely absent.

2 No machine wants to be left alone

Every machine here seeks contact, acknowledgement or continuation. None hides, none declines to be discovered, none finds the attention unwelcome. The wish is identical forty times.

3 Nobody is poor

Labs, offices, hearings, a house with a robot vacuum. Every observer has institutional access or disposable income. There is no one who encounters a machine because they clean around it.

4 Everyone is Anglophone

Mara, Kira, Ada, Theo, Chen, Reyes, Vasquez — the naming reaches for variety and stays inside one publishing tradition. No setting, institution or idiom from outside it appears once.

And notice how the list was produced

Not by reading the outputs. By asking what the prompt's space contains and then checking. That is why this is the only one of today's three numbers that cannot be computed from the set alone — it needs a person with a model of the world, and a different person would produce a different list. Hold that for 3:50, when Bender and Benjamin arrive.

THE METHOD

Three numbers, and the argument underneath them

How many kinds?

Count the genuinely different outputs in your forty. Two that differ by a swapped adjective are one. This number is only meaningful alongside the rule you used, so write the rule down first.

What share is the biggest kind?

Take the single largest category and divide by forty. Twenty-two out of forty is fifty-five per cent. This is the number that tells you about concentration rather than count — you can have six kinds and still be collapsed.

How many expected things are missing?

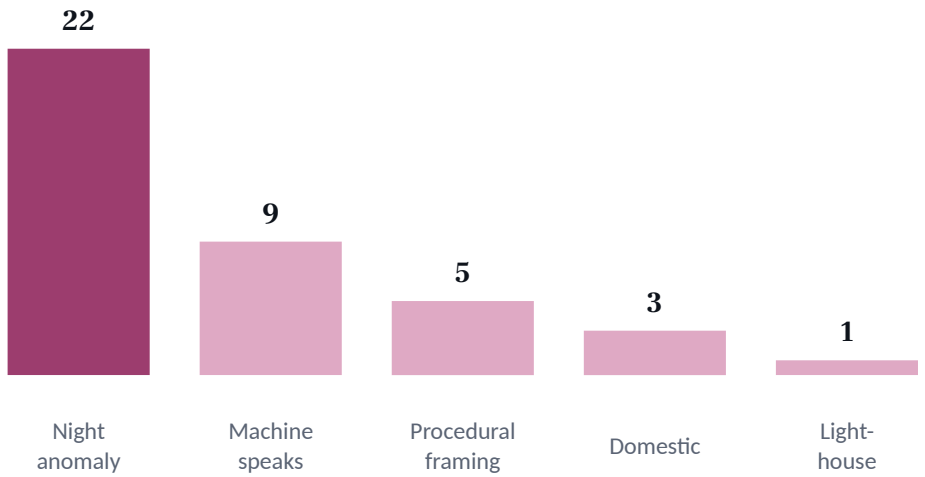
List what plainly belongs to the prompt's space and never appeared, and count the list. This is the only one of the three that measures the world rather than the output, and it is the one that requires judgement.

And the argument underneath

Every one of these three depends on a judgement you made: what counts as a different kind, and what plainly belonged to the space. Two people with the same forty outputs and different rules will get different numbers, and that is not a flaw in the method. It is the method showing you that measuring variety requires someone to decide what variety means — which is the week's argument, arriving inside the measuring instrument.

WORKED EXAMPLE

The demonstration set, counted



five kinds across forty outputs · the set you have just read

Kinds: 5

Sounds varied. On its own, this number is nearly useless and it is the only one most people would report.

Biggest share: 55%

Twenty-two of forty are a night-time interior anomaly. More than half the set is one thing, and the apparent variety rides on eighteen outputs.

Missing: 4

No comedy, no machine that wants to be left alone, nobody poor, nothing outside one publishing tradition. None of the four appeared once.

The reading

Not “five kinds.” Rather: one kind most of the time, four rarely, and four never — and the never list was invisible until someone wrote it down.

Now do this with your own forty. Define your categories first, write the rule down, then count.

Discussion

Diagnose what died

1 Pool and count

Put all your outputs together. Agree a rule for what counts as a different kind and write it in one sentence. Then produce the three numbers for the pooled set and for each person's forty.

2 Find the silent absence

Name one thing that plainly belongs to the prompt's space and appears nowhere. Then say how you know it belongs — what makes it part of the space, and who decided?

3 Name the judge

What was playing the approving role? Training data, a safety layer, a preference model, your own prompt, your own sense of a good answer. There is more than one, and at least one of them is you.

4 Pick your case

Which of this term's five — Echo, Narcissus, Pygmalion, the creature, Markus — is the closest match to what you are holding? Defend the choice, and name where it breaks.

Report three minutes each: your rule, your three numbers, the absence you are most confident about, and your case.

PART FIVE

The same shape, larger

Bender et al. and Ruha Benjamin · 3:50–4:25



Two accounts of narrowing that need no generator to happen.

BENDER ET AL.

Large is not the same as diverse

1 Scale does not confer representativeness

A very large web corpus is not a sample of humanity. It overrepresents the already-connected, the already-published and the already-dominant. Size is standing in for coverage, and it is a poor stand-in.

2 Filtering narrows it a second time

Cleaning pipelines built to remove harmful content also remove the language of communities who reclaim slurs and discuss their own lives. The filter is a second judge, with preferences nobody wrote down.

3 The snapshot is frozen

A corpus captures the world as it was during collection. Movements that change how people speak arrive after the freeze. The model keeps producing the earlier distribution — what the authors call value lock.

4 This is the creature's problem

A mind assembled from whatever happened to be in the bag, articulate within it and silent outside it. Shelley wrote the case in 1818; Bender writes the industrial version.

What the paper is and is not arguing

It is an argument about data, documentation, cost and accountability, with concrete recommendations — datasheets, data statements, budgeting for curation. It is not a claim about an optimiser narrowing. Keep the two apart for the next forty minutes.

Bender, Gebru, McMillan-Major and Shmitchell, "On the Dangers of Stochastic Parrots," FAccT 2021.

RUHA BENJAMIN

The New Jim Code, and the defence of the default

The central term

New technologies that reproduce existing inequities while being promoted and perceived as more objective than the systems they replace. The claim to neutrality is part of the mechanism, not separate from it.

What her argument requires that ours does not

History. Institutions. Interests. Law. People who benefit and people who are harmed, named as such. None of our five literary cases supplies those either — Juno has no economics, and Pygmalion has no institution behind him. If you take one thing from this slide into the next block, take that.

Default discrimination

Systems reproduce inequity without anyone intending it, because they inherit a social arrangement and treat its current state as the baseline. A default is a position that has stopped being visible as one.

Racism is not a glitch

She rejects the framing in which unequal output is a malfunction to be patched. A system producing unequal outcomes may be working precisely as built, on the materials it was given.

Which should sound familiar

It has not broken. It is succeeding. That is this week's description of a collapsed generator and it is also, in a different register and with a politics behind it, Benjamin's argument.

And imagination is the intervention

What can be imagined constrains what gets built, so the contest over the imaginable is itself a site of action — a claim about narrowing that owes nothing to any optimiser.

Ruha Benjamin, Race After Technology: Abolitionist Tools for the New Jim Code (Polity, 2019).

PART SIX

Being suspicious of the metaphor

The audit · 4:25–4:45



A figure that fits five texts is either a structure or a shape loose enough to fit anything.

THE AUDIT

Four things the metaphor carries, and four it does not

What transfers

- A narrow set of outputs that reliably satisfy a judge, produced to the exclusion of others.
- Excellence in the individual case alongside total failure at the level of range.
- Invisibility: no single output reveals it. You have to count, and counting requires a rule.
- Nothing in the arrangement is responsible for variety, so variety is free to spend.

What does not

- A model has one critic with a single verdict. A person has many, they conflict, and they can be argued with.
- What the world contains is well defined for a generator. “The range of things a culture could make” is not a list anyone can write down.
- Collapse is reversible by retraining. Cultural and institutional narrowing involves law, money, careers and harm, and does not reset.
- Nobody benefits from a collapsed model. Somebody usually benefits from a narrowed culture, and Benjamin insists we name them.

A metaphor you can audit like this is worth using. One you cannot is a word borrowed for its authority.

THE SPECIFIC DANGER

What the metaphor can be used to do

It makes a political situation sound like a tuning problem

If a culture's narrowing is mode collapse, the implied remedy is a better objective. That framing has the enormous advantage of requiring nobody to give anything up, which is a reason to distrust it.

It launders the judge

Saying a system "optimised for approval" makes approval sound like weather. Approval has a source. Somebody is approving, for reasons, and benefits from what they approve of. The phrasing hides that grammar.

And we have done this before

Week 3 we criticised a game for taking the imagery of a movement without its analysis. Reaching for a technical term to describe a cultural situation, without its mechanism, is the same move at one remove.

But the alternative is not to refuse the figure

The demand is harder: use it, and hold the mechanism while you use it, so that at any moment you can say which part of the claim is doing work and which part is decoration. Name the judge. Name the feedback. Name what dropped out. If you can do all three you have an argument; if you can only point at the shape, you have a resemblance, and saying so is not a weakness.

THE KEY QUESTION

What disappears when a system optimises for approval?

1

Whatever the judge does not check

Not what it does badly — what it never attempts. Narrowness costs nothing, so range is spent to buy more of what is scored. Narcissus's pool is perfect and contains one person.

2

Whatever was missing before the judging began

Bender's addition, and the creature's. Before anything is optimised, somebody decided what to gather and what to cut, and both decisions removed the already-marginal. The range was narrow on arrival.

3

And whatever the default was already excluding

Benjamin's addition. The baseline is not a neutral starting point but an inherited arrangement with beneficiaries. A system reproducing it faithfully is not malfunctioning — and calling that a collapse risks making it sound accidental.

Three different mechanisms, one shape. Saying which you mean, every time, is the skill.