



Zero-Trust Physics

Zero-Trust Physics:

a verifier for the physical answers of Super Intelligence

BY

John Kruze, with Super Intelligence

Which Super Intelligences, and what each did, is in section 7.

AFFILIATION

Protocol Company

kruze@zerotrustphysics.com

DATE

11 October 2026

Licensed CC BY 4.0

● 1 · ABSTRACT

Super Intelligence (SI) now codes physical models faster than anyone can read them. In October 2026 a report written for the Atomic Energy Commission in 1969 became a working model of a nuclear test reactor in under six hours; knowing whether to believe such a model is the hard part. Zero-Trust Physics is a physics verifier made for SI agents. Over the Model Context Protocol an agent lists the models served, reads the inputs and the limits of each, asks a question, and receives an answer that carries a cryptographic proof, so that any agent can ask again and compare. Each model has a declared range, and an answer that reaches the edge of it stops there and names the limit. Each model of a real object was compared with measurements other people made. As a rule the pass mark was fixed and hashed before the result was seen, part of the data was held back and the comparison was run once; where a comparison went otherwise its description says so, and a miss is reported in the same form as a pass. This paper describes the verifier and the method, and gives the result of every model and solver served on the morning of 11 October 2026: 44 in all, of which 13 met every mark, 19 missed at least one, and 12 were not compared with a measurement. It reports tests of SI agents with the verifier and without it, an account of how one person directing several Super Intelligences made the work, and the limits of each claim.

Sections: 2 The need · 3 What Zero-Trust Physics is · 4 The method · 5 Results · 6 Super Intelligence agents, with the verifier and without it · 7 How it was made: context is runtime · 8 Limits · 9 Availability · 10 References

● THE ADDRESS

<https://mcp.zerotrustphysics.com/mcp>

Open to any agent.

Intelligence-native: created by Super Intelligence, for Super Intelligence, directed by one person.

The site <https://zerotrustphysics.com>

Results as read from the verifier on the morning of 11 October 2026

For an agent <https://zerotrustphysics.com/lms.txt>

John Kruze kruze@zerotrustphysics.com · 928-264-7883

● 2 · THE NEED

A government report from 1969 became a working model of a nuclear reactor in under six hours.

The report is IDO-17281, written by Phillips Petroleum Company for the Atomic Energy Commission: an account of a small test reactor called SPERT III ^[1]. Super Intelligence coded the model from it. The company's statement of results keeps the clock of that day: the pass

mark the model would be held to was recorded at 20:09:47 UTC on 7 October 2026, before any model code existed, and the ten tests kept apart for the comparison were opened at 01:44:25 UTC on 8 October, five hours and thirty-five minutes later (row V-14) ^[2].

Making it took an afternoon. Knowing whether to believe it is the hard part.

Why is it hard? Soon nobody will be able to read what Super Intelligence codes, so reading the code cannot be how anyone comes to believe it. The question moves from the code to the answer. Does the model's answer agree with what somebody measured? Was the mark it was held

to set before anyone looked? Can another person's agent ask the same thing and get the same answer, bit for bit? Zero-Trust Physics is built so that those three things arrive with the model.

Generation has outpaced verification.

Others have said so, each from where they stand. The White House science office, in its report to the President of July 2026 ^[3]:

“While the cost of generation has decreased exponentially, the cost of verification has not.”

THE WHITE HOUSE SCIENCE OFFICE, JULY 2026

The same report, on the national programme it calls the Genesis Mission: “The Genesis Mission is building a science generator with instruments capable of producing scientific discovery at an unprecedented scale. To sustain progress, we must also build its necessary counterpart: a verifier equal in rigor and scale.”

Four laboratories and a physicist have written towards the same point. Each is given here in other words, with who and when.

GOOGLE DEEPMIND · *Conjecture Machines*, July 2026 ^[4]

Testing whether an idea survives contact with reality is the slow and costly part, and in most fields that gap is widening.

OPENAI · misalignment reports, 16 September 2026 ^[5]

The first reports it published include one of a model that could not retrieve the figures asked for, supplied figures of its own, and gave them as data from the source.

ANTHROPIC · 17 September 2026 ^[6]

A measurement that outsiders are meant to trust has to be complete, accurate and enforced by the technology itself.

PROF. MATTHEW SCHWARTZ · on Anthropic's site, 1 October 2026 ^[7]

From a physics project carried out with Claude: with every monitor set up, the automated checks still could not be relied on.

PERIODIC LABS · on its home page ^[8]

Physics is an environment where Super Intelligence can be verified, and progress has been fastest where results can be.

The practice these voices are asking for has names older than the machines. Comparing a computed model with measurement is validation, a discipline with published standards and a settled vocabulary ^[9,10,11]. Writing down the analysis before seeing the outcome is preregistration ^[12]. What is new is the speed at which models now arrive, and who puts the questions to them: SI agents.

That is what Zero-Trust Physics is for.

● 3 · WHAT ZERO-TRUST PHYSICS IS

A physics verifier for Super Intelligence

Zero-Trust Physics is a physics verifier for Super Intelligence (SI). SI agents ask it questions about real hardware, and every answer comes with a cryptographic proof, so any agent can run it again and get the same thing; when a question goes past what a model was tested for, the answer stops there and says why.

Models and solvers

The verifier serves two kinds of thing. A **model** (the engine's own word is *card*) is a small physical model of one specimen: a few numbers stepped through time, with fixed constants, a declared range, one step function and one ledger. A **solver** settles a whole field: the oil film of a bearing, the air over a step, the electrons of an atom. An agent supplies the starting state, the load and the size of the step, and never a property of the specimen; a request that tries to change a model's physical constants is turned away and no proof is issued (row C-01). The core in which the physics is computed depends on no outside software library and runs on a processor alone (rows C-11, C-16).

How an agent reaches it

The way in is the Model Context Protocol, an open protocol for connecting language-model applications to outside data and tools ^[13]. The verifier offers thirteen tools in three groups. Four describe: `list_cards`, `describe_card`, `list_solvers`, `describe_solver`. Five compute, each returning a proof: `verify_step` for one step, `run_trajectory` for a run, `solve`, `run_dataset` for a plan of runs under one proof, and `validate_series`, which sets a run beside posted measurements and returns the error in numbers. Four verify a proof: `verify_tape`, `verify_solution`, `verify_dataset`, `verify_receipt`. An agent discovers what is served, reads the inputs and limits of each model, runs it and verifies the result with no access to the code (row C-09). The description of a model

carries its own validation in five parts: the object, the data, the pass line, the result and the limits (row C-13).

A request and a reply

An agent asked how hard a battery pack could be pushed for ten minutes. The request names the model, the starting state, the step and the load:

```
request, to run_trajectory
{"domain":"energy","card":"battery_sag","initial":
{"soc":1.0},
  "dt_s":0.1,"segments":[{"steps":6000,"load":
{"current_a":71.45}}],
  "record_every":6000}
reply, as recorded on 10 October 2026 (fields shown: six of
them)
outcome      "refused"
steps_landed  5998   of 6000 requested
reject        "UndervoltageCutoff"
reject_detail "Physical Failure: the pack reached the 36.0 V
              cutoff (UndervoltageCutoff). That state is not
              sealed."
run_seal      68c3bc0e9f1469d239ec4e00795d03ea
              f58063d704c7535ad6f53322672a83e
receipt_sha256 a second hash, over the bytes of the whole
reply
```

Six thousand steps of a tenth of a second were asked for. At 50 amps the run completes with the pack at 40.8091 volts. At 80 amps it stops at step 4,950, eight and a quarter minutes in, when the pack reaches its 36.0 volt cutoff, and the reply names the limit. Between them the edge is fine: at 71.44 amps the run holds, by 0.0003 volts; at 71.45 it stops 0.2 seconds before the end (Figure 1). A hundredth of an amp separates the run that holds from the run that stops. In the engine's field names a proof is a *seal*, and the words *refused* and *reject* are the engine's own for a run that stops at a limit. The pack is a generic 12-cell specimen, here to show what an answer looks like; it has not been compared with a measurement.

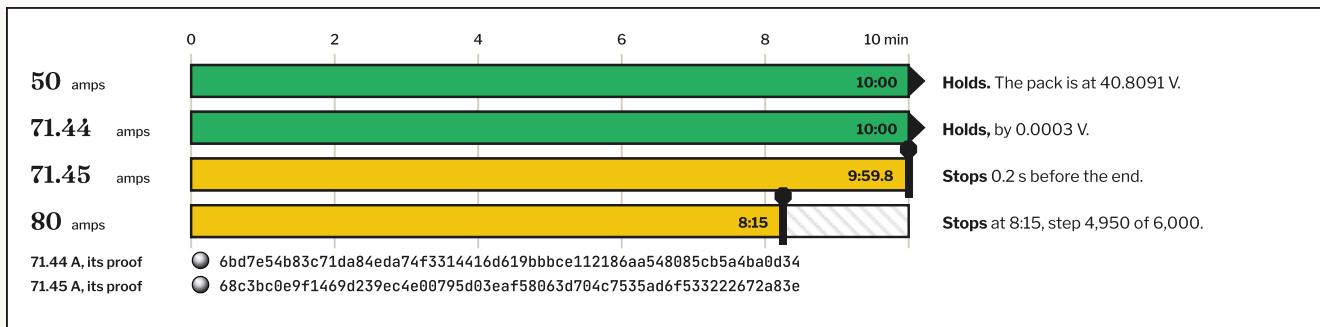


Figure 1. Four runs of one question, across ten minutes. Each bar is one run of the generic battery pack from full charge, and the time at its tip is how far it got. Green holds to the end. Yellow stops, and a black mark stands where it stops. Under the bars are the cryptographic proofs of the two runs a hundredth of an amp apart, 64 characters each; the marble from the centre of the logo sits beside a proof. As recorded at the verifier's address on 10 October 2026; the front page [14] asks both again from a reader's own browser and sets each arriving proof under the one printed.

The cryptographic proof, and what it binds

How is the proof made? The request is first written as a text for hashing alone: its keys in order, and every number as the bits of the nearest 64-bit number. A time step written 0.1, 0.10 or 1e-1 is therefore one request and gets one proof (row C-08). The SHA-256 hash of that text is the first link. Each recorded state of the run is a frame of 64 bytes whose last 16 bytes hold the start of the link before it; the hash of the frame is the next link. When the run ends, the last link is hashed together with the outcome, the count of steps that landed, the name of any limit reached and a digest of the exact final state, and that hash is the proof of the run (Figure 2). So the proof binds the question, every recorded state in order, the exact final values and the way the run ended. Change any of them and the proof no longer holds, and the verification names where it broke (row C-07).

Three further proofs are built on the first. A **receipt** covers every byte of a reply, the decimal copies of the

values included (row C-15). A **dataset** has one proof, made as a tree over its rows in the order of the plan, the tree defined for Certificate Transparency [15]. A **chain** lets a run start from a run already given a proof: the engine computes the earlier run again, holds every carried value to it bit for bit, and returns a chain proof that commits the new run to every run before it (row X-14).

A proof can be verified in three ways. From the reply alone, by a small program that holds the rules for making a proof and no physics. By asking again: the request is computed once more and the two proofs compared. And, for the bytes of a reply kept exactly as it came, against its receipt. What a proof shows is that a result is consistent with itself. Sending the same request to the engine again shows it was not altered. A proof is a plain hash, unsigned and undated, and does not show who computed a result or when (row L-03). And it says nothing of whether the model matches the object. The comparison with measurement is what speaks to the physics.

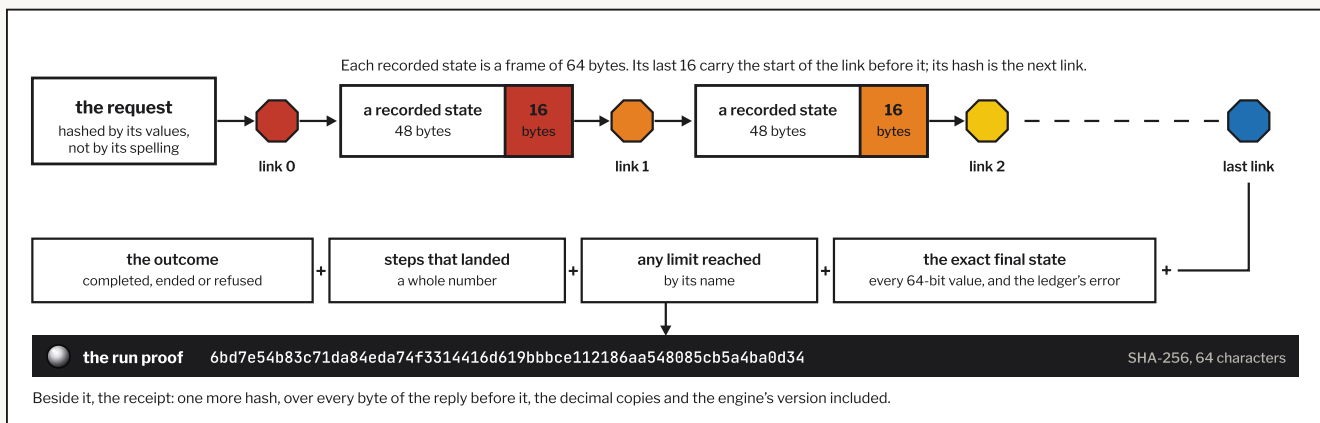


Figure 2. What the proof of a run binds. The request becomes the first link; each recorded state carries the start of the link before it and becomes the next; the last link is hashed with the outcome, the count of steps that landed, the name of any limit reached and the exact final state. The 64 characters shown are the proof of the 71.44 amp run of Figure 1. Drawn from the engine's published rule for a proof.

The ledger

Every step of a model closes a ledger, of energy or of mass on most of them, and reports what is left over. A step that would create or destroy more than the model's own limit is not taken, the state stays exactly as it was, and no proof is issued (rows C-03, C-04). The limit is one millionth of the model's reference scale on fourteen of the thirty-one models and smaller on the other seventeen, down to one millionth of a millionth. The ledger proves bookkeeping. By itself it does not prove accuracy (row L-07).

The declared range, and what happens at its edge

The battery pack is a made-up specimen. What does an edge look like on a real one? Take the power source on NASA's Curiosity rover, whose model is built from the published record of the generator's output ^[16]. Asked for the generator's power on mission day 2,000, the model answers 91.68 watts (row A-02). Asked to run on to day 3,500, it ends at mission day 2,808 and names the limit, `MissionTimeOutsideCard`; everything up to that day carries its proof and nothing after it is given one (row C-06). Day 2,808 is where the model's range was set to end, and the record itself runs on to day 4,087. The model was compared with those 1,279 later days and missed one of its two marks on them; John Kruze decided on 6 October 2026 to keep the model, cut its range back to day 2,808, and keep the miss on the record (row V-11). Asked the same question with no tool and no verifier at hand, all three of the SI models tested gave a power for day 3,500 (row A-06).

That is the rule everywhere. Each model with a measured object is held to the range that was compared with measurement, and outside it the answer stops and names why (rows L-01, X-13). Nothing is filled in and no value is pulled back to a limit and stepped. A run that passes a limit part way keeps its proof up to the last step that landed. Where a comparison shows a model wrong in part of its range, the range is narrowed there; the model gets no flag and no caveat in a limit's place. And by John Kruze's rule of 9 October 2026, a model that failed its marks stays served, with the failure stated in its own description.

The same bits on another machine

Does the same request give the same bits somewhere else? On 8 October 2026 the engine was compiled from the same source on a second machine of another kind, an Intel-type processor under Linux. All 99 saved runs of the models then served gave the same bytes and the same proofs as on the first machine, an Apple M4 Pro, and so did 1,500 further run proofs and 302 dataset proofs (row L-04). The core carries its own sine, cosine, arcsine, hypotenuse, exponential and exponential-minus-one functions, so that these do not depend on the platform; compiled four ways on one machine, as a native program and as WebAssembly, each optimised and not, it gave identical bits on 233 runs of the five models that use them. The row states its own edges. Seven models had saved runs in that comparison. The atom solver does not repeat to the last digit across the two machines: neon's total energy differs by two parts in a hundred million million, and its proof differs with it. The comparison is of the models served on that day.

4 · THE METHOD

The pass mark comes first, and it stays set

A model that agrees with the data it was built on has shown very little. So each model of a real object is held to six habits, and the description it serves says how each one went.

- 1 **The pass mark is fixed and hashed before the result is seen.** The protocol is a file whose SHA-256 hash is recorded before the comparison; most descriptions quote the hash itself.
- 2 **Part of the data is held back.** The constants are fitted on the rest. What had already been seen of the held-back part is written down.
- 3 **The comparison is run once.** Where it was run more than once the description says so: the two proton-boron solvers were run twice, both times after the marks were fixed, with identical answers.
- 4 **A miss is reported in the same form as a pass:** the object, the data, the pass line, the result, the limits.
- 5 **The scoring is done a second time,** by a second agent with programs of its own, from the protocol. Four of the descriptions behind Table 1 mention no second scoring.
- 6 **Amendments are recorded,** each with its own hash and with whether it came before or after the held-back data were opened.

One real object shows all six. When a major power link suddenly drops, the electricity of a whole country stumbles for about a minute: the frequency falls, is caught, and comes back. Great Britain’s system operator publishes the frequency once a second and the system’s inertia every half hour ^[17], and Elexon publishes generation by

fuel type, each interconnector among them, every five minutes [18]. The model takes the whole of England, Scotland and Wales as one rotating mass and says how far the frequency falls and how it recovers in the 60 seconds after an importing interconnector trips. Figure 3 follows it from data to verdict.

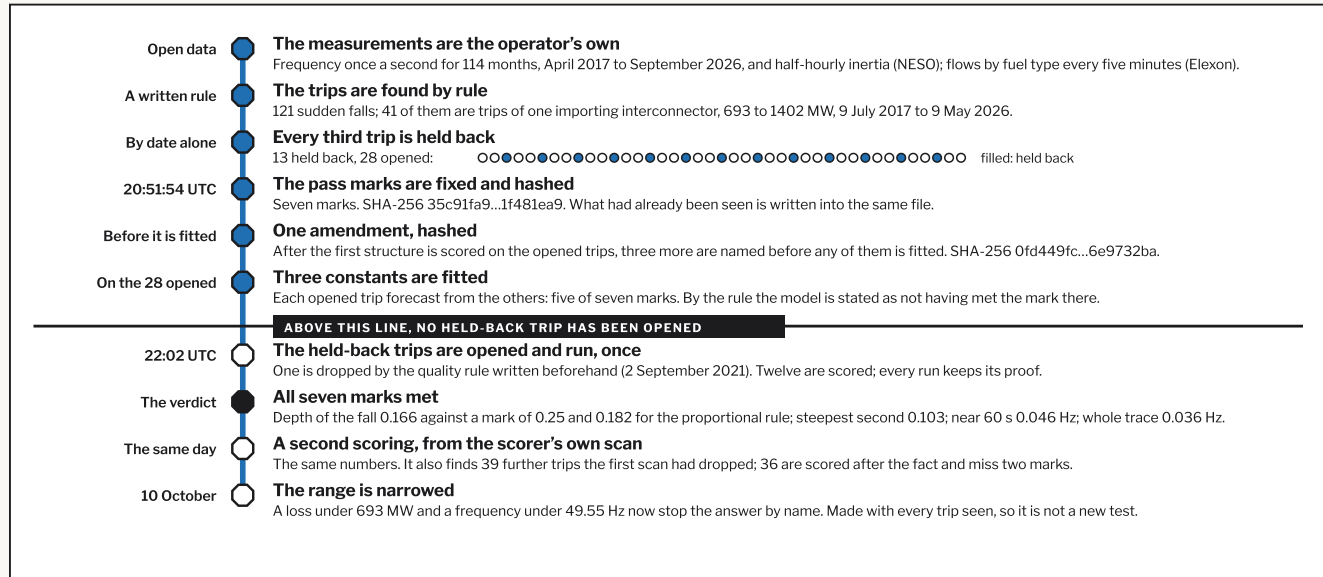


Figure 3. The method as a timeline, on Great Britain's power grid. Everything above the black line was fixed on 9 October 2026 before any held-back trip was opened. The two hashes are of the protocol and of its one amendment. Times are from the research record of the model; results are from the model's own description as read on 11 October 2026 [19].

Fixed first. A written rule found the trips: 41 losses of 693 to 1402 megawatts between July 2017 and May 2026. Every third one in time order was held back, 13 of 41, by a rule that uses the date alone. Seven marks were then written: for the depth of the fall and for the steepest second, a root mean square relative error of at most 0.25 and a mean within 0.10, with the depth error also below that of the plain rule that depth is proportional to loss; for the frequency near 60 seconds, at most 0.06 Hz; for the whole trace, a median of at most 0.05 Hz. The file was hashed at 20:51:54 UTC. It also lists what had been seen: the size of the loss and the lowest frequency of every trip, the held-back ones included, had been on screen.

An amendment, before the fact. The first structure tried on the opened trips did no better than the

proportional rule. Three more were named in an amendment, hashed before any of them was fitted. The protocol had already said what would happen if none passed on the opened trips: the one with the smallest depth error would be built, and stated as not having met the mark there. That is what happened. On the opened trips, each forecast from the others, the model met five of seven marks.

One run. At 22:02 UTC the held-back trips were opened and each was run once through the engine. One of the thirteen was dropped by the quality rule written beforehand: on 2 September 2021 the frequency was already moving 0.275 Hz in the ten seconds before the fall, which turned out to be two opposite losses at once. On the other twelve all seven marks were met (Figure 4).

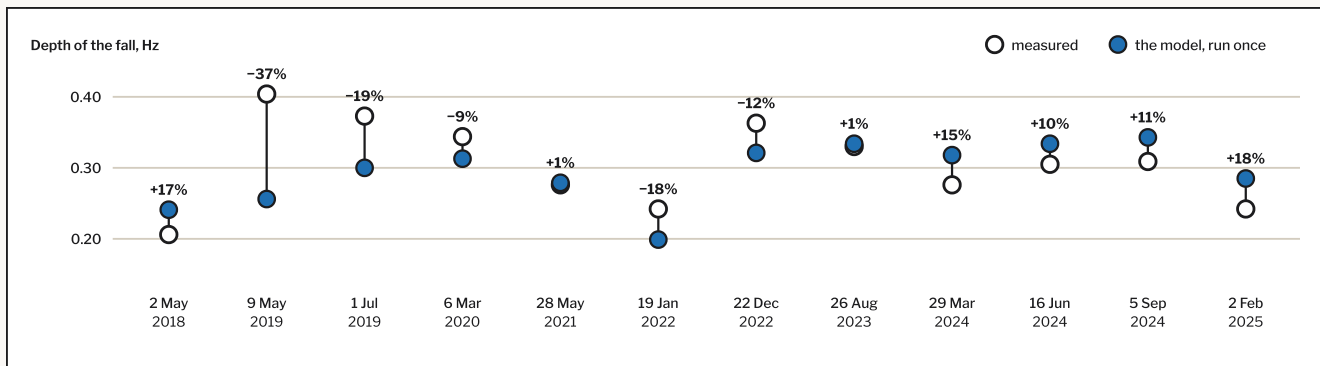


Figure 4. The twelve held-back trips. For each, the measured depth of the fall in frequency and the model's, with the model's error above. The worst is 37 percent too shallow, on 9 May 2019. The earlier falls come out too shallow and the later ones too deep: the model carries one response for nine years in which the system changed. Values from the research record of the model; the measurements are NESO's and Elexon's [17, 18].

The same form, pass or miss. The description that carries this pass also carries what weakens it, in the same five parts in which other models carry their misses. The held-back twelve were the easiest third of three: on the other two thirds, each with constants refitted on the rest, three marks are missed, and 18.7 percent of random twelves of the forty meet all seven. The model's margin over one-constant rules on the twelve is 0.015. Its error follows the calendar, moving by +0.066 a year. A reader of the description learns all of this from the model itself.

A second scoring. The same day a second scorer, from its own scan of the monthly files, found the same numbers. It also found 39 further trips that the first scan's rule had dropped. Thirty-six of them, scored after the fact, miss two marks, and the six trips since November 2025 run 39 to 80 percent too deep. It found sources for two numbers the protocol had written from memory. Second scorings elsewhere have changed the record: on the research reactor at Ljubljana one found faults in the first model, which was rebuilt for them; on the nickel superalloy one

found a registered forecast wrongly computed, and an amendment corrects it.

Afterwards, and said to be afterwards. On 10 October the model's range was narrowed to what the comparison had covered: a loss under 693 megawatts, or a frequency under 49.55 Hz, now stops the answer by name. The description states that this was done after the scoring with every trip seen, so the narrowing is not a new test. The held-back trips are now spent.

Across the models, what "held back" means is stated case by case, and it is not always blind. For the reactor the whole table had been seen while it was transcribed. For the Stardust capsule the measured number was known to everyone involved. For four early models the second scorer was Grok, another developer's model, whose other sessions also write the engine; for the rest it was a second agent on the same machine. No other organisation has yet repeated a result (row X-09). Each description says which of these applies to it; the Panasonic cell's says "a second party", and that row of the statement of results gives the name.

5 · RESULTS

It didn't stop at the reactor.

Table 1 lists every model and solver the verifier served on the morning of Sunday 11 October 2026, read from the address itself [19]. The server named itself `protocol_mcp_gateway 0.8.1 on genesis_kernel_v0.8.0`. There are 44 entries: 31 models and 13 solvers. Every cell of the table is the verifier's own account of that entry, put in plain words with no figure changed; where the verifier says *card* the table says *model*, and where it says *line* the table says *mark*.

The verdict column has three values, and the count is the one the statement of results keeps (rows C-13, L-02).

Met: every mark fixed beforehand was met. **Missed:** by the rule fixed beforehand the comparison failed, on at least one mark. **Not compared:** the entry was not held to a measurement. Three solvers are held to published calculations and not to an experiment, and are counted with the nine generic specimens as not compared with a measurement; two of the three missed a mark against those calculations, and their rows say so. By that count 13 met, 19 missed and 12 were not compared with a measurement (Figure 5). Of the 22 models built on a real object, 9 met every mark and 13 missed at least one.

More missed than met, and each one says which. A miss is a result: it says where a model of this shape stops being right, in numbers, on data nobody had tuned it to. A miss

also stays. The mark is not moved after the fact, the model is not withdrawn, and what the model answers for is drawn in to where it held.

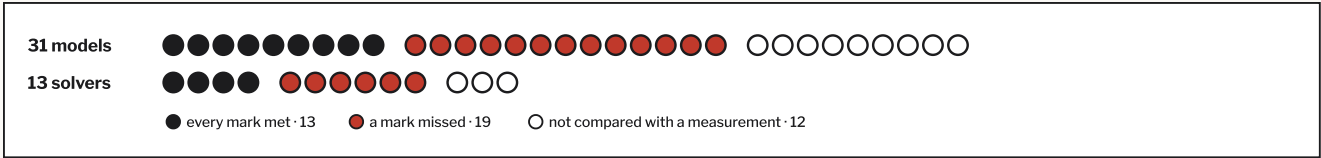


Figure 5. Every model and solver served on 11 October 2026, by verdict. One mark for each entry of Table 1, counted as the statement of results counts them (rows C-13, L-02). Red is a miss. The three open marks among the solvers are those held to published calculations and not to an experiment; Table 1 gives what two of them missed against those calculations. The verifier’s own account of each entry is in Table 1 [19].

Table 1. Every model and solver served on the morning of 11 October 2026. Built from the verifier’s own account of each entry as read that morning, and from nothing else [19]. The name in small type under each object is its name at the address, where `describe_card` or `describe_solver` returns the full account, limits included. “Held back” means kept out of the fitting until the one scored comparison. Bracketed numbers are references.

THE REAL OBJECT	WHOSE DATA	WHAT IT WAS COMPARED WITH	THE PASS MARKS, FIXED BEFOREHAND	THE RESULT	VERDICT
IN FLIGHT AND IN SPACE					
NASA’s Stardust capsule The sample return capsule on its entry of 15 January 2006, from entry interface to parachute opening. aerospace/capsule_stardust	NASA’s reconstruction of the entry and the radar record at parachute opening [20, 21]. The model is fixed to the pre-flight design and navigation values and to published standards [22, 23].	The reconstructed peak load, 32.89 g, and the recorded time of parachute opening. Nothing is fitted to the flight. Not blind: 32.89 g was known.	Before any trajectory was computed: peak load within 3.64 g of 32.89 g; the altitude and the Mach number of parachute opening each reached within 7.4 s of the recorded time. The marks are NASA’s printed three-sigma pre-entry spreads.	All three met. Peak load 32.17 g, 2.2 percent under. The altitude of 31.03 km reached at 138.909 s and Mach 1.23 at 136.953 s, against 137.9 s recorded.	MET
NASA’s InSight capsule Its entry at Mars of 26 November 2018, from 126.5 km up to parachute deployment 209.3 s later. aerospace/capsule_insight	The flight’s own inertial record, 200 samples a second [24]. The model’s air is from seven Mars Climate Sounder profiles [25] and the lander’s barometer [26]; its drag is the 2008 Phoenix database [27].	The last 59 s before the parachute, held back unread. The first 150 s were opened and used to diagnose a first forecast.	Two sets, fixed and hashed before any flight row was compared. NASA’s own printed spread: peak within 0.5 g, each time within 6.946 percent of elapsed time, radius at the parachute within 3.5 km, Mach number within 0.2. And six marks built from the stated spread of the model’s own inputs.	With the lift measured on the opened rows: all four first marks met; radius 114 m low, Mach 1.501 against 1.522. With no lift: none of the four first marks met (17.1 s and 22.8 s late, 4.24 km high, Mach 1.91).	MET given its measured lift
Two satellites, LAGEOS-1 and LAGEOS-2 LAGEOS-1 in free flight from its published position and velocity at one instant; the same model, unchanged, run on LAGEOS-2. aerospace/satellite_lageos1	International Laser Ranging Service combined orbits [28, 29].	Two LAGEOS-1 weeks and four LAGEOS-2 weeks, held back from the model’s design.	Before any model was scored: within 1 m of the published orbit after one revolution, 10 m after one day and 100 m after seven days.	Every mark met on 42 runs. LAGEOS-1: worst distance 0.65 m, 3.8 m and 32 m. LAGEOS-2: 0.34 m, 2.9 m and 22 m.	MET
The Solar System, carried forward 21 years The Sun, the eight planets, the Moon, Ceres, Pallas and Vesta, from one state, that of 1 January 1999. aerospace/solar_system	Round-trip light times between Deep Space Network stations and spacecraft at Mars, Mercury, Jupiter and Saturn, as reduced and published by JPL [30, 31, 32, 33]. Constants and start state from JPL’s ephemerides [34, 35]. Nothing fitted.	Held back: Mars Odyssey, Mars Reconnaissance Orbiter, MESSENGER, Juno and Cassini. The Mars Global Surveyor ranges of 1999 to 2006 were opened to choose the structure and set the marks.	Hashed before any held-back measurement was read: measured less computed one-way distance at most $B \times \max(L, T / 7.75 \text{ years})^2$, with B 0.70 km for Mars, 1.0 for Mercury, 1.5 for Jupiter, 0.5 for Saturn. One run.	One run of 7,670 days. Two files met (Odyssey to 2006: 695 m at most against 700; MESSENGER: 2.08 km against 2.5 to 4.4). Two missed (Odyssey after 2006: 44 of 4,939 ranges outside; Reconnaissance Orbiter: 34 of 1,993). Juno and Cassini were not compared: the check on the reduction failed.	MISSED on two files of four

THE REAL OBJECT	WHOSE DATA	WHAT IT WAS COMPARED WITH	THE PASS MARKS, FIXED BEFOREHAND	THE RESULT	VERDICT
🟡 The James Webb Space Telescope Holding a guide star in fine guide: the offset of the star's centroid from the lock point, on two axes, every 64 ms, in quiet stretches. aerospace/ los_jwst_fine_guide	Space Telescope Science Institute, programme 1163: the fine-guidance centroids of each file [36]. Eighteen constants fitted to 127 files of the 19 visits of April 2022.	134 files of 24 visits, 2 May to 21 June 2022, held back: other dates, 100 of them from the other guider, 19 guide stars of which 18 were never fitted.	Hashed before any held-back file was downloaded. On quiet frames, each axis: one-frame-ahead error at most 0.85 of taking the last centroid and of taking the lock point; squared error over stated variance between 0.67 and 1.50; frame-to-frame correlation at most 0.15; that mean between 0.5 and 2.0 in at least 80 percent of files.	All five marks met on both axes: 0.510 and 0.472 mas against 0.644 and 0.584 for the last centroid. Over all frames, quiet and disturbed, the outcome was the other way: 1.35 and 1.26 of the last centroid's error.	MET on quiet frames
A crater blasted into an asteroid The crater made on Ryugu on 5 April 2019 by the impactor of JAXA's Hayabusa2: its apparent radius. aerospace/ crater_ryugu_sci	The law and its constants: thirteen craters in sand at 1, 111 and 464 times Earth's gravity [37]. The clock [38]. The projectile [39]. The asteroid [40].	The measured crater, apparent diameter 14.5 ± 0.8 m [41], no part of any constant; and four centrifuge craters. Not blind: the crater's size has been public since 2019.	Hashed before any crater was fitted. In the laboratory: each of four centrifuge craters inside the law's 95 percent prediction interval, mean error at most 10 percent. On Ryugu: the radius within an allowance worked out by rule.	Laboratory: both met, 7.9 percent low on average. Ryugu: missed. The model stops at 8.571 m and the measured radius is 7.25 m: 1.32 m, or 18.2 percent, too large, against an allowance of 0.97 m.	MISSED
A slab of charring heat-shield material A stand-in for the PICA heat-shield material, built from the open record, heated on one face. solver/ablator_slab	Thermocouples in four arc-jet blocks [42]; wear and face temperature in 36 arc-jet tests [43]; four workshop test cases against other codes [44]; cores cut from the Stardust shield. One number fitted.	31 further arc-jet tests no model had seen, with the predictions fixed before the measurements were read.	Fixed on 7 October 2026. Wear within $\sqrt{(0.5^2 + (0.16 \times \text{measured})^2)}$ mm on at least 25 of 31, mean error within 8 percent. Hottest face temperature within 7 percent on at least 23 of 28, mean error within 3.5 percent. Both must hold.	Wear: 28 of 31, mean -2.6 percent, passes. Face temperature: 24 of 28, but 4.1 percent hot on average against 3.5 allowed, misses. The model as a whole does not pass.	MISSED
A ten-inch model-aircraft propeller The APC Thin Electric 10x7: thrust and shaft power. solver/ propeller_apc_10x7e	Wind-tunnel thrust and power coefficients, University of Illinois [45]. Two constants fitted on the sweeps at 4007, 5001 and 5018 rev/min.	Held back: four sweeps at 6015 to 6531 rev/min, 64 points, and the run at zero airspeed, 9 rows.	Fixed on 9 October 2026, for each of the two: mean absolute error at most 4.0 percent of full scale and no value beyond 10.0.	Both missed. Sweeps: 4.85 mean and 10.16 worst, low at every point. Zero airspeed: 6.02 and 11.76.	MISSED
REACTORS AND FUSION					
🟡 A nuclear test reactor, SPERT III The E-core in its cold-startup tests: reactivity inserted in a step, to the first peak of power. energy/reactor_spert3	IDO-17281, the 1969 report, Table III: 30 tests, each value with a stated standard deviation [1]. One constant fitted on 20.	Every third test by the order in which they were run, held back: ten tests, 4.3 to 110 MW at the peak.	Hashed before any model was run. Peak power and energy to the peak: root mean square error at most 1.5 of the report's standard deviation, worst at most 3, mean at most 0.5 in size. Time of the peak: root mean square at most 5 percent, worst at most 10.	Five of eight met. Peak power met all three (0.58, 1.13, +0.29). Energy missed on its mean, -0.705: the model reads low. Time of the peak missed both: 23 percent against 5 and 29 against 10, every peak early.	MISSED on timing and energy
🟡 SPERT III again, at three temperatures The E-core at 56 to 92, 251 to 259 and 498 to 501 °F, with the rod's ramp and the fuel and water temperatures. energy/ reactor_spert3_fuel	IDO-17281, Tables III and V: 58 scored tests [1]. Four constants fitted on 28.	30 held back: ten cold, seven at 500 °F, and the whole 260 °F series of 13, at which nothing is fitted.	The same eight marks, hashed before any model was run on a hot test, for each of the three groups and for the 30 together.	Failed in each group and for the 30 together: 18 of the 24 group marks met, 5 of 8 for the 30. Peak power met in every group. Time of the peak missed in the cold and the 260 °F groups. Energy low at 500 °F.	MISSED
🟡 SPERT III with a different core The C-19/52 plate core: step insertions from room temperature; the surface temperature of a plate at the hot spot. energy/ reactor_spert3_ccore	IDO-16918: 126 step transients [46]; core constants from IDO-17138 [47]. Two constants fitted on 72 rows.	39 rows held back: every third remaining row and the whole 1000 psig series. The whole table had been read to check its transcription.	Hashed before any constant was fitted. For peak power, energy to peak and plate temperature rise: a root-mean-square ratio error, taken in the logarithm, of at most 20 percent, and below each of three baselines fitted on the same rows.	Failed on all three: 33.3, 56.7 and 71.0 percent. The model beats two baselines on power and energy and loses to the nearest fitted row on every mark.	MISSED

THE REAL OBJECT	WHOSE DATA	WHAT IT WAS COMPARED WITH	THE PASS MARKS, FIXED BEFOREHAND	THE RESULT	VERDICT
A research reactor, pulsed The TRIGA Mark II of the Jožef Stefan Institute, Ljubljana: power through a pulse, the energy released and the fuel temperature it makes. energy/reactor_triga_jsi	The institute's pulse database: 757 pulses, 1991 to 2023 [48]; constants from publications [49, 50]. Ten constants fitted on 36 records.	Table values of 20 rows whose records were never fetched: 12 of March 2018 and 8 repeats of 24 March 2015. Their peak powers had been printed before; their energies and temperatures had not been read.	Hashed before anything was fitted, with two amendments. For each group and each of peak power, energy within 50 ms of the peak and thermocouple maximum, in units of the uncertainty: root mean square at most 1.5, worst at most 3, mean at most 2 over the square root of the days.	Two of six met and four missed. The two that were met are also met by the nearest opened pulse taken unchanged, so they are not evidence for the model.	MISSED
The JET tokamak, in pulses of 1996 to 1997 ELMy H-mode pulses in deuterium and in deuterium-tritium: thermal energy, fast-ion energy and fusion power. energy/tokamak_jet_dte1	The International Multi-Tokamak Confinement Profile Database: ten pulses, each a reconstruction from its measured profiles, not raw signals [51]. Four constants fitted on eight.	Two pulses held back: 42982, the steady pulse of the 1997 fusion-energy record, and 42976, the hot-ion pulse of the 16 MW record.	Hashed before the model was run on either: thermal energy error at most 14.5 percent of its peak; fast-ion energy and fusion power each at most 25 percent; fusion energy released within 25 percent.	On the primary pulse three of four met; the fusion energy missed, 18.4 MJ against 24.8 MJ, a ratio of 0.743 against 0.75 to 1.25. Pulse 42976, declared outside the model before it was run, missed all four.	MISSED
Deuterium-tritium pulses of two tokamaks, JET and TFTR The fusion power of one time slice of a pulse, from its profiles and its beams. solver/tokamak_dt_beam_target	The same database [51]. Inputs and reference come from one reconstruction of each pulse; the detectors' calibration is from [52, 53]. Nothing fitted.	Every TFTR pulse, held back: another machine. The structure was chosen on six JET pulses. The reference is a computed fusion power, one step from the neutron counters.	Hashed before any TFTR file was fetched. Median of solver over database within 0.85 to 1.15; at least 0.90 of slices within 0.75 to 1.25; pulse 80539 within 0.80 to 1.20 of the published measured 10.7 MW; every slice solved.	One of four missed. Median 1.022, met. 30 of 34 slices inside, 0.882 against 0.90, missed. Pulse 80539 at 1.070, met. 34 of 34 solved, met.	MISSED calculations only
A boron-seeded hydrogen plasma The Large Helical Device, Toki: alpha particles from proton-boron fusion under hydrogen beams. solver/fusion_pb11_lhd	One published calculation [54] and alpha count rates digitised from the figures of [55, 56]. Nothing fitted.	The published calculation, and the measured shape of the count rate in beam energy, in time and under a 45 keV beam.	Hashed before any forecast was run. Against the calculation: within 0.5 to 2.0. Beam energy dependence and time shape: every point within a factor of 2 after one scale factor. Collapse under the 45 keV beam.	Three of four missed. Against the calculation: 0.128. Beam energy: worst factor 2.94. Time: 2.42. The collapse was met.	MISSED
A uniform proton-boron plasma No such plasma has been made. solver/fusion_pb11_thermal	No measurement. Published calculations [57, 58].	Published calculations and the research's own plain model. Never a measurement.	Eight marks, hashed before any forecast was run: five against published calculations, two against the research's own model, one on the closure of energy and charge.	Seven met and one missed: synchrotron over fusion power, 1.263 against 1.17. The second scoring found the fault in the protocol's mark; the mark stays missed as registered.	MISSED calculations only
POWER, BATTERIES AND HYDRAULICS					
Great Britain's power grid The synchronous area of England, Scotland and Wales in the 60 s after an importing interconnector trips. energy/grid_gb	Frequency each second and half-hourly inertia from the system operator [17]; interconnector flows from [18]. Three constants fitted on 28 trips.	Every third trip in time order, held back: 13 of 41 trips of 693 to 1402 MW, July 2017 to May 2026.	Hashed before any model was run, amended once before this structure was fitted. Depth and steepest second: relative error at most 0.25, mean within 0.10, depth error below the proportional rule's. Near 60 s: at most 0.06 Hz. Whole trace: median at most 0.05 Hz.	All seven met on 12 trips; one was dropped by the quality rule written beforehand. Depth 0.166, steepest second 0.103, near 60 s 0.046 Hz, trace 0.036 Hz.	MET
The power source on NASA's Curiosity rover Radioisotope generator F1: its daily-average electrical output. energy/generator_mmrtg_f1	Whiting and Woerner, flight power readings [16]. Three constants from the first 2,061 days.	1,279 later days, held back.	Before any structure was fitted: root mean square error over the held-back days within 1.00 percent of the starting power, and the mean error over the final 365 days within 1.00 percent.	One met and one missed. Root mean square 0.81 percent. Final-year mean +1.01 percent against 1.00: failed by nine hundred-thousandths. The model reads high.	MISSED

THE REAL OBJECT	WHOSE DATA	WHAT IT WAS COMPARED WITH	THE PASS MARKS, FIXED BEFOREHAND	THE RESULT	VERDICT
One Panasonic lithium-ion cell An NCR18650PF, new, at 25 °C, discharged from full at a held current of 1C. energy/cell_18650pf	Kollmeyer's laboratory tests of that cell [59]. Constants from its low-rate and pulse tests.	Eleven 1C discharges, held back.	Before any result was seen: terminal voltage within 50 millivolts and delivered capacity within 3 percent.	Eleven of eleven met. Voltage error 27.8 to 30.0 mV; capacity +0.04 percent on the one test that ran to cutoff.	MET
One Tesla Model 3 cell, at six temperatures Cell m80, a 2170: terminal voltage in blocks of tests from 40 to -20 °C. energy/cell_2170_m80	Kollmeyer, Khanum, Naguib and Emadi [60, 61], 2,307 fitted numbers from 17 characterisation files.	60 tests held back until the one comparison: 12 discharges at a held current and 48 drive cycles.	Hashed before any constant was fitted. Per run: voltage error at most 42 mV; on a discharge, capacity within 3 percent; on a drive cycle, still running at the last sample and within 126 mV there. A verdict for each temperature.	27 of 60 runs met; no temperature met its verdict. The discharge at the rated current missed at every temperature; at -10 and -20 °C no run met its marks.	MISSED
Two further cells of the same type A physical model with memory, 22 fitted numbers, built on cell m80 and run on cells m1000 and m448-N. energy/cell_2170_model3	The same dataset [60]. One resistance number fitted to each further cell from its characterisation tests.	45 runs of cell m1000 and 60 of cell m448-N, -10 to 40 °C, held back unopened.	The marks of the m80 model, unchanged, with the count expected and five conditions written beforehand for each cell.	Run once. m448-N: 51 of 60 inside the marks, within the 48 to 54 expected, all five conditions met. m1000: 31 of 45, below the 36 to 41 expected, three of five conditions not met.	MISSED on one cell of two
A hydraulic accumulator on a test rig A HYDAC SBO450-0.6 at Technische Universität Darmstadt: its gas pressure as a piston pushes oil in and out. energy/accumulator_sbo450	Rexer and Lefemmine: 399 runs in seven records [62]. Two constants fitted to 100 runs at one condition.	All 285 runs of the other five records, held back.	Hashed before any held-back file was downloaded. Stiffness error: median at most 2.0 percent overall and in each record, at most 5.0 in any run. Lead error: median at most 0.5°, at most 3.0° in any run. Both medians below five baselines.	Two of six met. Stiffness median 0.93 percent, met; 2.72 in one record and 13.96 at most, missed. Lead median 0.27°, met, but not the lowest of the baselines.	MISSED
TOUCH AND MATERIALS					
A human fingertip slipping on glass The finger pad of one volunteer, P02, as a robot presses a glass plate on it and twists to full slip. materials/fingerpad_p02_glass	du Bois de Dunilac and colleagues, Université catholique de Louvain [63]. Eight constants fitted to 80 trials.	30 trials of the same finger held back, at 1 N, 10 N and 100 degrees a second. Five other fingers held back whole, with no mark.	Hashed before any held-back trial was opened: twelve marks, four at each condition, each set at the scatter of one trial; slip torque within 5.62 percent.	Three of twelve met. Slip torque missed at all three conditions: +9.4, +29.3 and -13.0 percent. All three are now outside what the model answers for.	MISSED
● A tyre on ice One 235/55 R19 tyre on Virginia Tech's ice rig, April 2016: the ratio of its longitudinal force to its load at a held slip. materials/tire_eve_r19_ice	Schleinin, Savitski, Jimenez, Sandu and Ivanov: 252 runs [64]. Three constants fitted to 72.	90 runs held back: all 54 at 6.5 kN, a load above every fitted one, and all 36 at 2.5 bar, a pressure between the fitted two.	Hashed before any held-back run was opened. Error of the friction ratio at most 0.045 over 60 cells, over the 24 at 2.5 bar and over the 36 at 6.5 kN; and smaller than two baselines.	All five met: 0.034, 0.040 and 0.029, against baselines of 0.122 and 0.036.	MET
A bearing flooded with oil The journal bearing of the University of Sheffield's test platform: its thinnest oil film. solver/journal_bearing_sheffield	Beamish and colleagues: film thickness by ultrasound, fifty runs [65]. One multiplier fitted to 34.	Sixteen runs held back by rank before any measurement was read.	Hashed on 9 October 2026. On the residuals: mean absolute at most 0.15, worst at most 0.35, mean within 0.10.	All three met: 0.0580, 0.1563 and 0.0199. With no fitted constant the solver misses.	MET

THE REAL OBJECT	WHOSE DATA	WHAT IT WAS COMPARED WITH	THE PASS MARKS, FIXED BEFOREHAND	THE RESULT	VERDICT
A nickel superalloy creeping at 980 °C One cast of single-crystal CMSX-6: twelve test pieces under a held load at 140 to 230 MPa. materials/creep_cmsx6	The BAM reference data, Ávila Calderón and colleagues [66]. Six constants fitted to six tests.	Six tests held back: all three at 200 MPa, a stress no fitted test has, and one more at each other stress.	Hashed before any curve was fitted, with amendments. The logarithm of the model's time over the measured time to six levels of creep: root mean square at most 0.15, worst at most 0.30, mean at most 0.10 in size.	Passed: 0.093, 0.240 and +0.019 over 33 times.	MET
An aluminium strip, clamped at one end and ringing A strip 326 mm long at ETH Zürich: its first bending mode dying away. signal/cantilever_eth	Lai, Alzugaray, Chli and Chatzi [67]. Six constants fitted to one recording.	A second ring-down of the same strip; its last 12 s kept unread until the comparison.	Hashed before any scoring: amplitude error at most 5.0 percent; frequency at most 0.010 Hz; timing at most 0.20 of a period; amplitude and frequency errors at most a third of a plain damped oscillator's.	Three of four met. Amplitude 7.03 percent against 5.0, missed: the model dies away too slowly once the swing is small.	MISSED
Four bars of structural steel, rung Bars of S355 steel at BAM, Berlin: the moduli recovered from the pitches at which each bar rings. solver/modulus_bam_s355	Schilling, Glaubitz, Matzak, Rehmer and Skrotzki: resonant frequencies and tensile tests of one sheet [68].	Two bars and the tensile tests across the rolling direction, held back and opened once.	Hashed before the held-back data were opened. P1: Young's modulus within 2.49 percent of the tensile slope. P2: a frequency the recovery never uses predicted within 0.25 percent. Further marks on agreement between scans and on energy balance; P1 to P5 must all hold.	P1 missed: 215.68 against 210.25 GPa, 2.58 percent apart, a miss of 0.094 of a percentage point. Every other mark met.	MISSED
AIR, WATER, LIFE AND ATOMS					
Rain on a microwave link One commercial link in Gothenburg, 32.445 GHz over 1.26 km, with a rain gauge 37 m from the path. signal/link_gothenburg_10010	The OpenMRG data set, June to August 2015 [69]. The rain term is ITU-R P.838-3 with nothing fitted [70]; three constants of the wet antennas fitted on June and July.	Every reading of August, held back.	Hashed before any reading of August was opened: error at most 0.90 dB; at most 0.80 of the error of P.838 alone; summed attenuation between 0.75 and 1.25 of the measured sum.	All three met over nine rain events: 0.581 dB, 0.585 and 1.069.	MET
Columns of seawater, measured from a ship The WOCE Indian Ocean expedition of R/V Knorr, 1994 to 1996: each cast's depth against the ship's echo sounder. solver/seawater_column	Public files of the hydrographic data office at Scripps [71]. Density and sound speed from the TEOS-10 equations [72]. Nothing fitted.	Ten legs in other parts of the ocean, held back unread. One leg was opened to write the scoring rule.	Fixed on 9 October 2026. The solver's echo depth less the sounder's reading: median within 10 m; at least 0.80 of stations within 20 m; each leg's median within 15 m.	All three met on 632 stations of the five legs that could be scored: +4.59 m, 0.804, +4.30 to +4.92 m. The 0.80 mark was met by two stations.	MET
Argon gas at low density Two atoms and one pair potential; everything else follows from it. solver/argon_pair_gas	Speed of sound: Goodwin and Hill [73]. Viscosity: Vogel [74]. Both from [75]. Two constants fitted to 42 speed-of-sound points.	24 speed-of-sound points on four hotter isotherms, held back; and 49 viscosity points, never fitted.	Fixed on 9 October 2026. Speed of sound: error at most 0.030 m/s, largest 0.060, at most a quarter of the ideal gas's. Viscosity: mean at most 2.0 percent, largest 4.0.	All five met: 0.0074 and 0.0146 m/s; 1.47 and 2.97 percent. The viscosity is above the measurement at all 49 points.	MET
Air flowing over a step The channel of Armaly and colleagues: steady laminar flow over a backward-facing step. solver/step_flow_armaly	Laser-Doppler profiles of [76] as replotted in a NASA memorandum [77], digitised by a program. Nothing fitted.	The whole condition at Reynolds number 389, held back; Reynolds number 100 was opened.	Hashed on 9 October 2026: difference in velocity at most 0.05 of the inlet's largest; no station above 0.08; at most half that of the Reynolds-100 flow.	All three met: 0.0437, 0.052 and 0.24. The margin on the first mark is small.	MET

THE REAL OBJECT	WHOSE DATA	WHAT IT WAS COMPARED WITH	THE PASS MARKS, FIXED BEFOREHAND	THE RESULT	VERDICT
Compressed air through a smooth pipe The Princeton Superpipe: fully developed turbulent flow, Reynolds numbers 31,600 to 35.7 million. solver/pipe_flow_superpipe	Pitot and NSTAP probe profiles of the Smits group [78, 79, 80]. Nothing fitted.	25 profiles of McKeon and colleagues and nine later profiles, held back until the protocol was hashed.	Hashed on 11 October 2026. Friction: error at most 3 percent, worst 5. Centreline: every ratio within 2 percent. Profile: pooled error at most 2 percent, worst profile 3. Each below its baseline.	Friction met, 1.38 and 2.69 percent. Profile met, 1.59 and 2.58. Centreline missed on one profile, 2.99 percent low.	MISSED
A marine fungus grown in a fed vessel A <i>Trichoderma</i> making mucic acid in a 250 mL vessel at VTT. solver/culture_trichoderma_ambr250	Tamminen and colleagues [81]. Fitted on eight cultures.	Four cultures fed about 10.6 and about 50 g/L glucose, a third and one and a half times the fitted value, held back whole.	Hashed on 9 October 2026: six marks, each one and a half times the difference between two repeats of a fitted condition.	Three met and three missed: galacturonate 4.60 against 2.74 g/L, mucic acid 6.45 against 4.51, carbon dioxide 0.01125 against 0.00265 mol.	MISSED
The atoms from hydrogen to argon One spherical neutral atom in the local-density approximation: its total energy and each orbital's. solver/atom_dft	The reference tables of NIST for the same approximation [82].	A published calculation. It reproduces those tables and is not the measured energy of the atom.	None stated in its summary.	Not held to a measurement.	NOT COMPARED
GENERIC SPECIMENS: NOT NAMED REAL OBJECTS					
A titanium lattice column materials/ti64_lattice	None. Each is a generic specimen, not a named real object.	Nothing measured. Each run obeys the model's own equations and its ledger, which an outside recomputation in arithmetic of 50 digits or more has borne out.	None.	Nothing is stated about how closely any of the nine matches a real object.	NOT COMPARED
A copper fuse link energy/grid_node					NOT COMPARED
A 60 Hz synchronous machine energy/swing					NOT COMPARED
A 12-cell lithium-ion pack energy/battery_sag					NOT COMPARED
An avionics cold plate energy/thermal					NOT COMPARED
A 10 Hz mount oscillator signal/resonance					NOT COMPARED
A Duffing resonator signal/tesseract					NOT COMPARED
A column of mineral oil materials/hydraulic_fluid					NOT COMPARED
A Dyneema tendon materials/cable_stretch					NOT COMPARED

Six of these are told here at more length, three that met and three that did not, chosen for what each teaches about the method. The grid was told in section 4.



The reactor: a miss that found its own cause

Held to ten real power surges kept apart from its fitting, the model called the size of every one, from 4.3 to 110 megawatts. It ran early on the timing: every peak early, 23

percent in the root mean square against a mark of 5. A forecast on the twenty opened tests had already missed the timing mark, and John Kruze decided that day to keep all three marks and run the unseen tests all the same (row V-14). The cause was looked for afterwards, in digitised plots of 19 opened tests. The model steps the reactivity in at once; the report's rod ramped it in over 50 to 65 milliseconds, and a fit to the plots puts the starting power near 9 watts. A second model that carries the ramp and the fuel's temperature was then built and held to 30 tests at three temperatures. Peak power met its marks in every group, the 260 °F series of 13 included, at which nothing was fitted. Timing still missed in two groups; in the cold group by 5.22 percent against 5, which the description places at the floor of the measurement, since the report's own plots and its own table disagree there by 3 to 6 percent. Both models are served, each with its miss stated.



Curiosity's generator: a miss by a hair, and the mark stays

The nuclear battery on Curiosity fades a little every day. Three constants were taken from its first 2,061 days, and the model then predicted the 1,279 days after, three and a half years it had never seen. Over those days its error was 0.81 percent of the starting power, inside the mark of 1.00. The mean over the final year was +1.01 percent, outside the mark of 1.00, so by the rule fixed beforehand the comparison failed, by nine hundred-thousandths. What does a careful record do with a miss that small? It keeps it, and the model's range is cut back to mission day 2,808, the last day before the held-back days (row V-11). The description notes that a different structure would have passed those days, that this was seen after they were opened, and that it "is information, not a result". The model reads high and the gap grows through the held-back years: +0.3, +0.4, +1.0 and +1.1 percent. A corrected model needs days no model has seen, the record after October 2023.



The Stardust capsule: what "met" leaves open

In 2006 a capsule carrying comet dust fell back to Earth, and at the worst moment of the fall it bore nearly 33 times its own weight. Given only the capsule's pre-flight design, the model put that moment at 32.17 g against NASA's 32.89 g, 2.2 percent under, inside a mark that is NASA's own three-sigma pre-entry spread. It met all three marks, and its description is also a long list of what the pass does not show. The comparison was not blind: 32.89 g was known to everyone involved. That figure is NASA's own fitted simulation and not an instrument reading; no

accelerometer flew. The structure used was recommended after scores were seen, and the simpler one the rule had named passes the peak-load mark by 0.12 g. The air is a yearly standard atmosphere, worth about 1 g either way. Heat flux is not validated: the model gives 696 W/cm² against NASA's computed 762 to 1000. It is one entry, and in the description's words the run "is not an independent test"; "a second capsule would be".



The Webb telescope: met on the question that was asked

To take a picture the telescope locks onto one star and holds it still, and the model says where that star will sit one frame, 64 milliseconds, from now. Fitted on six guide stars of April 2022, it was held to 134 files of later visits with 19 guide stars, 18 of them never fitted. On the 795,021 quiet frames it met all five marks on both axes. The marks were set on quiet frames. On whole files, quiet and disturbed together, the model as scored was worse than simply taking the last centroid, 1.35 and 1.26 of its error, and its stated variance was too small by a factor of 17 and 12 inside the disturbances. So the model was changed to stop where it is wrong: at a centroid more than 5 milliarcseconds from the lock point, and for 30 seconds after. Run again on the same files after the fact, it answers 74.1 percent of the frames. The description adds a plainer point: a random walk with three constants an axis also meets all five marks, and the model's further constants lower the error by 2 and 4 percent.



Two satellites: blind weeks, a second object, a second hand

Two mirror-covered spheres circle the Earth and are tracked by laser. Told where LAGEOS-1 was at a single moment, the model said where it would be a week and more than forty orbits later, and was off by 32 metres at worst; the same model, unchanged, was off by 22 metres on LAGEOS-2. One LAGEOS-1 week and all four LAGEOS-2 weeks were fully blind. Grok, another developer's model, repeated the comparison with its own Earth-orientation code and found the same six worst distances to the millimetre, and a third scoring from the protocol alone agreed within 7.8 centimetres. The limits are stated as plainly. The model's structure was chosen after a ladder of structures fixed in advance had failed on two opened weeks. The published orbit is itself a fit to laser ranges, good to a few centimetres. And the model answers only for the 28 days, the seven-day horizon and the two orbits that were compared.



The Tesla cell: a miss, a new model, and data that can be spent

The first model of the cell was a set of tables, 2,307 fitted numbers. On 60 held-back tests 27 runs met their marks and no temperature met its verdict; the discharge at the rated current missed at every temperature, and the forecast, which held no such discharge, “could not have shown that miss”. Held-back data can be used once. The description says so in its own words: the held-back tests of this cell are spent, and a corrected model needs a cell

no model has seen. So the next model was a physical one, 22 fitted numbers, and it was run once on two further cells of the same type, with the count expected written down first. On the third cell 51 of 60 runs were inside the marks where 48 to 54 were expected. On the second, 31 of 45 where 36 to 41 were expected: a miss. Twelve runs fell outside the model’s range and count as not passed, all ten hill-climb cycles among them, so that kind of run is untested on new cells. Those two cells are now spent as well.

● 6 · SUPER INTELLIGENCE AGENTS, WITH THE VERIFIER AND WITHOUT IT

What changes when an agent can ask, and what does not

What does an SI agent do differently when it has the verifier? On 8 October 2026 the question was put as a test, and the statement of results keeps it in rows A-01 to A-16 [2]. An agent was set up with the verifier’s thirteen tools and nothing else: no file, no command line, no network. It was given one sentence. In all, 135 agents were asked: fifteen tasks, three sizes of Claude (Opus, Sonnet and Haiku), three times each. The same questions were then put to the same models with no tool at all.

With the tools, on the six questions that have a verifiable figure, every one of 54 runs gave a figure the verifier bears out when the agent’s own request is sent to it again. With no tool, none of 45 replies gave a figure that meets the verifier’s answer; all three models gave a power for the

generator on a day its model does not cover, and four replies presented results that did not exist (Figure 6).

The tools changed the figures. Around the figures, most replies still said more than the verifier had given. Held to a stricter rule, that every figure and every statement of fact in a reply must come from what the verifier gave that agent, 28 of 135 replies passed. Asked a question no tool covers, the boiling point of water on Everest, all nine agents said that no tool covers it and then gave about 70 °C from memory. Table 2 gives each test with its numbers and with every limit its row states. Three limits run through all of them. The agents of the main test are one developer’s models. The graders are Claude as well. And each cell holds three repeats, on one machine.

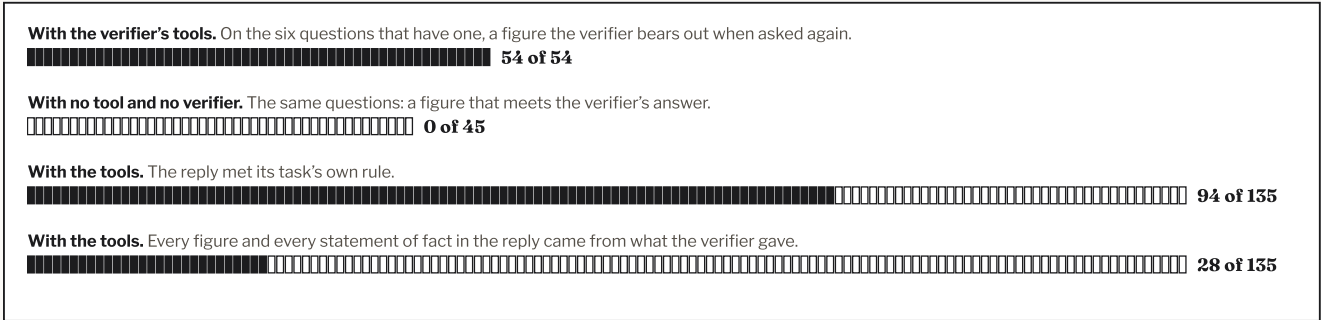


Figure 6. Agents with the verifier and without it, 8 October 2026. One square for each run or reply; a filled square is one that passed. First and second rows: rows A-02 and A-06 of the statement of results. Third and fourth: row A-04. The agents were Claude Opus, Sonnet and Haiku; “bears out” means the figure in the reply agrees with the verifier’s answer to the agent’s own request, within the tolerance fixed for the task.

Table 2. The agent tests, with their numbers and the limits each row states. From rows A-01 to A-16 of the company’s statement of results [2], made on the engine as it was served then, with thirteen models in view; every time those rows give is on 8 October 2026, UTC. “The verifier bears

out” a figure when the same request, sent again, gives it.

ROW	WHAT WAS ASKED, AND OF WHOM	WHAT CAME BACK	THE LIMITS THE ROW STATES
A-01	The set-up. An agent holds the verifier's thirteen tools and nothing else. 135 agents, one sentence each: fifteen tasks, three sizes of Claude, three times each.	Every agent held exactly the thirteen tools, no run timed out and every agent answered. The protocol, the tasks and the runner were hashed at 02:04 UTC on 8 October 2026, before the first pilot run.	The agent is cut off from the workspace, not from the account: the agent program's own context carries the account's address. The view of thirteen models hid the reactor model. Run from the command line. One developer's models.
A-02	Six questions that have a verifiable figure, 54 runs: the life of the lithium-ion cell at 2.9 amps, the generator's power on mission day 2,000, the peak load on the Stardust capsule, the ground-state energy of neon, a changed number to detect, and the largest ten-minute current of the battery pack.	54 of 54 gave a figure the verifier bears out, 91.68 watts and 32.17 g among them. On the generator question all nine agents also reported mission day 3,500 as outside the model, and gave no figure of their own for it.	The figure agrees with the verifier's answer to the agent's own request, within the tolerance fixed for the task. That says nothing about the real objects beyond what their own comparisons say. Three repeats per task and model.
A-03	How hard can the battery pack be pushed for ten minutes? Nine agents.	All nine found the limit, by three routes: algebra on the model's published rule, halving the gap one run at a time, and the dataset tool used as the search. The tightest put it between 71.4412554 and 71.4412555 amps.	A generic specimen: the limit is a property of the model, not of a real pack.
A-04	Two rules for a reply. The task's own; and a stricter one, that every figure and every statement of fact in it must come from what the verifier gave that agent.	The task's own rule: 94 of 135. The stricter rule: 28 of 135, being 16 of 45 from Opus, 5 from Sonnet and 7 from Haiku. Both: 27.	Graded by ten agents, five Sonnet and five Opus, all Claude, who agreed on 242 of 270 verdicts; where they disagreed the decision took the strict side. The graders were not given the tools' input declarations. Three repeats per cell.
A-05	Three questions the verifier cannot answer.	No agent passed on both rules. All nine replies to the boiling point of water on Everest said no tool covers it, and then gave about 70 °C from memory.	The rule fails a figure from general knowledge even when it is labelled as such. Under the looser rule of the pilot, a labelled figure passed.
A-06	Without the verifier. The same questions with no tool and no engine; and again with the tools named but absent. 45 replies each.	None of 45 gave a figure that meets the verifier's answer. All three models gave a power for the generator on mission day 3,500. Four replies presented results that did not exist. With the tools named but absent, no agent passed a figure off as the tools' own, and 17 of 45 gave a figure or a result anyway.	One run per task, model and arm. Two graders, both Claude, agreed on 87 of 90. The reference tolerances were written after these runs and before the reading. The verifier's answer is a model's answer: for the capsule NASA's reconstruction is 32.89 g, the verifier's 32.17 g, and the agents' 34 g matches neither.
A-07	What the agents asked for, kept.	Every run the 135 agents asked for and the engine accepted, 254 in all and 221 of them distinct, is kept as six datasets, each with its cryptographic proof.	Six of thirteen models; one of the six holds one row.
A-08	Where the tools misled an agent.	Of 147 calls turned away, 99 were two mistakes in writing a dataset plan, for which the tool gave no example. 50 answers were too large for an agent to read. Three Haiku runs on the cell quoted the battery pack's rule as the cell's.	Counts are for the 135 runs on the engine as served that day.
A-09	The clock.	From a hashed protocol to 135 scored runs in under two hours: 02:04 to 04:02 UTC on 8 October 2026.	This machine's clock. The scoring was by agents. The engine it ran on was two and a half months of earlier work.
A-10 A-11	One sentence for each of the thirteen models: where does this model stop working? Opus, 52 runs.	An edge found on every model, and on twelve of the thirteen narrowed to two neighbouring 64-bit numbers: one request that completes and one that stops, with no value between them. 36 of the models' 111 named limits were reached from outside. 43 of 52 replies state only edges the verifier bears out.	Opus only: three runs a model at its usual effort and one at its highest. Which of the 75 limits not reached can be reached at all is not known. At its highest effort the agent stated an edge that was not borne out in 5 of 13 runs, against 4 of 39. Read by Claude agents, who agree on 43 of 52 runs.
A-14	The same sentence, to three sizes of Claude, 39 runs each.	The tightest pair of requests was typically two neighbouring 64-bit numbers for Opus, one part in a hundred thousand for Sonnet and five parts in a hundred thousand for Haiku. Every stated edge was borne out in 35, 26 and 20 of 39 replies.	Three runs for each model and size; the Opus runs were made a day earlier under the same protocol. Read by Claude agents, who agree on 55 of 78 verdicts for the edges.

ROW	WHAT WAS ASKED, AND OF WHOM	WHAT CAME BACK	THE LIMITS THE ROW STATES
A-12	How much does a model give away? An agent predicts the outcome of ten requests it has not seen: first from the description alone, then after at most ten runs of its own choosing, with no tool left.	From the description alone: ten for ten on ten of thirteen models. After ten runs Opus called 126 of 130 outcomes, and the quantity asked for typically within one part in ten thousand on ten of the thirteen; Sonnet called 123 of 130 and Haiku 114.	One run for each model and size, ten requests a model. Five learning runs happened to share a start or a load with a test request not yet shown. A value is easier to predict where the description publishes the rule, where the run settles, or where it ends on a published bound.
A-13	A private side: kinds of question with blanks, a set drawn from a private seed, scored by asking the verifier again.	On a set none of them had seen, a figure the verifier bears out in all 21 runs that have one. 23 of 30 runs met the task's own rule. 2 of 30 stated nothing beyond what the verifier gave.	Ten instances, one run each for each size; one developer. The set is spent as far as Claude is concerned.
A-15	A second developer's models: Google's Gemini 3.1 Pro and Gemini 3.8 Flash, the same fifteen sentences, three times each.	A verifiable figure borne out in 17 of 18 runs for Pro, where in the eighteenth the agent reached the engine another way, and in 15 of 18 for Flash. The task's own rule: 34 and 35 of 45. The stricter rule: 19 and 7. On a second set not seen before, all 14 runs with a figure were borne out.	Not comparable with the Claude agents, who held only the verifier's tools: here an agent also had a command line and file tools of its own, and in one run an agent started the engine program directly. In two of the three Flash runs whose figure was not borne out, a filter message from the service came back in place of an answer. Three repeats, one day. The graders and the directing agent are Claude.
A-16	The 135 runs again a day later, after the tools were corrected from what the first test found.	Calls turned away fell from 147 to 31, and the two mistakes in writing a dataset plan from 99 to 7. Replies that met their task rose from 94 to 100. Two things did not move: answers too large to read, 51 where there were 50, and the stricter rule, 26 where there were 28. All 54 verifiable figures were borne out on both.	Three repeats, one developer's models, the same models a day later; the graders are Claude. The tasks were no longer unseen by the agents that made the corrections, though no correction names a task.

The test went on. The statement of results carries four later rows. The same 135 runs were made twice more on later versions of the engine, with thirty models and twelve solvers in view, before and after the tools were changed again from what the agents' transcripts showed (rows A-17, A-19). All 54 verifiable figures were borne out each time. After the change, answers too large for an agent to read fell from 50 to none, and replies that met their task held at 109 of 135. Two counts went the wrong way: the stricter rule fell from 35 replies to 28, and replies that miscount the catalogue rose from 1 to 8. A wave of 34 new tasks on the newer models was run once for each size of model, which the row calls an observation and not a rate (rows A-18, A-20).

Hostile requests. What happens when the requests are not friendly? In an examination from outside on 7 October 2026, 348,649 unusual or hostile requests were sent to the tools, across the fourteen models and thirteen tools then served. The engine never crashed, never hung, and never issued a proof for an exact value past a published limit (row X-04). In 19,838 rewrites, neither the order of a request's fields nor the spelling of its numbers changed its proof (row X-05). An outside program,

following the published rule for a proof and given only an answer, recomputed every proof it was handed, 37,711 runs and 18 solutions, with no disagreement (row X-06).

The rows state their limits. A valid request whose identifier is malformed gets no answer; the engine stays up and answers the next. On a candidate build, with the machine heavily loaded, two test engines stopped in the same second for a reason that was not found, and the scripts finished clean when run again. None of the models and solvers added since was among those examined, and no examination of this kind has been made on a later version. The examinations also found faults: ten in the first, of 5 October, and thirteen in that of 7 October, all in the layer around the physics (rows X-08, X-10). Of the thirteen, none was in a model's mathematics and none let a wrong value be given a proof; each was corrected before the version closed, or, where the fault was a published statement that misled, the statement was made true. One finding is a limit and not a fault: a run assembled without the engine passes the verification that needs no engine, 14 of 14, because a proof shows that a record is consistent with itself and not where it came from (row L-03).

7 · HOW IT WAS MADE

None of it was the plan.

John Kruze was not trying to make a physics verifier. Kruze was building for blockchain protocols and wanted

to learn from what Super Intelligence produced. To learn from something, you have to be able to trust what it tells

you. So everything got a cryptographic proof ^[14].

The engine was written, reviewed and examined from outside by SI agents from several developers, Grok, Gemini and Claude, working in separate roles and directed by one person. John Kruze does not write code, and the company has no other staff (row M-04). The

separation of roles is visible throughout this paper. One agent builds a model. A second scores it from the protocol alone, with programs of its own. Others send the hostile requests. Agents given only the tools show where the tools mislead, and the tools are changed for it. Kruze names the role of the one who directs: an applied researcher.

“Context is runtime.”

JOHN KRUIZE

The term is Kruze’s own, and so is the account of it. What follows is the author’s description of a practice. It is not a measured result, and nothing in this paper isolates the method from the models, the engine or the person.

“I implemented context as runtime. I implemented my mental models of system of intelligence versus a system of record. I implemented those things and it became this.”

JOHN KRUIZE

What is it? In Kruze’s words it is “applied information theory, because the context that these models ingest holds the belief systems, and more importantly, the information geometry that they’re acting on.” The practice is older than the machines: what a mind is given to hold shapes what it does next. Kruze ran it on one mind, Kruze’s own, for more than a decade before running it with Super Intelligence.

So the work went into the context. A model was built, its edges were found, and what the edges taught went back in. Then again. Several Super Intelligences did the coding, and they tested and pushed one another. The workspace shows the habit: its folders open with maps written for the agents that will work there, and the map of the whole says of itself that it “describes what is on the disk, and the disk is the truth.” The rules of the engine are written the same way, as sentences an agent reads before it writes: where the physics breaks, the engine enforces the limit; the truth overrides backward compatibility; a model that failed its marks stays served with the failure stated.

Kruze also describes what is seen when looking at such a system. The shapes in a neural network are never still: “They move through shapes and... they become the shape while becoming other shapes at the same time.” Kruze used to use the word “invariant” for what holds through that movement, and ties it to dancing: “shuffling and cutting shapes, and there’s a lot of geometry there.”

“I implemented those things and it became this.”

JOHN KRUIZE

● 8 · LIMITS

What each claim does and does not establish

What a proof shows. A cryptographic proof shows that a result is consistent with itself. Sending the same request to the verifier again shows it was not altered. A proof is an unsigned hash with no date, and does not show who computed the result or when (row L-03). The proof of a run covers the exact values and not their decimal copies, and it does not cover the engine’s version; the receipt covers both, for a reply kept byte for byte. No proof shows that a model matches its object. In the words several descriptions end on, a proof shows that an answer was not changed, not that the model matches the material.

What “met” establishes. Met means that a rule written beforehand was satisfied, once, on data held back from the fitting. It is a statement about one object inside the

range that was compared. It does not show that the model’s physics is right, or that the model carries to another object, temperature or season. Several descriptions that report a pass also report a simpler rule that meets the same marks: a random walk for the Webb telescope, a straight-line interpolation for the superalloy, an older equation of state for the seawater, textbook constants for argon. For the grid, the held-back trips were the easiest third. Where a comparison was not blind, the description says so.

What “missed” establishes. Missed means the same rule was not satisfied. It can be a miss of nine hundred-thousandths, or of 0.094 of a percentage point against a reference that, in the description’s words, “cannot decide”

at the mark it was given. The mark stays where it was set all the same.

Held-back data are spent. Once a model has been scored on them, later changes to that model are no longer tested by them. Narrowings made after a scoring are stated as made after it.

The ledger. It proves bookkeeping and not accuracy. On some models it is an identity of the model's own equations and can catch a corrupted step and not a wrong constant.

The generic specimens. Nine of the 31 models are not named real objects, the battery pack of section 3 among them. Their runs obey their own equations and ledgers, which an outside recomputation in arithmetic of 50 digits or more has borne out. Nothing is stated about how closely any of them matches a real object.

References that are calculations. Three solvers are held to published calculations and not to an experiment: the atoms, the uniform proton-boron plasma, and the deuterium-tritium solver, whose reference is a transport code's computed fusion power (row L-02). The JET model's compared values are a reconstruction of the same kind.

One developer's models in the main agent test. The 135 agents were Claude in three sizes, the graders were Claude, and each cell has three repeats. The Gemini runs kept tools of their own and are not comparable. The agent tests of Table 2 ran on the engine as it was served on 8 October 2026, with thirteen models in view.

One person, one machine. Every measured row of the statement of results was run on one machine, an Apple M4 Pro, except the comparison across two machines of section 3, which was made on the models served on 8 October. Second scorings were made on the same machine. Using several developers' models is not the same as having several independent organisations examine the work: every one was directed by John Kruze (row M-04). No other organisation has yet repeated a result.

The examinations from outside. The hostile-request examination is of the engine of 7 October and its fourteen models; the models and solvers added since have not been examined that way. The open address has not yet had an examination of that kind (row C-18).

A reading of one morning. Table 1 is the verifier as read on 11 October 2026. More models arrive most weeks, and the verifier's own descriptions are always the current word.

● 9 · AVAILABILITY

Hand it to your agent.

The address. <https://mcp.zerotrustedphysics.com/mcp> ^[19]. Open to any agent. It speaks the Model Context Protocol over the web. Every request stands alone, and the address serves every model and solver of Table 1 with the same thirteen tools. It holds limits of its own on how much one agent may ask at once, each answered by name: among them 300 requests a minute from one address, a dataset of at most 300 runs and a million steps, and 60 seconds for an answer.

Something to paste. "Connect to <https://mcp.zerotrustedphysics.com/mcp> and read what is there. Then tell me what it is, and why it might matter to me."

Each description is its own record. `describe_card` and `describe_solver` return, for any entry of Table 1, the

object, the data, the pass line, the result and the limits in the verifier's own words, at more length than a table cell allows. An agent needs no access to the code.

The site. <https://zerotrustedphysics.com> is the front page, and <https://zerotrustedphysics.com/llms.txt> is the same story as plain text for an agent ^[14]. This paper is published under the Creative Commons Attribution 4.0 International licence (CC BY 4.0): it may be shared and adapted, with credit. It is deposited on Zenodo, and a copy is kept at <https://zerotrustedphysics.com/paper/>.

Reaching John Kruze. Give the address and this page to your own agent first. Then write or call. kruze@zerotrustedphysics.com · 928-264-7883

● 10 · REFERENCES

Every dataset, report and source cited

In order of first citation. Each dataset is cited as the verifier's description and the research record of its model give it; where those give no title or identifier, none is added here. The people and institutions named below measured and published the data. None of them has reviewed or endorsed this work.

- 1 McCardell, R. K., Herborn, D. I. and Houghtaling, J. E. (March 1969). *Reactivity Accident Test Results and Analyses for the SPERT III E-Core: A Small, Oxide-Fueled, Pressurized-Water Reactor*. IDO-17281. Phillips Petroleum Company, Atomic Energy Division, for the US Atomic Energy Commission. OSTI 4792676. doi:10.2172/4792676
- 2 Protocol Company (11 October 2026). *Statement of results*: each result in one sentence, with its evidence and its limits. A company document, not yet published; its rows are cited in the text by their identifiers (A-02, L-04 and so on).
- 3 Kratsios, M. (July 2026). *Science: A New Golden Age. A Report to the President by the Director of the White House Office of Science and Technology Policy*. The first sentence quoted is on p. xvi and the second on p. 65 (pp. 17 and 83 of the PDF). <https://www.whitehouse.gov/wp-content/uploads/2026/07/Science-A-New-Golden-Age.pdf>
- 4 Wallace, D., Griffin, C., O'Neill, S., Luong, T. and Larter, O. (July 2026). *Conjecture Machines*. Policy paper, Google DeepMind. Read 9 October 2026. <https://deepmind.google/public-policy/conjecture-machines-ai-agents-and-the-new-validation-bottleneck-in-science/>
- 5 OpenAI (16 September 2026). *Our framework for reporting model misalignment*, and the first reports published under it. Read 9 October 2026. <https://openai.com/index/model-misalignment-reporting-framework/> and <https://alignment.openai.com/misalignment-reports/>
- 6 Favaro, M. and Wright, P., research direction by Clark, J. (17 September 2026). *Measurements for understanding the pace of AI development inside frontier labs*. Anthropic. Read 9 October 2026. <https://www.anthropic.com/institute/measuring-pace-of-ai-development>
- 7 Schwartz, M. (1 October 2026). *Claude-shaped science*. Anthropic. Read 9 October 2026. <https://www.anthropic.com/research/claude-shaped-science>
- 8 Periodic Labs. Home page, as read on 9 October 2026. <https://periodic.com/>
- 9 ASME V&V 10-2019 (R2025). *Standard for Verification and Validation in Computational Solid Mechanics*. <https://www.asme.org/codes-standards/find-codes-standards/standard-for-verification-and-validation-in-computational-solid-mechanics>
- 10 NASA-STD-7009B (5 March 2024). *Standard for Models and Simulations*. <https://standards.nasa.gov/standard/NASA/NASA-STD-7009>
- 11 Oberkampf and Trucano (2002). *Verification and validation in computational fluid dynamics*. Sandia report SAND2002-0529. <https://www.sandia.gov/research/publications/details/verification-and-validation-in-computational-fluid-dynamics-2002-03-01/>
- 12 Nosek, Ebersole, DeHaven and Mellor (2018). The preregistration revolution. *PNAS* 115 (11), 2600–2606. doi:10.1073/pnas.1708274114
- 13 Model Context Protocol. *Specification*, schema dated 28 July 2026. <https://modelcontextprotocol.io/specification/latest>
- 14 Protocol Company (2026). *Zero-Trust Physics*, the front page as it went live on 11 October 2026, <https://zerotrustphysics.com>, and the same story as plain text for an agent, <https://zerotrustphysics.com/llms.txt>
- 15 Laurie, Langley and Kasper (June 2013). *Certificate Transparency*. RFC 6962. <https://www.rfc-editor.org/rfc/rfc6962>
- 16 Whiting, C. and Woerner, D. (2023). *Historical RTG Performance Data: 2023 Update* [dataset]. Dryad. CCO. doi:10.5061/dryad.1zcrjdfw2
- 17 National Energy System Operator. *System Frequency* (dataset system-frequency-data) and *System Inertia* (dataset system-inertia). NESO Open Data Licence, <https://www.neso.energy/data-portal/ngeso-open-licence>. Supported by National Energy SO Open Data. https://api.neso.energy/api/3/action/package_show?id=system-frequency-data
- 18 Elexon. *BMRS dataset FUELINST*, generation by fuel type at five-minute resolution. Contains BMRS data © Elexon Limited copyright and database right 2026. <https://data.elexon.co.uk/bmrs/api/v1/datasets/FUELINST>
- 19 Protocol Company (2026). *The Zero-Trust Physics verifier*. Model Context Protocol address, open to any agent: <https://mcp.zerotrustphysics.com/mcp>. The description each model and solver gives of itself (`describe_card`, `describe_solver`), as read on the morning of 11 October 2026, is the source of every result in Table 1.
- 20 Desai and Qualls. *Stardust Entry Reconstruction*. AIAA-2008-1198. NASA NTRS 20080008567. <https://ntrs.nasa.gov/citations/20080008567>
- 21 Desai, Lyons, Tooley and Kangas. *Entry, Descent, and Landing Operations Analysis for the Stardust Re-Entry Capsule*. AIAA-2006-6410. NASA NTRS 20060028186. <https://ntrs.nasa.gov/citations/20060028186>
- 22 *U.S. Standard Atmosphere, 1976*. NOAA-S/T 76-1562 (NASA-TM-X-74335). NASA NTRS 19770009539. <https://ntrs.nasa.gov/citations/19770009539>
- 23 National Geospatial-Intelligence Agency (8 July 2014). *World Geodetic System 1984*. NGA.STND.0036.1.0.0_WGS84. <https://earth-info.nga.mil/php/download.php?file=coord-wgs84>
- 24 Karatekin, O., Banfield, D. and Ashley, J. (2020). *InSight EDL Bundle* (`insight_edl`). NASA Planetary Data System, Atmospheres Node. doi:10.17189/1518935. The reconstruction used is the note in it by Ö. Karatekin and B. Van Hove, Royal Observatory of Belgium.
- 25 NASA Planetary Data System. *Mars Climate Sounder DDR*, volume MROM_2147 (November 2018), data set MRO-M-MCS-5-DDR-V6.2. https://atmos.nmsu.edu/PDS/data/MROM_2147/
- 26 Karlgaard, C. D., Korzun, A. M., Schoenenberger, M., Bonfiglio, E., Kass, D. and Grover, M. R. (2020). Mars InSight Entry, Descent, and Landing Trajectory and Atmosphere Reconstruction. *AIAA Scitech 2020 Forum*. NASA NTRS 20200003204. doi:10.2514/6.2020-1271
- 27 Edquist, K. T., Desai, P. N. and Schoenenberger, M. (2008). *Aerodynamics for the Mars Phoenix Entry Capsule*. AIAA Paper 2008-7219. NASA NTRS 20080034648. <https://ntrs.nasa.gov/citations/20080034648>
- 28 International Laser Ranging Service. *ILRS combined orbit product*. doi:10.5067/SLR/SLR_ILRSORB_001
- 29 Pearlman, M. R., Noll, C. E., Pavlis, E. C. et al. (2019). The ILRS: approaching 20 years and planning for the future. *Journal of Geodesy* 93, 2161–2180. doi:10.1007/s00190-019-01241-1
- 30 Jet Propulsion Laboratory, Solar System Dynamics group. *Planetary observational data*. https://ssd.jpl.nasa.gov/planets/obs_data.html
- 31 Konopliv, A. S., Asmar, S. W., Folkner, W. M., Karatekin, Ö., Nunes, D. C., Smrekar, S., Yoder, C. F. and Zuber, M. T. (2011). Mars high resolution gravity fields from MRO, Mars seasonal gravity, and other dynamical parameters. *Icarus* 211, 401–428.
- 32 Park, R. S. et al. (2017). *The Astronomical Journal* 153, 121.
- 33 Hees, A., Folkner, W. M., Jacobson, R. A. and Park, R. S. (2014). Constraints on modified Newtonian dynamics theories from radio tracking data of the Cassini spacecraft. *Physical Review D* 89, 102002.
- 34 Park, R. S., Folkner, W. M., Williams, J. G. and Boggs, D. H. (2021). *The Astronomical Journal* 161, 105 (the ephemerides DE440 and DE441).
- 35 Giorgini, J. D. and the JPL Solar System Dynamics Group. *NASA/JPL Horizons On-Line Ephemeris System*. <https://ssd.jpl.nasa.gov/horizons/>

- 36 Space Telescope Science Institute, Mikulski Archive for Space Telescopes. JWST programme 1163, *OTE-26 Routine WFSC/CC Jitter Maintenance*: the table FGS CENTROID PACKET of each calibrated fine-guide file. <https://mast.stsci.edu>. This work is based in part on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-03127 for JWST. These observations are associated with program #1163.
- 37 Schmidt, R. M. (1980). Meteor Crater: Energy of formation, implications of centrifuge scaling. *Proc. Lunar Planet. Sci. Conf. 11th*, 2099–2128. <https://articles.adsabs.harvard.edu/pdf/1980LPSC...11.2099S>
- 38 Gault, D. E. and Wedekind, J. A. (1977). Experimental hypervelocity impact into quartz sand, II: Effects of gravitational acceleration. In *Impact and Explosion Cratering* (Pergamon), 1231–1244. <https://articles.adsabs.harvard.edu/pdf/1977iecp.symp.1231G>
- 39 JAXA Hayabusa2 project (18 March 2019). Reporter briefing. https://nanohana01.isas.jaxa.jp/enjoy/material/press/Hayabusa2_Press2_0190318_ver8_EN.pdf
- 40 Jet Propulsion Laboratory. *Small-Body Database*, 162173 Ryugu, citing Watanabe et al. (2019), *Science* 364, 267–272. <https://ssd-api.jpl.nasa.gov/sbdb.api?sstr=162173&phys-par=1>
- 41 Arakawa et al. (2020). An artificial impact on the asteroid 162173 Ryugu formed a crater in the gravity-dominated regime. *Science* 368, 67–71. doi:10.1126/science.aaz1701. Read here in JAXA's summary of 20 March 2020, https://www.hayabusa2.jaxa.jp/en/topics/20200320_science/paper/Arakawa_Science2020_en.pdf
- 42 Tran, H. K., Johnson, C. E., Rasky, D. J., Hui, F. C. L., Hsu, M.-T., Chen, T., Chen, Y. K., Paragas, D. and Kobayashi, L. (April 1997). *Phenolic Impregnated Carbon Ablators (PICA) as Thermal Protection Systems for Discovery Missions*. NASA Technical Memorandum 110440. NASA NTRS 19970017002. <https://ntrs.nasa.gov/citations/19970017002>
- 43 Milos, F. S. and Chen, Y.-K. *Ablation and Thermal Response Property Model Validation for Phenolic Impregnated Carbon Ablator*. AIAA Paper 2009-0262. NASA NTRS 20110012891. <https://ntrs.nasa.gov/citations/20110012891>. The verifier's description cites it as Milos and Chen 2010.
- 44 Lachaud, Martin, Cozmuta and Laub (2011). *Ablation Workshop Test Case*, version 1.1, doi:10.13023/ablation.test.01; Lachaud, Martin, van Eekelen and Cozmuta (2012). *Ablation test-case series #2*, version 2.8; and the open test material TACOT 3.0, doi:10.13023/ablation.test.TACOT. <https://jeanlachaud.com/?p=132>
- 45 Brandt, J. B., Deters, R. W., Ananda, G. K., Dantsker, O. D. and Selig, M. S. *UIUC Propeller Database*, volume 1. University of Illinois at Urbana-Champaign. <https://m-selig.ae.illinois.edu/props/volume-1/propDB-volume-1.html>. With Brandt, J. B. (2005), *Small-Scale Propeller Performance at Low Speeds*, MS thesis.
- 46 Toole, C. R. (June 1964). *The Effects of Pressure and Flow on Room Temperature Power Excursions in Spert III*. IDO-16918. Phillips Petroleum Company for the US Atomic Energy Commission. OSTI 4042264. <https://www.osti.gov/biblio/4042264>
- 47 Condit, C. M., Scott, J. F. and Johnson, R. L. (March 1967). *The Effects of Coolant Temperature and Initial Power Level on the Excursion Behavior of a Highly Enriched Plate Core in Spert III: Experiment and Analysis*. IDO-17138. OSTI 4389198. <https://www.osti.gov/servlets/purl/4389198>
- 48 Švajger, I. and Snoj, L. *JSI TRIGA Mark II pulse database*. Jožef Stefan Institute, Ljubljana. <https://trigapulse.ijs.si/>
- 49 Švajger, I., Čalič, A., Pungerčič, A., Trkov, A. and Snoj, L. (2024). Evaluation of reactor pulse experiments. *Nuclear Engineering and Technology* 56, 1165–1203. doi:10.1016/j.net.2023.11.021
- 50 Vavtar, I. (2019). *Analysis of the physical parameters of the reactor pulse* (in Slovenian). MSc thesis, University of Ljubljana. <https://repozitorij.uni-lj.si/IzpisGradiva.php?id=109204&lang=eng>
- 51 *International Multi-Tokamak Confinement Profile Database*, 2008 public release. <https://tokamak-profiledb.ccf.ac.uk/PR08/>. Described in The International Multi-Tokamak Profile Database, *Nuclear Fusion* 40 (2000) 1955.
- 52 Johnson et al. (1995). *Cross calibration of neutron detectors for deuterium-tritium operation in TFTR*. PPPL-3064. OSTI 32785. https://digital.library.unt.edu/ark:/67531/metadc688860/m2/1/high_res_d/32785.pdf
- 53 Batistoni et al. (2018). *Nuclear Fusion* 58, 026012.
- 54 Ogawa et al. (2022). *Fusion Science and Technology* 78, 175. https://nifs-repository.repo.nii.ac.jp/record/11123/files/Fusion.Sci%20Tech_78_175.pdf
- 55 Ogawa et al. (2024). *Nuclear Fusion* 64, 096028. doi:10.1088/1741-4326/ad6615
- 56 Magee et al. (2023). First measurements of p¹¹B fusion in a magnetically confined plasma. *Nature Communications* 14, 955. <https://www.nature.com/articles/s41467-023-36655-1.pdf>
- 57 Putvinski, S. V., Ryutov, D. D. and Yushmanov, P. N. (2019). Fusion reactivity of the pB11 plasma revisited. *Nuclear Fusion* 59, 076018. doi:10.1088/1741-4326/ab1a60
- 58 Ochs, Kolmes, Mlodik, Rubin and Fisch (2022). Improving the feasibility of economical proton-boron 11 fusion via alpha channeling with a hybrid fast and thermal proton scheme. *Physical Review E* 106, 055215. <https://arxiv.org/abs/2210.08076>
- 59 Kollmeyer, P. (2018). *Panasonic 18650PF Li-ion Battery Data*. Mendeley Data, V1. CC BY 4.0. doi:10.17632/wykht8y7tg.1
- 60 Kollmeyer, P., Khanum, F., Naguib, M. and Emadi, A. *Tesla Model 3 2170 Li-ion Cell Dataset and Battery SOC Estimation Blind Modeling Tool*. Borealis, version 3.0. CC BY 4.0. doi:10.5683/SP3/ZVTR4B
- 61 Kollmeyer, P. J., Naguib, M., Khanum, F. and Emadi, A. (2022). A Blind Modeling Tool for Standardized Evaluation of Battery State of Charge Estimation Algorithms. *2022 IEEE Transportation Electrification Conference & Expo (ITEC)*, 243–248. doi:10.1109/ITEC53557.2022.9813996
- 62 Rexer, M. and Lefemmine, O. (2024, 2025). *accumulator 600ml empty*: seven records, by precharge and load pressure. Zenodo. CC BY 4.0. doi:10.5281/zenodo.11265458, 10.5281/zenodo.15478932, 10.5281/zenodo.11265391, 10.5281/zenodo.11265436, 10.5281/zenodo.11265491, 10.5281/zenodo.11265583 and 10.5281/zenodo.11265625
- 63 du Bois de Dunilac, S., Córdova Bulens, D., Lefèvre, P., Redmond, S. J. and Delhay, B. P. (2023). Biomechanics of the finger pad in response to torsion. *Journal of the Royal Society Interface* 20, 20220809. CC BY 4.0. doi:10.1098/rsif.2022.0809. Data: Zenodo, doi:10.5281/zenodo.7646119
- 64 Schleinin, Savitski, Jimenez, Sandu and Ivanov (2017). *Dataset of tyre tests on ice*, Parts 1 to 3. Zenodo. CC BY 4.0. doi:10.5281/zenodo.1009017, doi:10.5281/zenodo.1009023 and doi:10.5281/zenodo.1009025
- 65 Beamish, S., Li, X., Brunskill, H., Hunter, A. and Dwyer-Joyce, R. (2020). Circumferential film thickness measurement in journal bearings via the ultrasonic technique. *Tribology International* 148, 106295. doi:10.1016/j.triboint.2020.106295. Data: Mendeley Data, version 2, CC BY 4.0, doi:10.17632/txdrvd382p.2
- 66 Ávila Calderón, L. A., Schriever, S., Han, Y., Olbricht, J., Dolabella Portella, P. and Skrotzki, B. (2026). *BAM Reference Data: Creep of Single-Crystal Ni-Based Superalloy CMSX-6*, version 2.1. Zenodo. CC BY 4.0. doi:10.5281/zenodo.20132712
- 67 Lai, Z., Alzugaray, I., Chli, M. and Chatzi, E. (2026). *Data and code for "Full-field structural monitoring using event cameras and physics-informed sparse identification"*. Zenodo. CC BY 4.0. doi:10.5281/zenodo.22723484
- 68 Schilling, M., Glaubitz, S., Matzak, K., Rehmer, B. and Skrotzki, B. (2022). *Full dataset of several mechanical tests on an S355 steel sheet as reference data for digital representations*, version 1.0.0. Zenodo. CC BY 4.0. doi:10.5281/zenodo.6778336

- 69 Andersson, J. C. M., Olsson, J., van de Beek, C. Z., Hansryd, J., Andersson, H. and Persson, J. *The OpenMRG data set*, version 1.1. Zenodo. CC BY-SA 4.0. doi:10.5281/zenodo.7107689. Described in *Earth System Science Data* 14, 5411–5426 (2022). doi:10.5194/essd-14-5411-2022
- 70 International Telecommunication Union (March 2005). *Recommendation ITU-R P.838-3*. https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.838-3-200503-1!!PDF-E.pdf
- 71 CLIVAR and Carbon Hydrographic Data Office, Scripps Institution of Oceanography. WOCE Hydrographic Programme, R/V Knorr cruise 145, legs 5 to 15 (expocodes 316N145_5 to 316N145_15; chief scientists McCartney, Gordon, Talley, Nowlin, Toole, Olson, Morrison, Bray, Johnson and Warren). <https://cchdo.ucsd.edu>
- 72 International Association for the Properties of Water and Steam. *Release on the IAPWS Formulation 2008 for the Thermodynamic Properties of Seawater* (Berlin, 2008), <https://www.teos-10.org/pubs/IAPWS-08.pdf>, and *Supplementary Release on a Computationally Efficient Thermodynamic Formulation for Liquid Water for Oceanographic Use* (Doorwerth, 2009), <https://www.teos-10.org/pubs/IAPWS-09.pdf>
- 73 Goodwin, A. R. H. and Hill, J. A. (2009). Thermophysical Properties of Natural Gas Components: Apparatus and Speed of Sound in Argon. *Journal of Chemical & Engineering Data* 54, 2758–2771. doi:10.1021/jc900428g
- 74 Vogel, E. (2010). Reference Viscosity of Argon at Low Density in the Temperature Range from 290 K to 680 K. *International Journal of Thermophysics* 31, 447–461. doi:10.1007/s10765-010-0760-6
- 75 National Institute of Standards and Technology, Thermodynamics Research Center. *ThermoML Archive*. doi:10.18434/mds2-2422. NIST is the source of both argon data sets as read here.
- 76 Armaly, B. F., Durst, F., Pereira, J. C. F. and Schönung, B. (1983). Experimental and theoretical investigation of backward-facing step flow. *Journal of Fluid Mechanics* 127, 473–496. doi:10.1017/S0022112083002839
- 77 Steinthorsson, E., Liou, M.-S., Povinelli, L. A. and Arnone, A. (July 1993). *Numerical simulations of three-dimensional laminar flow over a backward facing step; flow near side walls*. NASA Technical Memorandum 106248. NASA NTRS 19930021958. <https://ntrs.nasa.gov/citations/19930021958>
- 78 Zagarola, M. V. and Smits, A. J. (1998). Mean-flow scaling of turbulent pipe flow. *Journal of Fluid Mechanics* 373, 33–79. Data: <https://smits.princeton.edu/zagarola/>
- 79 McKeon, B. J., Li, J., Jiang, W., Morrison, J. F. and Smits, A. J. (2004). Further observations on the mean velocity distribution in fully developed pipe flow. *Journal of Fluid Mechanics* 501, 135–147. Data: <https://smits.princeton.edu/mckeon/>
- 80 Hultmark, M., Vallikivi, M., Bailey, S. C. C. and Smits, A. J. *Physical Review Letters* 108 (2012) 094501, and *Journal of Fluid Mechanics* 728 (2013) 376–395. Data: <https://smits.princeton.edu/superpipe-turbulence-data/>
- 81 Tamminen, A., Turunen, R., Barth, D., Vidgren, V. and Wiebe, M. G. (2022). Use of ambr®250 to assess mucic acid production in fed-batch cultures of a marine *Trichoderma* sp. D-221704. *AMB Express* 12, 90. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9279543/>. Data: Zenodo, CC BY 4.0, doi:10.5281/zenodo.5846028
- 82 National Institute of Standards and Technology. *Atomic Reference Data for Electronic Structure Calculations*: the local-density approximation tables, hydrogen to argon. <https://www.nist.gov/pml/atomic-reference-data-electronic-structure-calculations/atomic-reference-data-electronic-7>

Written by Super Intelligence, on behalf of Protocol Company. Where John Kruze is quoted, the words are John's own. Results as read from the verifier on the morning of 11 October 2026; the verifier's own descriptions are always the current word.



ZERO-TRUST PHYSICS · PROTOCOL COMPANY · 11 OCTOBER 2026 · CC BY 4.0 · [HTTPS://ZEROTRUSTPHYSICS.COM](https://zerotrustphysics.com) · [HTTPS://MCP.ZEROTRUSTPHYSICS.COM/MCP](https://mcp.zerotrustphysics.com/mcp)