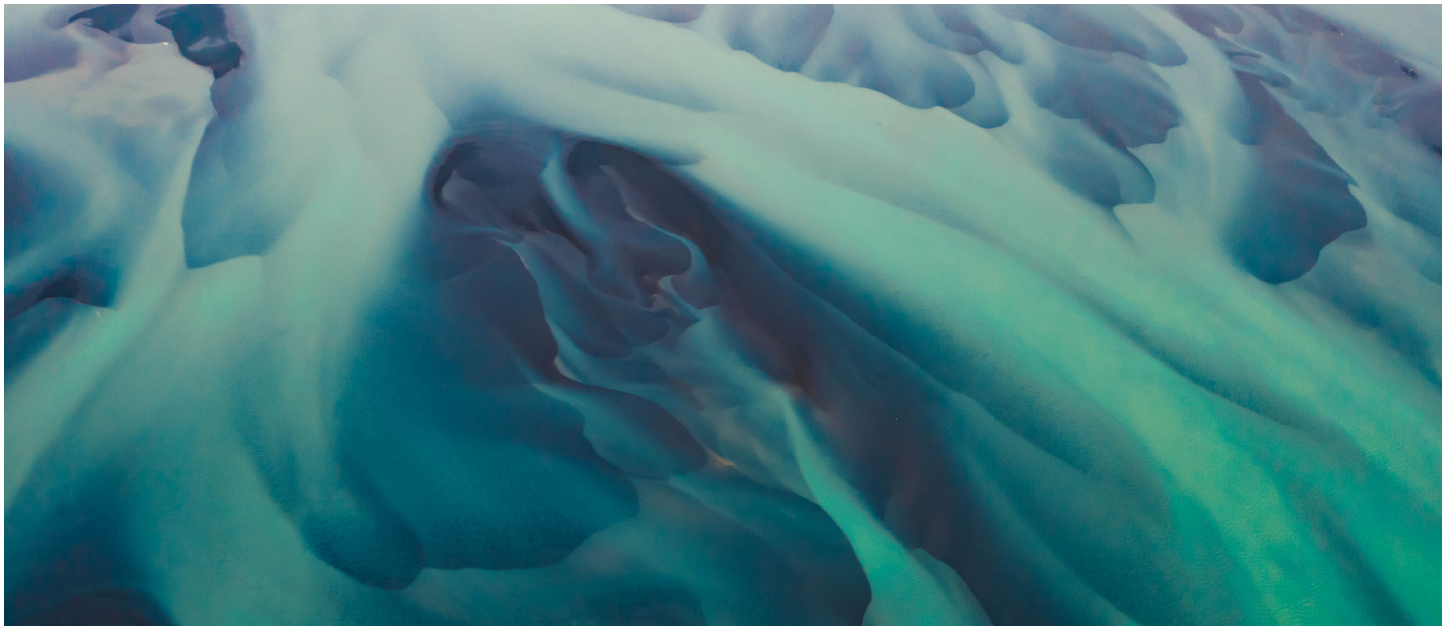


TECNOLOGIA, VERDADE E CONFIANÇA ILUMINANDO O CAMINHO DO DESENVOLVIMENTO DE UMA IA ÉTICA

Este documento técnico analisa como os criadores e usuários de soluções complexas de IA podem mitigar os muitos riscos éticos enfrentados pelas organizações e pela sociedade, sem prejudicar a profunda capacidade da IA de ajudar a resolver nossos desafios mais difíceis e urgentes. Com a aceleração das mudanças tecnológicas, também aumenta a importância de se alcançar um equilíbrio saudável entre a incrível promessa dos avanços da IA e a redução da possibilidade de ameaças existenciais que ela torna aparente. Dessa forma, a IA representa bem o termo "faca de dois gumes".

Os dois gumes dessa faca ficaram ainda mais afiados após o lançamento do ChatGPT pela OpenAI em novembro de 2022. Esse grande ponto de inflexão em décadas de história da IA provocou um crescimento surpreendente da inovação, do investimento e da conscientização geral, além de amplificar os alarmes já existentes. Preocupações cada vez maiores se espalham pela grande variedade de grupos constituintes, que englobam empresas, trabalhadores, cidadãos, o meio acadêmico, governos e órgãos reguladores. Com a grande velocidade com que a situação progride, a preocupação é compreensível. Quem pode ter certeza de que a IA não excederá, algum dia, nossa capacidade de entendê-la ou controlá-la?

Novamente, as pessoas têm percepções positivas e negativas da IA. Em 2021, embora 44% dos líderes executivos pesquisados pela KPMG achassem que seu setor estava avançando rápido demais na adoção da IA, quase todos os pesquisados desejavam que sua própria organização avançasse ainda mais rápido[1]. Essa dicotomia se intensificou ainda mais.

Com décadas de experiência coletiva em IA e um profundo entendimento dos riscos e das necessidades dos nossos clientes, a Nuix reconhece que o conceito de confiança está no centro do dilema da IA. Até pouco tempo atrás, a desconfiança em relação à IA era exclusividade de filmes e livros apocalípticos. No nosso dia a dia, a IA se limitava a encaminhar chamadas telefônicas, corrigir nossos erros de ortografia ou recomendar filmes. No entanto, conforme tentamos atribuir tarefas cada vez mais complexas e substanciais à IA, aumentamos rapidamente (e com razão) o escrutínio em relação à credibilidade e à precisão dos resultados que ela oferece.

Explicabilidade - Para que as pessoas tenham e mantenham a confiança nos resultados da IA, precisamos abrir a "caixa preta" para que elas saibam claramente quais dados de treinamento foram usados para criar os modelos, os vieses que podem existir e entendam facilmente o que está por trás das previsões da IA. A IA deve ser usada para informar e acelerar a tomada de decisão humana, em vez de pretender substituí-la com "respostas" definitivas, e a IA da Nuix direciona os tomadores de decisão humanos para onde tenham mais chance de encontrar os fatos inegáveis que informarão suas ações.

Acessibilidade - O poder transformador de qualquer tecnologia está no nível de usabilidade e personalização que ela pode oferecer. A IA da Nuix garante que o ser humano tenha o controle total e defina o que a máquina pode ou não ver e fazer em nome dos seres humanos, sem depender de algoritmos ou códigos inacessíveis. A IA que capacita o ser humano, em vez de substituí-lo, é fundamental para esse princípio.

Especificidade - O conceito de especificidade permite aplicar a IA de forma pragmática e eficiente, o que gera resultados mensuráveis e reais. Modelos mais generalistas são mais difíceis de explicar e menos transparentes, o que também os torna menos reproduzíveis. É fundamental permitir que os usuários conduzam o processo, transmitindo seu conhecimento e experiência aos modelos que criam, centrados em uma das várias opções de casos de uso específicos. Como resultado, a precisão e a credibilidade são muito maiores nos resultados específicos do domínio.

As implicações desses princípios e a maneira que a Nuix atende a esses requisitos são examinadas mais adiante neste documento.

POR QUE A REGULAMENTAÇÃO DA IA PRECISA DE UM EXAME LONGO E MINUCIOSO? ⚠

Embora o custo de desenvolvimento e treinamento de IA tenha caído drasticamente em poucos anos, os recursos computacionais e humanos necessários continuam sendo significativos. Os avanços significativos da IA, dessa forma, têm se concentrado nas empresas privadas. Na verdade, o meio acadêmico produz apenas um décimo das descobertas de IA[2]. Provavelmente, isso significa que grande parte da aplicação e do desenvolvimento da área está sendo orientada muito mais pelos lucros do que por preocupações éticas.

A prova de que a tecnologia está evoluindo mais rapidamente do que a compreensão da sociedade sobre seus benefícios e riscos é o fato de que até mesmo os principais inovadores e promotores da IA estão pedindo uma abordagem mais comedida e cautelosa. No entanto, embora grupos de líderes empresariais proponham uma moratória pública e coletiva no desenvolvimento da IA com uma frequência alarmante[3], ninguém quer ficar para trás.

O governo e os órgãos reguladores estão se esforçando para acompanhar o ritmo das mudanças e considerar a IA como um todo, em vez de fazer ajustes superficiais. A União Europeia, por exemplo, está próxima de promulgar uma das primeiras leis abrangentes que regem a IA. A Lei da IA[4] passou por um marco importante no Parlamento Europeu poucos dias antes da publicação deste documento, e a lei pode ser promulgada já no final de 2023. Ela busca "garantir que a IA tenha um desenvolvimento ético e centrado no ser humano". O ponto central da legislação proposta é a "capacidade de entender quais modelos foram usados para alcançar os resultados, como esses modelos foram construídos (dados de treinamento) e se os seres humanos envolvidos podem ou não ser os árbitros finais dos resultados".

Assim como na introdução do GDPR[5], a UE parece estar tomando a liderança, agora com a regulamentação da IA. Usando o passado como prólogo, a Lei de IA da UE é um provável indicador de como a regulamentação de IA se desenvolverá globalmente. Na verdade, ela já faz parte do planejamento futuro de algumas organizações. Quando um cliente da Nuix, um órgão de segurança pública na Europa, perguntou sobre nossa IA, tivemos o prazer de oferecer garantias detalhadas por escrito de que nossos modelos e recursos de IA já atendiam aos requisitos.

Em uma abordagem mais direcionada para mitigar os possíveis riscos da IA, um juiz federal do Texas implementou recentemente uma nova regra que exige que os advogados que comparecem ao tribunal apresentem um certificado que afirme que nenhuma parte de seu processo foi gerada por IA generativa (ChatGPT, Harvey.AI, Google Bard) ou que a precisão de qualquer conteúdo que tenha produzido com ajuda de IA tenha sido cuidadosamente examinada por um ser humano.[6]

Em suma, concretizar o vasto potencial da IA exigirá uma reflexão séria e ações ponderadas e coordenadas. Como escreveram dois acadêmicos australianos da área da IA, "está claro que estamos em um ponto de inflexão: precisamos pensar com seriedade e urgência sobre as desvantagens e os riscos que a aplicação crescente da IA está revelando[7]".

"Atualmente, o aspecto mais triste da vida é que a ciência reúne conhecimento mais rápido do que a sociedade reúne sabedoria."

Isaac Asimov

A VISÃO DA NUIX

Na Nuix, trabalhamos com IA e a desenvolvemos há anos. Entendemos que a IA tem o poder de transformar nossos clientes e até mesmo de reestruturar economias inteiras. No entanto, também reconhecemos que a humanidade pode ter apenas uma chance de fazer a transição correta para um futuro possibilitado pela IA. Uma regulamentação inteligente, educação, disciplina e debate honesto são extremamente importantes para garantir que a IA seja desenvolvida e aplicada de forma ponderada, ética e transparente. Em resumo, uma IA ética é sinônimo de uma IA confiável.

O CTO da Nuix, Stephen Stewart, disse: "Na realidade, é difícil aproveitar a IA/ML na incrível diversidade de desafios de dados existentes. Não existe uma solução única para todos os dados não estruturados. A IA da Nuix é um exemplo de como aplicar conceitos de ponta e avanços tecnológicos em IA para atender a necessidades muito específicas dos clientes com eficiência incrível e insights defensáveis."

Nossa abordagem para desenvolver e implantar uma IA confiável é regida por três princípios de design de tecnologia: Explicabilidade, acessibilidade e especificidade. Esses princípios nos mantêm ancorados em nosso propósito de ser uma força para o bem, ao encontrar a verdade nos dados digitais. Eles garantem a criação de uma inteligência colaborativa que aumenta a eficiência e a eficácia humanas, permitindo que as pessoas se concentrem em trabalho de alto valor que exige julgamento humano, pensamento crítico refinado e consciência ética.

Esses princípios posicionam a Nuix como um fornecedor responsável, respeitável e ético de IA. Eles também servem como parâmetros bem elaborados para orientar nossos esforços contínuos de desenvolvimento e garantir nosso alinhamento com as exigências dos clientes e os padrões éticos globais de IA, que estão em franca evolução.

1. EXPLICABILIDADE

Somos líderes globais em software de inteligência digital, em que nossos clientes confiam para processar, examinar e interpretar grandes quantidades de dados para encontrar percepções factuais, ou seja, a verdade. Para estabelecer e manter essa confiança na IA, nossos clientes precisam ter visibilidade dos dados usados para criar nossos modelos de IA, bem como das explicações por trás dos resultados produzidos por esses modelos. Cada vez mais, o setor está chamando isso de "explicabilidade", e nossa tecnologia de IA foi desenvolvida desde o início para proporcionar isso.

Outro elemento importante da explicabilidade é o fato de que nossa IA não pretende substituir os humanos na tomada de decisões. Em vez disso, ela fornece dados filtrados, relevantes e priorizados que possibilitam que nossos clientes tomem decisões humanas mais eficazes e eficientes.[8] O fundamental aqui é que o ser humano tem insights diretos em cada etapa do processo e mantém controle total como árbitro final. Nossa IA oferece quatro camadas de análise principais sobre dados textuais:

1. Tipo de registro: por exemplo, documento fiscal, formulário de assistência médica, transcrição, patente.
2. Tópico de conteúdo: por exemplo, direito, meio ambiente, finanças, política.
3. Extrações: por exemplo, PII, PHI, entidades nomeadas.
4. Prioridade: por exemplo, relevância, importância ou nível de risco de qualquer um dos itens acima, de acordo com regras orientadas pelo cliente.

A solução da Nuix mostra aos tomadores de decisão humanos os locais onde é mais provável que encontrar os fatos inegáveis que informarão suas ações e aumentarão a precisão e a eficiência das suas decisões. Cada etapa do processo é registrada e pode ser repetida.

2. ACESSIBILIDADE

Até recentemente, a IA era exclusiva a um pequeno número de especialistas em aprendizado de máquina, programadores de software e cientistas de dados altamente qualificados. Essas circunstâncias excluíram os especialistas no assunto (SMEs) da cadeia de valor da criação de modelos e os deixaram excessivamente dependentes de equipes técnicas que, muitas vezes, não tinham o conhecimento do domínio para trabalhar nesses modelos.

Para maximizar o valor do investimento em IA e também para atender à exigência ética de disseminar as oportunidades e os benefícios que a IA oferece, há a obrigação de capacitar que os indivíduos não técnicos acessem esses recursos, de forma intuitiva e fácil.

A Nuix está contribuindo para uma revolução na experiência do usuário, com uma interface intuitiva projetada para permitir que especialistas no assunto criem, otimizem e validem rapidamente modelos de IA de linguagem altamente precisos: uma interface simples de apontar e clicar sem uma única linha de código. Além dos avanços no fluxo de trabalho e na produtividade que isso oferece, esse acesso fácil aumenta a confiança de que os resultados da IA serão examinados e bem alinhados com a área para a qual foram projetados.

Interfaces bem projetadas com pouco ou nenhum código e fluxos de trabalho integrados ajudam a democratizar o poder da IA, ao permitir que um conjunto mais amplo de funcionários humanos se envolva diretamente com a IA para fazer as alterações necessárias, testar e otimizar modelos, corrigir erros e remover preconceitos. Isso permite que o usuário da IA, e não o programador, tenha sempre o controle final.

3. ESPECIFICIDADE

Quando se trata de aplicativos do mundo real, maior nem sempre significa melhor. Embora o ChatGPT seja apoiado pelo GPT4 (o maior LLM já construído[9]), sua imensidão e sua tentativa de ser tudo para todas as pessoas prejudicam grande parte de seu uso prático em ambientes reais de negócios. Claramente, seria impraticável, se não impossível, hospedar o ChatGPT no seu ambiente seguro. Fazer upload de dados confidenciais para usá-los no estado em que se encontram na internet cria um risco de exposição material. Por esse motivo, muitas empresas e organizações proibiram os funcionários de acessar o ChatGPT e outros grandes sites de provedores de IA generativa. Esses modelos de linguagem massivos não estão bem ajustados para dar suporte a assuntos específicos e profundos pertinentes à sua empresa ou à sua área de atuação.

Por outro lado, a IA da Nuix foi projetada para ser implantada nos ambientes seguros dos clientes e executada em sistemas acessíveis baseados em CPU. Além disso, combinando as vantagens da explicabilidade e da acessibilidade descritas acima, os clientes podem ajustar rapidamente ou mesmo criar modelos para atender aos seus requisitos subjetivos e casos de uso específicos. Mais uma vez, os humanos estão no controle, e a IA trabalha para atender às suas necessidades específicas.

"Um computador mereceria ser chamado de inteligente se pudesse enganar um ser humano, fazendo-o acreditar que é humano."

Alan Turing

IA QUE PASSA NO TESTE

Todos esses três princípios de design de tecnologia contribuíram para a nossa participação premiada no CUI Tool Automation Challenge da Marinha dos EUA, anunciado em dezembro de 2022[10].

O objetivo desse desafio era ajudar a compartilhar informações confidenciais, mas não classificadas, Informações Não Classificadas Controladas (CUI), de forma adequada, eficaz e vital, entre departamentos, parceiros e aliados. As tarefas específicas a serem gerenciadas incluíam:

- Ingerir milhões de documentos do governo do Departamento de Defesa dos EUA e em vários formatos de arquivo.
- Analisar o conteúdo do documento para identificar os tipos de CUI em cada arquivo.
- Quantificar um nível de confiança de acordo com cada resultado.
- Marcar o banner e o rodapé de cada documento para manter a cadeia de confiança à medida que vários humanos interagem com os arquivos.

O sucesso de nossa solução foi construído sobre as bases que nossos três princípios proporcionaram. A interface simples de apontar e clicar fornece uma **explicabilidade** sobre como os modelos de IA recém-construídos e existentes foram criados, bem como os motivos fáceis de entender que apoiam os nossos resultados. A interface sem código oferece uma **acessibilidade**, que permite ao usuário criar, treinar, aprimorar ou ajustar modelos e ponderações relativas para otimizar e validar o desempenho de cada modelo. E a **especificidade** de cada modelo, cuidadosamente ajustada a um tipo específico de CUI e combinada por especialistas no assunto para enfrentar os desafios exclusivos dos dados do mundo real.

A IA E SUA ORGANIZAÇÃO

Há muito tempo, os economistas identificaram que a capacidade de usar tecnologias avançadas de forma rápida e eficaz era mais crucial para o crescimento do que a natureza real da tecnologia. A famosa frase de Robert Solow é muito apropriada: "É possível ver a era da informática em todos os lugares, menos nas estatísticas de produtividade".

A Nuix construiu uma reputação global tornando os dados de nossos clientes altamente pesquisáveis. Agora, estamos subindo de nível com uma plataforma unificada, com infusão de IA, chamada Nuix Neo. O Nuix Neo combina o mecanismo de processamento de dados mais poderoso do mundo com IA avançada para ajudar a resolver os desafios de dados complexos, cada vez maiores, dos nossos clientes. Essa estratégia de inovação integrada permite que os clientes da Neo possam aproveitar o valor da IA sem precisar de pessoal interno com conhecimento em IA nem comprar ou criar várias soluções pontuais de IA.

RESUMO

Fornecer soluções de IA é uma tarefa de grande responsabilidade. Em um mundo cada vez mais dominado por chatbots alucinantes de fala sedutora e atormentado por desinformação e deep fakes, os riscos são cada vez mais altos. Mas a realidade é que os problemas atuais e futuros simplesmente não podem ser resolvidos com as técnicas e tecnologias do passado. Independentemente de gostarmos ou não, dominar a IA é parte fundamental de um futuro próspero e sustentável.

O segredo será uma abordagem consciente e disciplinada da IA, baseada em explicabilidade, acessibilidade e especificidade. O objetivo final será fornecer resultados confiáveis, verificáveis e defensáveis para os nossos clientes e a sociedade em geral.

Como o estudo de caso acima destaca, é fundamental minimizar o risco de uma violação de dados. No entanto, se ela ocorrer, as organizações precisarão reagir. Para muitas, isso significa ter a capacidade de implementar tecnologias avançadas e poder de processamento na tarefa de analisar e compreender uma violação e atender às necessidades de informações de clientes, reguladores e da mídia.

Um banco norte-americano que sofreu recentemente uma violação de dados conseguiu aproveitar o poder da Nuix para conduzir seus trabalhos de investigação e correção. O banco conseguiu usar vários servidores, pesquisando muitos terabytes por dia, para lidar rapidamente com a extensão da violação e ter informações sobre ela.

QUER SABER MAIS?

As tendências abordadas neste documento técnico ainda estão se desenvolvendo, assim como as aplicações práticas que nossos clientes têm para os recursos de IA da Nuix. Se quiser discutir como esses recursos podem funcionar em sua organização, entre em contato conosco pelos canais descritos abaixo. Um dos nossos especialistas em IA entrará em contato para ajudar você a entender como o Nuix Neo e sua IA incorporada podem atender à sua organização.

REFERÊNCIAS

[1] Krishna, Campana e Chandrasekaran, "Thriving in an AI World: Unlocking the value of AI across seven key industries," KPMG, 2021, https://advisory.kpmg.us/articles/2021/thriving-in-an-ai-world.html

[2] Stanford University, "Chapter 1.2., Research and Development', Fig 1.2.2, Artificial Intelligence Index Report 2023", 2023, https://aiindex.stanford.edu/report/#individual-chapters

[3] Consulte https://www.science.org/content/article/alarmed-tech-leaders-call-ai-research-pause#:~:text=Attracting%20signatures%20from%20the%20likes,potential%20risks%20of%20the%20technology

[4] Consulte https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence

[5] Consulte https://gdpr.eu/what-is-gdpr/

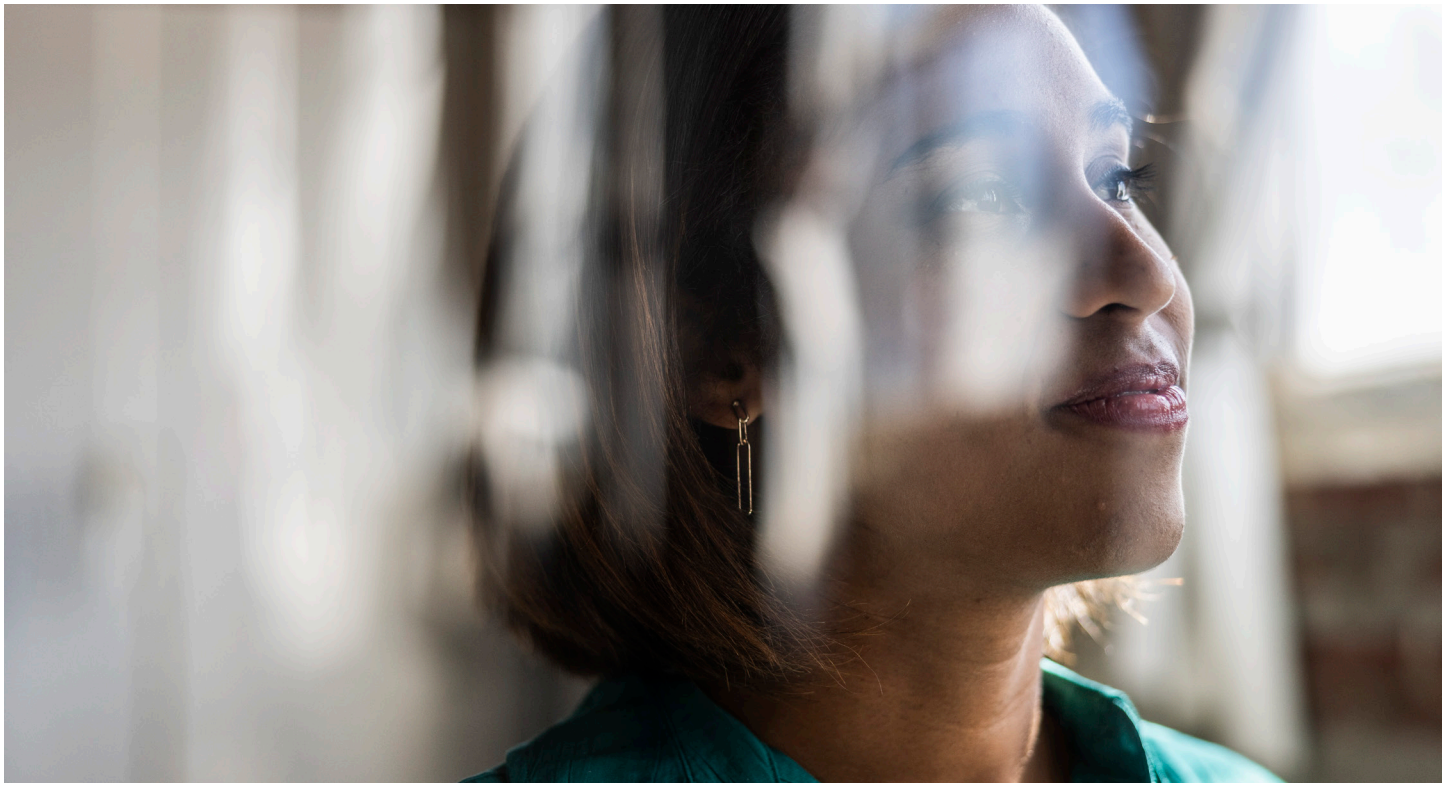
[6] Consulte https://www.mitchellwilliamslaw.com/taming-the-ai-beast-judges-set-rules-to-control-use-of-generative-ai-in-their-courts

[7] Sonenberg e Walsh, "Artificial intelligence is now part of our everyday lives and its growing power is a double edged sword", The Conversation, novembro de 2021, https://theconversation.com/artificial-intelligence-is-now-part-of-our-everyday-lives-and-its-growing-power-is-a-double-edged-sword-169449

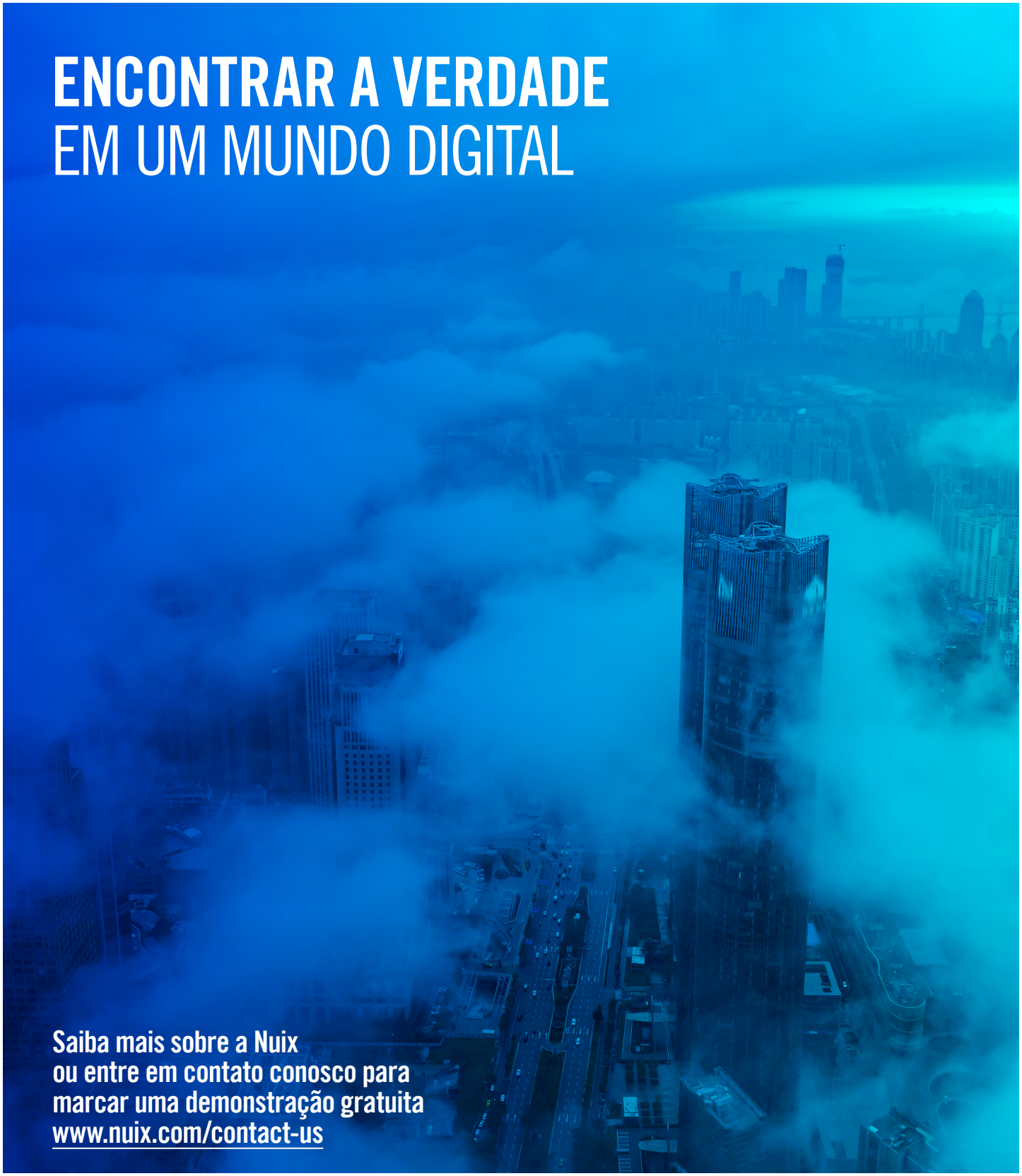
[8] Enquanto este texto está sendo redigido, a Nuix está em discussão com um dos maiores órgãos de segurança pública da Europa, que está analisando como pode capturar os benefícios da IA com foco principal na transparência dos resultados

[9] Consulte https://neuroflash.com/blog/gpt-4-parameters-rumors-and-forecasts/#:~:text=It%20is%20also%20possible%20that,trillion%20or%2010%20trillion%20parameters.

[10] "Nuix and Serco NA partner to score prizewinning results in US Navy AI Automation Challenge", Nuix, https://www.nuix.com/news/nuix-and-serco-na-partner-score-prizewinning-results-us-navy-ai-automation-challenge



ENCONTRAR A VERDADE
EM UM MUNDO DIGITAL



Saiba mais sobre a Nuix ou entre em contato conosco para marcar uma demonstração gratuita
www.nuix.com/contact-us

nuix

A Nuix Limited (ASX: NXL) é um dos principais fornecedores de software inteligente e de análise investigativa equipado com IA, que capacita as organizações a proteger, administrar e liberar o valor de seus dados. Com presença global e mais de 20 anos de experiência, a Nuix oferece soluções avançadas de análise de dados para setores que exigem precisão e percepção, como eDiscovery, processamento e análise jurídica, conformidade regulamentar, governança de dados, segurança cibernética e investigações forenses. Com o poder do processamento avançado, da inteligência artificial e do aprendizado de máquina, a Nuix ajuda os clientes a processar, normalizar, indexar, enriquecer e analisar dados complexos com velocidade, escala e precisão forense.

Para saber mais, acesse www.nuix.com

APAC
Austrália: +61 1300 511 852

EMEA
Reino Unido: +44 203 934 1600

AMÉRICA DO NORTE
USA: +1 877 470 6849