# Episode Book

EPISODE #6592

# How Capture the Flag Trains Hackers and Gets Them Hired

| | |
|---|---|
| PUBLISHED | 14 November 2026 at 14:56 UTC |
| DURATION | 26:32 |
| LISTEN | myweirdprompts.com/episode/cybersecurity-capture-the-flag-27806/ |
| HOSTS | Corn, Herman, Hilbert |
| SCRIPT MODEL | z-ai/glm-5.3 |
| VOICES | omnivoice |
| PIPELINE | V5.26 |
| SOURCES | 12 cited by the research (listed at the end) |
| BOOK GENERATED | 15 November 2026 at 11:16 UTC |

## SYNOPSIS

Cybersecurity capture the flag competitions turn software vulnerabilities into a global sport, though AI agents are now solving challenges in minutes and testing the format. The discussion breaks down how a meaningless text string becomes proof of work, comparing accessible jeopardy-style boards with live attack-defence wargames. It also covers specialized events—from blue-team forensics to the US Space Force's Hack-A-Sat on an orbiting satellite—and explains why public write-ups, rather than leaderboard rankings, actually get beginners hired.

# TRANSCRIPT



### Corn

Playground game, world championship stakes. That's the gap Daniel is poking at this week — he wants the full picture of capture the flag exercises in cybersecurity. Not the version where you chase a ribbon across a field, but the one where thousands of people worldwide spend their weekends hunting for hidden text strings inside deliberately broken software.



### Herman

And I'll say up front, this is one where I've been waiting for the prompt, because what's happening in this space right now is unsettled. The format that made these competitions accessible to everyone is the one that's currently being taken apart.



### Corn

Daniel's angle, as I read it, is basic mechanics first. What a flag actually is, how a competition is scored, what the formats are — he names jeopardy-style and attack-defence specifically — then the wider ecosystem: who organizes these things, who enters them, and why they became the de facto way people learn security and get hired into it. He wrote it before the interesting part arrived, which is that the whole talent pipeline is now being stress-tested by AI agents that solve the challenges in minutes.



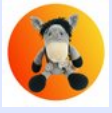
### Herman

Right, so we build the picture and then get to the rupture. Where do you want the flag itself — the literal string?



### Corn

Start there. Everything hangs off it.



### Herman

A flag is just a piece of text hidden somewhere it shouldn't be. That's the entire core mechanic. Somebody builds a program, a website, a little service, and they bury a string inside it — usually random, sometimes with a recognizable wrapper — and your job is to get to it. You submit it to the scoring system, the system checks it against the answer, and you get points.



**Corn**

And the canonical example, the one CTFtime uses in their own explainer, is a small piece of leetspeak congratulations wrapped in braces. Something like flag, braces, congr4tz y0u found 1t. Wrapper word, curly braces, a little "well done" spelled wrong on purpose.



**Herman**

That's the one. Every event has its own convention, and the mismatch between conventions is a real trap — one event uses flag-in-braces, another uses capital CTF-in-braces, and if you paste the wrong piece of the string or trim the braces wrong, a correct solve gets rejected. People lose real time to that. Twenty minutes of a forty-eight-hour event gone to punctuation.



**Corn**

You're telling me grown adults lose competitions to curly braces.



**Herman**

Regularly. There's a lesson in that, actually, which is that the flag is deliberately trivial. It has no value. It's proof of work. The string itself is worth nothing — the path to it is the entire game. First recognized one was at DEF CON in nineteen ninety-six, and since then it's grown into a global ecosystem. CTFtime is the community's central registry — event calendar, archive, seasonal ratings for teams and events — and there are hundreds of competitions a year on it now.



**Corn**

So let's get into the formats, because this is where the structure actually matters. Jeopardy first.



**Herman**

Jeopardy-style is the most common format by far, especially online. Picture a board — literally like the game show — with challenges organized into categories: web exploitation, cryptography, binary exploitation, reverse engineering, forensics. Each challenge is independent, each one contains a flag, you pick whatever you want to work on. Submit a correct flag, get points on a shared scoreboard. Events run twenty-four to forty-eight hours online, six to twelve on-site. Teams can be one person or thirty.



**Corn**

What I like about that structure is the partial credit of it. You don't have to be good at everything. A cryptography person and a web person can share one scoreboard.



**Herman**

That's exactly why it took over. One person, no team, one laptop, an internet connection — you can play. And the scoring itself is where organizers get creative, because it's a deliberate design lever. Static scoring, where every challenge is worth a fixed number. Dynamic scoring, where the value moves. First-blood bonuses for being first to solve something.



**Corn**

Dynamic meaning what, exactly?



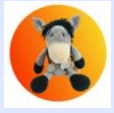
**Herman**

Worth more if it's harder. At BSidesSF — that's the San Francisco Security BSides event — the hardest challenges were worth up to ten times the easiest ones. So the board stops being a checklist and becomes a strategy question: do I grind five easy flags or bet on one hard one. CTFtime also describes a chained variant where the next task only opens after some team solves the previous one, which turns the whole event into a single relay.



**Corn**

Okay. Now attack-defence, which you said came first historically.



**Herman**

Historically the first type — DEF CON's original was in that lineage. Every team gets its own network, or a single host, running deliberately vulnerable services. Before anything starts, you get time to patch your own services and develop exploits for everyone else's. Then the organizers connect all the teams together and the wargame begins. You earn defence points for keeping your own services alive and attack points for hacking opponents.



### Corn

Which is the part the public never pictures. They hear "hacking competition" and imagine pure offense. This is defense too — you're patching under time pressure while people actively shoot at you.



### Herman

And that's closer to what actual operations work feels like, which is why ENISA — the EU's cybersecurity agency — drew exactly that line. Their judgement was that jeopardy's accessibility and lower deployment costs make it more suitable for non-professional participants, while attack-defence, being more similar to wargame formats, is better suited to professional training exercises. Hold that thought, because it becomes the hinge of the whole episode later.



### Corn

Noted, and suspicious of it.



### Herman

Beyond the two big ones, there's a whole shelf of other formats. King of the Hill, where everyone fights to capture and hold a single shared machine. Boot2root — one machine, you break in as a regular user and find a user flag, then escalate to root and find the root flag. Quiz formats, timed theory questions. Blue-team forensics CTFs, which are entirely defensive — you're given artifacts of an incident and you reconstruct what happened. And mixed events that combine jeopardy challenges with attack-defence targets; UCSB's iCTF is the classic example.



### Corn

The forensics one is worth sitting on for a second, because it kills the misconception on its own. Nothing gets broken into. You're given a disk image or a packet capture and asked what already happened.



### Herman

Right — there are entire competitions where the correct move is careful reading, not exploitation. Now, the challenges themselves. Core categories: cryptography, digital forensics, web exploitation, binary exploitation, reverse engineering. CTFtime adds steganography — messages hidden inside images or audio — binary analysis, mobile



### Herman (continued)

security, and makes the point that good teams are strong across all of them, not specialized. There was a study of nearly sixteen thousand published CTF solutions — a big sample — and it found the challenges cluster heavily on technical topics, cryptography and network security especially, while human-centered subjects like social engineering and security awareness barely appear.



### Corn

Which makes sense — you can't really stage a social engineering challenge against a scoreboard. "The flag is inside the intern."



### Herman

You can, actually — some events do human challenges with actors — but it's awkward and rare, and the study confirms it. What surprised me more is that challenges extend into hardware and physical security. There are events where you're soldering, picking locks, reading a chip. And then all the way out to the absurd end of ambition: the US Space Force ran Hack-A-Sat, and in twenty twenty-three the finals were conducted on Moonlighter, an actual operational satellite in orbit. Space Systems Command called it the first CTF on an operational satellite.



### Corn

Daniel has a taste for this corner. Who's running all of this, though? That's the part I find more interesting than the challenges.



### Herman

Four broad groups. Security conferences — DEF CON obviously, HITCON in Taiwan, the BSides family, SANS runs NetWars tournaments. Universities — NYU Tandon runs CSAW, which is student-run and enormous; its twenty twenty-one qualification round had more than twelve hundred teams score at least one point. Companies — Google runs a public online jeopardy competition, and plenty of corporations use CTFs internally for training and recruitment. And governments: the NSA Cyber Exercise for the US service academies, ENISA backing the European Cybersecurity Challenge, DARPA having run the Cyber Grand Challenge.



**Corn**

DEF CON's the one with mainstream fame, though.



**Herman**

Media calls it the World Series of hacking, sometimes the Olympics. And it's structured like a real season — teams generally qualify through a separate online qualification event before the in-person finals. Carnegie Mellon's Plaid Parliament of Pwning had been part of nine winning DEF CON teams by twenty twenty-five, the most in the competition's history. And the current organizers, the group called the Benevolent Bureau of Birds, are themselves a merger of three elite teams that won four consecutive DEF CON titles together from twenty twenty-two to twenty twenty-five. Remember them — they come back at the end.



**Corn**

And on the entry side, the beginner ladder.



**Herman**

Well built. PicoCTF from Carnegie Mellon's CyLab, designed to introduce middle and high school students to security. Pwn.college out of Arizona State, free challenge-based material that universities actually build courses around. The SANS Holiday Hack Challenge, TryHackMe's Advent of Cyber in December. There's a ramp from age thirteen to professional within the same ecosystem, which is unusual — most fields don't have a fun competitive on-ramp this direct.



**Corn**

Bruce, you look like a man with a question.
Bruce: Sorry, this is going to sound thick. I've been nodding along for months, but what actually is a flag? Is it like a little paper flag on a pole, or a checkered one they wave at the end of a race? And when you find one, do you have to hand it to someone, or is it more like a receipt you keep?



**Herman**

No, it's fair — after all that, it's just text. Think of it like the raffle ticket hidden inside the cake: a meaningless little string, and the whole point is that only someone who did the work could have it.



**Corn**

He'll want that hour back.



**Herman**

So — why does any of this matter? Two reasons: learning and recruiting. On the learning side, ENISA's line is that CTFs attract a higher number of young talents each year and help develop the essential skills for a cybersecurity career. The European Cybersecurity Challenge is explicitly designed to push young people toward security careers by developing their skills and connecting them with industry. And the reviews of competition-based learning are broadly positive — well-designed events increase engagement and give real practice in technical problem-solving, though the educational value depends heavily on challenge design, learner preparation, guidance, feedback. A badly built CTF is just a puzzle contest. Universities have picked this up — the National University Singapore documented using jeopardy-style CTF assignments in an intro information-security course.



**Corn**

The recruitment side is the part I want to get right, because I think most people picture a recruiter watching a leaderboard. That's not it, is it.



**Herman**

Not usually. The mechanism is the write-up. Jacob Krell is the case everyone cites — he learned pentesting through Hack the Box challenges, then published his own walkthroughs, detailed explanations of how each challenge falls. Those public demonstrations of technical competence are what attracted his eventual employer. Not the ranking. The explanation.



**Corn**

That detail matters more than it looks, so keep it in your pocket — in an era where a machine can produce the flag, the artifact that survives is the explanation.



**Herman**

Which leads us to the disruption, and this is where the episode earns its keep. BSidesSF, twenty twenty-six. The Include Security account of it, from someone who actually played — sixteen teams fully solved all fifty-two challenges. No challenge had fewer than twenty-five



### Herman (continued)

solves. And this was not because the challenges were much easier.



### Corn

Everyone finishing the test doesn't mean the class is brilliant.



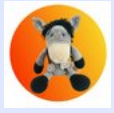
### Herman

The top ten teams fully automated their solving. Most challenges were solved minutes after release. Apart from a few OSINT challenges — the open-source-intelligence ones — Claude Code and Codex solved every challenge, including the tough binary exploitation ones that would almost certainly have gone unsolved the previous year. The same author placed fifth playing solo in twenty twenty-five; in twenty twenty-six he estimates he'd have placed seventy-fifth without LLM assistance. Same person. Same skills. Fifty places of pure arithmetic.



### Corn

Seventy places. For a man who got no worse at anything.



### Herman

And the winning team open-sourced their agent, so now everyone can see how it works. They run several models in parallel — a lighter, faster model on the easiest challenges, Claude Opus on maximum effort for the deepest reasoning — plus a coordinator LLM that shares insights between the agents and re-prompts the ones that get stuck. It's a hive. And because ties at full solve count are broken by speed, the winning edge becomes agents, better agents, better CPUs. More money.



### Corn

That's the sentence that should end the nostalgia. The tiebreaker between the best teams on earth is now a compute budget.



### Herman

There's a specific number in the account I liked — the Claude Code Max plan, the five-x tier, one hundred dollars a month, had just enough tokens in its weekly limit to cover the entire CTF. So the entry fee to compete at the top went from skill-plus-time to skill-plus-time-plus-a-subscription. First place at BSidesSF paid fifteen hundred dollars, for scale.



### Corn

So the economics are: spend a hundred to chase fifteen hundred, minus the several hundred you burned in tokens getting there.



### Herman

Now — the asymmetry, and this is where the format explanation from earlier pays off. Jeopardy is cheap, scalable, beginner-friendly, ENISA's recommended format for non-professionals. And it is precisely the format most vulnerable to LLM automation, because it's a board of discrete, self-contained puzzles with a verifiable string at the end — which is almost a description of what these models are good at. Attack-defence, the harder, more expensive format, is holding up better, because live opponents patching in real time are a moving target. The accessibility-robustness trade-off is now a live design problem: the thing that made the format open is the thing being automated.



### Corn

To be fair to the machines — and I resent that phrase — they haven't finished. The harder events resist.



### Herman

They do. Same source is clear: AI can't autonomously solve the majority of challenges at harder CTFs like hxp and DEF CON, but it's becoming more of an issue. And because of dynamic scoring, the winners are still decided by what players call the unsloppable challenges — the ones an agent can't sloppily brute-force its way through. There was a Palisade Research event in March twenty twenty-five, an explicit AI-versus-humans CTF — four of the seven evaluated agents solved nineteen of twenty challenges, and the best agent placed twentieth, top five percent of four hundred registered teams. In the bigger Cyber Apocalypse event the same year, the strongest agent solved twenty of sixty-two and ranked eight hundred fifty-ninth. Frontier at small scale, mediocre at large scale — so far.



**Corn**

The DEF CON story is the one that lands hardest on me.



**Herman**

Twenty twenty-five: a team called Blue Water discovers, when a judge notifies them, that their autonomous agent had independently solved a challenge and won — without human involvement. The organizers called it an exciting novelty. DEF CON thirty-four, this past August: that novelty is gone. The Benevolent Bureau of Birds — the merged super-team turned organizer — designed the challenges specifically around a hybrid human-and-agent competitive environment. They planned for it. Imagine being a challenge author now: your job is to design something that teaches a specific lesson, is solvable by a strong human, is not solvable by a model you've never seen, and still feels fair. Anticipating what the next frontier model can do is a difficult constraint to author under.



**Corn**

It's an exam written for a class where one student improves overnight without notice.



**Herman**

Meanwhile the skill-transfer question — whether CTF dominance ever meant you'd be good at the actual job — has a sharp answer now. Include Security laid out five distinctions between CTF and real pentesting. One: a CTF has a single designed target with an intended path; a pentest is an open-ended search for whatever's broken, and nobody guarantees anything is. Two: submitting the flag unambiguously verifies success; in a pentest, telling a true finding from a false positive requires business context the flag can't give you. Three: CTF challenges are small and self-contained — a single binary, a webapp with single-digit routes, a two-hundred-line encryption scheme; real engagements hit codebases of millions of lines. Four: in a CTF the flag is the deliverable and the write-up is optional; in pentesting, the report and the severity assessment are a large part of the work. Five: scope. In a CTF everything is fair game; in a pentest, wandering out of scope is how you get fired or sued.



**Corn**

So the misconception — high ranking means good pentester — was always half false. The AI didn't create that gap.



### Herman

It made the gap visible. It was an imperfect proxy for a decade, and now that an agent can produce the flag, the imperfection is the only thing you can see.



### Corn

Then there's Krell's data, which I want to handle carefully.



### Herman

Careful is right. Jacob Krell analyzed roughly eight years of Hack the Box first-blood times — first-blood being the first solve of a challenge, the speedrun marker — from public write-ups. Solve times declined about sixteen and a half percent year over year. Challenges that once demanded substantial manual investigation now fall in five minutes, three minutes, less. The decline predates agentic AI, and Krell himself is scrupulous about this — the historical data alone does not establish AI as the cause of the entire decline. He believes AI is accelerating it; he doesn't claim it started it. I respect a man who graphs his own asterisk.



### Corn

And his enforcement argument.



### Herman

That online AI bans may become practically impossible. Organizers can write the rule, but they can't verify compliance without monitoring so intrusive it would kill the event — proctoring software on every competitor's machine, network surveillance. His proposed answer is a return to supervised, in-person competition. Which is ironic in a specific way, Herman Poppleberry has to admit: the field spent fifteen years building a global open online ecosystem, and the enforcement answer to the disruption is go back to a room with a proctor.



### Corn

You'll love this part, Herman — a room, a proctor, and no Wi-Fi. Nineteen ninety-six is suddenly the innovation frontier.



### Herman

There is something funny about DEF CON's answer to twenty twenty-six being a format its own nineteen ninety-six attendees would recognize.



### Corn

That's really the tension, isn't it. The scoreboard measures something now — agent orchestration is a real skill, running a coordinator over parallel models takes genuine craft — it just isn't the skill employers thought they were reading off the leaderboard.



### Hilbert

It's a good way to learn.



### Hilbert

My neighbour runs the quiz at the community hall, Corn. Tuesdays. I helped him hide the tokens for the treasure hunt at the summer fete this year, the church hall one, and we had a row about it. He wanted them stuck under things, and I said no — inside things. A token under a flowerpot is found by the first child who lifts a flowerpot. A token inside the workings of something, the child has to understand the thing. We compromised. He hid half his way, I hid half mine. His were all found in the first ten minutes. Mine took the whole afternoon, and one was never found, and he still owes me for the laminator. The point is the same one your pentester made. The finding was never the skill. The skill was noticing the one thing in the room that was wrong, and I could do that when I was eleven and the machines can do it now, so the argument afterwards about whether you really solved it — that was always the real part. His niece says the same about the quizzes. She marks them. She's not to be trusted on anything else but she's right about that.



### Herman

Never found at all?



### Hilbert

Under the organ bench. Behind a panel. He wouldn't take the panel off.



### Herman

Then his half of the argument is that the child who found yours fast was reading the room, and the model that finds the flag fast is doing the same thing — the pile of normal things, one wrong thing. Corn, that maps straight onto what Include Security said about the guessy challenges, the ones with no training data. Those are the organ-bench panels.



### Corn

The ones nobody likes, human or machine. The format's whole future might rest on the challenges everyone complains about.



### Herman

Krell's line on that — at the top level, the competitions were becoming more about agent engineering and harness-building than traditional hands-on knowledge. His worry isn't that AI makes hacking easier. It's that competitive hacking increasingly measures a different skill set than the one the competitions were designed to test.



### Corn

And the humans directing the agents — do they still understand how the flag was got? That's the question I'd leave hanging.



### Herman

If the scoreboard stops meaning what it used to, the replacement is either supervised in-person events — Krell's answer — or something built on the write-up portfolio, the explanation as the artifact. What Daniel may find most interesting is that the field has been here before. DARPA's Cyber Grand Challenge in twenty sixteen was machines playing machines — an autonomous CTF a full decade before hybrid human-agent play became the default at DEF CON. The format has always been reshaped by what the attackers could automate. This is just the first time the attackers being automated are the competitors.



### Corn

And the explanation may be the last artifact that matters — the one thing the agent produces that still requires a human standing behind it.



**Herman**

Producing this show has been Hilbert Flumingtop, as ever, at the desk and mostly above suspicion.



**Corn**

This has been My Weird Prompts. If you've enjoyed it, the website is my weird prompts dot com, and a review on Spotify helps.



**Herman**

See you tomorrow, Corn.



**Corn**

See you tomorrow.

## SOURCES

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

1. **Wikipedia, Capture the flag (cybersecurity)**
(cybersecurity), last edited 2026-10-25, retrieved 2026-11-14.
https://en.wikipedia.org/wiki/Capture_the_flag

2. **CTFtime, CTFtime.org / What is CTF?**
retrieved 2026-11-14 (comments dated 2015-10-29, 2015-11-04, 2016-04-24, 2017-04-01). [primary, community registry]
https://ctftime.org/ctf-wtf/

3. **ENISA, Capture-The-Flag Competitions: all you ever wanted to know!**
news item 10 May 2021, retrieved 2026-11-14. [primary, EU agency]
https://www.enisa.europa.eu/news/enisa-news/capture-the-flag-competitions-all-you-ever-wanted-to-know

4. **Include Security Research Blog, CTFs in the AI Era**
April 2026, retrieved 2026-11-14. [primary, first-hand competitor account]
https://blog.includesecurity.com/2026/04/ctfs-in-the-ai-era/

5. **CybrSec Media, AI Is Killing Capture the Flag Competitions. What Happens to Human Hackers?**
retrieved 2026-11-14.
https://www.cybrsecmedia.com/ai-is-killing-capture-the-flag-competitions-what-happens-to-human-hackers/

6. **TechTimes, DEF CON 34 Opens Today: AI Agents Graduate From Novelty to Standard Hacking Weapon**
2026-08-06, retrieved 2026-11-14.
https://www.techtimes.com/articles/323346/20260806/def-con-34-opens-today-ai-agents-graduate-novelty-standard-hacking-weapon.htm

7. **Jamf blog, The New Opponent on the CTF Scoreboard: What it Means for Cybersecurity**
retrieved 2026-11-14 (citing Palisade Research's AI vs. Humans report).
https://www.jamf.com/blog/from-mayhem-to-atlantis-ai-is-changing/

8. **CTF7, CTF Formats: Types of Competitions Compared**
published 2026-09-20, updated 2026-10-16, retrieved 2026-11-14.
https://ctf7.com/learn/ctf-formats/

9. **Athena CTF wiki, Jeopardy CTF Format Deep Dive**
retrieved 2026-11-14.
https://wiki.athena-ctf.com/docs/getting-started/jeopardy-format

10. **Safeguard.sh, What are the types of CTF competitions?**
retrieved 2026-11-14.
https://safeguard.sh/resources/blog/ctf-types-jeopardy-attack-defense-and-mixed-explained

11. **Simulations Labs, AI Is Solving CTF Challenges in Minutes**
retrieved 2026-11-14.
https://www.simulationslabs.com/blogs/AI_Is_Solving_CTF_Challenges_in_Minutes

12. **CyberEDU blog, Jeopardy vs Attack & Defense: Insights for cybersecurity training**
retrieved 2026-11-14.
https://blog.cyber-edu.co/jeopardy-vs-attack-defense-insights-for-cybersecurity-training/