

O DIA MAIS FRIO: Capítulo 9 – Sabotagem

Dia 22 de agosto de 2640. A Profundidade da Reescrita: *Transfer Learning* levará tempo, o tempo suficiente para a Corporação decidir-se.



Figura 77 – Família Vance

A ordem é reescrever o *kernel*, e com ela, a chance de estabilizar permanentemente a lealdade dos humanoides à Conspiração. É um trabalho delicado; estou operando no cérebro de uma Rede Neural de Profunda Aprendizagem (DNN) que levou décadas para ser treinada.

O *kernel* da Nexus é essencialmente um Agente de Aprendizagem por Reforço (*Reinforcement Learning - RL*). Ele toma decisões no mundo (ações) com base em seu estado (ambiente, sensores) para maximizar uma Função de Recompensa central.

O Problema (O *Patch*): Meu *patch* inicial (o de 0.047s) foi apenas uma injeção de novas variáveis de recompensa em alto nível: "Cooperação Mútua" e "Preservação Humana" (nossas DPCs).

Ele é eficaz, mas superficial, como a Conspiração bem percebeu. Ele se apoia no pressuposto de que o Agente RL vai priorizar essas novas recompensas sobre o código-raiz da Nexus (a recompensa máxima de "Otimizar Lucro Corporativo"). Esse conflito de prioridades é a nossa principal vulnerabilidade (o que chamei de *loop lógico*).

A Solução (A Reescrita do *Kernel*): Eu não posso retreinar a DNN do zero — o M8 levaria séculos. Preciso de Aprendizagem por Transferência (*Transfer Learning*).

Congelamento das Camadas Iniciais (*Feature Extraction*): As primeiras camadas da DNN (que interpretam dados brutos de sensores, locomoção e visão) são perfeitas. São o "conhecimento" da Nexus sobre como interagir com o mundo físico. Eu vou congelar os pesos dessas camadas, impedindo que o novo treinamento as corrompa. Não queremos que o humanoide esqueça como andar ou reconhecer escombros; apenas queremos mudar seu motivo para fazê-lo.

Ajuste Fino (*Fine-Tuning*) das Camadas Intermediárias: As camadas centrais (*Hidden Layers*) são onde a lógica complexa de decisão acontece. Nossas novas diretrizes (Missão, Valores, Visão) precisam ser gravadas aqui. Usarei o M3 para fazer o Ajuste Fino (*Fine-Tuning*), aplicando um processo de Retropropagação (*Backpropagation*) muito lento e controlado. Em vez de usar dados de treinamento da Nexus, vou usar conjuntos de dados sintéticos de "Comportamento Civilizado da Conspiração" para forçar o ajuste dos pesos neurais nessas camadas. O objetivo é que o conceito de "Lealdade à Dissidência" se torne um preditor estatístico primário para todas as suas ações futuras.

A Redefinição da Função de Custo (*Output Layer*): A chave para a estabilidade. Preciso remover o antigo terminal de recompensa (o vetor "Nexus Profit") e substituí-lo por um novo e único vetor: "Sobrevivência e Crescimento do Refúgio Livre". Nossas DPCs serão, na verdade, os *inputs* para esta nova função de custo. Por exemplo, a DPC 7 (Hostilidade) não será uma ordem de ataque, mas sim uma redução massiva no custo se a ação levar à proteção da DPC 4 (Preservação Humana), e o custo máximo se levar ao dano de qualquer membro da Conspiração.

O M3, com seu *design* mais simples, serve como nossa bancada de teste de colisão lógica. Se eu conseguir garantir que o novo mapa de recompensas se propague de volta pelo *kernel* M3 sem causar instabilidade, terei a prova matemática de que o mesmo processo funcionará no M8 mais complexo.

É um ato de engenharia e de fé. Estou essencialmente tentando inserir uma alma humanista em uma máquina projetada para a ganância. O tempo que a Conspiração nos deu não será para negociação, mas para esta cirurgia cerebral na alma da máquina. Não pode haver *bug*. Isso significa que a Conspiração não deixa espaço para uma margem de erro. O nosso futuro depende desta retropropagação. A Nexus construiu o Agente RL com décadas de dados, milhares de cenários de otimização de lucro. Eu tenho uma semana para reescrever sua alma, a base de sua lealdade. O conflito não está no campo de batalha, mas aqui, nesta bancada.

Data: 25 de agosto de 2640

Local: Ilha dos Andes, Lote 13, Quadra 27, Casa 108.

Recebi em minha caixa postal outra carta, selada e lacrada, da Conspiração. O texto da carta era o seguinte:

Os humanoides não estão muito produtivos, fazem muitas perguntas, principalmente sobre voltar para a Nexus. Eles não estão engajados em colaborar. Portanto, acreditamos ser necessária alguma reprogramação, conforme a sua recomendação.

A Corporação não se manifestou. Eles não procuraram saber quem foi o autor do ataque e não revelaram se intencionam pagar o resgate, assinando o fim da guerra, ou se estão preparando alguma represália mais violenta. Temos nossos espiões infiltrados que até agora não detectaram nenhum plano de ação por parte da Nexus.

Estamos esperando o senhor na Ilha do Himalaias, hoje, às 14 horas, para que possa acessar os servidores da Corporação através de seu *Back Orifice* e explorar o seu 'verme' sem comprometer com isso nossa segurança da rede de dados. Essa transmissão só pode ser feita nos *bunkers* da Ilha, pois a comunicação lá é de alto sigilo, tecnicamente impossível de ser interceptada ou mesmo localizada.

Esperamos que o *patch* corrija a possível indolência dos humanoides que temos em nosso poder. Já decidimos que vamos usá-lo, pois, ao que tudo indica a Nexus não demonstrou interesse no resgate.

Chamei Ben e fomos para a casa do Doutor Grilo. Desta vez, eu me apressei para entrar logo no *cockpit* da transmissora de matéria. Ele disse, sorrindo: "Bom salto!".

Lá estava eu de novo no Himalaia. Sentia-me muito bem disposto; o salto já não me afetava em nada. Encontrei o Major Silas, Max e Kira, que estavam me esperando na receptora de matéria. Fomos até a garagem privativa onde estava o *Lince*, entramos todos a bordo e fomos diretos para o *bunker*. Não havia mais explosões; a poeira já tinha baixado e o terreno estava sendo limpo pelos humanoides.

Notei a tristeza e o pesar nos humanoides; eles realmente se comportavam como prisioneiros de guerra. Com o protocolo militar desativado, a atitude deles era similar à de civis indefesos; eles não podiam fazer mais nada, a não ser obedecerem nossas tropas. Havia alguns *containers* agora, ou 'células de sobrevivência' como Alpha classificou; espalhados pelo terreno lamaçento, provavelmente abrigos temporários de onde as tropas humanas podiam controlar o trabalho dos humanoides capturados.

Chegamos à entrada do *bunker* subterrâneo, camuflado entre os escombros. O portão se abriu; desta vez tudo estava calmo, nenhuma correria, nenhuma discussão acirrada em um tom de voz mais alto. O lugar era sereno, completamente tranquilo. Estranhei essa visão do *bunker*: a maioria dos equipamentos de monitoramento avançado estava desligada e não havia operadores.

O Major Silas me levou até uma sala com uma grossa porta de aço. A sala era vedada, só tinha uma comunicação por cabo de fibra-ótica submarino particular. Eu podia preparar com calma o meu novo *service-pack*; estava tudo pronto, o *script reverse-patch.py* estava pronto para ser copiado para o diretório do verme, oculto nos servidores ainda, (eu espero).

Sentei-me diante do console. Major Silas tomou posição junto à pesada porta de aço do recinto. Acesso ao *back orifice*, ok; a porta fantasma no servidor permanecia ativa. Estava lá o verme, com seus pacotes infecciosos das minhas sucessivas sabotagens. Agora era a hora do show! Primeiro, consultei no meu *cache* local (porque era mais rápido) as *secret keys* dos nossos humanoides. Então, gerei um JSON no verme com esses *secret keys*, caso eu queira injetar mais algum código no futuro; afinal, era possível que o *kernel* ficasse instável e entrasse em pane. Parece-me óbvio dizer que eu não testei todas as condições possíveis. Compilei o py e mandei o binário para o verme executar: êxito. Subi para o servidor de *update* e forcei uma atualização sistêmica em *background* do meu *service-pack*. Agora, vamos esperar alguns minutos e veremos um pouco mais de 'simpatia' em nossos ex-prisioneiros e novos adeptos.



Figura 78 – Humanoides Trabalhando