The Hydra's First Head

"Listen, I don't say this lightly." The video on the screen showed a slanted, low-angle shot of the President's face. The audio was muffled but the voice was unmistakably his. The movement of his lips matched the words perfectly. *"China will stab us in the back if we don't act first. Going nuclear is the only option. Let's drop one on Chengdu, tell them the next one will be Shanghai. The kid gloves are off."*

Dr Harry Shah, the White House's newly-appointed 'Truth Guru', as both Congress and the press insisted on calling him, paused the video.

"Allegedly a secret recording of a clandestine meeting with the Secretary of Defense," he said. "Forty million views and counting."

The five meeting attendees shifted uncomfortably in the overheated, windowless room.

"It's fake, obviously," said Harry finally. "I assumed that didn't need stating."

Eliza Stevens, the CIA's Head of Cybersecurity, snorted dismissively. "Just more deepfake bull. My team sees hundreds of these every day. We'll get the social media networks to slap a disclaimer on it and it'll be forgotten in a week."

"Unfortunately, the Generative AI behind it is fooling all our AI-detection tools," replied Harry. "It's our word against the video. That's not going to fly with platforms, journalists, or China. Speaking of which?" He turned to Doug Hardiman, the Secretary of State's Chief of Staff, who was sweating in a too-tight suit.

"No luck getting hold of China so far," said Doug. "Our embassy in Beijing says their lines have gone dead too. And satellites have picked up an uptick of activity around China's known nuclear warhead sites."

"Can we be sure this video isn't a disinformation attack from China itself?" asked Annette Baker, the Department of Defense's Chief Information Officer.

"Unlikely," answered Doug. "They're still negotiating terms with us after their Taiwan blockade. Escalation to the nuclear level is in neither party's interests."

"So what's the plan, 'Truth Guru'?" said Eliza. "What does your 'disinformation doctorate' say we should do?"

"Well, that's what this meeting is for," answered Harry measuredly, not rising to the jibe. His new role had led to him inadvertently stepping on Eliza's toes on several occasions. "I have a plan, but it'll need coordination. This video is just the first. It's more sophisticated than anything we've seen, and the algorithms behind it have potentially already been spread far and wide. Right away, we need manual monitoring from the CIA of all social media and news outlets, and rapid takedowns of anything suspected to be disinformation."

“Surely you know that we can’t regulate the information put out by private entities, despite your administration’s best efforts,” said Eliza. “The best we can do is ask them nicely.”

“When it’s an issue of national security, I imagine you can get more leverage,” said Harry. “Someone is trying to end the world as we know it. We can’t just sit back.”

“How dramatic. But even if we can take down these posts, we can’t sustain the manpower it’ll demand. My team is already over-stretched fending off Chinese cyberattacks. We need a surefire way of detecting and deleting these posts automatically.”

“Which is why Professor Guo is here.” Harry gestured to the tall bespectacled woman sitting next to him. “She was my supervisor at MIT. The mind behind many revolutionary machine learning advances.”

Professor Guo gave a nervous wave. It had taken some effort from Harry to persuade her to enter the Capitol.

“This new deepfake technology calls for a new adaptive detection system,” she said, speaking as if from a script. “We can build on our existing machine learning models to develop this within a week or two, with the right resources. We request the CIA provide us with the training data – all the footage recognized as false but not picked up by existing detection software.”

“That’s quite the ask, given the volume of deepfakes we could be talking about,” said Eliza.

“There’s more. Location of the deepfake algorithm itself would also accelerate the process. I recommend seeking the original developer, if they can be found. And finally, we need as many AI-capable semiconductors as possible.”

“Computer chips?” Eliza said. “Surely MIT has plenty of them at its disposal.”

“Not the kind we need, unfortunately,” said Professor Guo. “Given Taiwan’s, er, situation, global computer chip supply is at a bottleneck.”

“The state will fund whatever you need,” said Doug brusquely. “Just put in the order with Harry and we’ll make it happen.”

“It’s not about funding. The chips we order are tailor-made, but the few US foundries capable of making them are too over-subscribed to take any new orders. We’re improvising, cannibalizing as many chips as we can from other projects and institutes, but given they’re not optimized for our purposes, we need a lot of them. So we’re hoping the government can step in to prioritize our orders.”

“Not possible, unfortunately,” said Annette, folding her arms. “The Department of Defense has requisitioned all US foundries’ capacities, with the President’s sign-off. We need the chips for our own technologies. Anything that might give us the edge over China, if

it comes to that. And before you ask, we can't divert any capacity to this project. You have your arms race, we have ours."

"It's the same race," said Harry, trying to rein in his frustration. "Videos like these have the power to precipitate nuclear war. Aside from how they corrupt public opinion, if just one person with the right amount of power believes these videos – here, China, wherever – it's game over. Are there really no strings you can pull?"

Annette rubbed her forehead. "Well, this information stays within this room, but – the folks at NASA are sitting on a bunch of supercomputers left over from the last administration's canned Space Force project. We're trying to requisition the tech but NASA is annoyed with us after the ISS decommissioning debacle, so they're stalling. A request from MIT might be met more positively."

Professor Guo nodded pensively.

"And if you do secure the chips, the DoD would be very interested in purchasing them once you're done with them," said Annette. "At a discount, preferably."

"Well, then," said Harry, standing up. "Guess we'd best get those chips put to good use before they're shipped off to war."

#

"Reports of US citizens being taken hostage by the Chinese government, even allegations of torture…"

"Apparent Sinophobic attacks continue to rise, as a Vietnamese business owner in Garden Grove was found unconscious…"

"A nuclear attack from the USA will be met with overwhelming retaliation…"

#

More videos flooded the online world, muddying fiction and reality. A small screen in a corner of Harry's D.C. office showed newsreels of missile test launches, gesticulating politicians, protests in the streets of New York, Shanghai, Taipei, Moscow, Berlin. Even Harry didn't know which were real and which were fabrications.

As soon as Harry ended a call with Professor Guo, who confirmed that NASA were cooperating, his phone rang again. It was Eliza.

"The posts spring up faster than we can pressure the platforms to take them down, and every time one does get deleted the mob cries government censorship," she said. "And similar videos are breeding agitation on Chinese social media, too, which we can't do anything about at all. But there's progress on another front."

"Yes?" Harry latched on to the glimmer of optimism.

"The original source. The President video was first released publicly by a bot, naturally, but one of the deepfake companies we contacted found the original video prompt

on their servers. Their program's output was a fairly obviously AI clip – nothing groundbreaking, but likely the foundation of the video we've all seen."

"And who submitted the prompt?"

"Another bot. Source untraceable. But fortunately for us, this deepfake company doesn't quite play by the rules, and had a tracker placed in their product code, to monitor copyright infringement apparently. The tracker points to a computer in St Louis."

"Great! So you have a team heading over there?"

"I certainly do. And you'll be with them."

#

Though Harry hated to admit it, Eliza's logic made sense: as 'Truth Guru', he had become something of a public figure, and his presence at the debunking of the President video could help boost legitimacy. Still, as the plane began its stomach-lurching descent, he suspected Eliza's motives had an edge of spite; surely the CIA's file on him had the words 'Fear of Flying' printed large.

He wobbled off the plane to find a CIA van waiting on the runway, occupied by three agents and a couple of carefully-vetted journalists. The ensuing drive was quiet. Harry wondered what the agents knew. Probably more than him; Eliza had relayed the bare minimum.

When they reached the unassuming suburban house, Harry and one agent ventured forth from the van. He noted the journalists discreetly filming from the van's rear window. Perhaps they were hoping for a booby trap.

The agent rang the doorbell. Moments later the front door was opened by an elderly woman, who took a startled step back at the sight of the duo. The agent flashed his badge; Harry gestured lamely to his Capitol lanyard. The woman didn't look very criminal-like, cyber or otherwise.

"You're Arno Jackson's grandmother?" asked the agent.

The woman nodded warily.

"May we search the property? Beginning with Arno's room."

The old woman stood aside, rubbing her knuckles fretfully, and gestured upstairs. Harry gave her a cheerless smile as he passed her in the hallway.

The bedroom bore the hallmarks of an occupant who had rarely left it. The smell was stale and salty, and the cloudy window permitted little daylight. A single bed huddled behind the technological monstrosity that dominated the space. Six screens loomed over a worn mechanical keyboard. Below the desk, green LEDs illuminated a hulking black cuboid.

The agent approached the computer. "What do you think, Doc? Capable of creating that President video?"

“Not from scratch, but powerful enough to modify or manipulate an existing video, make it look more believable,” said Harry. “Do we have a location for Arno?”

“He’s no longer with us.” The old woman hovered in the doorway. “He took his own life last week.”

Harry paused. A trifling detail that Eliza had decided not to pass on, it seemed.

“I’m sorry for your loss.”

Arno’s grandmother looked down at her shoes. The agent, seemingly oblivious to the grieving woman, began taking photographs of several large textbooks on neural networks. Harry looked around, feeling he should help or at least distract the old lady somehow, and picked up a framed photograph from a shelf.

“That’s his ex-girlfriend,” Arno’s grandmother said. “It was an online relationship. Very common these days, he said. She ended things with him recently, which maybe was why he…” She trailed off.

“Where’s she from?”

“Beijing, I think.” She looked imploringly at Harry. “What’s this about? Did Arno do something?”

Harry saw the agent retrieve a smashed-up electronic component from the wastepaper basket and give a signal; time to go.

“Sorry, ma’am. We’ll be in touch.”

They exited the house, laden with hardware. The cameras clicked and flashed.

#

“So, the leading theory is that this social recluse is engineering nuclear war – from beyond the grave, no less – to get revenge on his Chinese ex?” Doug was understandably incredulous.

“Seems so,” said Harry. The taskforce of five had gathered again in the stuffy briefing room. “His grandmother confirmed that he’s always been a tech head, with the brains to match. He’d also emptied her retirement fund to purchase some very high-spec components.”

“But even so, is it really possible that one guy built this from his bedroom?”

“With some clever resourcing, yes,” said Professor Guo. “The data recovered from the damaged hard drives, though incomplete, gave us an idea of his process. He essentially cobbled together open-source algorithms, alongside some ingenious additions of his own, to build two programs. One takes existing deepfake materials and makes them vastly more convincing, and the other disseminates these next-gen deepfakes online at a phenomenal rate.”

“And he’s definitely dead?” Annette asked. “He’s not hiding in a storm drain somewhere, plotting his next attack?”

“The body’s prints and DNA all match up,” said Eliza. “Presumably he wanted to avoid life in prison. But the algorithms are self-propagating. They can’t be contained.”

“But they can be countered,” Professor Guo said, smiling faintly. “We’ve gathered the ingredients we need and development has been swift. We can release an early version of the program tomorrow.”

Doug grunted. “Good. The President has an executive order ready to go this evening. In short, it demands that any organization that broadcasts information needs to integrate this program or face a nationwide ban. Radical, but necessary.”

“And China?” asked Eliza.

“They’re talking to us again, and have even agreed to use the algorithm on their own platforms, pending their own cybersecurity checks.”

“So we did it?” said Professor Guo. “We’ve potentially stopped nuclear war?”

“Don’t speak too soon, but it seems that way,” Harry replied. “We just need to be ready for next time.”

#

Far away, a dim, low-ceilinged room reverberated with a two-pitched hum. One tone emanated from banks of mainframe computers, the other from the air conditioning units that kept them cool. An operative leaned back in his chair, surveying the screens in front of him.

The operation ran itself these days. One program scoured the internet for likely targets: mostly solitary, chronically online, but highly intelligent young men with a fascination for AI. Thousands from the USA; the country had done such a terrific job of isolating and alienating its own people, the operative mused, that he sometimes wondered if it had been intentional.

The second program then molded these targets, nudging them towards their purpose. In a blink, AI-authored scripts created convincing forums on whatever topics best galvanized the targets: fascism, misanthropism, and anarchism were all popular. Bots messaged the targets on social media and lured them in. There, forum posts revealed the destructive potential of AI, shared resources for building potent algorithms.

The final program, the operative’s favorite, fabricated a romance. A kindred spirit met online. Appearance and personality drawn from the target’s search history. Foreign, preferably; the AI chatbots still weren’t perfect, but any confusing outputs could be attributed to the language barrier. Through the power of love, push the target’s half-baked ideology deeper. Encourage them to use their formidable intelligence to create something incredible, something world-changing.

Then, switch off the love. Tell them they will never amount to anything. Will never change the world. Push them over the edge. Let chaos unfold. Delete all evidence, target included.

Arno Jackson was the first major success, but there would be many more. The algorithms improved themselves every time. Though others might create counter-algorithms, one factor remained infinitely vulnerable and pliable: the human mind.

The operative turned off the monitors and headed for the door. The two-tone hum continued unabated.