

WHAT THE BOX KEEPS

Why the size and shape of a container decides what a machine deletes from your writing

Michael Diener

CRCP 6390 Data Expression · Southern Methodist University · Professor Burke Jam

August 2026 · Working edition

Status. Every claim in this paper is a prediction with a stated failure condition. None of it has been measured. Read it as an instrument to run, not a result to cite.

01

The ground rules

Euclid opened his geometry book by defining his terms, listing the moves he was allowed to make, and stating the rules of reasoning he would accept. Only then did he prove anything. He did it in that order so a reader could check his work rather than take his word.

This paper follows the same order. Here is what the words mean, what I am allowed to do, and what would count as proof.

The words

Container. A rule about how much writing may occupy a position, and how those positions relate to each other. One hundred words is a container. So is "two halves of equal length." So is "three parts, and the third must disagree with the first two."

Forced cut. What happens when the contents exceed the container. Something has to go.

Fingerprint. The pattern in what gets cut. It is not random, and that is the whole claim.

Reshaping. When the machine does not delete the awkward part but quietly rewrites it to fit. This is the dangerous one, and almost nobody looks for it.

Works. A stranger with no background reads the output and can say what changed and what it cost. If they cannot, the design failed. Not the reader.

What I am allowed to do

1. Change only the instructions I send. No retraining the machine, no touching its internals.

2. Use only rules that can be counted. If I cannot verify compliance by counting, it does not count as compliance.
3. Ignore sound. Rhyme and rhythm do not survive the way a machine reads text, so nothing here rests on them.
4. Work with sixteen containers, not ninety-four. A small set that demonstrates the routing beats a catalogue that lists entries.

What counts as reasoning

5. When capacity drops below contents, selection is forced, and forced selection follows the shape of the container.
6. A thing missing from the historical record tells you about the record before it tells you about the thing.
7. Building something is not proving it works. These are two separate claims and I will not run them together.
8. If a non-expert is confused about the purpose, that is a fact about my design.

The claim itself

Euclid states the target before he starts building, so you can tell at the end whether he hit it. Here is the target.

Poetic metres are rules about how much information fits in a position. Because they are, applying one as an instruction to a machine changes which parts of your writing survive, and the losses fall in a pattern you can name before you press send.

Two claims live in that sentence and Euclid forbids collapsing them.

I built a game. It exists, it runs, and a stranger can play it and come away naming six specific ways a machine ruins a rewrite. That was to be done.

Metres work as containers with predictable fingerprints. That was to be shown — and the game does not show it. The game teaches the prediction. It does not test it. Anybody who reads the game as evidence has confused a thing that was made with a thing that was proved, and the rest of this paper exists to stop that.

02

The argument, with every joint exposed

Stephen Toulmin noticed that real arguments are not tidy syllogisms. They are structures with six named parts, and most of them fall over because the invisible bridge between the evidence and the conclusion was never built. So here is the whole frame with every part named, including the two I

cannot fully carry.

The claim, stated with its hedges attached

In most cases, above a certain machine size, giving a length-and-shape rule to a language model changes what it keeps and what it drops, in a direction you can predict from the rule.

The evidence

Jason Wei and colleagues ran the decisive test almost by accident. They wanted to know whether "think step by step" works because it gives the machine more room to work in. So they replaced the reasoning steps with a row of dots, matched to the same length. Same extra room, same number of chunks, no reasoning. Performance did not move. Then they tried generating the reasoning after the answer instead of before. Performance fell back to baseline again.

Two findings there, and both matter. Chopping text into pieces does nothing by itself. And the structure has to be doing the work during the writing, not decorating it afterward.

The human-side version comes from Christina Bartl. She found that thinking out loud does not improve problem-solving. Structured self-questioning does. The pattern of the talking is the active ingredient, not the talking.

Erik Altmann's memory research supplies the decay curve. Instructions are not stored in a perfect stack that holds until the job is done. They are ordinary memories that fade and get crowded out. That is why forms built on repetition survive long documents and terse ones drift.

And then there is the catalogue: ninety-four verse forms from every literate and oral tradition on record. Different continents, no contact, and the same handful of moves keep appearing. Three thousand years of people solving a storage problem under a memory limit instead of a token limit.

The bridge

Here is the load-bearing sentence, stated so it can be attacked directly.

If capacity is fixed below what the contents need, selection is forced. If selection is forced under a stated rule, the outcome follows from the rule.

That is why this is a claim about mechanism rather than resemblance. A metre and an instruction are not two similar things that remind me of each other. They are the same operation running on different material.

What holds the bridge up

Euclid gives the deep form in Book Ten. Take away more than half of what remains, repeat, and the leftover eventually falls below any limit you name. Bounded reduction with a guaranteed rate and a stopping condition. Compression with a receipt.

Herbert Simon gives the second beam: solving a problem means representing it so the answer becomes visible. He also observed that when you push a designed thing past what it can handle, the properties of its internal machinery start showing through. That is exactly what the fingerprint is.

Norbert Wiener gives the third: in any communication system, the constraint is not decoration on the channel. The constraint is the control.

The hedges, and every one of them is load-bearing

Presumably. In large machines only. In the counting-and-position channel only. And not yet measured.

Wei's data show the effect appears only above roughly a hundred billion parameters. Run any of this on a small local model and you get fluent nonsense that proves nothing. Nothing in my corpus has been through a controlled test. These are predictions with named failure conditions, and calling them findings would break the discipline the whole program was built on.

What would prove me wrong

Four things, and I would rather name them than have someone find them.

One. The plain-prose control wins. Take the same text, chop it into the same number of pieces at the same length, no special rules. If it performs identically, then chopping explains everything and my claim dies. This is the cheapest test available and it is the one I have not run.

Two. The name does the work, not the shape. Strip the historical label off and keep only the counting rules. If the effect vanishes, the machine was responding to cultural memory in its training data, not to structure.

Three. Small machines. A failure below the size threshold is a fact about the machine, not about the method.

Four. The machine cannot see it. Machines read text in fragments that do not line up with syllables. Any rule touching characters or syllables can fail mechanically while looking like it disproved the idea.

The Houston move

Toulmin's best example is Charles Hamilton Houston, who did not dodge the precedent that stood against him. He took *Plessy v. Ferguson* and forced compliance with it until it broke.

The strongest objection to everything above is that containers distort data. Squeeze text into a shape and the machine will invent balance where none exists, manufacture a disagreement when the evidence is one-sided, and delete the condition before it deletes the claim.

That objection is correct. It is also the finding. Reshaping is the failure nobody demonstrates, because every demonstration counts only what disappeared. This paper is built on the objection

rather than around it.

03

The build, graded honestly

Dieter Rams wrote ten principles for good design. Used as a grading sheet rather than a compliment, they produce deductions. This section is written to produce them.

The thing under review is MetreQuest. One HTML file, nothing to install. A knight's wife and daughter have been taken and the road to them has twelve stops. At each stop the player picks one of six tools to fix a piece of writing. English and Spanish, phone or laptop, online or off.

Principle	Verdict	Finding
Innovative	Pass	Not a slideshow about poetry, not a chatbot in costume. The argument is the game mechanic rather than the tutorial text. Small technical novelty, real structural novelty.
Useful	Pass	Twelve stops, six tools, one decision each. A player who never opens the glossary still leaves able to name six things that go wrong when a machine rewrites something.
Aesthetic	Pass	A sprite is content allocated across a fixed grid and a palette is a controlled inventory, so the art runs the same logic the argument describes. Sprites scale to seventy percent on phones and redraw on rotate.
Understandable	Pass	Drove the third rebuild. Version two had twenty-five encounters and twenty-two cards, and a playtester could not say what it was for. Now: one choice per stop, tools named after actions, glossary always open.
Unobtrusive	Pass	No experience points, no party members, no equipment slots. Every game convention that survived is carrying an argument.
Honest	Partial	Three deductions, itemised below. This is the largest loss in the report and it is named here rather than discovered by a reader.
Long-lasting	Pass	One file, no dependencies, no build step, nothing to go stale. It opens from an email attachment.
Thorough	Pass	Breakpoints at 900, 620 and 400 pixels. Fourteen-pixel tap targets. iOS landscape text inflation suppressed. Fixed-width canvases removed. State saved, so closing a tab does not cost the run.
Environmentally friendly	Pass	One file, no content network, no asset pipeline, no tracking. The only network cost is the model call, and the offline path means it does not require one. The game does not show that cost, which it could.

As little design as possible	Pass	Nine screens, then twenty-five encounters and twenty-two cards, then twelve stops and six tools. Each cut removed something I liked. Arrived at by subtraction rather than by plan.
------------------------------	------	---

Where it fails the honesty test

The metres are hidden. Six tools carry action names and the underlying forms never appear. A player can win the game without meeting the word "metre." Readability went up and the academic claim went under the hood.

The compliance check did not ship. The design called for the game to count, after each rewrite, whether the rule actually fired, and post the number to the player, declaring the result meaningless below eight in ten. It does not. It says the fault was fixed. It does not show the arithmetic. That gap is the difference between an instrument and a poster.

Output quality is two-tier, and the disclosure is incomplete. Online, the model returns real prose. Offline, the fallback chops words, and the Spanish fallback chops them worse. The build does name the mode — a fixed badge reads MODEL: LIVE or MODEL: OFFLINE FALLBACK, the HUD carries the same status bilingually, and an offline rewrite is annotated as having run on the player's machine. What it never says is that the offline result is worse. The player is told which mode is running, not what that mode costs.

04

Why one deletion becomes a problem

Start small. Somebody shortens a paragraph and one qualifier disappears. "Sales rose twelve percent, though two of four regions declined" becomes "Sales rose twelve percent." The result reads better. Nobody notices anything left, including the person who asked, because fluent output feels accurate.

Now multiply. The same operation runs across a company's meeting notes, a newsroom's summaries, a hospital's discharge instructions, a court's briefs. Each individual cut is defensible. The direction of the cuts is not random and it does not average out. Every summary leans the same way: toward the confident claim, away from the condition that made the claim true.

After a year you have a body of documents more certain than the evidence underneath it, with no single decision you can point at as the error.

The fix is not better instructions. The fix is two columns instead of one. The first holds what survived. The second holds what left, split into two kinds: things that vanished, and things that got bent to fit. Almost nobody counts the second kind.

Do this and the world changes in one modest way. Every automated rewrite arrives with a receipt. You still lose things. You lose them on purpose, you know what you lost, and you can argue about whether the trade was worth it. That is a low ceiling and it is achievable now, which is more than most claims in this field can say.

05

What I would do next, in order

Ship the compliance check. The number is the difference between an instrument and a poster. I cut it for time.

Run the fifteen-cell test. One source passage, five conditions, three containers, one afternoon. The plain-prose control either kills the claim or clears it. Right now the strongest honest thing I can say is that I have not looked.

Surface the metres somewhere optional. A glossary entry naming the form behind each tool costs nothing in readability and puts the thesis back where a curious reader can find it.

The thing I would not change is the order: argument locked, then defended, then code. Three rebuilds happened because the argument kept sharpening, and every rebuild was cheap because nothing structural depended on the last version's code.

The open question, stated plainly. Does the shape of the container do the work, or does merely chopping text into pieces explain everything? One afternoon and one passage settles it. Until somebody runs that test, including me, this is a well-supported guess with a clear way to be proven wrong. That is the strongest thing I am entitled to say, and I would rather say it than say more.

06

Sources and links

The project

MetreQuest, playable build <https://michaeldiener.ai/play/metrequest/>

Source library, NotebookLM, 59 sources

<https://notebooklm.google.com/notebook/e3192af6-faaa-49b4-839c-e4fbb9301d3b>

Prison-yard haiku artifact <https://littlebleenhat.github.io/>

Metre page <https://littlebleenhat.github.io/meter.html>

Framing

Dieter Rams, Ten principles for good design, Vitsoe <https://www.vitsoe.com/rw/about/good-design>

Euclid, Elements, Fitzpatrick translation of the Heiberg text

<https://farside.ph.utexas.edu/Books/Euclid/Elements.pdf>

Stephen Toulmin, The Uses of Argument, Cambridge University Press, 1958. Print.

Evidence

Wei et al., Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

<https://arxiv.org/abs/2201.11903>

Madaan et al., Self-Refine: Iterative Refinement with Self-Feedback

<https://arxiv.org/abs/2303.17651>

Bartl, C. Towards the Relation between Language and Thinking. Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society, 1999.

Altmann, E. M., and Trafton, J. G. Memory for Goals: An Architectural Perspective. Same proceedings, 1999.

Simon, H. A. The Sciences of the Artificial, 3rd edition. MIT Press.

Wiener, N. Cybernetics: or Control and Communication in the Animal and the Machine. MIT Press.

Kelleher, J. D., and Tierney, B. Data Science. MIT Press. Course text.

Coeckelbergh, M. AI Ethics. MIT Press. Course text.

Program documents

FM 0-1 Mechanism Operating System. FM 0-2 Mechanism Operations. FM 6-2026

Metre-Mechanism Prompt Context Loop Engineering. FM 7-2026 Evidence Integration and Model Review Board. Metre-Mechanism Matrix v3. Cross-Cultural Atlas of Poetic Metre. Global Poetic Metres Index. Interrogation Battery III: A Euclidean Synthesis of Ninety-Four Poetic Metres. The Metre-Mechanism Prompt Engineering Guide.

Course

CRCP 6390 Data Expression, Southern Methodist University.

Professor Burke Jam bjam@smu.edu · burkejam.com

Set in one typeface at three sizes. No decoration, no rules that do not separate something, no colour that does not mean a link. Ten principles were the grading sheet for the game; they were also the grading sheet for this document.