A IA Proibida Vazou e os Governos Perderam o Controlo

written by Mário Portela | 28 de Abril, 2026

Ah, foi 25 de Abril. Dia da Liberdade em Portugal. Há uma ironia quase poética no facto de estarmos a celebrar a queda das amarras e a devolução do poder ao povo numa semana em que, a nível global, a narrativa dominante nos corredores do poder é uma só:

“Os governos perderam o controlo da Inteligência Artificial”.

Como costumo discutir no podcast «IA&EU», **o medo infundado é tão tóxico como o *hype* desmedido.** Mas confesso que os últimos sete dias testaram os limites do meu pragmatismo. Entre modelos de IA classificados como “demasiado perigosos para existirem”, guerras de versões lançadas à pressa e reguladores a agirem como dinossauros a tentar legislar meteoros, a semana que passou provou que o fosso entre quem cria a tecnologia e quem a tenta domar se tornou um abismo intransponível.

Sirvam-se de um café forte *(vão precisar para a azia digital que se avizinha)* e vamos desembrulhar este caos. Porque, no meio desta tempestade geopolítica e tecnológica, há lições vitais sobre como deves proteger o teu negócio e a tua sanidade.

O "Modelo Proibido" e a Caixa de Pandora da Anthropic

A grande bomba atómica da semana *(embora já se ouviam zumzuns antes)* atende pelo nome de **Mythos**, o novo modelo super-secreto da Anthropic. Se bem se lembram, já tínhamos falado nas semanas anteriores sobre o *Project Glasswing*, a iniciativa da empresa para avisar as grandes tecnológicas de que tinham criado uma IA capaz de encontrar falhas de cibersegurança críticas.

Pois bem, a coisa escalou. O Mythos foi oficialmente rotulado nos bastidores como o "modelo proibido". É um sistema com capacidades agênticas tão avançadas que consegue navegar autonomamente pela internet, analisar infraestruturas e encontrar *zero-days* (vulnerabilidades até então desconhecidas) em sistemas operativos, bases de dados governamentais e redes corporativas. Em teoria, a Anthropic criou isto para fortalecer as nossas defesas. Na prática, construíram a chave-mestra capaz de abrir qualquer cofre digital no planeta.

O problema? A caixa forte não aguentou. Surgiram esta semana múltiplos relatos e fugas de informação sugerindo que **já houve acesso não autorizado ao modelo**.

Isto é o equivalente digital de deixar os códigos nucleares esquecidos no balcão de um café. A Anthropic, que construiu toda a sua reputação a vender a imagem de "IA Segura e Constitucional", enfrenta a sua maior crise de sempre. O incidente ateou fogo ao debate mais feroz do momento: *Open-source vs. Closed-source*.

Os defensores do código fechado dizem: *"Vêem? Isto é demasiado perigoso para andar aí solto na rua*". Os defensores do código aberto (onde a China tem apostado as fichas todas) ripostam: *"Se o código fosse aberto, milhares de olhos já teriam encontrado os buracos na segurança antes dos hackers"*. A verdade? Estão ambos certos e ambos errados. Fechar um modelo não garante segurança (como o vazamento do Mythos prova), mas abri-lo indiscriminadamente quando tem capacidades destrutivas é suicídio colectivo. Para o utilizador comum e para as empresas, a lição é brutal: não confiem a 100% nas plataformas. Se o Mythos está aí a farejar, a vossa higiene de cibersegurança básica (aquelas actualizações chatas que andam a adiar) nunca foi tão crítica.

A Guerra dos Agentes: GPT-5.5 vs. Opus 4.7

Enquanto a Anthropic lida com apocalipses de segurança, a **OpenAI** fez o que o Sam Altman faz melhor: atirou uma distracção brilhante para o ar. Para combater a tração mediática do *Claude Opus 4.7 (que estava a dominar os benchmarks de produtividade)*, a OpenAI decidiu antecipar o lançamento do **GPT-5.5**.

A fasquia subiu vertiginosamente. Já não estamos a falar de modelos que escrevem *e-mails* bonitos; estamos a falar da maturação da **Agentic AI** (IA Agêntica). O GPT-5.5 não responde apenas às tuas perguntas; ele planeia, executa, corrige erros e orquestra outros pequenos agentes para cumprirem tarefas complexas de ponta a ponta.

Mas vamos ser frios e cínicos por um momento: este lançamento tresanda a pânico corporativo. A OpenAI percebeu que as empresas estão a migrar para o Claude no que toca a fluxos de trabalho reais, e precisava de estancar a sangria. O GPT-5.5 é formidável, mas está cheio de arestas por limar. As alucinações diminuíram, mas a tendência para os agentes

entrarem em *loops* infinitos quando não percebem uma instrução mantém-se.

No final do dia, esta guerra de boxe entre o GPT-5.5 e o Opus 4.7 é excelente para nós, os consumidores, porque empurra os preços para baixo e a inovação para cima. No entanto, recuso-me a alinhar no delírio. Como mentor empresarial de IA, o conselho que dou é sempre o mesmo: não mudem toda a infraestrutura da vossa empresa a cada atualização de versão. Sejam pragmáticos. Usem a ferramenta que vos resolve o problema de hoje, e não a que vos promete a utopia de amanhã.

O Pânico Geopolítico: Dinossauros a Legislar Meteoros

Para rematar a insanidade, temos de olhar para Washington, Bruxelas e Pequim. O sentimento generalizado de que *"os governos perderam o controlo da IA"* motivou uma reacção alérgica e atabalhoada dos legisladores.

Os Estados Unidos apertaram ainda mais os controlos de exportação de chips de alta performance para a China, e a Europa, como habitual, seguiu o rebanho com as suas directivas pesadas e lentas *(uma vez mais a perder dinheiro em nome dos senhores do ocidente)*. Há listas negras (*blacklists*) a voar de um lado para o outro. A própria Anthropic continua sob o olhar desconfiado do Pentágono, num jogo de gato e rato sobre até que ponto a IA militar pode ser "ética". Aliás, os rumores de que sistemas de IA agêntica estão a ser ativamente usados em operações táticas (como alvos de *drones* e guerra cibernética) deixaram de ser rumores para passarem a segredos mal guardados.

Mas a tentativa de travar a China retendo peças de *hardware* é como tentar apanhar o vento com uma rede de borboletas. A Ásia respondeu abraçando o *open-source* com ferocidade, desenvolvendo modelos hiper-eficientes que correm em *hardware* mais antigo ou em infraestruturas locais descentralizadas.

A tragédia cómica disto tudo é assistir a comités governamentais repletos de políticos que precisam de ajuda para ligar o *Wi-Fi* do telemóvel, a tentarem redigir regulamentações sobre "Agentes Autónomos Auto-Replicáveis" ou "Alinhamento Estocástico". A regulação está a ser feita por quem não percebe minimamente de tecnologia, o que invariavelmente resulta em leis que estrangulam a inovação local e entregam a liderança de bandeja a blocos económicos que não jogam com as mesmas regras morais.

Conclusão: Quem Segura o Leme?

Esta semana em que celebramos a liberdade por aqui, o panorama tecnológico devolve-nos uma reflexão dura: **somos livres para usar as melhores ferramentas já criadas pela humanidade, mas estamos perigosamente reféns de quem as controla.**

A Anthropic criou um monstro e deixou-o destrancado. A OpenAI continua a lançar software não testado apenas para esmagar a concorrência. E os governos, em vez de educarem a população, reagem com proibições inúteis e pânico geopolítico.

A IA não vai destruir o mundo por si só, mas a incompetência humana que a rodeia está a fazer um esforço admirável para ajudar. A vossa única defesa? **Literacia tecnológica.** Compreender o que a IA faz, o que não faz, e onde residem os vossos dados. A IA é uma ferramenta poderosa SE souberes usar. E se não a souberes usar, prepara-te para ser ultrapassado por quem sabe.

Se gostas deste tipo de análise sem tretas, que separa o ruído da realidade sem cair em dramatismos de Hollywood, subscreve esta *newsletter*… ou acompanha as minhas crónicas no jornal "O CIDADÃO". Aqui analisa-se a tecnologia com a frieza que ela exige, e com o humanismo que ela não tem.

Descobre como integrar a tecnologia no teu dia-a-dia a sério, e não apenas para fazer imagens engraçadas de cães a andar de

bicicleta.
Até para a semana, e vão lá mudar a vossa palavra-passe. Nunca se sabe se o Mythos não está à escuta.

Artigo publicado simultaneamente n' O Cidadão e no substack do autor