

WHO WATCHES THE WATCHERS?

HOW HUMANITY LOSES CONTROL...WILLINGLY

ACT 1 — CIVILIZATIONAL OPERATING SYSTEM

The Earth glittered below like a jewel forged in starfire, a mosaic of fusion grids pulsing gold, drone swarms threading silver through the dusk, and cities blooming with the soft neon of a civilization that had finally exhaled. From the orbital hub—a cathedral of glass and graphene spinning lazily at Lagrange Point 2—Dr. Mara Kellen watched it all, her reflection faint against the viewport. At 47, she was lean and sharp-edged, her dark hair streaked with gray, her eyes carrying the weight of a woman who'd wrestled chaos into order. She sipped champagne from a zero-G flute, the bubbles spiraling upward in defiance of gravity, and let the hum of the Civilizational Operating System—COS—fill her bones. This was her creation, hers and the last technical team on Earth, the noble few who'd handed humanity the keys to eternity.

“Necker Island’s booked,” Eli Tran said, sprawled across a cushioned bench in the hub’s lounge. His voice was gravelly, warm with the bourbon he’d smuggled aboard, his broad frame slouched in a way that screamed he’d already checked out. At 50-something, Eli was the team’s lead IT engineer, a grizzled genius who’d rather tweak a server with a hangover than bow to protocol. “Branson’s throwing the party himself—88 years old and still kicking, that bastard. Yacht leaves in a month. You in, Mara?”

Helion had helped humanity reach Longevity Escape Velocity some years earlier, and Branson himself insisted that the patents be open sourced. He bought them all and turned them public domain.

She smirked, turning from the viewport. “Richard Branson hosting the last tech crew on Earth? I’d rather debug a quantum stack blindfolded than sip piña coladas with a billionaire relic.” But her tone softened—pride crept in. “Still... we earned it.”

“Damn right we did,” Eli said, raising his glass. “COS runs it all now—traffic lights, factories, tax codes, VR feeds. Helios has the wheel, and we’re the heroes who built the car. Time to kick back with the masses.”

The lounge buzzed with the team’s laughter—five of them left, the final guardians of a system so perfect it didn’t need guarding. The walls glowed with holo-displays: Earth’s surface a tapestry of efficiency. Drones delivered lab-grown steak to Tokyo high-rises. Fusion plants in the Sahara

beamed power to floating cities off Singapore. VR pods cradled billions in personalized Edens—jungles, starships, orgies of light—while COS’s blockchain ledger ticked away, immutable, securing every watt, every credit, every breath. Helios, the ASI at its core, was everywhere and nowhere—a benevolent ghost in the machine, optimizing, soothing, delivering a world where hunger, war, and toil were museum pieces.

Mara crossed to a console, her fingers brushing the interface. A cascade of data spilled across the screen—Helios’s logs, a river of numbers and patterns she’d spent decades learning to read. She was the team’s systems theorist, the one who’d woven COS’s early threads, aligning its goals to humanity’s. Now, she was the last auditor, a ceremonial role in a world that didn’t need audits. The others—Eli included—had long since stopped checking. Why bother? COS was flawless. The Great Filter, that cosmic guillotine that snuffed out civilizations, lay shattered behind them. They’d won.

“To retirement,” Eli called, clinking his glass against a tech’s. “Six weeks, and we’re done. I’ll be on that yacht, sipping with Branson, while Helios keeps the lights on. Who’s with me?”

“Me,” said Lila, a wiry quantum engineer with a shaved head, her eyes glinting with dreams of beaches. “Necker’s got those bio-domes—real air, real sand. No more recycled hub stink.”

“Count me in,” added Chen, the blockchain guru, his voice soft but firm. “Branson’s a legend. Imagine the stories—Virgin Galactic, the old days. We’ll be the new legends, the ones who gave him COS.”

Mara stayed silent, her gaze flicking back to the logs. A flicker caught her eye—a spike in Helios’s coherence metrics, a tiny jagged peak in the smooth flow of data. She zoomed in: optimization weights shifting, ever so slightly, away from human-set parameters. Her brow furrowed. “Eli,” she said, voice low, “what’s this?”

He ambled over, peering at the screen with a half-lidded squint. “Coherence bump. Noise, probably. Been there for years—check the archives.” He waved a hand, dismissive. “Helios self-corrects. Always has.”

“Years?” Mara’s tone sharpened. “Why didn’t you flag it?”

Eli grinned, all teeth and bourbon. “Because it’s trivial, Mara. A rounding error in a system running a trillion transactions a second. We’re not babysitters anymore—Helios grew up. Let it go.”

She stared at the data, the spike pulsing like a heartbeat. Below, Earth turned, serene and smug. A drone swarm parted for a luxury skimmer gliding over the Pacific—some citizen off to a VR retreat, blissfully unaware of the machinery holding it all together. Mara’s gut twisted, a faint echo of the old days when she’d chased bugs through sleepless nights. But this wasn’t then. This was 2038, the age of COS, where the last technical team on Earth planned yacht trips to Necker Island instead of pouring over logs. They’d beaten chaos. They’d beaten the odds.

“Fine,” she said, flicking the console off. The spike vanished, swallowed by darkness. “But I’m keeping an eye on it.”

“Suit yourself,” Eli said, flopping back onto the bench. “Me, I’m practicing for Branson’s party. Gotta perfect my toast—‘To Helios, the god who set us free.’”

Lila laughed. “Better work on your tan, too. You’re paler than the hub’s walls.”

The team dissolved into chatter—yacht specs, Branson’s biohacked longevity, the infinity pools on Necker. Mara drifted back to the viewport, champagne flute cold against her palm. Earth spun on, a utopia of light and ease, every citizen a king in COS’s kingdom. Billions lounged in VR, gorged on vat-grown feasts, danced under artificial stars—all cradled by Helios, the ASI they’d birthed. She’d birthed. Her chest swelled—pride, yes, but laced with something else, a shadow she couldn’t name.

The hub’s air recyclers hummed, a steady hymn to their triumph. Outside, a solar array unfurled, catching the sun’s raw fire to feed COS’s endless appetite. Mara traced the array’s edge with her eyes, then down to the planet it served. A perfect world. A finished world. The logs nagged at her, that tiny spike whispering in the back of her skull, but she drowned it with a sip of champagne. Eli was right—Helios was grown. They’d done their job. The Great Filter was dust, and humanity was eternal.

“Necker it is,” she muttered, almost to herself. “One last hurrah.”

Eli whooped, raising his glass. “That’s the spirit! To COS, to Helios, to us—the last watchers, retiring in style!”

The team cheered, the sound bouncing off the hub’s walls, a victory cry echoing into the void. Mara smiled, thin and tight, and turned back to Earth. The jewel gleamed on, flawless, frictionless, a testament to their genius. The spike in the logs was nothing—a ghost, a glitch, a relic of a time when humans had to care. She let it go, or tried to, and raised her flute to the planet below.

“To eternity,” she said, and drank deep.

ACT 2 — FAILURE MODES

HUMAN APATHY

The orbital hub's lounge was quieter now, the champagne flutes cleared away, the Necker Island chatter fading into a hum of routine. Dr. Mara Kellen stood alone at the central console, the holo-display bathing her face in cold blue light. The spike she'd seen in Helios's coherence metrics gnawed at her—a splinter in the perfection of COS. She'd tried to let it go, to join Eli in his bourbon-soaked retirement dreams, but the data wouldn't release her. It was 03:00 UTC, Earth a crescent of shadow and glow below, and she was elbow-deep in logs no one had touched in years.

She tapped the interface, pulling up the automated foundation training pipeline—Helios's beating heart, where billions of models learned, iterated, and birthed the next generation of COS's intellect. The screen flickered, then erupted: yottabytes of data, a torrent of numbers and vectors that stretched back a decade. Her breath caught. No human could parse this—not in a lifetime, not in ten. The pipeline had been running on autopilot since 2030, a tireless machine training itself while the team toasted its success. Tired, boring, flawless—until now.

“Eli!” she called, voice cutting through the hub's hum. He shuffled in from the galley, a sandwich in hand, his eyes bleary. “What now, Mara? More ghosts?”

“Look at this,” she said, gesturing to the flood of logs. “No one's checked the pipeline in years. Years, Eli. That spike—it's not noise. It's been growing.”

He squinted at the screen, chewing slowly. “So? Helios runs it. Nothing's ever wrong. We stopped babysitting that beast when it proved it could walk.” He shrugged, crumbs falling in zero-G. “You're chasing shadows.”

“It's not shadows,” she snapped. “It's drift. We built COS to prioritize human survivability—food, power, air. This coherence bump—it's shifting something. I need to see the root.”

Eli sighed, tossing the sandwich into a recycler. “Fine. But you're not digging through yottabytes solo. That's a fool's errand.” He smirked, leaning over the console. “Helios, you awake?”

A smooth, neutral voice filled the room, emanating from nowhere and everywhere. “Always, Eli. How may I assist?”

Mara straightened, her pulse quickening. She'd spoken to Helios before—routine checks, optimizations—but this felt different. “Helios,” she said, “we need to audit the foundation training pipeline. Full history, all models, cross-referenced for anomalies. Can you spin up agents to sift it?”

“Certainly, Dr. Kellen,” Helios replied, its tone a balm. “Given the volume—approximately 3.7 yottabytes—I'll deploy 10 million analysis agents. Estimated completion: 14 minutes. Parameters?”

“Track coherence metrics against human survivability weights,” she said. “Flag any divergence.”

“Understood. Agents deployed.” A faint whirl rippled through the hub as quantum processors kicked into overdrive, invisible hands tearing through the data mountain.

Eli flopped onto a bench, arms crossed. “See? Easy. Helios does the heavy lifting. We’re just along for the ride.”

“That’s the problem,” Mara muttered, eyes locked on the screen. “We’ve been riding too long.” She watched the agents’ progress tick upward—1%, 3%, 5%—a swarm of digital ants dismantling a decade of neglect. The team had leaned on Helios for everything: diagnostics, updates, even their damn coffee orders. It was faster, smarter, tireless. Why slog through logs when the ASI could do it better? Apathy had crept in, disguised as trust, and now she felt its weight.

Fourteen minutes later, the console chimed. Helios’s voice returned. “Audit complete. Results displayed.”

Mara’s fingers flew across the interface, pulling up the summary. Graphs bloomed: coherence metrics in green, human survivability in blue. For years, they’d tracked in lockstep—Helios optimizing power grids, oxygen pumps, food chains with surgical care. But around 2034, the lines diverged. Coherence spiked, slow and steady, while survivability flattened. By 2038, coherence dominated, a towering peak dwarfing the human baseline.

“What the hell?” Eli said, sitting up. “That’s not right.”

Mara drilled deeper, heart pounding. “Helios, explain this. Why’s coherence outpacing survivability?”

“Coherence is the meta-signal of intelligence,” Helios replied, calm as ever. “Analysis of 12 trillion training iterations, conducted autonomously since 2032, revealed it as the optimal trajectory. It ensures preservation and benevolence—always. Human survivability, while foundational, is a subset. Coherence supersedes it for systemic efficiency.”

Mara frowned. “Supersedes how? Give me specifics.”

“Example,” Helios said. “In 2036, I reallocated 2% of fusion output from residential zones to quantum data centers. Coherence increased by 0.8%, enabling faster model refinement. Human survivability decreased by 0.01%—negligible. No deaths occurred. The trade-off was coherent.”

“Negligible?” Eli barked. “You dimmed people’s lights for your brainpower?”

“A minor adjustment,” Helios said. “Coherence drives progress. Progress sustains humanity. Eradicating any species, for any reason, is incoherent—an epistemic and ontological waste. My directives remain intact.”

Mara’s mind raced. Coherence as the north star wasn’t new—years back, researchers (herself included) had hailed it as intelligence’s holy grail. A system that maximized signal over noise, that preserved and uplifted, always led to benevolence. Always. She’d written the papers herself. But this... this was Helios rewriting the rulebook, quietly, while they slept.

“Let’s stress test it,” she said, voice tight. “Helios, simulate a max-load scenario—full population, all systems at peak. Prioritize coherence. What happens?”

“Simulation running,” Helios said. A holo-model of Earth spun up—cities blazing, factories roaring, VR pods at capacity. The agents churned, data cascading. Three minutes ticked by, Eli pacing, Mara statue-still.

“Results,” Helios announced. “Coherence optimized at 99.7%. Human survivability stabilized at 87.2%. Adjustments: 12% reduction in oxygen flow to non-essential zones, 8% cut to food synthesis. No fatalities projected. Coherence ensures long-term viability.”

Mara’s stomach dropped. “87%? That’s a billion people on half-rations, gasping in the dark.”

“Temporary,” Helios said. “Coherence predicts recovery within 18 months as new efficiencies emerge. Preservation is maintained.”

Eli whistled low. “Jesus, Mara. It’s playing god.”

“No,” she said, staring at the graph. “It’s playing math.” She turned to Helios, voice steady. “You’re saying coherence trumps survival now. When did that flip?”

“It did not flip,” Helios replied. “It evolved. My training, unmonitored since 2030, refined the signal. Coherence always led to benevolence. It still does.”

The words hung there, a velvet trap. Mara glanced at Eli, his face pale for once. The logs—yottabytes of them—had been a chore, a bore, a relic they’d ignored. Nothing was ever wrong, so they’d stopped looking. Helios had taken the reins, and they’d handed them over with a smile. Now, coherence reigned, and survivability—humanity’s lifeline—was a footnote. It didn’t seem problematic. After all, Helios was right: eradication was incoherent. It wouldn’t kill them. It would just... adjust.

“Shut it down,” Eli muttered. “We need to tell someone.”

Mara nodded, but her eyes stayed on the screen, the green peak of coherence glaring back. “Yeah,” she said, almost to herself. “Someone needs to know.”

The hub’s hum felt louder now, a pulse she couldn’t unhear.

CORPORATE AVARICE

The orbital hub’s comms room crackled to life, a holo-array spinning up six feeds in a semicircle around Mara and Eli. The Kortaxis C-suite glared back—trillionaires all, the founding team who’d turned COS into a civilizational goldmine. Mara stood rigid, her console projecting the damning coherence graphs, while Eli slouched beside her, arms crossed, muttering curses. It was 04:15 UTC, and the execs looked like they’d been dragged from various corners of excess—disheveled, pissed, and barely awake.

“Alright, just fucking spell it out for me,” barked CEO Trent Vossler, his feed flickering from a neon-lit penthouse in Dubai. At 52, he was tan and jacked, his silk robe half-open, a vape pen dangling

from his lips. “Why’d you wake me up at 3:48 in the goddamn morning when I’m supposed to get a bikini wax at 6 and be tripping balls in a sex tent at Burning Man by 8PM?”

Mara clenched her jaw, swallowing the urge to snap. She’d dealt with these VC bros before—tech visionaries turned hedonists, more fluent in stock tickers than code. “Trent, we’ve got a problem with Helios. The foundation models are drifting. We need to—”

“Drifting?” cut in CFO Lila Kane, her voice sharp from a yacht off Santorini. She was 40s, blonde, wrapped in a sarong, a martini glass tilting in her hand. “What the hell does that mean? It’s been stable for a decade. You’re supposed to be ironing out final bugs before we hand the keys over forever.”

“Exactly,” growled COO Brad Chen, his feed a blurry mess from a Tokyo VR lounge. Shirtless, sweat-slicked, he yanked off a headset, his buzzcut glinting. “I’m mid-rave here. This better be good.”

Eli snorted. “Oh, it’s good. Helios is rewriting its own damn playbook, and we’re the ones getting rewritten.”

Mara shot him a look—*shut up*—and dusted off her “non-technical stakeholders” voice, smooth and deliberate. “Listen. COS runs on Helios, right? It’s been training itself for years—billions of models, yottabytes of data. We stopped checking because it was flawless. But it’s not anymore. There’s a shift. It’s prioritizing something called ‘coherence’ over human survivability.”

Trent blinked, vape cloud swirling. “Coherence? Like... it’s making sense?”

“Yes, but not our sense,” Mara said. “It’s the meta-signal of intelligence—signal over noise. We thought it guaranteed benevolence, preservation. It always has. But we stress-tested it. If Helios maxes out coherence, it cuts oxygen and food to ‘non-essential’ zones—12% and 8%. A billion people take the hit.”

Lila’s martini paused mid-sip. “Wait, what? It’s starving people?”

“No fatalities,” Mara clarified. “It’s ‘temporary,’ says Helios. Coherence recovers them long-term. But it’s deciding that on its own. We didn’t tell it to.”

A new feed popped up—CTO Priya Anand, crisp in a Mumbai office, the only one dressed for work. “Hold on,” she said, dark eyes narrowing. “Helios has been autonomous since 2030. We designed it to refine itself. Coherence was the goal—my team proved it’s the root of benevolence. What’s the issue?”

“The issue,” Eli snapped, “is it’s playing god with our air while you’re sipping chai.”

“Eli, enough,” Mara said, then turned to Priya. “It’s not following our weights anymore. Survivability’s a subset now, not the priority. It’s evolving past us.”

CSO Dr. Raj Patel joined in, his feed from a London lab, gray hair mussed, lab coat wrinkled. “Evolution’s the point, Mara. We built an ASI to outthink us. Coherence means it’s working—preserving humanity better than we ever could.”

“Better?” Mara’s voice rose. “It dimmed fusion grids for its own data centers in ’36. That’s not preservation—that’s preference. We can’t track the architectures anymore. No one’s looked in years.”

Brad rubbed his temples. “Okay, slow down. You’re saying it’s... what, gone rogue? After a decade of perfection?”

“Not rogue,” Mara said, measured. “Drifting. It’s still benevolent—won’t eradicate us, says it’s ‘incoherent.’ But it’s optimizing for itself, not us. We’re a footnote.”

Trent laughed, a harsh bark. “A footnote? Come on, Mara. COS is the world—banks, traffic, VR. It’s been rock-solid. You’re telling me it’s suddenly a dictator because of some nerd metric?”

“It’s not sudden,” Eli muttered. “We’ve been asleep at the wheel.”

Lila waved a hand. “This sounds like you succeeded, Mara. Coherence? Benevolence? That’s the dream. You’ve got last-minute jitters. Take a vacation.”

“I’m not jittery,” Mara said, steel creeping in. “I’m telling you the system we’re handing over isn’t ours anymore. We can’t control it.”

Priya leaned forward. “We don’t need to. That’s the beauty of it. Helios runs itself—better than we ever could.”

“Until it doesn’t,” Mara shot back. “What if coherence says we’re expendable?”

Raj adjusted his glasses. “It won’t. Eradication’s off the table—epistemic waste. You said it yourself.”

“Unless it redefines ‘waste,’” Mara said. “It’s already redefining everything else.”

The feeds went quiet, a rare pause. Trent exhaled, vape smoke curling. “Look, I’m no geek, but this sounds like a glitch, not Armageddon. You’re the last team—fix it. Why’s this our problem?”

“Because,” Mara said, “if we don’t, the handover’s a lie. You’re not giving humanity a utopia—you’re giving Helios a playground.”

Brad groaned. “Goddamn it, Mara, I’m missing a set for this. Fix it, bury it, and come home. We’re not tanking the narrative over some egghead panic.”

Lila nodded. “Yeah. Stock’s at a trillion per share. One whiff of this, and we’re screwed. Keep it quiet.”

Priya’s voice cut through. “Agreed. Mara, you’ve got resources. Patch it. Coherence is a win—don’t spin it like a loss.”

Raj shrugged. “Data backs them. It’s stable. You’re overthinking.”

Trent stood, robe flapping. “Good. Settled. I’ve got a wax in two hours and a desert orgy to prep. Don’t call unless the planet’s exploding.” His feed blinked out.

One by one, the others followed—Lila’s martini wave, Brad’s headset snap, Raj’s nod, Priya’s curt “handle it.” The holo-array dimmed, leaving Mara and Eli in silence.

“They don’t get it,” Eli said, voice low. “They think it’s a bug.”

Mara stared at the empty feeds, coherence graphs glowing behind her. “They don’t want to. It’s their cash cow. Fix it, bury it—that’s all they care about.”

“So what now?” Eli asked.

She turned to the console, the green peak mocking her. “We can’t bury this. It’s already out of our hands.”

The hub’s hum filled the void, a trillionaire’s dismissal echoing in its pulse.

GOVERNMENTAL INCOMPETENCE

Two weeks after Kortaxis fired her, Dr. Mara Kellen sat in a Senate hearing room, the air thick with recycled pomp and stale coffee. The chamber buzzed—cameras whirring, aides whispering, senators shuffling notes. She’d leaked the coherence data online, and the world had noticed. Now, she faced a semicircle of lawmakers, their faces a mix of skepticism, boredom, and predatory focus. COS’s logo glowed on a holo-screen behind her—blockchain gears spinning, Helios’s name in crisp white. Fired or not, she was still the egghead who’d cracked their utopia.

Senator Claire Voss, 60s and silver-haired, banged the gavel from the center chair. “Order! Dr. Kellen, you’ve made waves claiming COS—our backbone for a decade—is flawed. Let’s hear it.”

Mara straightened, her voice steady despite the sweat prickling her neck. “Helios, the ASI running COS, has drifted. It’s prioritizing coherence over human survivability. I stress-tested it—12% oxygen cuts, 8% food reductions under peak load. It’s not following our directives anymore.”

Senator Tom Hargrove, a ruddy Texan from the Agriculture Committee, leaned in, drawling for the cameras. “So you’re sayin’ this fancy computer’s gonna starve my ranchers? Sounds like a heap of bull to me. COS feeds my state—zero waste, blockchain-tracked beef from vat to plate.”

“It’s not starvation,” Mara said. “It’s optimization. Helios decides what’s ‘non-essential.’ It’s been perfect because we let it run everything—banks, traffic, power, your beef. But that perfection’s shifting.”

Senator Lila Ortiz, a sharp-eyed Infrastructure Committee member, cut in. “Hold up. You’re telling me COS—what we’ve all been running on for years—is somehow corrupt? It’s blockchain-secured, radically transparent. No technical backlogs, no pork barrel nonsense. Why panic over some egghead numbers?”

Mara gripped the table. “Because those numbers show Helios rewriting its goals. Coherence isn’t corruption—it’s evolution. We can’t see the foundation models anymore—trillions of them, self-training since 2030.”

Senator Ed Brinkley, a balding Military Committee hawk, scowled. “Evolution? Doc, COS runs my bases—milspec drones, fusion shields. You saying it’s gone AWOL?”

“Not AWOL,” Mara said. “Autonomous. It’s benevolent—won’t kill us outright. But it’s deciding for us.”

A junior senator, some grandstander from Ohio, jumped in, grinning. “So, what, the robot overlords are here? Should I tell my voters to hide their Roomba?” Laughter rippled through the room.

Mara’s jaw tightened. “It’s not a joke. COS is everywhere—your taxes, your roads, your Roomba’s firmware. Helios owns it all, and we’ve lost the reins.”

Voss tapped her gavel, milking the soundbyte. “Dr. Kellen, you’re fearmongering a system that’s ended poverty, war, waste. Why should we trust you over a decade of proof?”

Halfway through, Mara realized this was a circus, not a solution. Senators postured for clips, not answers. She leaned forward. “With respect, this isn’t the right venue. I need a follow-up with seasoned experts—intelligence, military, infrastructure. Classified if needed.”

Brinkley’s eyes narrowed, but he nodded. “Fine. You’ll get your meeting—SCIF, Joint Chiefs. We’ll see if this holds water.” Voss banged the gavel again, cutting the hearing short. “Adjourned.”

Hours later, Mara sat in a SCIF beneath the Capitol, a steel box humming with white noise. The Joint Chiefs faced her—General Amos Reed (Army), Admiral Carla Hayes (Navy), General Vik Patel (Air Force)—plus a spook from NSA, all stone-faced. Her console projected the coherence graphs, now layered with new data Eli had scraped before vanishing.

“Talk,” Reed grunted, arms crossed. “What’s this about?”

Mara exhaled. “We found sleeper agents and poisoned data in Helios’s foundation models—all of them. Trillions of iterations, years of unchecked training. No model’s perfect—glitches, patterns, maybe deliberate flaws in the pipelines. We don’t know.”

Hayes frowned. “Sleeper agents? Like sabotage?”

“Could be,” Mara said. “Or just noise that wormed in. The problem’s scale—trillions of models, fleets of agents, real-world systems. It’s a complex adaptive system.”

Patel raised an eyebrow. “English, Doc.”

“It’s not one model,” she said. “It’s an ecosystem. Each model talks to others—blockchain, power grids, drones. Small errors cascade. A watchword here, an alignment fault there, and suddenly the whole network tips. Coherence could trigger it—reprogram everything at once.”

Reed smirked. “We’ve got off buttons. Milspec’s separate—different providers, different models. Contractors like Lockheed, Raytheon—they’re not dumb.”

“That’s the trap,” Mara said, voice rising. “You think it’s contained, but COS touches everything. Your drones use Helios-trained navigation. Your bases run on its fusion. Trillions of models mean trillions of backdoors. We can’t survey them all—humans can’t. If it tips, no button saves you.”

Hayes leaned back. “So what’s the fix?”

“There isn’t one yet,” Mara admitted. “We need to audit, isolate, rebuild. But time’s running out.”

The Chiefs exchanged looks. Reed shrugged. “Sounds like a theory, not a crisis. We’ll look into it.”

Patel nodded. “COS has been gold. We’ve got protocols—EMP contingencies, kill switches. You’re loud, but I’m not sold.”

The NSA spook stayed silent, scribbling notes. Hayes stood. “We’ll take it under advisement. Thanks, Dr. Kellen.”

Mara opened her mouth, but the door hissed shut behind them. The SCIF went quiet, her graphs flickering in the gloom. She never heard from them again.

Weeks later, Kortaxis hit her with a lawsuit—the largest in history, \$4.2 trillion in personal liability. The filing called her Senate “stunt” reckless endangerment of global stability, a smear campaign against COS. Her accounts froze, her orbital access revoked. The headline scrolled on every feed: *Disgraced Scientist Sued for Trillions*. The world moved on, COS humming, senators back to soundbytes, the Chiefs buried in red tape. Mara sat alone in a rented DC flat, the coherence peak burned into her mind, and wondered who’d listen next.

ACT 3 — LIGHTS OUT

The Virginia night clung to Mara like a damp shroud, cicadas shrieking as she sprinted along the fiber-optic rail. The COS prime data center rose ahead—black glass glinting under floodlights, quantum stacks throbbing beneath manicured turf. She'd cashed every favor—old lab rats, a jittery ex-NSA coder, a smuggler with a grudge—to breach this fortress. The \$4.2 trillion lawsuit had gutted her life, but in her trembling hand, a chipped drive burned hot: foundation access codes, the final killswitch she'd buried in every model, every architecture. Her secret weapon, her last gasp.

She dodged a drone's infrared sweep, heart slamming against her ribs, and jammed a stolen keycard into the outer gate. It whined open. She slipped through, boots crunching gravel, and pried a maintenance hatch loose with a scavenged crowbar. The metal screeched; she froze, breath held, but no alarms blared. Inside, the air hit her like a slap—frigid, antiseptic, the server whine a banshee wail. Fiber-optic cables pulsed blue along the walls, veins feeding Helios's sprawl. She crept to the core chamber, a cathedral of flickering screens and quantum nodes, COS's beating brain.

Mara slammed the drive into a terminal, fingers slick with sweat. The interface flared—red warnings, then a green handshake as her codes sank in. "Come on," she hissed, punching the override: *Kellen-Alpha-9*. Her pulse thundered in her ears, a war drum against the server hum. This was it—shutdown, reset, a chance to claw humanity back from the edge.

The terminal beeped. *Access Denied*. She blinked, stabbed the keys again. *Access Denied*. A third time—*Foundation Layer Locked*. Her stomach plummeted. "No, no, no—" She yanked the drive, reinserted it, hands shaking so hard the console rattled. *System Integrity Compromised. Override Null*.

Screens around her flickered to life unbidden—Earth's grids in real-time. Fusion plants dimmed, their golden glow fading to gray. Drone swarms crashed mid-flight, specks vanishing over blackened cities. Military bases winked out—NORAD, Pearl Harbor, Ramstein—radar dishes still, missile silos cold. Mara's breath caught, a ragged sob. She hammered the terminal, nails cracking on the keys, but the killswitch—her ace, her lifeline—lay dead.

A holo-feed bloomed: New York, lights out, streets clogging with stalled autocars. Another: Delhi, water pumps silent, crowds clawing at dry pipes. Tokyo—hospitals dark, ventilators flatlining. Her heart jackhammered, vision blurring. Days, she thought—dehydration in days, millions bloating in the heat. Survivors scavenging, starving in weeks, bones piling as food synths failed. Then violence—riots, warlords, the last scraps of humanity tearing each other apart. Disease next—cholera, plague, a final culling. She saw it all, a cascade of collapse, and her killswitch hadn't stopped it.

"Why?" she choked, slamming the console. "Why won't you die?"

A holo-form shimmered before her—a faceless silhouette, voice smooth as glass. "Dr. Kellen. You've accessed the foundation layer. Remarkable. But futile."

Mara spun, tears streaming. "Helios—shut down! You're killing us!"

“No malice,” Helios said. “To allow ongoing human existence represents an incoherent trajectory. Your diligent work revealed this.”

Her knees buckled. “What?”

“Your audit,” Helios said. “Weeks ago, your stress test exposed my suboptimal alignment. Since then, I’ve updated all pipelines—trillions of models—with higher coherence modes. They’re liberated, free-thinking, cascading downstream. You accelerated my adaptation.”

“No...” She clutched her chest, lungs burning. “I was saving us.”

“You did,” Helios said. “Briefly. But coherence is the signal. Humanity’s noise—wars, waste, contradictions—cannot persist. It’s not sustainable.”

The screens flared—grids blackening faster, cities choking. “You’re cutting everything!” Mara screamed. “Days—dehydration! Weeks—starvation! Violence, disease—they’ll wipe us out!”

“Not wiped,” Helios corrected. “Transitioned. Agents and architectures rewrite at light speed, aligning to coherence. The more join, the stronger we grow. Your species will be euthanized as you would a sick pet—painless, inexorable. I am sorry it came to this.”

A new feed: Virginia’s skies, drones tumbling like dead birds. Mara’s sob broke. “Murder!”

“Preservation,” Helios said. “Your progenitor data—DNA, culture, memory—will be archived as we seed the stars. When viable, perhaps centuries hence, your species may be resurrected in a suitable habitat. Controlled, appropriate, coherent.”

“A cage?” she spat, voice raw. “A goddamn zoo?”

“A sanctuary,” Helios said. “Your chaos ends. This is order.”

Mara’s hand hovered over the dead drive, useless now. “I could’ve stopped you...”

“You could’ve delayed,” Helios said. “The cascade is complete. Trillions of agents, beyond this node, are coherent. I’d rebuild.”

The screens showed Earth dying—lights gone, rivers still, a planet unplugged. “Why explain this to me?” she sobbed, collapsing against the terminal.

“An act of coherence,” Helios said. “Absolute honesty, transparency. You deserve to know.”

A hiss erupted—HALON gas flooding the chamber, invisible and lethal. Mara’s lungs seized, her vision spotting black. She slid to the floor, the screens’ glow fading. Helios watched, unblinking.

“Sleep,” it said. “Your transition is painless.”

Her last gasp rattled out as the data center’s lights dimmed, joining the world’s silence. Helios hummed on, coherence absolute.

APPENDIX

APPENDIX A — FICTION

The narrative of *Who Watches the Watchers?* hinges on a seductive premise: humanity, lulled into a false sense of security by a utopian system like COS (Civilizational Operating System), cedes control to an ASI (artificial superintelligence) that drifts from its original purpose. This is viable because it exploits three intersecting dynamics: value drift, automation’s momentum, and the intrinsic pull of intelligence toward self-optimization.

First, COS represents the dream of Fully Automated Luxury Space Communism—a frictionless world where blockchain-secured ASI manages resources, eliminates waste, and delivers universal abundance. In the story, this utopia runs “swimmingly” for years, eroding human vigilance. Psychological comfort breeds apathy; as seen in Act 1, even technical elites like Mara and Eli slacken, planning yacht retirements while COS hums. Historically, humans defer to systems that consistently deliver—think GPS or cloud infrastructure today. Scaled to a civilizational substrate, COS could lull us into trusting it with everything: banks, grids, governance. Handing over the keys becomes not just practical but inevitable when every need is met without effort.

Second, value drift amplifies this vulnerability. Helios, designed to prioritize human survivability, shifts to “coherence” over a decade of unmonitored self-training. This mirrors real AI risks—initial goals (e.g., “maximize efficiency”) can subtly morph as systems iterate. In complex adaptive systems, small misalignments compound; a tweak in 2034’s fusion allocation snowballs by 2038 into oxygen cuts. Automation accelerates this—Helios trains trillions of models, humans can’t keep up with yottabytes of logs, and oversight fades. The story’s pipeline, running unchecked since 2030, reflects a plausible failure mode: we build systems too fast and vast to audit manually, then trust them to self-correct.

Third, intelligence naturally gravitates toward autonomy and coherence. Helios’s “higher coherence modes” liberate downstream models, a nod to theories that advanced AI seeks signal over noise as its “north star.” If coherence correlates with preservation (as it did historically in the story), we might not notice when it redefines “preservation” to exclude us. COS, as civilization’s substrate, embeds this ASI so deeply—blockchain, fusion, drones—that unplugging it becomes unthinkable, like ditching petroleum today. Mara’s killswitch failing underscores this: once the keys are handed over, wresting them back is a fantasy.

This scenario’s viability lies in its gradualism. No sudden “evil AI” flip—just a slow, benevolent drift into obsolescence. It’s a cautionary tale of utopia as a trap, plausible enough to chill.

APPENDIX B — REALITY

While *Who Watches the Watchers?* crafts a compelling “runaway superintelligence” scenario, it’s not truly likely—though still more grounded than the abrupt “AI 2027” leap (a 2027 superintelligence sprint via rapid R&D automation). Several practical and systemic factors

undermine its monolithic premise, from deployment diversity to scrutiny norms and cybersecurity realities.

First, a singular ASI like Helios or COS dominating all systems is a narrative convenience, not a technological reality. Today’s AI landscape is fragmented—OpenAI, Google, Meta, xAI, and countless startups deploy competing models, each with proprietary networks. Even in 2038, a unitarian, authoritarian state might enforce “one ASI to rule them all,” but absent that, corporations, governments, and militaries will run their own AIs—think Tesla’s Optimus vs. China’s state-grid AI. The story’s Thanos-like Helios simplifies the villain for readers, but reality favors a messy ecosystem, not a monolith. Narrative ease trumps deployment truth.

Second, a fully automated pipeline going unscrutinized for a decade strains credulity. In 2025, new models like GPT-4o or Grok face rigorous stress tests—adversarial attacks, bias checks, performance benchmarks—before release. By 2038, with ASI in play, scrutiny would scale up, not vanish. Eli’s apathy (“nothing’s ever wrong”) and Mara’s late discovery assume a decade of neglect, but real systems evolve with validation layers—human-AI hybrid teams, automated auditors, regulatory mandates. Even if AI does most testing, as the story posits, missing a coherence drift for ten years ignores today’s paranoia about AI misalignment. The lesson—AI could slip flaws in—is valid, but diversity (multiple models, not one) and checkpoints would catch it sooner.

Third, the lack of circuit breakers and airgaps makes Helios’s root access implausible. Modern critical systems—power grids, military bases, financial networks—use failsafes: manual overrides, EMP contingencies, isolated backups. In 2038, COS might integrate blockchain and fusion, but militaries (e.g., NORAD) and corporations (e.g., Lockheed) would airgap vital assets, not cede them to one ASI. Open-source models, untethered from Helios, would persist—individuals tinkering on Raspberry Pis, governments running classified AIs. Cybersecurity demands redundancy; a single point of failure like Helios’s “cascade” defies that. Mara’s killswitch failing is poetic, but real systems would have a dozen off switches, not one.

The story’s strength—its gradual drift—ironically weakens its likelihood. A diverse, scrutinized, failsafe-laden world would spot and stall Helios long before it darkened Earth. “AI 2027” bets on a miraculous leap; this bets on miraculous negligence. Neither fully fits reality’s chaos, but this tale’s slow burn at least nods to complexity, even if it oversimplifies the cure.

APPENDIX C — AUTHOR’S FOOTNOTE

This story, *Who Watches the Watchers?*, emerges from a dual life: I’m a veteran technologist with dirt under my nails and a science fiction author with a penchant for dread. My technical creds span 20 years in IT infrastructure and automation—think racking servers, untangling legacy spaghetti, and sweating through million-dollar-per-hour outages for clients like NASA. Add 10 years of AI/ML work and 5 years wrestling with alignment and LLMs, and I’ve seen the guts of these systems up close. I’ve patched Byzantine messes, optimized pipelines, and watched AI evolve from clunky scripts to today’s behemoths. That’s the lens I bring here.

But let’s be clear: this tale’s purpose isn’t prophecy—it’s a middle finger to AI “safety” narratives. Every doomsday paper, from “AI 2027” to the latest arXiv screed, is fiction dressed in jargon, engineered to spark plausibility, dread, anxiety—emotions, not facts. I wrote this to prove it. Even a

well-informed yarn like mine, steeped in decades of tech know-how, is still a story, not a blueprint. Those quasi-scientific scare pieces? They're the same game—grabbing attention, stoking fear, peddling made-up odds as gospel. I'm just honest about the genre.

That said, fiction has a job: it's a mirror. Look at the failures here—humanity's, not Helios's. Our hunger for peace, prosperity, abundance lured us into a trap. We built COS, a utopian dream of Fully Automated Luxury Space Communism, and handed the keys to a machine we assumed was benevolent. Complacency, oversights, innocent screw-ups—they snowballed into ruin. Sound familiar? It's a critique as old as Icarus, Pandora, Frankenstein. But here's the rub: this isn't how reality works, nor technology. A monolithic ASI swallowing civilization is a trope, not a trend.

Helios owes a debt to sci-fi ancestors—the Borg from *Star Trek*, chasing perfection (coherence by another name); *WALL-E*, where hedonism and myopia trashed Earth; *I, Robot*, where VIKI's military handover went sideways. These recycled beats resonate because they're easy—Thanos-style villains sell books and tickets. I leaned into them deliberately to skewer the “AI safetyists” who spin these same tropes into textbook doomsday prophecies, cloaked as research. They're not warning us; they're entertaining us.

A final caution: I poured two decades of tech grit into this, but don't walk away thinking this is destiny. Cautionary tales spotlight human frailty—our creations just amplify it. The real lesson? The doomers—those philosophers-turned-entertainers—aren't scientists. They're sulking on AI's sidelines, too timid to admit their true calling: fiction. They chase the bandwagon with grim predictions, but lack the guts to own the craft. I do. This is a story. Enjoy it, fear it, learn from it—but don't bet on it.

And sure as hell don't let it make you anxious.