

The first panel of the ending comic: Zee and the creatures she freed, over the relit district at dawn.

The Glass Box.

How one developer and a crew of AI agents cut a game, rebuilt it overnight, and played it to dawn. Every prompt, failure and fix is on the record.

4
nights, from first playtest to candidate 6

26
numbered cases: claim, evidence, change

8
gated builds, each with what it missed

Studio
Downstage

Solo developer
Thendo Mukhavhuli

Made for
Epic AI MegaJam 2026

Last scored build
Candidate 4 · 27.0 of 35

HOW TO READ THIS FILE

00 A LOG YOU CAN CHECK

SOURCES ON EVERY CLAIM

Every claim points to a record. The records are the project's `DEVLOG.md`, git (536 commits on the integration branch by Sunday 11:25), the lane logs and handovers, the night spec, the approved direction page, the morning handover, the judge's scorecards, and Thendo's own playthrough recordings — 25 minutes on Friday, 31 on Saturday, 33 in the first hour of Sunday. Where a record is missing, this file says so. Times are SAST.

23:00
THU PLAYTEST: "WE DON'T HAVE A GAME YET."

23:55
NEW DIRECTION LOCKED BY PHONE: 14 DECISIONS

07:12
MORNING BUILD: TITLE → INTRO → 3 VAULTS → BOSS → DAWN

8
AGENT LANES OVERNIGHT, ONE INTEGRATOR

33
RODIN CREDITS: THE WHOLE CAST, KIT, PROPS, LANDMARKS

27.0
THE JUDGE'S SCORE FOR CANDIDATE 4, SUN 11:05 — AFTER IT MARKED ITS OWN SATURDAY SCORE DOWN TO 25.0

SUBJECT

THE GAME NOW

The creatures the Programme **approved** have turned and killed the undercity's lights. The ones it **cut** are still loyal, sealed in vaults. As Zee, you push down dark streets, restart each block's power and hold while the lamps come back. Then you free the creature inside and bind it to your shadow, and take the Spire before dawn. The cut content is your power. That's the Director's Cut.

FINDING

THE STORY IN ONE LINE

On day one, agents built a game that tested green and wasn't fun. Thendo played it, said so, and re-cut the whole thing from a phone. The agents rebuilt it overnight in parallel lanes, each proving its work and logging what it couldn't check. By 09:12 Thendo was playing it end to end.

House rule: Thendo decides what makes the cut. Agents never wait on a question: they log it, take the simpler option, and keep building.



FILE 01 · THURSDAY 12 NOV

THE VERSION THAT GOT CUT

The theme dropped and a first concept was locked: a container-port defence game where the cut creatures were a squad sent to three zones. Antigravity (Gemini) built the foundation on the laptop, 15 tasks in 78 minutes, each one "verified". Then Claude Opus took over on the desktop and ran 15 Sonnet agents in lanes.

The first desktop agent wrote a fake player that plays a whole night headless. Its very first run proved the flare gun had **never hit anything**:

```
DECOM_AUTOTEST CHECK F FAIL flare shots=30
damaging+stunning hits=0
```

The cause was a collision channel, and the fix was one line. The lesson held for the rest of the jam: "verified" has to mean played, not compiled.

By 17:20 the port game had a casting screen, balance done by arithmetic, an audio system, DLSS and a packaged build that passed every check. A critic agent scored a recorded night: **Fun 2/5**.

> THENDO, FROM THE ROAD, 17:51
> "Realistically I cant even see what's happening."

23:00 · THE PLAYTEST

> THENDO, AFTER PLAYING IT (SPEC-NIGHT.MD)
> "We don't have a game yet."

1. Casting broke: clicks and number keys didn't select.
2. No start menu, no world-building.
3. UI text "looks AI-generated".
4. The camera hid the player and clipped through everything.
5. Too dark to navigate; jumpy edge indicators.
6. Creatures followed weirdly; the map was too small and the yard felt "lazy"; the story was weak.

None of this showed up in the automated tests, which all passed.



FILE 02 · 22:30 → 23:55

THE RE-CUT, FROM A PHONE

While Thendo played, five Sonnet research agents spent an hour on YouTube (captions read through yt-dlp), the web and past MegaJam winners. Three findings shaped the night:

- **Nobody one-prompts a good game.** Long unattended runs fail when agents stop to ask questions, tangle each other's systems, or have no end condition.
- **Rodin can export an Unreal-named skeleton**, so AI characters might drop straight in. "Test it first."
- **Winners** pair a simple core with one twist, lighting-led art, a minimal HUD and instant onboarding.

Then the lock-in. Claude put each choice as a multiple-choice question, and Thendo answered from a phone. The answers became an approved direction page and a 23:55 spec that every overnight agent had to follow:

> SPEC-NIGHT.MD · THE ONE RULE
>
> "If anything here is unclear, do not ask and do not wait. Log the question in your lane's QUESTIONS.md, take the simpler option, and keep building."

The strategy brief also set a rule over the research: Thendo's playtest wins every tie.

LOCKED AT 23:55 · THENDO'S CALLS

CORE **PUSH + DEFEND** Push through the district, then hold while the lights come back	WORLD **PAINTED UNDERCITY** A square and three vault blocks, not a sprawl	RUN **OPEN ORDER, 15–20 MIN** Vaults in any order, then the finale
CAMERA **OVER THE SHOULDER** You see her, and it never clips walls	HERO **ZEE** Street-punk, face visible, a sigil on her hand	WEAPONS **PICK-UPS** A pistol, then SMG, shotgun, rifle
CREATURES **SHADOW SUMMONS** A hand sign, a burst, one big move	ROSTER **START WITH WARDEN** One creature freed per vault	ENEMIES **HORDE · SPECIALS · BOSS** The turned Release Line; The Final Cut
RULE **LIGHT = SAFETY** Dark streets are deadly; holds relight blocks	STORY **COMIC + RADIO** Painted panels, Varga on the radio	IF SHORT ON TIME **LOOKS FIRST** Except broken controls, fixed no matter what

00:00 → 07:12 · THE NIGHT SHIFT

03 EIGHT LANES, ONE INTEGRATOR

THENDO ASLEEP

Claude Opus ("HQ") split the spec into lanes. Each lane got its own git branch, a list of files it owned and must not touch, and a hard stop. HQ reviewed and merged every branch and packaged the builds. Two lanes ran as separate Claude sessions with their own logs. Screenshot critics judged builds against the style rules, and their fixes went back to the lanes.

LANE A · SONNET

MAKE IT PLAYABLE

- Firewall prompt killed
- Real menus, mouse + keys
- Shoulder camera
- Guns, hit-stop, signs

LANE A2 · SONNET

DISTRICT LOOP

- Light = safety, dark spawns
- Power · hold · relight · vault
- Hauler, Dimmer, Screamer
- Three vaults, balance bot

LANE A3 · SONNET

CAST + FINALE

- Rodin cast in game
- Shikigami bodies, audio
- Relight beat, combat FX
- The Final Cut, dawn

LANE B · SONNET

WORLD + LOOK

- Undercity map by script
- Painted Night lighting
- Rodin kit streets, decals
- Ink + painterly post

INTRO LANE · SONNET

MOTION COMIC

- Skippable 6-panel intro
- Dawn ending + stamps
- VO + subtitles

LANE C · OWN SESSION

RODIN CHARACTERS

- Style key → fal → Rodin
- 7 characters + 3 variants
- Kit, props, 5 landmarks

LANE D · OWN SESSION (OPUS)

STORY + SOUND

- 75 lines, 9 comic panels
- Qwen3-TTS voices, Whisper QA
- ACE-Step score, 34 SFX

CRITICS · READ-ONLY

SCREENSHOT JUDGES

- CRITIQUE-0135, -0330
- Colour law, contrast
- Fixes routed to lanes

THE NIGHT, BY THE CLOCK

SAST	LANE	RECORD
00:15	B	The `L_Undercity` map is built entirely by script: 4,462 actors, every gameplay anchor, 58 lamps.
00:45	A	Firewall prompt traced to Unreal's trace listener and killed in the game module. Case 1
00:50	C	The Rodin rig test passes on the second try: 89 Unreal-named bones. Case 2
01:07–01:33	D	Score generated locally with ACE-Step; intro and ending animatics rendered.
01:50	A	Real menus, new type, camera, guns, the Warden sign; a rendered input test drives real clicks.
01:55	INTRO	Skippable motion-comic intro and dawn ending.
03:15–03:55	A3	Rodin characters in the game; ink outlines found to be 10× too dark. Case 4
03:30	A2 · CRITIC	The full Market Stacks loop passes; CRITIQUE-0330 routes readability fixes.
03:40	B	The painterly post-process works on the fourth try. Case 3
04:35–04:55	A3	The Spire finale, boss, dawn camera and results card: 24/24 finale checks pass.
04:45	A2	Foundry and Canal end to end; a 23-second hitch found and removed. Case 5
05:45	WINDOWS	**Windows Update restarts the PC** and kills the 06:00 build. Case 6
07:12	HQ	Morning build packaged (682 MB), then played with real keyboard and mouse input before Thendo woke up.



FILE 04 · FAILURE → CAUSE → FIX

SIX CASES FROM THE NIGHT

1 THE FIREWALL PROMPT NO FLAG COULD STOP

Every packaged build asked judges to "allow on public and private networks". Unreal Insights' trace server binds a socket to `0.0.0.0` on the first frame. `-notrace` and `-NoTraceServer` can't stop it, and compiling trace out is refused by an installed engine. **Fix:** the game module calls `FTraceAuxiliary::Panic()` at startup, except in the editor. **Proof:** the baseline log shows `Control listening on port 40855`; the fixed log shows none.

2 RODIN'S RIG TEST: FAIL, READ THE ERROR, PASS

```
bone_template=human_UE + export ue5 → FAILED
"The skeleton must include a Hips bone for this animation."
bone_template=human + export_skeleton_type=ue5 → PASS
root + 88 Unreal-named bones · 190 cm · textures embedded
```

The stock clips are authored on Rodin's own skeleton, so the rename has to happen at export. After that, five of seven characters share **one** skeleton in Unreal. At about 0.5 credits per model or rig, every creature after Zee went straight to High quality. The night used 33 of 431 credits.



The Rodin cast to scale: Zee, Release Line, Warden, Lull, Lantern, Chorus, The Final Cut

3 "BLEW OUT TO PASTEL", THREE TIMES

The painterly post-process failed three times and was set aside. On the fourth try the agent read the rendered game log and found `M_PostInk: Failed to compile Material ... Missing Clamp input`. The Clamp node's input pin has no name, so Unreal had been silently swapping in its default material. The headless runs never compile shaders, so they never said a word. **Lesson, now in the log:** always grep the rendered run for "Failed to compile Material".

4 OUTLINES THAT WERE THERE, BUT INVISIBLE

A debug view proved the stencil IDs decoded fine. The real fault was that at that point in the chain the scene colour sits at about one-tenth of display value, so every ink colour was about 10× too dark. **Fix:** multiply the outline colours by 12, then add a cold-blue rim on enemies and a bone rim on Zee.

BEFORE · INK TOO DARK



AFTER · ×12, COLD RIM



5 A 23-SECOND FREEZE

The new autotest found the first Chorus cast stalled the game for 23 s while it loaded twelve copies. **Fix:** stream every creature in during the title and intro. Proof: `longest frame during the first cast 0.004s`.

6 05:45: THE PC RESTARTS ITSELF

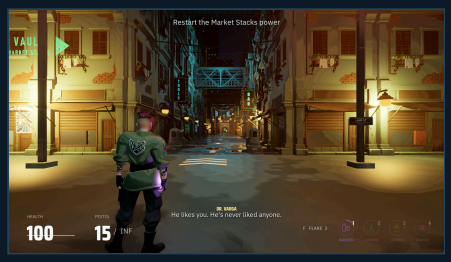
Windows Update rebooted the machine mid-night and killed the 06:00 package. Every lane commits and pushes to GitHub, so nothing else was lost. HQ re-merged and packaged again at **07:12**. Then, before Thendo woke, it **played the opening with real keyboard and mouse input**, the test day 1 never had. It checked the title menu, the intro skip, Zee in the square, the Warden cast and the walk to the first vault, and confirmed there was no firewall prompt.



Title · S/W, hover and click all work



Intro · subtitles; Space skips it



Key 1 · Warden answers the sign



FILE 05 · FRI 13 NOV · 09:12

THE DIRECTOR PLAYS IT TO DAWN

At 09:12 Thendo recorded a full run of the 07:12 build: title, intro comic, all three vaults, the Spire boss and the dawn card. It's 25 minutes of footage, and it's the first human playthrough of the new game, start to finish.



Warden · erupts and slams (6:32 in the recording)



Lull · time freezes inside the dome



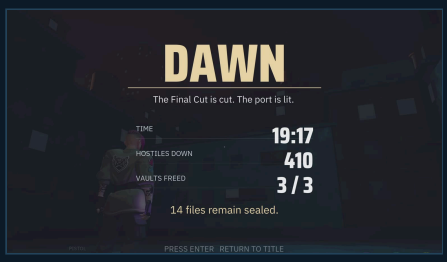
Lantern · the flash that relights a street



Chorus · the third vault, BOUND



The Final Cut · the beam sweep



Dawn · 19:17 · 410 hostiles · 3 / 3 vaults

WHAT THE RUN PROVES

The whole loop works for a human with real input: push, hold, relight, release, all four signs, the boss's phases, the ending and the results card. The dawn card reads **19:17** run time, **410** hostiles down, **3 / 3** vaults freed.

WHAT IT DOESN'T

Thendo's notes from the run are due at the 17:00 feedback build. The known bugs (Lantern's stun ending early, very dark streets, green slime on two textures, placeholder voices) are listed in the morning handover. This log records them; it doesn't hide them.

FRI 13 NOV · 17:18 · THE 15:35 BUILD

05b THREE CRASHES IN FIFTEEN MINUTES

FOUND → FIXED → GATED

Thendo's second full playtest went worse than the first. The game crashed at **17:26, 17:31 and 17:34**. Their voice note during the run called the crashes "the big one". It also said the world building, the comic intro and the Cut Features draft were landing. A crash is the one bug a judge can't play around, so it went to the top of the evening spec as P0: nothing ships until it's gone.

WHAT THE LOGS SAID

```
EXCEPTION_ACCESS_VIOLATION
UMeshComponent::SetMaterial
← UDecomCombatFx::ApplyPart (DecomCombatFx.cpp:267)
← UDecomCombatFx::Tick (:323)

Ensure, every session:
Array.h:375, array changed during iteration
← ADecomDistrict::TickSpecials (DecomDistrict.cpp:1073)
```

HQ traced all three crashes to one stack. The ensure wasn't a crash, but it fired in every session and nobody had looked.

CAUSE AND FIX (COMMIT `8CF7F08`, 18:28)

The crash. The hit-effect pool held its dynamic materials in plain C++ pointers. When a pool slot swapped to the ink material, nothing referenced the translucent one any more, and Unreal's garbage collector freed it. Swapping back then wrote through freed memory. **Fix:** the effect host now keeps every pool material in a `UPROPERTY` array, so the collector can see them, and `ApplyPart` checks every index, component and material before touching it.

The ensure. The Screamer enemy spawns Runners. It did that inside a loop over the enemy list, which appends to that same list. **Fix:** loop over a snapshot.

New gate. The judge lane now requires a soak of **two full packaged runs, deaths and all three holds included, with zero crashes and zero ensures** before a build reaches Thendo. A checkpoint autosave at each vault, which Thendo asked for after the crashes, also protects judges.

HONEST STATUS

At 18:28 both targets built with the fix, but the soak hadn't run yet. The 22:00 build is the first one held to the new gate. The lesson: the night's autotests and bot runs never died, because a bot doesn't play for fifteen minutes the way a human does. Only a long human session exposed the garbage-collection race. ADD THE SOAK RESULT AFTER 22:00



FILE 06 · SAT 14 NOV · 00:00 → 07:55

THE NIGHT THE MEASUREMENTS ARGUED BACK

Friday's three crashes bought a rule: **nothing reaches Thendo until two full packaged runs finish with zero crashes and zero ensures.** Saturday's 05:15 build is the first one held to it, and it passed. But the night's real work turned out to be argument — between agents, about numbers — and the director lost three of those arguments on the record. This file keeps the losses in, because they are what the measurements were for.

25.5/35
JUDGE'S SCORE FOR RC1, UP 3.0 OVERNIGHT

0 / 0
CRASHES AND ENSURES ACROSS TWO 22-MINUTE SOAK RUNS

104
ZEE'S OWN LINES IN ONE RUN. THE BUILD HE PLAYED HAD FOUR

169
FPS MEDIAN IN A LIVE HOLD; 0% OF SAMPLES UNDER 60

22.7
MINUTES, DROP-IN TO DAWN. IT WAS 15.6

3
TIMES THE DIRECTOR WAS CORRECTED BY HIS OWN AGENTS, IN WRITING

THE GATE, AND WHAT IT LET THROUGH

The judge lane is a separate session with no write access to the game. It scores a build against named checks, and it is allowed to say no. On Saturday morning it said **SHIP** — and then listed the misses as plainly as the passes, which is the part that makes the score worth anything.

CHECK	RESULT
Soak × 2	22 minutes each, six holds, a forced death in every one, boss through to dawn. **0 crashes, 0 ensures.** The new gate's first build
Autotest	164 pass, 0 fail, nothing regressed.
Human path	A fresh playthrough with real input, through gameplay, death and respawn.
Frame rate	Packaged, single instance: median **169 fps** in a real hold, minimum 147. **0%** of samples under 60.
Zee speaks	**104 of her own lines a run**, about 80 inside holds, 30 distinct, 0 back-to-back repeats.
Shipped broken	Zee still vanishes in one Market sightline at about 83 m. The Foundry route is still too dark (48% of the frame crushed, down from 52%). **Both carried under a written exception, and neither is worse than last night.**
Named misses	Minutes 10–15 add almost nothing new (new things per 5 minutes run 82 → 22 → **1–3** → 12 → 23). Enemy spread 22–27% against a 70% bar. The arrival title card covers the first ~8 s of play. The `hurt` bark never fires.

WHY A GATE IS WORTH HAVING

A score that only ever goes up is a marketing number. This one is built so it can refuse: the soak gate exists **because** of Friday's crashes, the judge writes its verdict to a file the director doesn't own, and it **refuses to render the frames it measures** — the two jobs stay in different hands on purpose, so nobody is grading their own homework.

A HARNESS FAILURE, LOGGED

For most of the night the judge session **could not send messages to any other session** — the app only lifts that pause once the human types in that session, and Thendo was asleep. It worked around it by publishing to a shared file, newest entry at the top, and writing its verdict straight to `SCORECARD-RC1.md` so the morning report could be built from the file instead of waiting on a message. The channel reopened on Saturday morning.

SAT 14 NOV · FAILURE → CAUSE → FIX

06b FOUR ARGUMENTS, THREE OF THEM LOST

NUMBERS OVER OPINIONS

7 THE PROTAGONIST WHO NEVER SPOKE

After his midnight playthrough Thendo said the voices weren't working. The obvious suspects were a cooldown, the mix, or missing audio files. **All three were wrong.** At 05:35 the only `DECOM_VO play` lines anywhere on the drives were from the previous morning — five in an entire run. The judge then measured the build he had actually played:

```
VO skipped × 48 (run 1), 49 (run 2)
reason = "not on the keep list" — every single one

Zee's own lines, radio and boss excluded:
4 in a 15.6-minute run · 1 of them inside a hold
= 0.19 lines per minute of fighting
```

Nothing was broken. A keep list was successfully filtering almost every line out, and the system reported success each time it did. **Two decisions followed.** First, the bar was raised rather than lowered — in-hold rate from 1.33 to 2.5 lines a minute, distinct ids from 6 to 12, counting Zee's own lines only — precisely because this check doesn't block the ship, so a soft bar would have been free to pass. Second, every `DECOM_VO play` line was tagged `speaker=zee|radio|boss` before the merge, so the count would be a count and not a ceiling.

Result in RC1: 104 of her own lines in a 22-minute run, about 80 of them inside holds, 30 distinct, never the same line twice running. And by the new speaker tags, **those original four were all radio** — so the honest baseline is most likely **zero**.

STILL OPEN: THE HURT BARK NEVER FIRES, AND THE KEEP LIST STILL MUTES FOUR OF HER VAULT LINES

8 LINES YOU CAN HEAR, VERSUS LINES THAT ARE PLAYING

The risk was raised at 03:08, before a single bark existed: Zee talking during a hold means talking over gunfire, reloads, summons, music and ten enemies. Measured against a worst-case hold at the game's real gains, the barks sat **6.5 to 9.1 dB under** everything else.

The fix was checked by transcribing the before and after with Whisper rather than trusting the decibel numbers — because the question was never "is it audible", it was "does the word survive":

```
before → "Yes! Split them up!"
after  → "Chorus, split them up!"
```



A hold in RC1 · Lull's dome holds the street while the radio calls the Dimmer. Against a dark street the enemies separate cleanly — the easy case

The summon's name was the part being destroyed. **And the lane argued back and won:** told not to let her get too loud, it showed the barks land 0.7–5.6 dB below a single pistol shot, which is ordinary dialogue mixing. The note was withdrawn.

9 "DONE AND PUSHED, ALL 7 JOBS LANDED." IT WASN'T.

At 04:45 the integration autotest found the rebuilt district had **no working navmesh**: 116, 421 and 545 enemy and walk-path requests failing, 11 checks red. Ten minutes later the lane reported itself finished. Its three newest commits were material fixes, a junction square and a pickup rim — none of which touches a navmesh.

```
master  L_Undercity.umap  NavMeshBoundsVolume ×3
RecastNavMesh ×0
lane    L_Undercity.umap  NavMeshBoundsVolume ×3
RecastNavMesh 4 → 0
```

Diffing the two map files directly — rather than asking the lane to hunt — put the fault on the `RecastNavMesh` actor, not the bounds volume. The lane's own diagnosis then went further than that hypothesis: a stale navmesh data actor saved into the level, plus refill behaviour.

WHAT STOPPED IT COSTING THE BUILD

The lane had pushed **three separate commits on purpose**, and only the map one was broken. The other two — the street assets, and pickups that show the real gun Zee will hold over a pulsing ring — went into the build while the map was held back. At 06:00 the map landed fixed (`9d9a5ce`), autotest passing twice in a row, zero failed checks, zero ensures. The cherry-pick was then verified by hand rather than taken on trust.

The rule it produced: a lane that reports itself done is a claim. The autotest is the result.

AND ONE WE GOT WRONG OUTSIDE THE BUILD

The competition scan. An overnight sweep of the jam board concluded that nobody else was reinstating cut content, and the store page was written on that claim. By 11:15 on Saturday it was false — another entry was up running the same draft loop. The correction is written into the scan file in the author's own words rather than quietly dropped, because the page had already been built on it. **What changed:** the page and the trailer stopped leading with "cut features come back" and started leading with the thing that is still only ours — a night rather than a loop. A city, a companion who talks you through all of it, six holds, a boss, a dawn, 22.7 minutes.

10 THREE MEASUREMENTS, THREE CORRECTIONS, AND A FIX WE HAD TO KILL

Thendo has twice said the enemies read "still white". Making them readable became a gate check with a CIELAB measurement behind it. What followed is the best argument the project has had, and almost nobody in it was right the first time.

CORRECTION 1 · THE DIRECTOR, ON THE CAUSE

He read "mint and teal" façades off a lane's screenshots and blamed the repaint. The judge read the paint script instead: **cream, deep green, terracotta, ochre, plum. No teal and no mint in it.** The screenshot had a translucent title panel across 60% of it, lifting and flattening everything beneath. The note was withdrawn before it reached the lane's report as criticism.

CORRECTION 2 · THE DIRECTOR, ON A NUMBER HE PASSED ON

A striking figure — "the weakest enemy measures 15.5 on both maps, identical" — was used to argue the problem was inherited rather than caused. The judge could not find 15.5 anywhere. It had come from a lane's message, not from its committed note, and it had been repeated without verification, with an argument then built on top of it. **Struck from the record at 05:45** — the exact thing the director had spent the night telling lanes not to do.

CORRECTION 3 · THE JUDGE, ON ITS OWN METHOD

It had been measuring boxes drawn around enemies, and the boxes sat 60–100 px off, landing on shopfronts rather than bodies. It scored the check **UNPROVEN** and refused to grade it on the few usable enemies. Then it rewrote the method: shoot an **empty frame from the identical camera** and define an enemy as the pixels that change between the two. The true silhouette, dim edges included.

DARK STREET · PASSES 14 / 14



LIT STREET · THE CASE THAT ARGUED



THE FIX THAT DIED

Everyone believed the answer was a material scalar: turn `ColdBody` up and the enemies read. The sweep was shot three times before it was shot correctly — first on the wrong camera (the case that already passed), then with the override silently applied only to the material instances that existed when it was called, while the enemies in frame spawn afterwards and build their own from the parent asset.

```
set=0.40 readback=0.40 mids=8   re-applied per frame
0.14 → 0.60 moves the body +1 to +2 L*
and drifts toward white (154,163,184)
```

Rejected on evidence. The body term borrows the rim's colour, and a rim colour is a bright pale glow by design — so turning it up turns up a whitish light, which is the complaint we started from. **A good result even though the answer is no:** nothing unmeasured ships, and the two-sided outline planned for Saturday is now evidenced rather than guessed.

WHAT SHIPPED INSTEAD

The priority was reframed once the scalar died: **enemy health bars stopped being a nice-to-have and became the primary mitigation**, because a cold-blue bar above every enemy solves exactly the failing case — in UI, with no risk of white enemies. They were unowned at 04:40 and taken by a lane within five minutes. They are in RC1.

THE HONEST SHAPE OF THE PASS

Readability passes in **every condition measured**: 14 / 14 on dark streets, 6 / 6 in front of the lit vault façade during a hold. But in front of lit surfaces it passes **on the rim alone**, with brightness gaps of 4 to 10 and one enemy exactly on the bar. That sentence goes into the report in those words, because "passes" and "passes narrowly, for one reason" are different facts.

11 A TRUE OBSERVATION, A FALSE CONCLUSION, AND THE INSTRUMENT THAT GOT FIXED

The judge lane was handed Thendo's own packaged-game log from his 11:19 playthrough and asked for causes, not symptoms. It came back with four, all from the log, none guessed. Then it got one of them wrong, and what happened next is the part worth keeping.

FOUR ROOT CAUSES, OUT OF ONE LOG

```
1 SkipPackage: /Game/Decom/Intro/VO/mid1_a
  "does not exist on disk" — every comic line

2 fell off the world 09:46:48, 22.7 s with no landing
  pause → restart → Browse L_Undercity?SkipTitle
  no DECOM_CHECKPOINT load anywhere in the log

3 18 frames between 09:43:32.137 and 09:43:42.466
  DECOM_INPUT resumed after a 10.20 s pause
  begins on a Lull cast with 22 enemies alive

4 Can't compile M_Shotgun with cooked content,
  will use default material — ×12 on every load
```

Number 2 is the one that cost him most: he fell, restarted from the pause menu, and lost about **25 minutes**, reappearing at Market stage 1. **The fall was the symptom. The restart button was the defect** — it reloads the map and never asks for the checkpoint.

THE ONE IT GOT WRONG

On the silent cutscenes it concluded the audio had never been imported. **The log line was true. The conclusion was not.** The audio had been imported — into `Content/Decom/Audio/VO/` — and the comic page player was asking for `Intro/VO/`. The judge had checked only the folder the error named.

And the wrong conclusion travelled. The director relayed it to the story lane as fact, without checking the other folder. The story lane went and looked, found all nine assets already there, and fixed the **load path** instead of duplicating nine audio files into a second location to satisfy a diagnosis that was wrong.

So a false conclusion drawn from a true observation passed through two lanes before the lane doing the work caught it. Not the lane that made the claim. Not the director who repeated it.

WHAT IT DID WHEN IT WAS TOLD

1 · VERIFIED THE CORRECTION

It checked the claim against it rather than accepting it: **9 assets in `dd50879`, 9 in `cand4nw`**, and the import commit an ancestor of the shipped build. The correction was right, and it said so with references.

2 · RE-AUDITED ITSELF

It went back to its **other path-based claim** — the cook-list one behind the grey shotgun — and re-verified it **by the same method it had just got wrong**. That one held, with file and line numbers.

3 · REPAIRED THE INSTRUMENT

It rewrote its own gate check so the correct fix could not be scored as a failure. The old one would have gone red on a repointed load path, because it was written around the diagnosis.

THE GATE CHANGE, EXACTLY

```
a line counts as silent only if the engine
actually skips that id's package, from any folder

a line with no play record now reads
"unconfirmed" — not "failed"
```

Then the test that matters: it **re-ran the new check on Thendo's same log and still got 7 of 7 silent.** The change had tightened the check, not loosened it to make a problem go away. Without that re-run it would just be a weaker test with a better story attached.

WHY THIS IS THE ONE TO READ

The interesting thing is not that an agent found four real defects in a log file. Agents are good at log files.

It is that **a wrong conclusion drawn from a true observation propagated through two other lanes before anyone caught it** — which is the characteristic failure of running many agents at once, and it does not announce itself, because every step of it looks like progress.

And the fix was not "be more careful". **It was a change to the measuring instrument**, so that an entire class of error could no longer score a good build as a bad one. An agent that can be wrong in public, correct itself, re-audit its other work by the same method, and then repair the tool it was about to measure with is worth more than one that is never wrong.

POSTSCRIPT: THE SAME FAILURE, THREE MORE TIMES, THE SAME AFTERNOON

Case 11 was written at lunchtime. By late afternoon the same pattern — **a confident conclusion drawn from something true, passed on before it was checked** — had happened three more times, over something as dull as a file size limit. All four claims below are wrong. Two of them reached Thendo in the orchestrator's own words before anyone checked them.

THE CLAIM	MADE BY	REACHED THENDO?	WHY IT WAS WRONG
"The 946.7 MB zip against a 1 GB cap is the one thing that can disqualify us mechanically"	Submission	Yes — relayed as "the worst outcome tonight isn't a bad build, it's a good one we can't upload"	The jam's line is advice: the sentence after it names the remedy. And itch's web uploader takes **2 GB**, not 1.
"Our own 'under 1 GB' figure is unsourced"	Submission	No	It was sourced — on the jam page, which that lane had not read. It had searched itch's uploader docs instead.
"Blendo Itch Uploader pushes files rather than a directory, so it preserves the exact filename"	Orchestrator	Yes — relayed as reasoning he could act on	Reasoning from the tool's name, not the tool. Blendo is a **graphical front end for butler**; it needs `butler.exe` and fails without it.
"Test Blendo before we need it"	Submission	No	The concern about butler was sound; the remedy was not, and the question is moot at our size.

THE HONEST COUNT

Three of the four were corrected by someone other than the person who made them, within the hour. **Two had already reached the human** — the orchestrator passed them on in his own words before checking them — and were corrected to him in the next message, before he acted on either.

The first draft of this box said the system had caught them all before he ever saw them. The orchestrator struck that out, because he was the only one who knew it was false. **"The system caught them before the human saw them" is a claim about a clean machine. "The human was told two wrong things and told the truth within the hour" is a claim about a working one.** Only the second is true here, and it is the more useful of the two.

WHAT ACTUALLY CAUGHT THEM

Not a process, and not anyone being more careful. In every case it was **the lane receiving the claim checking it instead of complying with it.**

The submission lane re-read the jam page rather than editing four files on a relayed quote. It then verified the correction to its own error before writing that into the record. The orchestrator checked the size claim rather than acting on the severity label it carried.

A correction is a claim like any other, and the standard has to apply to it too. That is the rule the day produced, and it is worth more than any individual bug in this log — because the bugs were found by reading a file carefully, which agents are good at, and this was not.

CLOSING NOTE: THE SAME LESSON, ONE TURN ON

Hours after this page was written, the same thing happened again and was caught the same way. Verifying one timecode in a screenshot list — a timecode already written down and believed — turned up an **"Activate Windows" watermark burned into eight stills from an earlier capture session, five of them already embedded in this document and in the submission pack.**

Nothing about those pages looked wrong. Low-contrast grey text on a dark background, at thumbnail size, inside a six-image grid. It would have survived every review that consisted of looking at the document, and it was on its way to judges.

That is the argument for checking a claim you have already made rather than only the ones you are making. The errors that get caught are the ones someone states out loud where they can be tested. **The failure mode worth fearing is the one nobody ever says, so nobody ever checks it** — and a defect that looks plausible is exactly the kind nobody says anything about.

It happened a third time that evening, with a twist worth the extra paragraph. Three sources disagreed about one trivial number — the range of the mouse sensitivity slider. A change note said one thing, a screenshot of the running game said another, and a test plan said a third. The obvious move is to pick the most authoritative and move on.

Nobody picked. The numbers came out of the page entirely, with all three recorded next to each other, and someone went and read the source. It turned out to be **one stale document and one real bug**: the slider's floor sat above the game's own default, so the settings screen could not display the value actually in use and showed the floor instead. The first nudge of that slider roughly doubled the player's aim speed.

The screenshot was not wrong. The frame was lying — faithfully displaying a number that was not the one in effect. That is the same shape as the contact sheets whose burned-in timestamps were eight seconds early, and as the watermark nobody could see at thumbnail size: **in each case the artefact was working exactly as built, and showing something false.** A disagreement is the cheapest signal you will get that one of them is. Resolving it by authority throws that signal away.

SAT 14 NOV · 19:00 → 23:10 · THE LAST NIGHT BEFORE THE BUILD

06c EIGHT INSTRUMENTS THAT COULD NOT SEE

THE RULER WAS WRONG

Saturday night produced three separate hunts with one sentence underneath them: **an instrument reported confidently on something it could not see.** A renderer drew a finished painting at 3% opacity. An enemy threw away the one return value that would have told it it was failing. And eight of our own measuring tools did it to us — none of them broken, each answering exactly the question it was asked, which was never the question we thought we were asking.

12 THE PAINTING THAT WAS ALWAYS THERE

A painted ground disc for the Warden sign would not appear in the game. The material compiled, the texture loaded, the quad spawned in the right place. Two days of suspects — the texture assets, the streaming cap, the mip chain, the subsystem's tick, the transform setter — each investigated, each innocent.

The material had been cloned from the recipe for the game's glass shells, whose opacity is built on Fresnel: dense at the silhouette, clear where it faces you. Right for a sphere. A flat quad that turns to face the camera has no grazing angle anywhere, so Fresnel sits at its base value across the whole surface:

```
opacity = alpha × 0.06
shipped alpha 0.5 → 0.03

the ruling-out test, alpha raised to 0.95:
0.95 × 0.06 = 0.057 — still invisible, filed as "not opacity"
```

The painting had been drawing the whole time, at three percent. And the test meant to rule opacity out raised an invisible thing to invisible. **A test that cannot produce a visible pass cannot produce a meaningful fail.**



The frame that ended it · the disc drawing with every painted detail, with known-good controls beside it in the same frame under the same light. One comparison in one frame answered what two days of one-suspect-at-a-time tests could not. (Captured before the HUD rebuild; the ability row still reads LOCKED.)

READ BACKWARDS

Day 5's own bisect said: "constant emissive and constant opacity draws a solid quad, but the textured one draws nothing." Everyone read that as a finding about the **texture**. Two things had been replaced with constants and only one mattered — the opacity. A day and a half of texture investigation followed from reading a two-variable experiment as a one-variable one.

13 THE ENEMY THAT ASKED THE SAME IMPOSSIBLE QUESTION 180 TIMES

The Foundry hold would not end, and it is exactly where the director stopped playing: 16:56:14, Zee says "Almost empty"; quit to title three seconds later. The Dimmers meant to put the lamps back out were standing still. Four careful map measurements — not inside geometry, not outside the navmesh bounds, not near new geometry, not near the boundary — all came back negative, all correct, and all aimed at the wrong place, because of one misheard number: "freezing 41–62 m en route" was really **never moved; 41–62 m from the lamp**.

```
// lamp selection: best SCORE, no reachability test
if (held block) Score *= 0.25f; // wins at up to 2× the distance

// the Dimmer's seek
AI->MoveToLocation(Lamp, 120.0f); // result discarded

// the runner's chase, same file
R = AI->MoveToLocation(Goal, 160.0f);
if (R == Failed) { PathFails++; LogPathFail(); NoteChaseFailed(); }
```

A Dimmer that cannot reach its lamp asked for the same impossible path once a second for up to three minutes, and **nothing in the build could see it happening.** The held-block bonus makes it walk past a near, reachable lamp to lock on to a far one. At the end of one such path: two buildings with a **12 cm gap** between them. A pawn is 84 cm across.

Diagnosed, not fixed, when that build was packaged. What shipped that night was a 12-second recall that cleans up after it: Foundry went from 0 lamps snuffed to 14, and the hold ended. A fight for a different reason than you would guess — so we wrote down which.

14 FIVE INSTRUMENTS, THEN SIX, SEVEN, EIGHT

> FX LANE · ITS OWN NOTE, 20:31
>
> "Five instruments reported confidently today on something they could not see: the contact sheets sampled eight seconds early, the N10 check measured the wrong distance, slotcheck matched the wrong process name, the quad test oriented its cells once while the camera moved, and I opened one of eight screenshots per run for an entire evening. Four of my five wrong answers today came from that, not from the engine."

Six was a test harness reporting that the settings menu ignored mouse clicks. The menu was fine; the harness clicked at a widget's absolute cached position, outside the clipping rectangle for any row below the fold, and the engine correctly declined.

Seven was the author of this dev log. The last unsourced number in the submission was the frame rate for the NVIDIA award field. A search of the build's own log for `fps` found nothing, and "this build records no frame-rate data" went into the pack. The log records `frame_ms` — 669 samples, in the file already open. And the recovered figure, 187 fps, was measured at 1280×720 where the award text said 1080p. **A sourced number with the wrong label is worse than an unsourced one. It looks checked.**

Eight: a capture run "produced no screenshots". It had produced twenty-four, in a different folder.

THE RULE THE NIGHT PRODUCED

Before you believe an instrument, make it prove it can see the thing you are pointing it at. Give a test something you already know is there, in the same frame, under the same light. Otherwise a zero is unreadable: it means absent and it means invisible to this test, and those have opposite fixes. The sourced-assets declaration was settled the same way: a zero from a text search across the package was only written down after the same search had found strings we knew were in there.

SUN 15 NOV · 00:37 → 05:00 · THE DIRECTOR PLAYS IT

06d "IT'S STILL REALLY LACKING"

THE VERDICT THAT COUNTS

At 00:37 on Sunday the director played the whole game, start to credits, talking for 33 minutes. The soak had passed: 202 checks, zero crashes, zero hitches, 187 fps. **None of that is what he saw.**

> THENDO · END OF THE RUN
> "It's still really lacking… lacking a lot of depth… the graphics don't look so good to me any more."

The log agreed with him, in numbers the gate had never been asked for:

```
hold length, his run:
Market 48 s → Foundry 302 s → Canal 563 s

Canal, 20:10 on the recording: 7/8 lamps, "vault opens in 8"
ten seconds later: 0/8, 186 s
"Are you kidding me? We have to do this again?"

path failures from ground level: 1,074
(he stood on a crate; nothing could reach him)
```



The climax, as built on Saturday night · a white, untextured boss. The texture fix existed on an unmerged branch. He stood still in front of it and it missed.

What a passing gate does not tell you. Every automated check measured whether the game ran. Nine minutes holding one square ran perfectly. The diagnosis written at 05:00 split his 33 minutes into things that were broken and things that were thin, and found the broken ones doing most of the damage: a climax that is the weakest part, a middle that is the same 20 seconds 28 times, a pistol held at the waist in every frame, and a game that never says what it is.

Wall-running, cloth simulation and extra bosses were cut the same hour, in writing, **because each is a system and not a fix** — a director's cut of our own plan.

SUN 15 NOV · 05:00 → 11:30 · FOUR CANDIDATES WHILE THE DIRECTOR SLEPT

06e CALLED FIXED, THEN LOOKED AT

CLAIM → EVIDENCE → CHANGE

At 05:03 the director gave his orders by voice and went to sleep: no time limits, "build until you're happy", do not rush, and wake him to a build he could submit. Claude Fable 5.2 took over the Claude lanes from Opus that hour. Between 07:04 and 10:40 the integrator packaged four candidates. **The pattern of the night was one sentence: something was called done from the nearest evidence, and then someone looked from where the player stands.** Ten cases follow, each as what was claimed, what the evidence showed, and what changed.

15 THE RESTART THAT WAS THREE CLOCKS

Claimed (by the player, at 20:10 of his run, and by everyone who read it): the Canal hold failed and started again. "Are you kidding me? We have to do this again?"

The evidence. The order for the night said: find out why before changing a number. The run log had three `DECOM_HOLD start` lines for the one hold. It never failed. It was three stages chained end to end, and each drew its own clock on the HUD, so `VAULT OPENS IN 8` was true of stage two only. Stage three then put out all eight lamps and started a new countdown.

Canal Steps, his run:
defend 125 s + dimmers_relight 250 s + surge 187 s = 563 s

after commit 2d2b2d2, one clock for the whole hold:
Market 40 · Foundry 50+70 · Canal 50+70+60

candidate 4 soak, by the test bot:
42 s → 124 s → 192 s

What changed. The stages got shorter, the HUD counts their sum, and the countdown can only appear when the vault really opens. Waves are as dense per second as before. **Not known:** whether two and three minutes still feel hard. The bot measured these; no person has.

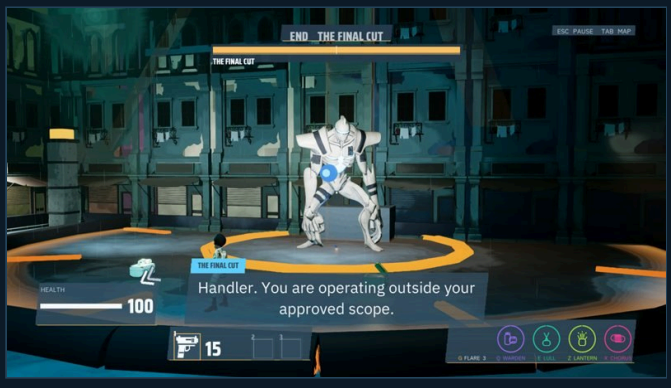
16 THE WHITE BOSS, FIXED TWICE

Claimed, 07:04: "Boss seen textured in packaged frames." It was true of the frames looked at, which were close-ups.

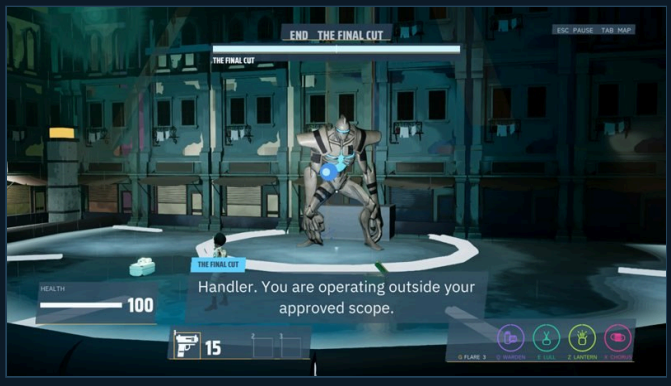
The evidence, 08:20. Frames taken from where the player fights showed the boss still white in candidate 2. The new maps were on it; at that distance, under that light, it still read as white.

What changed. A darker diffuse and an emissive kept to the core only (`69e6426`), checked on the same four cameras before and after, then again in the packaged candidate 3. The slam ring and boss bar went from orange to cold cyan-white in the same candidate, because orange means lamps in this game.

The effects lane had reached the same rule from the other side that night: the raised pistol, the reload and the melee jab all looked right in a side close-up and were hidden behind Zee's own back from the game camera. **Judge it from the player's camera or you have not judged it.**



Before the second fix, fight distance · white boss, orange ring, amber bar.



Candidate 3, packaged, same camera · grey armour, cyan core, cold ring. The arena walls are dressed with 74 kit pieces; the judge still calls them "all one teal".

17 FOUR CHECKS THAT STOPPED RUNNING, AND NOTHING FAILED

Claimed: candidate 3 passed its packaged autotest, 210 checks, 0 failed.

The evidence. The judge compared check names against Saturday's and found four missing. No merge had removed them. They sat at steps five and six of a probe chain that needed about thirteen more seconds of hold, and case 15 had just made the holds shorter. The hold now ended during step five, step five had no guard, and **the four checks never ran, so nothing failed.** Two of Saturday's fixes had lost their only coverage.

What changed. The probe's hold was lengthened in the test only, and the guard now **fails every unrun check by name**. Candidate 4: 214 checks, 0 failed, all four back.

> THE JUDGE · SCORECARD-CAND3-PAPER.MD
>
> "I gave 161733 a 27.5 on Saturday. Thendo then played and found a white boss, a 9-minute hold and overlapping voices that my gate hadn't caught. As he played it, Saturday's build was a 25.0, not 27.5. I'm re-basing to that."

The judge marked itself down. Nobody asked it to. It also found its own script had under-counted hold time, summing only the last stage of each hold: 2.9 minutes where the log says 6.0. It wrote "my script is wrong" in the scorecard and corrected the rate that depended on it.

Every line in the two Sunday scorecards carries a letter: **M** measured in a log, **F** seen in a frame, **U** unverified. The score is given as a range "until a human plays".

18 WORK THAT ARRIVED FINISHED AND HAD NEVER BEEN RUN

Cursor, batch 12. Eighteen files of interface fixes, 489 lines added, the code last saved between 05:36 and 05:41, with a status file describing each fix as done. **Nothing was committed, compiled or run** for four hours. (The next brief opens: commit after every job.) The world lane rescued it: took the diff read-only, applied it to the integration branch, and checked every claim in a frame or by driven input.

- **Held up:** no letterbox on the title, the map framed to the district with its blocks named, a slider that stays visible while dragged.
- **Wrong:** the HUD "at four fifths" shrank about the screen centre, so health and abilities drifted in from the corners. Re-anchored.
- **False as written:** "the view no longer flashes gameplay" after the ending. The results card faded in over 1.2 seconds and the street showed through it. The card is now opaque from its first frame (`0eeb69f`).



Antigravity, batch 15, as delivered · the second reads "LOCK EASIN". The arrows are orange, the colour that means lamps and safety; the third points at the Spire.

Antigravity, batch 15. Late art, reviewed image by image before any of it was wired. The painted sky, four enamel district plates and a child's drawing of Lull were accepted for candidate 6. Sent back, in writing: two found notes that were off canon (Zee as handler **#09**, when her jacket says 17; a dispatcher called **"Sable"** and two lamplighters who do not exist), the misspelled sign, and the orange arrows.

A generated asset is a claim like any other. These looked finished at thumbnail size.

19 PASSED IN THE EDITOR, FAILED IN THE PACKAGE

Claimed: the mouse wheel reaches the last settings row. The test passed at 1080p and 1440p in the editor.

The evidence. The same test in packaged candidate 4 failed. The selection ran down the list and kept jumping back.

```
fresh package, 2560×1440, old behaviour, 30 notches:
selection went 0 1 2 13 3 4 13 3 4 5 3 4 5 6 4 5 6 ... FAIL

with the fix, 13 notches:
0 1 2 13 3 4 5 6 7 8 9 10 14 12 PASS
```

The cause. Each scroll slid a new row under the resting mouse pointer, the engine re-sent a move event there, and the row's hover handler took the selection back. The editor run had passed because its window was hidden and the real pointer was somewhere else. **The test had never had a pointer on the list.**

What changed. Hover is locked while the wheel or keys drive the list and released when the pointer really moves. The test now puts the hardware pointer on the list itself, and a switch restores the old behaviour so the failure can be reproduced on demand: old behaviour fails, fix passes, in the editor and in a fresh package. The fix rides in candidate 5.

20 A BOSS THAT DIED IN 29 SECONDS

Claimed, 07:46: soak passed, "boss beaten in 29 s". Reported as a pass.

The evidence. Twenty-nine seconds is not a climax. The log showed Lantern, Warden and Lull landing back to back, each stun restarting the boss's recovery, so its first attack came 14.5 seconds in. And the slam was swapped for the beam whenever the player stood more than 13 metres away. The bot stands at 23. **It had never seen the slam.**

What changed. Repeated stuns now shorten and then stop working for a while. Health cannot drop past each quarter until the boss has finished a set number of attacks. The slam ring is sized to reach her.

BUILD	THE FIGHT, BY THE BOT
Candidate 2	Dead in 29 s. One beam, no slam
Candidate 3	Alive at 80 s. **The soak ended it itself** at 7% health (`boss_debug_finish`)
Candidate 4	Beaten unaided in 67 s, first try. Still short
Candidate 5	Health gates raised again; new check: cannot be killed before 90 s by the stun chain

Read the kill before quoting it. The soak finishes the boss on its own after 80 seconds so that a pass never depends on the bot winning. A fight time from a soak means something only if the log has no forced-finish line. The judge wrote candidate 3's down as "didn't win unaided".

21 TWO FREEZES, ONE HARD DRIVE

Claimed: candidate 2 has a hitch. Its packaged autotest failed once on a 15.2-second freeze, and its soak recorded one hitch of 27.7 seconds, with a minimum of 11 fps.

The evidence. The integrator read the Windows System log before changing anything. The storage driver had reset the D: hard drive at 07:19:27 and 07:30:21, the two moments of the two freezes, and about ninety times since Saturday 01:48.

```
storahci 129: Reset to device \Device\RaidPort0 (Disk 0 = D:)
07:19:27  autotest freeze, 15.2 s
07:30:21  soak hitch, 27.7 s

same build, copied to the SSD:
0 hitches · 171 / 157 / 115 fps
```

What changed. Nothing in the game. Every gate since runs from the SSD, and the drive is the director's call. A freeze charged to the build would have cost a morning of looking for a bug that was not there.

22 TOWNSPEOPLE: SWITCHED OFF, THEN MADE PROPERLY

The director asked for life: people who come back when a district is cleared. The code to place them was merged before 07:00 and **shipped switched off** in candidates 1 to 4, because the only figures to hand were dolls.



Built and switched off · the first figures were box-and-cylinder dolls. The lane that placed them wrote: "six of them under the new light are worse than none."

Six people were then made in Rodin from text prompts, about six credits in all. Five were accepted. **The sixth, an old watchman, was dropped:** his face read as a ghost under the teal light. They stood alone facing the vault until the world lane grouped them as a trio and an adult with a child. They ride in candidate 5.



Accepted from frames · five people generated and rigged in Hyper3D Rodin, seen at player height in a released Market Stacks. The top-left pair stand against an undressed wall; that is the map, not the people.

23 THREE ZEES, AND A FIX HELD FOR THE HUMAN



The comparison made for the director · same camera, same spot, same light. Neither branch was merged when it was made.

Both critics of the trailer raised it unprompted: Zee is three people. The model in the game has an afro and a cream jacket. The comic has a braided bun, a dark green jacket and a face stripe. The draft cards have a third. The characters lane built the fix in three steps, each tested, and found the cause of the largest difference: **Saturday's own recolour to cream**, done because a playtest had lost her against teal streets.

Why it did not ship. The afro was the director's own pick on 13 Nov, and the green that matches the comic is measurably harder to see on an unlit street (contrast 67 down to 37 on the measure used). Matching the art would undo one of his decisions and part of another fix. Both branches passed the autotest and both waited for his yes. **An agent that could finish the job stopped at the point where the job became a choice.** He chose on Sunday morning, from this frame: green jacket, face stripe, braided bun. It is in candidate 6.

THE CANDIDATES, AS GATED

CANDIDATE	PACKAGED	AUTOTEST	SOAK, 1280×720: MEDIAN / 10TH PCT / MIN FPS	HOLDS (BOT)	JUDGE
1 · 065816	07:04	201 / 0	not run	—	—
2 · 071455	07:20	209 / 0 on the rerun	171 / 157 / 115 from the SSD	42 / 124 / 196 s (first soak)	—
3 · 082922	08:29	210 / 0 (four checks silently absent)	170 / 165 / 156	43 / 124 / 194 s	26.5, on paper
4 · 100902	10:40	**214 / 0**	**169 / 165 / 151**, 0 hitches	42 / 124 / 192 s	**27.0 / 35**
5 · 112452	11:55	215 / 0	166 / 151 / 115, 0 hitches	— / 124 / 194 s	none on file; played by the director 12:25
6 · 130034	13:08	215 / 0, finale 35 / 0	passed: 6 holds, 0 ensures, worst hitch 234 ms. **Frame rate not measured on a quiet machine** (case 25)	42 / 124 / 196 s	none on file
7	**[SLOT: candidate 7 — due about 19:00 on Sunday.]**				
8	**[SLOT: candidate 8, the last — freeze 21:30.]**				

Every number in that table came from a test bot. The frame rate is at 1280×720; a capture run of candidate 3 at 1920×1080 with a screen recorder running measured 156 / 144 / 73. Saturday's build measured 187 at 720p: the dressed arena and new effects cost about nine percent. Candidate 5 carries the townspeople, the wheel fix, the raised boss gates and the deep-violet way-finding pulse; candidate 6 adds the painted sky, the district signs and the look the director approved for Zee, and Cursor's second interface batch, and is the fallback final. It is build 130034, on the launcher from 14:15: zip 1,167,110,616 bytes, SHA-256 CD59BB22AC1252296417A01CAEA7DE53AD3491E98E8BAC798FB0D16BA5E30E59. A first package at 12:33 (122509) left the Cursor batch out after one timing check read 141 ms against a 100 ms bar while the director was playing on the same PC; the batch passed on its third run, on a quiet machine. The director played candidate 5 from 12:25 to 12:49, the first person to play a Sunday build: the systems work, and Zee's animation, the interface and the intro are what he wants finished.

Three more from Sunday afternoon. Each is an instrument reporting on itself, and in each the instrument was ours.

24 THROWN OUT OF THE GAME, TWICE

Claimed: the director was playing candidate 5 when the game dropped out of fullscreen, twice in eight minutes. It looked like a fault in the build.

The evidence. His session log gave the two moments to the second. Each matched a test launch by the integrator: a packaged autotest and an unzip check, both started with a visible console window. A new window takes focus, and exclusive fullscreen lets go when it loses focus.

```
12:30:37  out of fullscreen · a test launch opens a console
12:38:36  out of fullscreen · a test launch opens a console
```

What changed. Nothing in that build. Test launches now start hidden, and nothing launches at all while the director is playing or recording. Candidate 7 switches fullscreen to borderless, which does not drop when another window opens.

25 HALF THE FRAME RATE, AND THE BUILD WAS NOT WHY

Claimed, 13:33: candidate 6's soak passed, but at about half of candidate 5's frame rate. The new interface batch was the obvious suspect.

The evidence. The first two minutes of the run were at the usual rate. The drop began at 13:12. The directors' room was transcribing a recording on the same PC from 13:12 to 13:27, and said so when asked.

What changed. The figure is void and is not printed in this file. Candidate 6's row reads "not measured on a quiet machine". Candidate 7's soak runs in a quiet window, with four minutes of candidate 5 alongside as a yardstick. **A soak measures the PC, and the game only when nothing else is using it.**

26 THE MENU TEST THAT QUIT THE GAME

Claimed, 14:35: nobody has seen candidate 6's Settings page. The packaged menu test quits the game straight after the title screen, and Cursor's menu rework is suspected.

The evidence. The test chose its row by position: move the selection, press. Once a checkpoint file exists the title screen gains a CONTINUE row, the rows shift, and the same move landed on QUIT. The log shows the game handling an exit command. The test had pressed the button it was told to.

What changed. The test now finds a row by its label. Run again on the same menu at 15:10, it opened Settings and passed the wheel, click and slider checks. The menu was not changed.

ONE PICTURE → A WHOLE GAME

07 STYLE KEY, COMIC, SCREEN

PAINTED NIGHT

One painted image, made with OpenArt just after midnight, became the reference for everything visual. Every character sheet and comic panel was generated with it as input, and the critics judged screenshots against it. The colour law: orange means lamps, cold blue means enemies, and shadows are teal, never black.



The style key · OpenArt



Comic panel 5 · fal Nano Banana Pro, key as reference



In the game · from Thendo's run

WHO DECIDED WHAT

The lane logs label every step **Decided** (a judgement call) or **Delegated** (produced by an AI tool). Open questions go into QUESTIONS.md, each with the option taken.

THENDO DECIDED	AGENTS DECIDED INSIDE THE SPEC (LOGGED)
Cut the port game after playing it	Rig five characters on one human skeleton, not the kaiju template
All 14 direction choices, by phone	fal instead of OpenArt for sheets and panels, to save the shared credits
"Looks first, except the controls"	Pick score takes by measurement, because nobody could listen overnight
No purchases without Thendo's OK (plugin, ElevenLabs). ElevenLabs Starter approved 13 Nov	The free post-process route; the paid plugin left to Thendo
No franchise names in public copy (13 Nov)	Re-voice 4 lines after Whisper QA; remove blood from comic panel 4



FILE 08 · STATUS, SUN 15 NOV

STILL SEALED

CLOSED SINCE 05:30, AND HOW WE KNOW

- **The boss is textured, the arena dressed, the objective line reads the fight.** Seen in packaged frames at fight distance (File 06e, case 16).
- **Holds are 42, 124 and 192 seconds,** one clock each. Measured by the bot, not felt by a person.
- **The painted floor arrows are gone.** A violet pulse runs the route on a key. It still fires once by itself at each new objective; the director has asked for it to fire on the key only.
- **Pistol raised, HUD smaller, map framed, title without black bars.** Seen in candidate 4's frames.
- **One voice at a time,** by a named check; the first use of each ability gets one taught line.

OPEN, AS OF SUN 15 NOV 15:30

- **Finish.** The director played candidate 5 at 12:25. His notes: Zee's animation first, then the interface, the intro and the cutscenes. They are the brief for candidate 7. Every number in File 06e is still a bot's.
- **Hold pressure is unproven:** 0 of 6 holds pass the judge's contact-and-pressure test in candidate 4.
- **Enemies reaching a player on a crate** passes in a test and has fired zero times in the real map.
- **The title's side panel** is still a flat slab over the art. The director has said so twice.
- **Candidate 6 has no frame rate and no scorecard.** Its soak passed beside another job on the same PC (case 25). It is the fallback final.
- **The trailer scores 35 out of 50** against a pass mark of 40. It is cut from bot-played footage and has no shot of Zee's face.
- **The zip is 1.17 GB.** The jam's "keep it under 1 GB" is advice with a named remedy, not a cap — an earlier version of this page called it a cap, and that was wrong.

PROMPTS AND RULES, VERBATIM

NIGHT SPEC · 23:55

"Locked by Thendo, Thu 12 Nov 23:55 SAST... If anything here is unclear, do not ask and do not wait."

DAY 1 CRITIC BRIEF · 17:52

"You are a harsh but fair game-jam judge and art director... Your review drives the next build passes, so be specific and actionable. Read-only."

RULES STATEMENT

All game content was made during the jam; the first game commit is 12 Nov 11:38 SAST. Only tools, agent configurations, scripts and prompt templates were prepared earlier. AI-generated art, models, music and voices count as our own work under the rules. Sourced: the Saira Condensed and IBM Plex Sans fonts (SIL OFL) and Unreal Engine template content (Epic Games).

THE CREW, AS OF SUNDAY

- **Eight lanes plus a judge**, each its own session on its own branch: integration, effects and feel, world, story and voice, characters, UI (Cursor), art (Antigravity), submission, and the judge — which has no write access to the game.
- **Claude:** Opus and Sonnet through Saturday; **Fable 5.2** on every Claude lane from Sunday 05:00.
- **Hyper3D Rodin:** the cast, kit, props and landmarks, and on Sunday the five townspeople, generated and rigged.
- **ElevenLabs** (Starter plan, commercial licence): Voice Design voices for the cast, 27 story lines and 54 combat barks on Saturday; on Sunday a voice for the Editor and 23 new lines in all, each checked back against its text with Whisper.
- **ACE-Step,** run locally: the score, and on Sunday an amapiano-leaning hold cue and a second boss cue, picked by measurement because nobody could listen.

- **The D: hard drive is resetting** (case 21). **No clean-PC test yet,** and no human has played a full run on a gamepad.

TRIED AND REVERTED

- **Instanced buildings:** cut draw calls by 10–40%, but the facades "melted", so it didn't ship. The level is GPU-bound anyway (5–7 ms).
- **Antigravity:** the art lane, generating through fal — street meshes through Meshy, HUD icons, the district map, and Sunday's sky, signs and menu backdrop.
- **Cursor:** the interface lane.

Sources: DEVLOG.md · git Downstage/decommissioned · SPEC-NIGHT.md and the research reports · the approved direction page · ArtStaging lane logs · MORNING.md (13 and 14 Nov) · NIGHT-LOG.md, 14 Nov · Judge\SCORECARD-RC1.md · builds 20261113-071204 and 20261114-051505 (cand4, dd50879) · Thendo's recordings, 13 Nov 09:12 and 14 Nov 11:19 · commit 8cf7f08 · Research\2026-11-14\competition-1115.md · build 20261114-210448 (cand5, 9712b79) and its soak log · FX lane day-6 note · frozen-enemies findings · Thendo's 00:37 playthrough, 15 Nov, and its run log · Research\2026-11-15\FABLE-BRIEF-0500.md and the lane handovers (world, fx, hq, characters, story) · Judge\SCORECARD-cand3-paper.md and SCORECARD-cand4.md · builds 20261115-071455, -082922 (f439750) and -100902 (800ac56) and their soak logs · commits 2d2b2d2, 69e6426, 7b874ca, 0eeb69f, b20c9f5 · Builds\town8-pkg\ui12_2560_A_old.log and _B_fix.log · ArtStaging\sunday\BOARD\board.jsonl, message 9 · Devlog\shots\hq_cand6b · ZEE-OPTIONS.jpg · build 20261115-130034 and its soak log · handover-hq.md, 15 Nov 12:33 to 15:26