
"DEEPFAKES"
E O COLAPSO DA AUTENTICIDADE


"DEEPFAKES"
E O COLAPSO DA AUTENTICIDADE



Vivemos num momento de viragem na democratização da IA generativa. Apesar de todo o bem que a tecnologia pode trazer, a verdade é que os deepfakes vão ser mais comuns e parecer cada vez mais reais. A mesma tecnologia que promete transformar a produção de conteúdo pode ser utilizada por cibercriminosos para ataques cada vez mais sofisticados.

RUI DAMIÃO

O LANÇAMENTO do Google Veo 3, em maio de 2025, marca potencialmente um momento de viragem na democratização da Inteligência Artificial (IA) generativa. Esta ferramenta, disponível por 249 dólares mensais, permite a qualquer utilizador criar vídeos com áudio nativo de qualidade cinematográfica – incluindo diálogos de personagens, efeitos sonoros e música ambiente – a partir de simples *prompts* de texto.

Com capacidades que "excedem em física, realismo e aderência aos *prompts*", como menciona a própria Google, o Veo 3 representa um salto qualitativo que coloca ferramentas outrora exclusivas de estúdios profissionais nas mãos de milhões de utilizadores.

Esta revolução criativa traz uma realidade mais sombria: a mesma tecnologia que promete transformar a produção de conteúdo está a ser apropriada por cibercriminosos para ataques cada vez mais sofisticados contra as organizações.

Os deepfakes corporativos já não são uma ameaça teórica. Em fevereiro de 2024, a empresa britânica Arup perdeu 25 milhões de dólares numa videoconferência onde todos os participantes, incluindo o diretor financeiro, eram falsificações criadas por inteligência artificial. Meses antes, a WPP foi alvo de um ataque similar: cibercriminosos criaram uma conta WhatsApp falsa do CEO Mark Read, marcaram uma videochamada e utilizaram clones de voz e imagens manipuladas obtidas do YouTube para enganar os colaboradores.

Estes casos, longe de serem isolados, fazem parte de uma tendência. De acordo com o relatório

IM. | Copyright licensed by Visapress | P. 3 of 17



Global Threat Report da CrowdStrike, os ataques de vishing aumentaram 442% entre a primeira e a segunda metade de 2024, impulsionados precisamente por ferramentas de IA generativa.

A evolução técnica dos deepfakes nos últimos dois anos é "substancial", nas palavras de Bruno Castro, Fundador & CEO da VisionWare. Se em 2023 ainda era possível detetar falhas visuais como movimentos labiais pouco naturais ou iluminação inconsistente, hoje esses sinais tornaram-se "cada vez mais raros e subtis".

Um relatório de setembro de 2024 da Security.org revela que mais de 10% das empresas já enfrentaram tentativas de fraude com deepfakes, com danos que chegaram a representar 10% dos lucros anuais das organizações afetadas. Ainda mais alarmante, 80% das empresas não possuem protocolos para lidar com estes ataques e apenas 5% dos líderes empresariais afirmam ter prevenção abrangente contra deepfakes.



- Bruno Castro -
Fundador & CEO da VisionWare

Para as organizações, e à semelhança dos ciberataques 'tradicionais', esta não é uma questão de "se" vai acontecer, mas de "quando". O vazio legal – onde não existe legislação específica em Portugal para deepfakes – combina-se com uma crescente sofisticação dos ataques para criar um cenário de vulnerabilidade crítica. As organizações que não se prepararem agora, enfrentam não só perdas financeiras diretas, mas danos reputacionais potencialmente duradouros numa era onde a distinção entre real e artificial se esbate a cada dia.

> "OS VÍDEOS E ÁUDIOS DEEPFAKES SÃO AGORA MAIS CONVINCENTES, NATURAIS E DIFÍCEIS DE DISTINGUIR DE CONTEÚDOS AUTÊNTICOS, MESMO POR OBSERVADORES MAIS ATENTOS"
>
> BRUNO CASTRO, FUNDADOR & CEO DA VISIONWARE



A EVOLUÇÃO SILENCIOSA DE UMA AMEAÇA REAL

Bruno Castro é direto na sua avaliação: "a diferença de qualidade entre os deepfakes de há dois anos e os atuais é substancial". O especialista em cibersegurança e investigação forense explica que, "em 2023, muitos deepfakes apresentavam ainda falhas visuais subtis, como movimentos labiais pouco naturais, piscadelas descoordenadas ou iluminação inconsistente. Atualmente, os avanços nos modelos generativos tornaram essas falhas cada vez mais raras e/ou subtis".

Esta evolução técnica representa uma mudança de paradigma fundamental. "Os vídeos e áudios deepfakes são agora mais convincentes, naturais e difíceis de distinguir de conteúdos autênticos, mesmo por observadores mais atentos", refere Bruno Castro. Assim, as organizações já não podem confiar na capacidade humana de deteção como primeira linha de defesa.

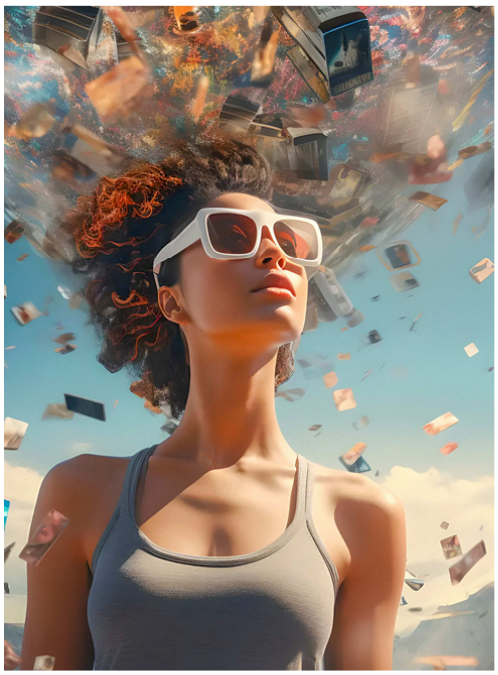
A velocidade desta evolução surpreende, também, os próprios especialistas. Onde antes era necessário equipamento sofisticado e conhecimentos técnicos avançados, hoje qualquer utilizador com acesso a ferramentas como o Veo3 pode criar conteúdo manipulado de qualidade cinematográfica. Esta democratização tecnológica transformou os deepfakes de uma curiosidade técnica numa arma acessível a qualquer atacante com motivação.

Os tipos de deepfakes corporativos mais comuns mostram uma clara evolução tática dos cibercriminosos. Bruno Castro identifica três categorias principais: "vídeos falsificados de executivos a anunciar decisões empresariais; manipulações de voz em chamadas telefónicas que instruem transferências bancárias fraudulentas; e ainda conteúdos que visam desinformar ou influenciar o valor de ações na bolsa".

O caso mais sofisticado documentado até à data envolveu a manipulação da identidade de Mark Read, CEO da WPP, como mencionado. Este ataque revela uma metodologia refinada dos cibercriminosos: começam por recolher material público – como fotografias, vídeos do YouTube e entrevistas –, criam perfis falsos em plataformas de comunicação empresarial e utilizam tecnologia de clonagem de voz para conduzir reuniões convincentes. A sofisticação técnica combina-se com engenharia social para criar ataques que exploram tanto as vulnerabilidades tecnológicas, como as humanas.

A crescente frequência destes ataques demonstra que os cibercriminosos já dominaram o processo. "Nos últimos anos, os deepfakes corporativos têm vindo a tornar-se numa ameaça crescente para empresas de diversos setores", alerta o Fundador & CEO da VisionWare, acrescentando que estas não são ocorrências isoladas, mas sim parte de uma tendência estrutural.





O FIM DOS VESTÍGIOS TÉCNICOS

A realidade técnica da deteção de deepfakes é mais complexa do que muitos líderes empresariais assumem. Bruno Castro explica que "ainda existem vestígios técnicos que permitem identificar deepfakes", incluindo "artefactos visuais microscópicos, inconsistências de iluminação e de sombreamento, ausência de batimentos cardíacos visíveis em áreas faciais, ou falhas na sincronização entre áudio e imagem".

No entanto, a utilidade prática destes sinais está a diminuir rapidamente. "Com a rápida evolução da tecnologia, muitos destes sinais estão a tornar-se cada vez mais difíceis de detetar, especialmente quando os conteúdos são criados com ferramentas avançadas e posteriormente sujeitos a pós-produção profissional", alerta Bruno Castro, da VisionWare.

Mais preocupante é a limitação dos sistemas de deteção automática existentes que "dependem principalmente das assinaturas/classificadores de deteção de conteúdo deepfake. Contudo, estes classificadores têm limitações que podem ser facilmente contornadas", refere Bruno Castro. "Por norma, a maioria dos modelos só consegue identificar os 'cheapfakes'", alerta.

Deste modo, as organizações não podem depender exclusivamente de soluções tecnológicas de deteção, uma vez que existe uma janela de oportunidade para os atacantes. Enquanto os deepfakes de baixa qualidade são facilmente identificados, os conteúdos sofisticados – aqueles utilizados em ataques corporativos de alto valor – passam despercebidos aos sistemas de defesa tradicionais.

O LABIRINTO LEGAL

Ricardo Henriques, Sócio da Abreu Advogados e Co-Coordenador da Área de Prática de Propriedade Intelectual e Tecnologias de Informação, é claro sobre o estado atual da legis-

IM. | Copyright licensed by Visapress | P. 6 of 17



lação portuguesa sobre este ponto: "à data, a ordem jurídica portuguesa não possui uma lei que regule especificamente os deepfakes, sem prejuízo, naturalmente, das referências a este fenómeno no Regulamento de Inteligência Artificial", nomeadamente no que diz respeito às "falsificações profundas", ou até "no Regulamento de Serviços Digitais no contexto de avaliações de risco".

Daniel Reis, Sócio da DLA Piper, oferece uma perspetiva complementar e explica que "as deepfakes deram origem a novas formas de criminalidade através de meios informáticos, carecendo, todavia, ainda de um enquadramento legal em matéria penal óbvio em Portugal". Esta lacuna regulatória coloca as organizações numa posição particularmente vulnerável, onde os danos podem ocorrer sem que existam mecanismos legais específicos de resposta.

Contudo, ambos os especialistas concordam que esta ausência de legislação específica não significa uma falta de proteção total. Ricardo Henriques esclarece que "existem já hoje normas legais no domínio do direito civil, como sucede em relação à responsabilidade por ato ilícito, ou penal, como o caso do crime de burla".



- Daniel Reis -
Sócio da DLA Piper

A abordagem atual baseia-se numa adaptação da legislação existente. Daniel Reis detalha que

> A ATIVIDADE DE PRODUÇÃO DE CONTEÚDO DEEPFAKE COM BASE EM DADOS DE ENTRADA ILÍCITOS (...) PODERÁ ENCONTRAR ENQUADRAMENTO LEGAL ENQUANTO CRIME DE FALSIDADE INFORMÁTICA, PREVISTO NA LEI DO CIBERCRIME"
>
> DANIEL REIS, SÓCIO DA DLA PIPER



"a atividade de produção de conteúdo deepfake com base em dados de entrada ilícitos (não existindo consentimento para a utilização de determinada imagem ou vídeo) poderá encontrar enquadramento legal enquanto crime de falsidade informática, previsto na Lei do Cibercrime, ou crime de desvio de dados agravado à luz da Lei de Execução do RGPD".

A questão da responsabilidade em casos de deepfakes corporativos revela-se mais complexa do que pode aparentar inicialmente. Ricardo Henriques explica o princípio geral e diz que, "salvo nos casos de responsabilidade objetiva, aquele que gera e dissemina o deepfake será, em regra, o responsável civil e/ou penal pelos danos causados a terceiros. Isto é, exigindo a prática de um ato ilícito de forma dolosa que provoca danos a terceiros, o seu agente terá a obrigação de indemnizar o lesado".

Porém, a responsabilidade não se limita aos perpetradores diretos e Ricardo Henriques refere que "tal não afasta o risco de a organização poder também ser responsabilidade, por exemplo, através da inexistência ou da insuficiência de medidas de cibersegurança para fazer face à crescente ascensão e sofisticação destes conteúdos maliciosos".

Daniel Reis confirma esta perspetiva multifacetada. "Na eventualidade de uma deepfake do CEO da uma empresa causar danos a terceiros, a responsabilidade por esses danos poderá recair sobre diferentes intervenientes dependendo das circunstâncias do caso em concreto: entre outros, poderá ser responsabilizado o autor da deepfake, quem divulga a deepfake ou mesmo a própria empresa no caso de negligência, caso não tenha implementado medidas adequadas para prevenir ou mitigar a disseminação dessas imagens ou vídeos manipulados".

Esta realidade legal cria um imperativo claro para as organizações: não basta ser vítima e é necessário demonstrar que foram tomadas medidas adequadas de prevenção. Ricardo Henriques refere que "deverão ser adotadas medidas de deteção e resposta a incidentes provocados por deepfakes como, por exemplo, mecanismos de deteção para identificar conteúdos deepfake e distingui-los dos reais, tal como devem ser adotados mecanismos de autenticação e verificação do conteúdo produzido ou partilhado".

"A IMAGEM CONSTITUI UM DADO PESSOAL"

A proteção de dados pessoais introduz uma camada adicional de complexidade legal nos casos de deepfakes. Catarina Mascarenhas, Consultora da Abreu Advogados, relembra que "o RGPD não foi concebido para regular diretamente os deepfakes", mas "aplica-se, de facto, a vários aspetos da



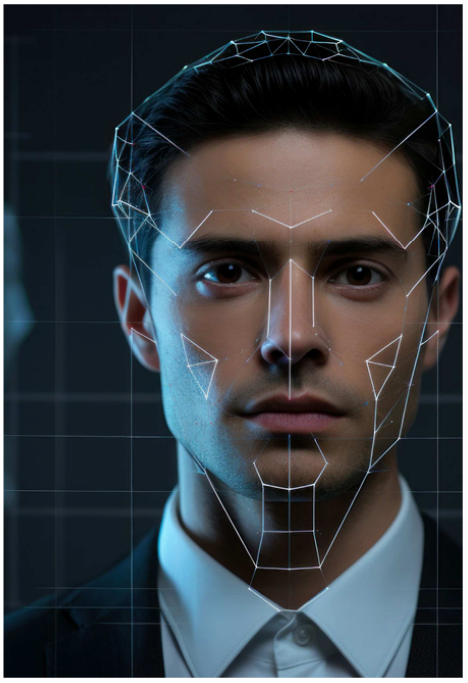
sua criação e utilização, especialmente quando envolvem o tratamento de dados pessoais".

"A imagem do executivo constitui um dado pessoal e, como tal, a sua utilização está abrangida pelo âmbito de aplicação do RGPD", explica Catarina Mascarenhas. Assim, "a utilização não consentida ou sem outro fundamento de licitude de tratamento, como a execução de um contrato ou o prosseguimento de um interesse legítimo da empresa, constitui um tratamento ilícito e, consequentemente, uma violação do regime de proteção de dados pessoais".

Daniel Reis reforça esta perspetiva e refere que "tanto o direito da União Europeia como a Constituição e o Código Civil consagram o direito à imagem. Em particular, o Código Civil prevê que o retrato de uma pessoa não pode ser exposto, reproduzido ou lançado no comércio sem o consentimento da mesma".

Já no que às consequências legais diz respeito, Catarina Mascarenhas alerta que, "no caso de deepfakes, à partida, e se tiver um intuito malicioso, além das consequências penais, civis e outras medidas sancionatórias, pode também originar uma contraordenação por violação de dados pessoais".

Uma questão particularmente relevante para as organizações prende-se com a proteção da imagem dos executivos e, neste ponto, Daniel Reis reforça que "a imagem e a voz são direitos das pessoas singulares, não podem ser propriedade de uma empresa. As pessoas singulares poderão dar o seu consentimento para a utilização da sua imagem e voz pela empresa". Ricardo Henriques acrescenta que a imagem e voz, nos termos legais, "são à partida irrenunciáveis e pertencentes ao seu titular, que, no entanto, pode autorizar o seu uso em determinados contextos".





O tempo é um fator crítico na resposta legal aos deepfakes corporativos. Catarina Mascarenhas salienta que "poderão existir diferentes prazos a considerar pelas organizações" que dependem "do resultado do uso indevido de um deepfake". Para crimes particulares, o prazo é restritivo: "caso esteja em causa um ilícito criminal e o crime aplicável seja um crime particular (*i.e.* um crime que dependa de queixa, como por exemplo o crime de ofensa a organismo, serviço ou pessoa coletiva) o direito de queixa extingue-se no prazo de seis meses a contar da data em que o titular tiver tido conhecimento do facto e dos seus autores", diz Catarina Mascarenhas.

Já no âmbito civil, "estando em causa o direito de indemnização, este prescreve, regra geral, no prazo de três anos a contar da data em que o lesado tenha dito conhecimento do direito que lhe compete", acrescenta a Consultora da Abreu Advogados.



- Ricardo Henriques -
Sócio da Abreu Advogado



- Catarina Mascarenhas -
Consultora da Abreu Advogados

> DEVERÃO SER ADOTADAS MEDIDAS DE DETEÇÃO E RESPOSTA A INCIDENTES PROVOCADOS POR DEEPFAKES COMO, POR EXEMPLO, MECANISMOS DE DETEÇÃO PARA IDENTIFICAR CONTEÚDOS DEEPFAKE E DISTINGUI-LOS DOS REAIS"
>
> RICARDO HENRIQUES, SÓCIO DA ABREU ADVOGADOS



Os prazos mais apertados relacionam-se com violações de dados, refere: "se, por força da utilização de um deep fake, a organização sofrer uma violação de dados pessoais, em determinadas circunstâncias, a mesma deve ser comunicada no prazo de 72h à entidade competente". Para as organizações em setores críticos, os prazos são mais exigentes: "se a organização estiver abrangida pelo regime de incidentes de segurança com impacto significativo, deve notificar a entidade competente no prazo de 24h a 72h".

Daniel Reis, por sua vez, esclarece que "não existe um prazo específico para tal na lei portuguesa", referindo-se especificamente à descoberta de ser alvo de um deepfake, sublinhando, no entanto, que as obrigações de notificação dependem das consequências específicas do ataque, não do ataque em si.

> "OS DEEP FAKES NÃO SÃO APENAS UM TEMA DE TECNOLOGIA. REPRESENTAM HOJE SÉRIOS RISCOS OPERACIONAIS, LEGAIS E REPUTACIONAIS PARA AS ORGANIZAÇÕES"
>
> CATARINA MASCARENHAS, CONSULTORA DA ABREU ADVOGADOS

A GUERRA PELA CREDIBILIDADE

Elsa Lemos, Especialista em Comunicação de Crise, é categórica sobre a nova realidade temporal: "já não vivemos na era da *golden hour* [da resposta a uma crise de comunicação] – estamos nos *golden minutes*. O ritmo das redes sociais e a volatilidade da atenção pública exigem uma resposta com outra velocidade, proporcional ao impacto do conteúdo que circula e impacta as organizações".

Esta aceleração temporal não é apenas uma questão de rapidez, mas de sobrevivência reputacional. "No caso de um deepfake viral com potencial de gerar danos reputacionais ou financeiros sérios, a atuação deve ser ainda mais rápida. Mas, rapidez sem preparação pode comprometer a credibilidade", alerta.

A especialista defende uma abordagem estratégica



onde "os cenários que envolvem deepfakes devem constar nos manuais de comunicação de crise. Temos de encarar crises geradas com deepfakes como um cenário equivalente a tantos outros, como uma cibercrise".

A resposta eficaz exige métricas claras de ativação. Elsa Lemos explica que "a resposta a conteúdos manipulados e altamente virais, como os deepfakes, deve ser proporcional ao impacto real ou percecionado da crise. Há algo que posso garantir: quanto mais elevada for a exposição mediática, o potencial de dano reputacional, o envolvimento de figuras públicas ou o impacto financeiro ou outro, mais rápida e estratégica tem de ser a comunicação de crise".

Para operacionalizar esta rapidez, as organizações precisam de "discussões estratégicas que ajudem a definir métricas claras para ativação da resposta comunicacional", como por exemplo, o "grau de viralidade, sensibilidade ou tipo do conteúdo, desinformação associada, envolvimento de *stakeholders* críticos ou impacto financeiro que possa gerar".

Os deepfakes transcendem os ciberataques tradicionais e acabam por entrar numa categoria única de crise. Elsa Lemos explica que "uma crise provocada por deepfake não segue a lógica clássica de comprometimento de sistemas" uma vez que "compromete algo mais simbólico: a credibilidade, a integridade da imagem pública e a confiança".

A especialista em comunicação de crise detalhe o mecanismo de ação: "diferencia-se por atuar diretamente sobre o espaço social 'da verdade'. Ou seja, desestabiliza os fundamentos que as pessoas usam para distinguir o que é real e confiável do que é falso e repleto de desinformação. Simula com tal grau de realismo que provoca confusão, faz as pessoas duvidarem de fontes antes credíveis e, muitas vezes, leva à erosão da confiança coletiva".

Assim, esta é, nas palavras de Elsa Lemos, "uma forma de ataque à própria infraestrutura da confiança social". O poder dos deepfakes está na sua capacidade sensorial única: "Ao simular discursos, gestos, atitudes e vozes com alto realismo, pode induzir à formação de crenças falsas com efeitos reais – seja sobre uma liderança, uma marca ou uma instituição pública, privada ou Organização Não Governamental".

A vantagem dos deepfakes sobre a desinformação tradicional é, no entanto, significativa. "O poder dos deepfakes está na sua riqueza sensorial que integra múltiplos sentidos, que ativa o realismo heurístico – um atalho mental que faz com que as pessoas julguem algo como verdadeiro só porque parece real", explica Elsa Lemos.

Esta capacidade confere "uma capacidade de cultivar perceções erradas com mais eficácia do que a desinformação textual tradicional". A diferença é, assim, fundamental: ", a desinformação tradicional (como um texto enga-





- Elsa Lemos -
Especialista em Comunicação de Crise

nador) depende apenas da leitura e da interpretação racional, ao passo que o uso de imagem, som e até emoções associadas a gestos, tom de voz ou expressões faciais, proporcionam que os deepfakes criem uma experiência sensorial completa e enganadora".

Elsa Lemos refere que "alguns estudos revelam que os humanos falham com frequência ao detetar deepfakes de voz, mesmo quando são treinados. Isto revela que estamos a lidar com uma ameaça tecnicamente sofisticada e socialmente desafiadora".

"A MENTIRA É MAIS INTERESSANTE DO QUE A VERDADE"

Um dos aspetos mais desafiantes da resposta a deepfakes é a assimetria natural entre verdade e falsidade. Elsa Lemos identifica o problema central: "a mentira é muitas vezes mais interessante, emocional e viral do que a verdade – porque vem embalada com tensão, drama ou indignação. A 'verdade', por sua vez, tende a ser mais racional e sem apelo narrativo".

Para contrariar esta desvantagem estrutural, "é preciso refazer a narrativa de forma que ela seja ouvida e retida. Isso exige criar uma narrativa com intencionalidade estratégica (quem conta, o que conta, por que conta); usar *storytelling* factual;

> "É IMPORTANTE INCENTIVAR O PENSAMENTO CRÍTICO (DENTRO E FORA DAS ORGANIZAÇÕES) E ENSINAR AS PESSOAS A COLOCAR EM CAUSA CONTEÚDO DUVIDOSO OU A RECONHECER FONTES/CANAIS CREDÍVEIS"
>
> ELSA LEMOS, ESPECIALISTA EM COMUNICAÇÃO DE CRISE



responder sobre a forma de vídeo curto que mostre que aquilo que parece aborrecido é, na verdade, relevante".

A estratégia também passa pela mobilização de terceiros credíveis, ou aquilo a que Elsa Lemos chama de aliados, "ou seja, figuras externas credíveis", como "especialistas, jornalistas ou cidadãos afetados", para "reforçar a legitimidade da verdade.

Ao mesmo tempo, Elsa Lemos adverte contra a tentação de competir diretamente com o espetáculo. "Não se trata de competir com a mentira em espetáculo, mas de dar à verdade uma oportunidade de ser escutada com atenção e credibilidade", diz.

No entanto, para além da resposta reativa, é fundamental investir em prevenção através da educação e "deve ser realizado todo um trabalho de *awareness* e de literacia: é importante incentivar o pensamento crítico (dentro e fora das organizações) e ensinar as pessoas a colocar em causa conteúdo duvidoso ou a reconhecer fontes/canais credíveis", explica Elsa Lemos.

PREPARAÇÃO E RESPOSTA CORPORATIVA

Bruno Castro, da VisionWare, defende uma estratégia multifacetada: "para se protegerem contra deepfakes, as organizações devem adotar uma abordagem integrada de cibersegurança e literacia digital. Tal inclui a formação contínua de colaboradores para o reconhecimento de sinais de manipulação, a implementação de mecanismos de verificação em comunicações sensíveis, o reforço de processos de autenticação (como biometria ou múltiplos fatores), e o investimento em soluções tecnológicas de deteção automática de deepfakes".

Catarina Masracenhas, da Abreu Advogados, complementa esta visão com uma perspetiva organizacional: "os deep fakes não são apenas um tema de tecnologia. Representam hoje sérios riscos operacionais, legais e reputacionais para as organizações. Nesse sentido, as empresas devem, no contexto da sua estrutura organizativa, e a par de outras medidas, como formações, auditorias ou de gestão de terceiros, incluir o tema na sua estratégia".

A preparação eficaz vai além das medidas técnicas. Bruno Castro enfatiza que "é essencial que as empresas desenvolvam planos de resposta a incidentes que contemplem este tipo de ameaça emergente, de modo a reagir com rapidez e eficácia em caso de ataque". Esta perspetiva é reforçada por Daniel Reis, da DLA Piper, que sublinha que, "de forma genérica, o importante é o reforço da segurança dos sistemas de informação das empresas".

Ricardo Henriques, da Abreu Advogados, especifica as medidas técnicas necessárias: "deverão ser adotadas medidas de deteção e resposta a incidentes





provocados por deepfakes, como, por exemplo, mecanismos de deteção para identificar conteúdos deepfake e distingui-los dos reais, tal como devem ser adotados mecanismos de autenticação e verificação do conteúdo produzido ou partilhado. Estas medidas revelam-se de particular importância se estivermos no domínio de setores essenciais, como o financeiro ou saúde".

A documentação adequada não é apenas uma boa prática – é uma necessidade legal e operacional. Catarina Mascarenhas especifica os requisitos e refere que as empresas devem "manter documentação que estabeleça procedimentos de autenticidade e verificação de conteúdo, incluindo a verificação de fontes, bem como procedimentos de identificação e validação de conteúdo gerado ou divulgado pela empresa de modo a garantir a respetiva integridade e a mitigar os riscos legais associados aos deepfakes".

Esta documentação deve abordar aspetos práticos do dia-a-dia empresarial, como "avaliar se são colocadas imagens em redes sociais, em que termos são disponibilizados áudios", por exemplo, principalmente "se os conteúdos envolverem informações corporativas, garantindo a sua conformidade com as leis aplicáveis", explica Catarina Mascarenhas.

Daniel Reis identifica dois tipos fundamentais de documentação. "De forma a garantir a autenticidade do conteúdo corporativo, uma empresa deverá, entre outros, manter um registo dos consentimentos recolhidos em confor-



midade com o RGPD, assegurando a legitimidade do tratamento dos dados pessoais associados ao conteúdo capturado e/ou divulgado, bem como implementar políticas internas claras sobre a criação e gestão de conteúdo, nomeadamente no que toca à utilização de IA e, em particular, sistemas de IA generativa", diz.

CULTURA DE COMUNICAÇÃO E SALA DE CRISE

Elsa Lemos identifica a preparação como fator crítico de sucesso. "Uma das melhores formas de combater os deepfakes e a desinformação é antecipar este cenário. Este tipo de crise, provocada por deepfakes, é particularmente sensível ao fator preparação. É uma daquelas situações em que, se houver um trabalho prévio bem feito, a probabilidade de uma resposta bem-sucedida aumenta significativamente".

A preparação deve ser abrangente e prática. Elsa Lemos detalha os componentes essenciais, que incluem "análise de cenários em reuniões estratégicas, definição clara de critérios de ativação da resposta; inserir o deepfake nos manuais de comunicação de crise; construção de mensagens-chave base que possam ser adaptadas a qualquer instante; treino de porta-vozes; simulações realistas com deepfakes".

A tecnologia também desempenha um papel preventivo e o "investimento em ferramentas de IA que possam identificar e sinalizar conteúdo manipulado, ou seja, um *crisis scanning* que faz um varrimento do que possa gerar uma situação de potencial crise", como indica a especialista de comunicação em crise, deve ser feito pelas organizações.

O elemento cultural é fundamental para o sucesso. "Se existir uma cultura de comunicação de crise bem enraizada, a organização terá maior capacidade de ativar rapidamente uma resposta robusta, adaptada e coerente com os seus valores. A confiança já está construída internamente, o que permite que as decisões não sejam tomadas no calor da surpresa, mas na frieza da estratégia já no ADN da organização", afirma Elsa Lemos.

Também a preparação técnica é importante e Bruno Castro refere que as organizações habitualmente "descobrem que foram alvo de campanhas de deepfake através de denúncias de colaboradores, clientes ou parceiros, da monitorização de redes sociais e media, ou através da deteção de comportamentos anómalos nas comunicações internas ou externas. Em alguns casos, apenas após a ocorrência de um dano concreto é que se revela a existência do conteúdo manipulado".

A resposta eficaz a deepfakes exige coordenação entre múltiplas áreas especializadas. Elsa Lemos

IM. | Copyright licensed by Visapress | P. 16 of 17



explica que "essa articulação exige integração estratégica" e, numa crise de deepfake, "a resposta eficaz resulta na integração e várias linhas de resposta, assim como qualquer outra cibercrise". A especialista em comunicação de crise detalha as responsabilidades de cada área: "a técnica deteta, analisa, identifica falhas de veracidade, usa análise forense digital, IA para confirmar a manipulação; a jurídica avalia as implicações legais (calúnia, difamação, fraude, usurpação de identidade), preserva provas, aciona queixas e regula a proteção de imagem; e a comunicacional traduz tudo isso para os públicos – com clareza, empatia e sentido de urgência". Aqui, a coordenação é o fator diferenciador e a "eficácia só acontece quando estas três áreas trabalham em sala de crise, com acesso mútuo à informação, alinhamento de mensagens e gestão do tempo".

Bruno Castro reforça a importância da rapidez na análise técnica: "a confirmação de que um vídeo altamente convincente é falso pode demorar desde algumas horas até vários dias, dependendo da complexidade do conteúdo e dos recursos técnicos disponíveis. A análise forense digital requer equipamentos especializados e conhecimentos avançados, sendo por vezes necessário recorrer a equipas externas ou laboratórios forenses".

O IMPERATIVO DA AÇÃO IMEDIATA

A democratização de ferramentas como o Google Veo 3 marca o fim da era em que os deepfakes eram domínio exclusivo de especialistas técnicos. Hoje, qualquer pessoa com 249 dólares que tenha motivação pode criar conteúdo manipulado capaz de enganar até os observadores mais atentos.

Para as organizações, esta realidade exige uma resposta que vai muito além das medidas tradicionais de cibersegurança. Não se trata apenas de proteger sistemas informáticos, mas de blindar a própria credibilidade empresarial numa era onde a distinção entre real e artificial se esbate a cada dia.

Mesmo sabendo que tudo no mundo atual muda rapidamente, o momento de agir é agora. As ferramentas legais existem, as melhores práticas estão identificadas e os casos de sucesso e falha oferecem lições claras. As organizações que integrarem a proteção contra deepfakes na sua estratégia não só de cibersegurança, mas também de comunicação de crise, não vão apenas proteger-se melhor, como vão ganhar uma vantagem competitiva numa era onde a confiança se tornou num dos ativos mais valiosos e vulneráveis no mundo empresarial. ■