ARTIGO 20 **PT**

Discriminação algorítmica

No final de fevereiro, nem um mês depois do lançamento, após a divulgação de erros crassos de representação histórica, a Google foi obrigada a suspender temporariamente a funcionalidade de geração de imagens do seu modelo mais avançado de inteligência artificial (IA) que veio substituir o *'Bard'* agora ***'Google Gemini'***. As imagens mais divulgadas mostram soldados nazis alemães e vikings retratados como asiáticos e afrodescendentes, papas mulheres, entre outros exemplos na mesma linha. E foi assim que um tema que não é novo, em torno dos **riscos da IA e dos vieses (*'bias'*) algorítmicos**, perante o silêncio ensurdecedor da empresa sobre as especificidades da programação, regressou em força aos tópicos de discussão.

Relembro que há relativamente pouco tempo esta mesma empresa já tinha sido alvo de críticas relacionadas com a forma sexualizada como as mulheres estavam a ser representadas na funcionalidade de imagens do seu motor de busca. Não sendo um exclusivo desta plataforma, este incidente permitiu chamar a atenção para a permeabilidade das ferramentas de geração de imagem aos vieses algorítmicos, assim como, é importante realçar, para a acrescida dificuldade que existe em os ultrapassar.

No caso *Google Gemini*, muito provavelmente, e à semelhança da abordagem assumida pela OpenAI na ferramenta DALL-E-2, a incorreção das imagens deve ter resultado da própria tentativa de neutralização dos vieses, nomeadamente com a utilização de *'prompts'* de diversidade, que consiste em adicionar palavras, como 'mulher' ou 'asiático', a pedidos gerais de imagens. Uma 'espécie' de discriminação positiva, tipo quota, incorporada no sistema de modo a compensar situações reconhecidas de desvantagem que afetam desproporcionalmente determinados grupos.

Começo por aqui porque este é um exemplo que me parece ilustrar bem a importância de conhecer e trabalhar mais o **impacto, positivo e negativo, da IA nos direitos fundamentais, sob a ótica dos princípios**

ARTÍCULO 20 **ES**

Discriminación algorítmica

A finales de febrero, solo un mes después de su lanzamiento, Google se vio forzado a suspender temporalmente la funcionalidad de generación de imágenes de su modelo más avanzado de inteligencia artificial (IA), que reemplazó a "*Bard*" y ahora se llama "***Google Gemini***". Las imágenes más difundidas muestran a soldados nazis alemanes y vikingos retratados como asiáticos y afrodescendientes, mujeres en el rol de papas, entre otros ejemplos en la misma línea. Y así fue como un tema que no es nuevo, en torno a los riesgos de la IA y los sesgos algorítmicos, ante el silencio ensordecedor de la empresa sobre las particularidades de la programación, volvió con fuerza a los temas de discusión.

Recuerdo que hace relativamente poco tiempo, esta misma empresa ya había sido objeto de algunas críticas por la forma sexualizada en que las mujeres estaban siendo representadas en la funcionalidad de imágenes de su motor de búsqueda. Sin ser exclusivo de esta plataforma, este incidente ha permitido llamar la atención sobre la permeabilidad de las herramientas de generación de imágenes a los sesgos algorítmicos así como, y es importante subrayarlo, la creciente dificultad que existe para superarlos.

En el caso de *Google Gemini*, con toda probabilidad, y de forma similar al planteamiento adoptado por OpenAI en la herramienta DALL-E-2, la inexactitud de las imágenes fue debida al propio intento de neutralizar los sesgos, concretamente con el uso de *"prompts"* de diversidad. Esto consiste en añadir palabras como "mujer" o "asiático" a las solicitudes generales de imágenes. Una especie de discriminación positiva, similar a un sistema de cuotas, incorporada en el sistema para compensar situaciones de desventaja reconocidas que afectan desproporcionadamente a determinados grupos.

Y comienzo con este ejemplo, que me parece ilustrativo acerca de la importancia de conocer y trabajar más sobre el **impacto, positivo y negativo, de la IA en los derechos fundamentales, desde la perspectiva de los**

da igualdade e da não discriminação. E tanto pela perspetiva dos utilizadores finais, como das empresas que, num ambiente maioritariamente desregulado e ainda com conhecimentos muito limitados quanto às potencialidades destes modelos, são cada vez mais chamadas a intervir em áreas tradicionalmente do domínio público.

Para melhor se compreender, é importante, ainda que a título sumário, saber como opera a **discriminação algorítmica**. Maioritariamente sob a forma de discriminação indireta, os vieses normalmente presentes ora se reconduzem ao próprio algoritmo (*'algorythmic bias'*), ora aos dados (*'data bias'*).

Quer isto dizer que, ao contrário do que muitas vezes se pensa, os enviesamentos não resultam apenas dos dados, sendo importante analisar as **variáveis** presentes, as quais podem simplesmente reproduzir os preconceitos, conscientes ou inconscientes, dos programadores.

Menos evidente, mas talvez por isso mais relevante, é a chamada **discriminação por *proxy***, quando variáveis aparentemente neutras, por via do processamento de grande volume de dados, são automaticamente associadas a certos grupos protegidos, passando a funcionar como fatores indiretos de exclusão. É conhecido o exemplo do código postal, que sendo uma mera informação de contacto, também pode levar à identificação de uma pessoa com determinado grupo étnico, residente naquela localidade. Ou da remuneração para indicar o sexo da pessoa, tendo como referencial as desigualdades salariais entre homens e mulheres.

Já os **vieses nos dados**, mais não são do que um reflexo do preconceito e da intolerância instituídos em sociedade. Se dados enviesados são utilizados para treinar um algoritmo, naturalmente que os resultados produzidos dificilmente não serão discriminatórios. Sendo que o mesmo se pode passar com um algoritmo em pleno funcionamento, mesmo que submetido a uma rigorosa fase de validação. Desta vez, o enviesamento será fruto não dos dados de testagem, mas das interações com os utilizadores, que automaticamente alimentam e desenvolvem os sistemas de autoaprendizagem. É o caso de um algoritmo que, tendo por base o histórico de interações do sistema, acaba a apresentar anúncios de vagas de engenheiria maioritariamente aos homens, desconsiderando eventuais

principios de igualdad y no discriminación. Y tanto desde la perspectiva de los usuarios finales como de las empresas que, en un entorno mayoritariamente no regulado y aún con un conocimiento muy limitado del potencial de estos modelos, se ven llamadas cada vez más a intervenir en ámbitos tradicionalmente de dominio público.

Para comprender todo mejor, es importante, aunque sea brevemente, saber cómo funciona la **discriminación algorítmica**. Principalmente en la forma de discriminación indirecta, donde los sesgos que normalmente están presentes se deben al propio algoritmo ("sesgo algorítmico") o a los datos ("sesgo de los datos").

Esto significa que, al contrario de lo que a menudo se piensa, los sesgos no son solo el resultado de los datos, sino que es importante analizar las **variables** presentes, que pueden simplemente reproducir los prejuicios, conscientes o inconscientes, de los programadores.

Menos evidente, pero quizá por ello más relevante, es la denominada **discriminación por *proxy***, cuando variables aparentemente neutrales, a través del tratamiento de un gran volumen de datos, se asocian automáticamente a ciertos grupos protegidos, convirtiéndose en factores indirectos de exclusión. Un conocido ejemplo es el código postal, que, siendo un mero dato de contacto, puede llevar también a identificar a una persona con un determinado grupo étnico, residente en esa localidad. O también la remuneración para indicar el género de una persona, teniendo como referencia las desigualdades salariales entre hombres y mujeres.

Por otra parte, los **sesgos en los datos** no son más que un reflejo de los prejuicios y la intolerancia que se han instaurado en la sociedad. Si se utilizan datos sesgados para entrenar un algoritmo, es probable que los resultados producidos sean discriminatorios. Lo mismo puede ocurrir con un algoritmo en pleno funcionamiento, incluso si ha sido sometido a una rigurosa fase de validación. Esta vez el sesgo no será el resultado de los datos de prueba sino de las interacciones con los usuarios, que alimentan y desarrollan automáticamente los sistemas de autoaprendizaje. Es el caso de un algoritmo que, basándose en el historial de interacciones del sistema, acaba presentando anuncios de empleos de ingeniería mayoritariamente a

mulheres interessadas, que acabam assim por ficar prejudicadas.

A **diversidade dos dados** é fundamental para garantir a ausência de um padrão discriminatório. Uma tecnologia de recrutamento que utilize as redes sociais para avaliar candidatos, vai certamente acabar por afetar pessoas menos ativas nestes meios, como por exemplo os mais velhos, cuja sub-representação levará, por sua vez, à sistematização de um tratamento diferenciado potencialmente discriminatório.

Ou seja, de facto, tal como as pessoas, também os algoritmos discriminam, pelo que importa agora esclarecer qual a diferença que justifica a urgência do debate, nomeadamente em torno da necessidade de adoção de novas políticas públicas, com mecanismos reforçados de controlo e responsabilização.

O risco mais imediato é o da escalabilidade destes sistemas, que pode levar a uma sistematização dos vieses e, consequentemente, da discriminação. Ainda assim, talvez o risco mais diferenciador esteja associado ao chamado efeito ***'black box'*** do algoritmo. Se em princípio um programador deveria conseguir explicar o que é suposto que um *software* esteja a fazer, com os modelos de autoaprendizagem (*'machine learning'*), que acabam por se auto programar, pode acontecer que os resultados não sejam assim tão fáceis nem de compreender, nem de interpretar, nem de explicar. Como vimos pelo caso inicial, continua a haver uma enorme falta de progresso na interpretabilidade dos algoritmos, um fator que, associado à evolução exponencial da tecnologia, muito tem contribuído para a defesa de um caminho de regulamentação.

Em suma, pouco transparente e muitas vezes protegido pelas regras do segredo comercial, a deteção e prova de situações de discriminação é extremamente difícil. Tudo no quadro de um ecossistema digital frágil em matéria de prevenção e proteção, que frequentemente coloca o ónus da responsabilidade sobre as partes mais vulneráveis.

A acrescer a isto, há também um **problema de expetativa e de qualificação do conteúdo**. É o chamado preconceito da automação e da confirmação, que se traduz não apenas numa confiança cega nos algoritmos, percecionados como certezas matemáticas, como também numa diminuição do espírito crítico

hombres, descartando a las mujeres interesadas, que acaban saliendo perjudicadas.

La **diversidad de los datos** es fundamental para garantizar la ausencia de un patrón discriminatorio. Una tecnología de contratación que utilice las redes sociales para evaluar a los candidatos acabará afectando sin duda a las personas menos activas en estos medios, como las personas mayores, cuya baja representación dará lugar a su vez a la sistematización de un trato diferenciado y potencialmente discriminatorio.

En otras palabras, al igual que las personas, los algoritmos discriminan, por lo que es importante aclarar cuál es la diferencia que justifica la urgencia del debate, especialmente en torno a la necesidad de adoptar nuevas políticas públicas con mecanismos reforzados de control y rendición de cuentas.

El riesgo más inmediato es la escalabilidad de estos sistemas, que puede conducir a una sistematización de los sesgos y, en consecuencia, a la discriminación. Sin embargo, quizá el riesgo más diferenciador esté asociado al denominado efecto "***black box***" del algoritmo. Mientras que, en principio, un programador debería ser capaz de explicar qué se supone que hace un *software*, con los modelos de aprendizaje ("*machine learning*"), que acaban por autoprogramarse, los resultados pueden no ser tan fáciles de entender, interpretar o explicar. Como hemos visto en el caso inicial, sigue habiendo una enorme falta de progreso en la interpretabilidad de los algoritmos, un factor que, combinado con la evolución exponencial de la tecnología, ha contribuido en gran medida a la defensa de un camino hacia la regulación.

En definitiva, al carecer de transparencia y a menudo estar protegidos por normas de secreto comercial, detectar y probar situaciones de discriminación resulta extremadamente difícil. Todo ello en el marco de un ecosistema digital frágil en términos de prevención y protección, que a menudo coloca la carga de la responsabilidad en los más vulnerables.

A esto se añade el **problema de las expectativas y la calificación de los contenidos**. Es lo que se conoce como sesgo de automatización y sesgo de confirmación, que se traduce no solo en una confianza ciega en los algoritmos, percibidos como certezas matemáticas, sino también en una disminución del espíritu

sempre que os resultados confirmam as nossas expetativas, indo ao encontro de ideias pré-concebidas ('**assessment bias**'). No contexto de uma cultura laboral cada vez mais focada em objetivos de produtividade, que faz do tempo um fator de pressão, dificilmente a utilização de ferramentas de IA para certos processos decisórios, como por exemplo o recrutamento, serão objeto de dúvida, ou mesmo escrutínio, por parte dos utilizadores.

Ora, se os **desafios são claros, o mesmo não se pode dizer das soluções**, tanto no quadro das estratégias e políticas empresariais empreendidas, como do respetivo enquadramento legal.

Exemplificando. Uma das abordagens da OpenAI para reduzir o risco de acidente e mau uso é o ***'reinforcement learning from human feedback'*** (RLHF), que basicamente consiste em perguntar a opinião dos utilizadores sobre a adequabilidade da resposta do sistema ao pedido formulado. A fragilidade da estratégia é evidente, não só porque dependente do espírito crítico do utilizador, cujos níveis de consciência e participação são normalmente baixos, como acaba por remeter para avaliações subjetivas e não uniformes.

Outra abordagem é o chamado ***'red-teaming'***, em que as empresas contratam equipas, ou suscitam a participação de pessoas, cuja função é tentar quebrar os algoritmos e antecipar potenciais erros ou problemas. Sendo uma boa prática, também tem as suas limitações, nomeadamente porque não há qualquer obrigatoriedade nem de correção nem de transparência na sequência dos resultados alcançados por estas equipas. Algo dificilmente compreensível, principalmente quando o sistema já está em uso e as pessoas totalmente expostas ao risco.

Uma outra solução, cada vez mais estudada e desenvolvida, é colocar **modelos de IA a vigiar outros modelos de IA**. Alguns investigadores têm chamado a isto de 'IA Constitucional' (*'Constitutional AI'*), onde modelos secundários têm por função avaliar a conformidade constitucional dos resultados dos ditos modelos principais. A verdade é que, sendo positivo, também aqui é preciso alguma precaução. Estamos numa fase muito inicial do desenvolvimento destas tecnologias e, como já se percebeu, a correção de um algoritmo, ainda que bem-intencionada, pode acabar

crítico cuando los resultados confirman nuestras expectativas y responden a ideas preconcebidas ("**sesgo de evaluación**"). En el contexto de una cultura laboral cada vez más centrada en los objetivos de productividad, que convierte el tiempo en un factor de presión, es poco probable que el uso de herramientas de IA para determinados procesos de toma de decisiones, como la contratación, esté sujeto a dudas o incluso sometido a escrutinio por parte de los usuarios.

Si bien los **retos están claros, no puede decirse lo mismo de las soluciones**, tanto en el marco de las estrategias y políticas empresariales emprendidas, como del respectivo marco legal.

Por ejemplo. Uno de los enfoques de OpenAI para reducir el riesgo de accidentes y usos indebidos es el "**aprendizaje reforzado a partir del feedback humano**" (RLHF), que consiste básicamente en pedir a los usuarios su opinión, sobre la idoneidad de la respuesta del sistema a la petición realizada. La fragilidad de esta estrategia es evidente, no sólo porque depende del espíritu crítico del usuario, cuyos niveles de concienciación y participación suelen ser bajos, sino también porque acaba derivando en evaluaciones subjetivas y no uniformes.

Otro enfoque es el denominado "***red-teaming***", en el que las empresas contratan equipos, o fomentan la participación de personas cuya función es intentar descifrar los algoritmos y anticiparse a posibles errores o problemas. Siendo una buena práctica, también tiene sus limitaciones, sobre todo porque no hay obligación de corregir ni de ser transparente sobre los resultados obtenidos por estos equipos. Esto es algo difícil de entender, sobre todo cuando el sistema ya está en uso y las personas están completamente expuestas al riesgo.

Otra solución, cada vez más estudiada y desarrollada, consiste en **encargar a modelos de IA la supervisión de otros modelos de IA**. Algunos investigadores lo han denominado "*IA constitucional*", donde los modelos secundarios se encargan de evaluar la conformidad constitucional de los resultados de los modelos principales. Lo cierto es que, si bien esto es algo positivo, también debemos ser cautos al respecto. Nos encontramos en una fase muy temprana del desarrollo de estas tecnologías y, como ya se ha comprovado, la corrección de un algoritmo, aunque sea con buena

por se traduzir na sobrecompensação de um grupo protegido. Ou seja, uma forma de discriminação positiva que, sob pena de violação do princípio da igualdade, tem de obedecer a um conjunto rigoroso de requisitos, como o de se basear sempre numa justificação objetiva e razoável. No final, a atuação corretiva pode simplesmente deslocar a discriminação de um grupo protegido para outro grupo protegido.

Com alguma dose de ironia, se considerarmos o caminho da dogmática da igualdade, começa agora a falar-se da necessidade de se recorrer às categorias proibidas de discriminação exatamente para corrigir os enviesamentos. Todavia, é importante refletir se isto é desejável, tendo nomeadamente em conta o risco de discriminação direta, potenciais abusos e até a eficácia desta intervenção, que nunca é garantida. Se olharmos para o enquadramento legal em vigor, também não ficamos mais esclarecidos. O RGPD, que parte de uma regra de proibição de tratamento destes dados pessoais, admite algumas exceções, mas muito restritivas, como a obrigação de consentimento e o interesse público, que mesmo assim deve ser proporcional ao objetivo visado. A recém-aprovada Lei europeia de IA também não me parece poder vir a resolver o problema.

Mas nada melhor do que analisar um **caso prático e real.**

No final do ano passado uma empresa holandesa a explorar uma aplicação de encontros online dirigiu-se ao Instituto Holandês de Direitos Humanos para saber se podia ajustar o algoritmo de autoaprendizagem, de modo a que pessoas com pele escura e de origem não holandesa pudessem passar a ser apresentadas a outros usuários com a mesma frequência que pessoas com pele clara e de origem holandesa. A resposta não se fez tardar e o Instituto, sem nunca apresentar uma solução, confirmou o diagnóstico e determinou que a empresa não só estava autorizada a reajustar o algoritmo, como recaia sobre si uma verdadeira obrigação de atuação. E isto independentemente de não se conhecer a verdadeira causa para aqueles resultados.

Ora, partindo das informações à nossa disposição, tenho alguma dificuldade em acompanhar esta resposta, seja pela necessidade de sustentação do diagnóstico de discriminação, que sempre exigiria o uso de dados sensíveis, seja pela imputação total da responsabilidade da correção à empresa. Para além do problema

intención, puede acabar compensando en exceso a un grupo protegido. En otras palabras, una forma de discriminación positiva que, so pena de vulnerar el principio de igualdad, debe cumplir una serie de requisitos estrictos, como basarse siempre en una justificación objetiva y razonable. Al final, las medidas correctoras pueden simplemente trasladar la discriminación de un grupo protegido a otro.

Con cierta dosis de ironía, si tenemos en cuenta la trayectoria de la dogmática de la igualdad, ahora se habla de la necesidad de utilizar las categorías prohibidas de discriminación precisamente para corregir los sesgos. Sin embargo, es importante reflexionar sobre si esto es deseable, sobre todo teniendo en cuenta el riesgo de discriminación directa, los posibles abusos e incluso la eficacia de esta intervención, que nunca está garantizada. Si nos fijamos en el marco legal vigente, tampoco lo tenemos más claro. El Reglamento General de Protección de Datos (RGPD), que parte de una norma que prohíbe el tratamiento de estos datos personales, permite algunas excepciones, pero muy restrictivas, como la obligación de dar el consentimiento y el interés público, que debe seguir siendo proporcional al objetivo perseguido. La recién aprobada Ley Europea de IA tampoco parece capaz de resolver el problema.

Pero nada mejor que analizar un **caso práctico y real.**

A finales del año pasado, una empresa holandesa que gestionaba una aplicación de citas online se dirigió al Instituto Holandés de Derechos Humanos para consultarle si podía ajustar el algoritmo de autoaprendizaje de modo que las personas de piel oscura y origen no holandés pudieran ser presentadas a otros usuarios con la misma frecuencia que las personas de piel clara y origen holandés. La respuesta no se hizo esperar y el Instituto, sin llegar a presentar una solución, confirmó el diagnóstico y dictaminó que la empresa no sólo estaba autorizada a reajustar el algoritmo, sino que tenía la obligación real de actuar. Y ello con independencia de que se desconociera la causa real de los resultados.

Ahora bien, partiendo de la información de la que disponemos, me resulta algo complicado compartir esta respuesta, tanto por la necesidad de fundamentar el diagnóstico de discriminación, que exigiría siempre el uso de datos sensibles, como por la imputación total de la responsabilidad de la corrección a la empresa.

técnico, que em seguida referirei, de ajustar o algoritmo sem recorrer a categorias proibidas, como é o caso da etnia, qualquer intervenção resultará provavelmente numa situação de tratamento diferenciado, que para não constituir uma violação do princípio da igualdade, como já se referiu *supra*, deve fundar-se numa justificação objetiva e razoável.

Inclino-me para uma resposta negativa. Nomeadamente se considerarmos que o suposto resultado discriminatório pode decorrer simplesmente de uma sub-representação de dados, associado ao facto de o algoritmo trabalhar sobre um universo reduzido de pessoas com pele escura e não holandesas que, na Holanda, não recorrem, também por escolha, a esta aplicação. E como já se referiu, existe o risco real de a correção se limitar a deslocar a discriminação para outro grupo protegido.

E o que diz a lei ao nível do RGPD e, já agora, a recém-aprovada Lei de IA? No caso do RGPD, o processamento dos dados exigiria consentimento por parte dos utilizadores, dado de forma voluntária e explícita, o que se imagina difícil de acontecer, inviabilizando desde logo a eficácia da dita correção. No caso da Lei da IA, há de facto uma exceção, mas dentro de muitas condições restritivas e, o mais importante, apenas para sistemas de alto risco, entre os quais não constam este tipo de aplicações. O legislador europeu parece ter adotado uma posição de prudência que, mesmo nos sistemas de alto risco, como os sistemas de *'credit score'*, pode vir a suscitar dificuldades de aplicação, exigindo-se mais orientação.

Resultado: todos ficámos a conhecer o problema, mas ninguém sabe qual foi a solução. E é neste ambiente que se suspendem aplicações, faz-se um *mea culpa* público e prometem-se correções, até ao próximo problema.

Nunca desvalorizando o pilar da responsabilidade das empresas, não podemos claramente ficar dependentes da autorregulação. Há importantes decisões normativas sobre como o sistema se deve comportar, as quais não devem ser transferidas para os programadores, fornecedores e usuários, nomeadamente sobre o que é justo e discriminatório, se uma certa desvantagem é grave o suficiente que justifique intervenção ou se a discriminação indireta tem uma justificação objetiva e razoável.

Aparte del problema técnico de ajustar el algoritmo sin utilizar categorías prohibidas, como la etnia, cualquier intervención probablemente dará lugar a una situación de trato diferenciado, que, para no constituir una vulneración del principio de igualdad, como se ha mencionado antes, deberá basarse en una justificación objetiva y razonable.

Me inclino por una respuesta negativa. Sobre todo si tenemos en cuenta que el supuesto resultado discriminatorio podría ser simplemente fruto de una infrarrepresentación de los datos, unido al hecho de que el algoritmo funciona sobre un pequeño universo de personas de piel oscura y no holandesas que, en los Países Bajos, no utilizan esta aplicación por elección propia. Y como se ha mencionado existe un riesgo real de que la corrección simplemente traslade la discriminación a otro grupo protegido.

¿Y qué dice la ley a nivel del RGPD y la recientemente aprobada Ley de IA? En el caso del RGPD, el tratamiento de datos requeriría el consentimiento de los usuarios, otorgado de forma voluntaria y explícita, lo que se prevé difícil que se produzca, haciendo inviable la efectividad de dicha corrección. En el caso de la Ley de IA, sí existe una excepción, pero bajo muchas condiciones restrictivas y, lo más importante, solo para sistemas de alto riesgo entre los que no se encuentran este tipo de aplicaciones. El legislador europeo parece haber adoptado una postura cautelosa que, incluso en sistemas de alto riesgo como los de "*credit score*" o calificación crediticia podría generar dificultades de aplicación, requiriéndose más orientaciones.

El resultado: todos conocemos el problema, pero nadie conoce la solución. Y es en este contexto cuando se suspenden las aplicaciones, se hace un *mea culpa* público y se prometen correcciones hasta el siguiente problema.

Sin devaluar nunca el pilar de la responsabilidad empresarial, es evidente que no podemos depender de la autorregulación. Hay importantes decisiones normativas respecto de cómo debe comportarse el sistema que no deben transferirse a programadores, proveedores y usuarios: qué es justo y qué discriminatorio, si una determinada desventaja es lo suficientemente grave como para justificar una intervención o si la discriminación indirecta tiene una justificación objetiva y razonable.

De facto, são estes os exemplos que nos obrigam a refletir sobre o mundo que queremos e qual o nosso papel. Se desejamos as vantagens do muito que a IA pode facultar, temos que, em primeiro lugar, admitir que haverá sempre uma certa dose de insegurança e risco. Em segundo, decidir em consciência um modelo de *governance* mais adequado, o qual, a meu ver, requer seguramente mais níveis de escrutínio, públicos e privados, bem como abordagens multifacetadas que baseadas nos direitos humanos, de forma equilibrada, sejam capazes de promover a inovação, sem deixar nunca de colocar a pessoa ao centro.

En definitiva, estos son los ejemplos que nos obligan a reflexionar sobre el mundo que queremos y sobre nuestro papel. Si queremos aprovechar todo lo que la IA puede ofrecernos, debemos admitir, en primer lugar, que siempre habrá un cierto grado de inseguridad y riesgo. En segundo lugar, debemos decidir conscientemente cuál es el modelo de gobernanza más adecuado, algo que, en mi opinión, requiere sin duda mayores niveles de escrutinio, tanto público como privado, así como enfoques polifacéticos basados en los derechos humanos, que, de forma equilibrada, sean capaces de promover la innovación, sin dejar nunca de poner en el centro a las personas.

Ms Teresa Anjinho

Former Deputy Ombudsman Portugal, Independent Human Rights Expert and Chair of the Supervisory Committee of European Anti-Fraud Office - OLAF

Dña. Teresa Anjinho

Ex Subdefensora del Pueblo de Portugal, Experta Independiente en Derechos Humanos y Presidenta del Comité de Supervisión de la Oficina Europea de Lucha contra el Fraude - OLAF