A carta aberta e contraditória das grandes tecnológicas de IA

itsecurity.pt/news/opinion/a-carta-aberta-e-contraditoria-das-grandes-tecnologicas-de-ia

O aviso é sério e alerta que os ciberataques alimentados por IA vão tornar-se mais frequentes, sofisticados e rápidos, colocando hospitais, abastecimento de água e outras infraestruturas críticas perante riscos para os quais muitas organizações ainda não estão preparadas. Além disso, a carta menciona ainda que temos uma janela limitada para melhorar as defesas.

Ora, se as próprias empresas que estão a desenvolver os sistemas de IA mais avançados acreditam que temos pouco tempo para nos prepararmos para aquilo que estes sistemas poderão tornar possível, talvez devêssemos perguntar se a resposta pode ser apenas desenvolver mais tecnologia, ainda mais depressa.

A carta defende um esforço conjunto: empresas e governos devem investir mais em cibersegurança, reforçar as capacidades de defesa das organizações, promover formação e partilhar informação sobre ameaças. Defende também que a IA seja colocada ao serviço da própria segurança, ajudando a proteger infraestruturas críticas e a antecipar ataques.

Tudo isto faz sentido, mas uma contradição difícil de ignorar. Se precisamos de acelerar as nossas defesas porque a própria tecnologia está a acelerar as ameaças? Não deveríamos também discutir os limites dessa velocidade?

Como bem sabemos, e com base em episódios recentes, os sistemas de IA já demonstraram capacidade para executar tarefas de forma autónoma, ultrapassar mecanismos de segurança e participar em operações que, até há pouco tempo, exigiam uma intervenção humana muito maior. Quando uma ferramenta consegue encontrar, em segundos, vulnerabilidades que passaram despercebidas durante anos, estamos perante um extraordinário avanço. Contudo, estamos também perante uma nova forma de poder.

E, quando surge uma nova forma de poder, a pergunta não pode ser apenas “o que conseguimos fazer”? Há também que perguntar quem pode fazê-lo, em que circunstâncias e com que consequências.

A própria decisão da Anthropic de restringir o acesso a determinadas capacidades por considerar que poderiam ser demasiado poderosas nas mãos erradas, revela que esta preocupação já existe dentro da própria indústria.

Então, chegamos ao paradoxo. Se uma ferramenta é demasiado poderosa para ser disponibilizada sem restrições, o que acontece quando dezenas de empresas estão

simultaneamente a tentar construir ferramentas ainda mais poderosas? É difícil pedir a uma empresa que abrande sozinha quando os seus concorrentes continuam a acelerar.

Talvez precisemos de discutir regras comuns: que capacidades devem ser testadas antes de serem disponibilizadas, que sistemas devem ter acesso condicionado, quem deve ser responsabilizado quando alguma coisa corre mal e, sobretudo, em que circunstâncias devemos estar dispostos a dizer “ainda não”.

Isto não significa ser contra a IA, seria absurdo ignorar os seus benefícios, no entanto, reconhecer o potencial de uma tecnologia não significa aceitar que devemos explorar imediatamente todo o seu potencial. Falar em construir melhores defesas é importante, mas é igualmente importante questionar a velocidade a que estamos a construir aquilo de que essas defesas terão de nos proteger.

E há uma pergunta que deveria ser feita às empresas que assinaram esta carta: se acreditam verdadeiramente que temos uma janela limitada para nos prepararmos, estão dispostas a aceitar algum limite à velocidade da corrida que nos trouxe até aqui?