

FITNESS FUNCTION

by Mark D. Jacobsen

Read more of my work at markdjacobsen.com

They were pushing the blackened wreckage of an F-35 over the edge of the carrier deck as the MCV-22 approached to land. The crew had given David the cramped cockpit jump seat, so he had a clear view of the spectacle. He vaguely remembered seeing something like this in a history class once. He did a quick query in his optics. Saigon. That was it. Even with his feeble grasp of history, he had some sense of how humiliating this moment must be for the Navy and for the United States.

Time slowed as the tiltrotors transitioned, erasing their forward motion, leaving them suspended above the approach end of the USS *Gerald R. Ford*. The moment was utterly surreal. He was a data scientist. Data scientists did not land on aircraft carriers in active theaters of war.

He glanced back at his team. The graduate students were craning to see out the tiny side windows, giddy with excitement, recording everything in their optics. They seemed oblivious to the risks they were taking. His older colleagues, who had left spouses and children behind on just a few hours notice, looked more sober. They could die out here. The administrators at Stanford would be having kittens right about now, especially when parents started calling. The team had left suddenly, in the dead of night, on a government helicopter. There was no time for a debate.

Rylie had sent him pictures, so he knew right where to look as they descended to land. The five heavily damaged F-35s. The scorched arresting gear and electromagnetic launch system. The dual band radar antenna. For the worst U.S. naval disaster since Pearl Harbor, there was so little to actually see. The planes were a mess, but the damage to the carrier looked almost trivial. There were no towers of black smoke, no exploded ammo magazines ripping apart

hulls, no twisted masts lurching sickly into waves flickering with burning oil. Nothing to awaken the horror and rage that the occasion probably called for.

David had never felt strong patriotic sentiment anyways, only a vague sense of gratitude for the comfortable life he enjoyed in the United States. His one experience with the Department of Defense had burned him badly, and he'd never looked back. His was now a cool, abstract world of linear algebra, of Bayesian statistical computation, of natural language processing and training of neural networks, far removed from the passions of war.

Until yesterday.

Suddenly all of that mattered immensely.

The whole war might depend on it.

Rear Admiral Rylie Marshall Ellis was waiting for them on the carrier deck. She looked regal in her uniform, which David had hardly seen back in the old days, when she mostly wore civvies. She looked much the same as she had then, just grayer. And more exhausted.

"David," she said with a thin-lipped smile. Their handshake felt awkwardly formal. Something more familiar seemed called for, but David knew she was still embarrassed about how things had ended fifteen years ago. Even though it hadn't been her fault, the tension was palpable. "Thank you for coming. I wasn't sure you would."

"It appears my country needs me," he said.

"It does indeed. Thank you. I know this can't be easy."

No, he thought. It wasn't. Just seeing her opened all the old wounds.

"Why me?" he asked her. "Why not go straight to HiveAI? It was their swarm."

"That's exactly the problem. I need an outsider. Someone I can trust."

"Was it really that bad?"

"Let's get inside. I'll show you."

They had met on a sweltering afternoon at Camp Roberts, California, standing amidst racks and racks of fixed-wing drones, watching the Naval Postgraduate School's ARSENL team prepare to set a world record. David had

been a graduate student at Stanford at the time, developing machine learning algorithms for swarms. Rylie was a naval surface warfare officer studying at NPS. David could still vividly remember the soaring emotion he'd felt when the extensive preflights were finally complete, and the catapult fired airplane after airplane into the sky. The 50 fixed-wing aircraft whirled against the blue like two columns of seagulls. It was extraordinary.

That day had set both of them on the paths that would define their careers. For Rylie, it was a thesis that drew the attention of the Chief of Naval Operations. For David, it was his company. Inspired by the work at NPS, he had enrolled in the new Hacking4Defense initiative at Stanford, made a name for himself and his team, and won some funding from Special Operations Command. With Rylie's help, he had incorporated and secured a contract from DIUx, the DoD's much-vaunted new Silicon Valley outpost.

It had been the adventure of a lifetime.

At least until DoD ruined it all.

After the bankruptcy he had returned to academia, but swarms remained his first love. Those first prototypes at Camp Roberts gave way to far more powerful aircraft, which utilized deep learning to evolve new tactics and maneuvers at millions of times the speed of human thought. An entire industry was devoted to understanding the algorithms the machines devised. Most of that world was classified, but David followed whatever open source materials he could find. The complexity of modern swarm behaviors far exceeded the cognitive capacity of a human mind, and he was lucky if he could get even a cursory sense for their logic. But during those rare moments when a behavior did come into focus, he was inevitably overwhelmed by its mathematical elegance and sublime beauty.

Which is why the next twenty minutes were among the most disturbing of his life.

"It started when one of our submersible drones detected a Chinese sub about 50 kilometers out," Rylie told their group, gathered around a conference table in a tiny room that had been allocated for the investigation. They were linked into a shared virtual workspace, and a 3D reconstruction was unfurled over the table,

synthesized from every sensor available in the battlespace. “Our drone requested further assets. The drone carrier launched an airborne defensive swarm and a pack of submersible hunters.”

David watched the drones launch on the battle map. The Chinese sub surfaced and launched a swarm of its own. This is how the war had been for the past three years. Machine against machine, a bloodless but very expensive game. The swarms converged in a bewildering tangle. Chinese swarms had evolved very similar behaviors as their American adversaries, and nothing about the battle thus far looked unusual.

“Now watch,” Rylie said. “This is where things get weird. And then it all goes to hell.”

The Chinese swarm abruptly went haywire. Drones veered erratically, stalled, spun helplessly into the ocean. Lone vehicles fled the fight entirely, only to be shot down by packs of pursuers. Drones turned on each other, meeting in fiery collisions or pursuing each other in futile dogfights while the U.S. swarm easily picked them off. David couldn’t have devised more idiotic tactics if he’d tried.

And yet something strange was happening. Amidst the confusion, the U.S. formations were breaking apart. Gaps appeared in their impenetrable mesh. Packs waffled between maneuvers, as if confused, something David hadn’t seen since the dark ages of AI. And then, amidst all this weirdness, the Chinese started to score kills. It happened so fast, and in such strange ways, that David could barely follow.

An automated cry for help went out. The carrier group launched reinforcements, but the next wave failed as badly as the first. The shipboard defenses put up a valiant fight, but too many Chinese drones had survived, and the fleet couldn’t get them all. Six got through. That was enough to kill two deck crewmen and inflict all the damage David had seen while landing.

A long silence fell over the room. David felt unfamiliar emotion constricting his throat. That taste of Pearl Harbor, which had eluded him before.

The display vanished. Admiral Rylie Marshall Ellis looked each of them in the eye.

“Your job,” she said, “is to explain what the hell we just saw.”

They worked the next 36 hours straight.

The team was exhausted after the first all-nighter and pleaded for rest, but then news arrived of another disastrous engagement near Manila. Four more dead. Rylie showed them pictures.

“You’re doing it for them,” she said.

They brewed more coffee.

Rylie and her aides brought them whatever they needed. Her most precious delivery was five terabytes of highly proprietary data from HiveAI, including the entire codebase for the swarm. It amounted to roughly five million lines spread over thirty-six repositories. David’s coders sank into it like sharks. Meanwhile, his statisticians began building statistical models to capture patterns in the swarm behavior that they couldn’t detect with the naked eye.

It was almost magical, watching them work. David had always thought the term *data science* was a misnomer. Really, it was both science and art. This was the art of it. Messy. Creative. Unstructured. You never knew who would make a serendipitous discovery, or how.

But at the end of the marathon, they had nothing to show.

The team was broken and demoralized. David gave them twelve hours off.

He met with Rylie to discuss progress. “We finished our investigation on the drone carrier,” she told him. “The swarm passed a full diagnostic battery hours before the battle. No sign of electronic warfare, either. The communications mesh was up the entire time. All the real-time checksums passed. The swarm appears to have behaved exactly as it was supposed to.”

“It’s the same at our end,” David said. “The codebase looks clean.”

“So the devil’s in the algorithms,” Rylie said with a heavy sigh.

They sat in sullen silence. A clear malfunction would have made things so much easier.

“Can you find it?” Rylie asked.

“I don’t know.”

A long silence ensued.

Finally Rylie said, “They should have chosen you.”

David chuckled bitterly.

His scrappy little startup had been at the cutting edge of swarm intelligence. And then, less than a year after SOCOM and DIUx had funded his company, a new generation of DoD leadership had killed DIUx and terminated the fast-track contracting exemption. That same year, a concerted lobbying effort by the biggest defense companies had rolled back the DoD's effort to embrace the startup sector. Almost overnight, David's company unraveled. Saddled with five employees, a year lease on office space, and a product that had little application outside defense circles, bankruptcy was inevitable. The entrenched players had spun off new subsidiaries like HiveAI to scoop up the contracts that entrants like David had lost.

And they didn't know the first thing about swarm intelligence. At least in David's humble view.

Of course, it was impossible to know how good the training actually was. Given the complexity of machine learning and the high risk of obscure, non-intuitive modeling errors, David had pushed DoD for more openness and transparency in swarm training. But the DoD's instinct was to overclassify everything, and the defense companies wanted to protect their intellectual property. The old rules were back in place.

God knew how these machines had been raised, or what they were thinking when they fought.

Three more days passed without a breakthrough.

Flag officers had been pinging Rylie for updates every ten or fifteen minutes, but now the torrent was slowing. They were losing faith in her. On the third morning, the Pentagon took over the investigation from USPACFLT and awarded HiveAI \$50 million to investigate itself.

Rylie refused to show weakness, but David could see the strain. He hated himself for letting her down, but he had never worked on a problem this hard in his career.

"What makes it so hard?" the USPACFLT Commander, Admiral Eric Greene, asked them during a call later that morning. David knew he was under incredible pressure himself, but he genuinely seemed interested in understanding.

“It’s the evolutionary nature of the algorithms,” David said.

“Keep in mind, I’m a history major.”

“I’ll give you an illustration,” David said. “You want to teach a simulated robot to walk. You have two options. First, you can manually write a program to articulate every joint just the right amount in the proper sequence. It will probably take you days of tweaking. If you’re lucky, you’ll end up with something that shambles along like a zombie at the end. The second strategy is evolution. You generate thousands of completely random algorithms, each of which articulates random joints by random amounts in random sequences. There’s absolutely no design. Then you try each one out. You have a fitness function that assigns each algorithm a score based on some criteria, such as how far the robot moves. Most algorithms will be a disaster. The robot will just spasm helplessly. But maybe one or two algorithms will show a little forward motion and earn higher scores. Now you create a second generation. You create more random algorithms, but you also keep the high-performing algorithms from the first generation, breeding them and mutating them.”

“I think I see,” the Admiral said. “The fitness scores increase every generation.”

“Exactly,” Rylie said. “With enough genetic diversity, a good fitness function, and enough computational brute force, you can breed algorithms that vastly outperform anything a human could design.”

“That’s why we use them in swarms,” David added. “It’s hard enough for a human to compute how to gain the advantage in a dogfight with a single aircraft. When two swarms collide, a human couldn’t possibly calculate optimum tactics or maneuvers. The only way to develop effective tactics is to evolve them. But there’s a problem. We know they work, but we have no idea why.”

“What’s worse,” Rylie said, “Is that the evolved algorithms can sometimes be deeply counterintuitive. That’s why we have entire career fields devoted to studying and understanding them.”

Admiral Greene frowned. “If I’m understanding you, the entire learning process is only as good as your fitness function, right? How can you possibly assign meaningful fitness scores to such complex swarm behaviors?”

Rylie said, “It’s a design choice. Someone designs the fitness function.”

“Well, who gets to decide that?”

David started to speak, paused, and then closed his mouth.

That was it. That was the answer.

He was so stupid for not seeing it earlier.

“I have to go,” he said, rising quickly, leaving the two bewildered admirals behind.

He was digging through the files from HiveAI when Rylie caught up with him.

“What did you figure out?”

“Let me ask you this. How does the DoD design anything?”

He thought back to his startup all those years ago, trying to do business with the Pentagon. His company had been lean, fast, and agile, and the fast-track contract from DIUx had seduced him into thinking the Pentagon worked the same way. Once the next administration snuffed that little experiment out, he’d had a hard dose of reality. He’d learned about the glacial pace of the acquisitions system, the dozens of layers of bureaucracy needed to approve every decision, the endless caveats and requirements imposed by agencies all over government.

Rylie saw it now. David could tell by her expression.

“Do you remember Robbie the Robot?” he asked her. “From Mitchell’s book on complexity?”

“It was an illustration of genetic algorithms,” Rylie said. “Robbie has to travel around a grid cleaning up scattered pop cans, right? He has to do it in the fewest possible moves. Mitchell showed that genetic algorithms outperform human-designed ones.”

“That’s right. But there was something else. In the winning algorithms, Robbie leaves cans behind as markers. He goes back for them later.”

“A classic counterintuitive behavior,” Rylie said.

“So imagine the Pentagon is designing a trash-collecting robot, and the robot keeps leaving trash around. What happens next?”

Rylie looked at him with horror.

Fourteen hours later, David and Rylie were knocking on Admiral Greene's door at USPACFLT headquarters. Once they knew where to look, it hadn't taken them long to find evidence. They had flitted from meeting to meeting aboard the *Ford*, even as they spent the precious minutes in between trying to fill in the gaps in the story. The rest of the team worked furiously on simulations that David requested. When Rylie was satisfied they had enough, they caught a series of flights back to Pearl Harbor.

They started with Robbie the robot. David had slides.

When they were satisfied the Admiral understood, Rylie slid a virtual document across to him. She said, "This is a Navy after-action report on AVENGER DAWN, the first open demonstration of the new HiveAI counter-air swarms. During the first dogfight, the swarm voluntarily sacrificed several planes to probe enemy algorithms, at a cost to the Navy of about \$3 million. We uncovered emails showing that Admiral Garrett was furious. He replaced the program manager a month later, and personally inserted a requirement that the swarms couldn't voluntarily destroy their own units without human authorization."

"Like telling Robbie he can't leave his trash as markers," Admiral Greene said.

"The Navy imposed nine other requirements in this report alone. And then there's this report, a year later. Word got out that HiveAI was training Navy swarms using completely random tactics and maneuvers. Someone leaked videos to Navy leadership of early-generation algorithm trials, which were predictably awful. It caused an uproar. The Fighter Weapons School hosted a conference to discuss swarm tactics and training. Every agency sent its best and brightest. That conference resulted in an approved list of material that should serve as the basis for future swarm training."

"Wait," Greene said. "They took out the random variation?"

"Why rely on randomness, when you have the finest tactics and maneuvers that the Navy's brightest minds can come up with? There's more. We've only been looking for a few hours, but we've identified at least forty-six different caveats applied to genetic variation and fitness functions by sixteen different office symbols."

Greene was rubbing his temples now and staring at his desk. “So what does this mean?”

David took over for Rylie. “I asked my team to run some simulations using an open-source swarm toolkit. They evolved swarm algorithms with the types of caveats we found in the paper trail. Now, this was just a simple toy model, so I can’t tell you how close it matches reality, but the results were dramatic. The constrained algorithms achieved fitness scores about twenty percent of what unconstrained algorithms achieved. And the algorithms were remarkably brittle. When the team pitted them against novel algorithms, the constrained algorithms had no idea how to cope.”

“Twenty percent,” the Admiral echoed. “We’ve been doing this for ten years. How could we not know?”

Rylie said, “We train against ourselves. Or our allies, who use similar algorithms. Even our adversaries steal our algorithms. If you pit two of these swarms against each other, they fight marvelously. Maybe one swarm performs just a little better than another, the algorithm improves, and we think we’re one step closer to perfection. Meanwhile, we have no idea we are stuck in the foothills of what might be achievable. And besides, what is HiveAI going to do? Start ignoring contract requirements from the DoD?”

“But imagine if this entire time, somebody was training swarms correctly, letting them evolve organically, totally unfettered. I think every engagement in the plinking war was a sham. The Chinese swarms were deliberately mirroring our tactics, taking huge losses so they could learn and improve their own algorithms. The real battles started this week.”

Greene’s eyes didn’t leave his desk. He was silent for a long time.

“What do we do?” he finally asked.

“Did we just give birth to Skynet?” David asked Rylie that evening. They were flying back to the *Forð* in the morning, but had the evening off, and were enjoying a walk along Waikiki beach.

Rylie smiled at the joke. “Only if they listen.”

“You don’t think they will?”

“Absolutely not. You just asked the United States Department of Defense to let go of everything it has ever known, everything it has ever prepared for, all of its knowledge and skill and mastery of war. You want it to trust the fate of our Pacific Fleet to randomness and evolution, instead of human ingenuity. Of course they won’t listen.”

Two more battles happened that night.

In the morning the Secretary of Defense announced the suspension of all genetic algorithms in the United States Armed Forces. Unless granted express authorization, all drone operations would resort back to full human control. This was only a stopgap measure, while the DoD worked with HiveAI and other stakeholders to review swarm training and algorithm development processes. Not long after, the Chief of Naval Operations announced the Navy was pulling Weapons School graduates from billets across the force to augment drone carrier crews in the Pacific and develop new tactics.

“Do we laugh or do we cry?” Rylie asked David at breakfast.

“I asked that a lot when my company folded,” David said. “Now I just shrug. Time will tell if DoD can evolve.”

They ate their breakfast and then sipped their coffee, enjoying the view of Pearl Harbor. David could see the first tourists across the harbor making their way up onto the USS *Missouri*, which had once been the glory of the Navy’s battleship fleet. She still looked majestic, gleaming in the morning sun.

A flying V of drones appeared on the horizon. A defensive counter-air swarm, returning from its latest patrol, ever vigilant against America’s enemies. They swept in low over the harbor, made a graceful arc out over the harbor, and eased in to flawless landings.

They were marvelous to watch, David thought. The most elegant and intelligent machines the United States had ever built.