Deepfakes empresariais: o CEO pediu-me 25 milhões

digitalinside.sapo.pt/deepfakes-empresariais-o-ceo-pediu-me-25-milhoes

Bruno Castro

A utilização de deepfakes em ciberataques representa hoje uma das maiores ameaças às organizações, em particular, quando o objetivo é o de impersonificar líderes empresariais. As ferramentas de inteligência artificial generativa permitem criar vídeos, áudios e até chamadas em tempo real que imitam com realismo a voz, a imagem e, em alguns casos mais sofisticados, as especificidades de comunicação de executivos de topo. O resultado é um campo fértil para fraudes financeiras, manipulação de equipas e roubo de informação sensível, impulsionado pela crescente acessibilidade das ferramentas de criação de deepfakes, disponíveis a baixo custo e de fácil acesso.

Nos últimos anos, têm-se multiplicado os casos em que colaboradores receberam instruções falsas, aparentemente vindas de CEO's ou diretores financeiros, para efetuar transferências urgentes ou partilhar dados estratégicos. Em várias situações, estas mensagens foram acompanhadas de deepfakes audiovisuais que reforçaram a credibilidade do pedido, levando a perdas que, em alguns casos, ultrapassaram centenas de milhões de euros. Esta tendência demonstra que não estamos perante um risco futurista, mas sim uma realidade concreta, já a afetar empresas de diferentes dimensões e setores.

Segundo o The Wall Street Journal, em 2024 foram reportados mais de 105 000 ataques relacionados com deepfakes nos EUA, representando um ataque de deepfake a cada cinco minutos, com prejuízos financeiros superiores a 200 milhões de dólares, apenas no primeiro trimestre desse ano. Empresas prestigiadas como a Ferrari, Wiz, WPP e a consultora britânica de engenharia Arup foram também alvos destes esquemas cada vez mais sofisticados. Uma situação particularmente dramática foi o caso da Arup que perdeu 25 milhões de dólares, após um colaborador do escritório de Hong Kong, convencido por uma reunião virtual que incluía deepfakes realistas dos seus executivos, cumprir as instruções dadas pelos criminosos.

Estes casos demonstram que não se trata de um cenário distópico: são ameaças atuais, concretas e com impacto elevado e cuja eficácia se baseia na manipulação de três vetores psicológicos fundamentais: autoridade, urgência e confiança. A familiaridade com a voz ou imagem falsificada é suficiente para quebrar resistências individuais e abrir caminho a ações impulsivas. Quando uma voz idêntica à do CEO dá uma instrução direta, a predisposição natural para agir quase de imediato é muito elevada. É neste ponto que a cibersegurança da modernidade deve ser complementada com medidas humanas e processuais.

Entre as principais estratégias de mitigação destes riscos destacam-se a adoção de políticas de autenticação multifatorial para qualquer pedido sensível, a criação de protocolos internos claros para validação de transferências e o reforço da literacia digital

das equipas, incluindo formação para identificar sinais subtis de manipulação digital. Paralelamente, os centros de operações de segurança (SOC) e a crescente adoção de soluções de strategic intelligence assumem um papel preponderante na monitorização de padrões suspeitos e na deteção precoce de conteúdos manipulados.

À medida que os deepfakes se tornam mais acessíveis e realistas, a capacidade de resposta das organizações depende da conjugação entre tecnologia, processos e cultura interna. A proteção contra os esquemas de deepfakes de líderes empresariais não é somente uma questão técnica, mas sim, uma questão de confiança institucional e de resiliência organizacional. Num contexto em que a reputação e os ativos financeiros podem ser gravemente comprometidos por um único ataque bem-sucedido, investir em cibersegurança e preparar as equipas para este novo tipo de ameaça é uma prioridade estratégica inadiável.