

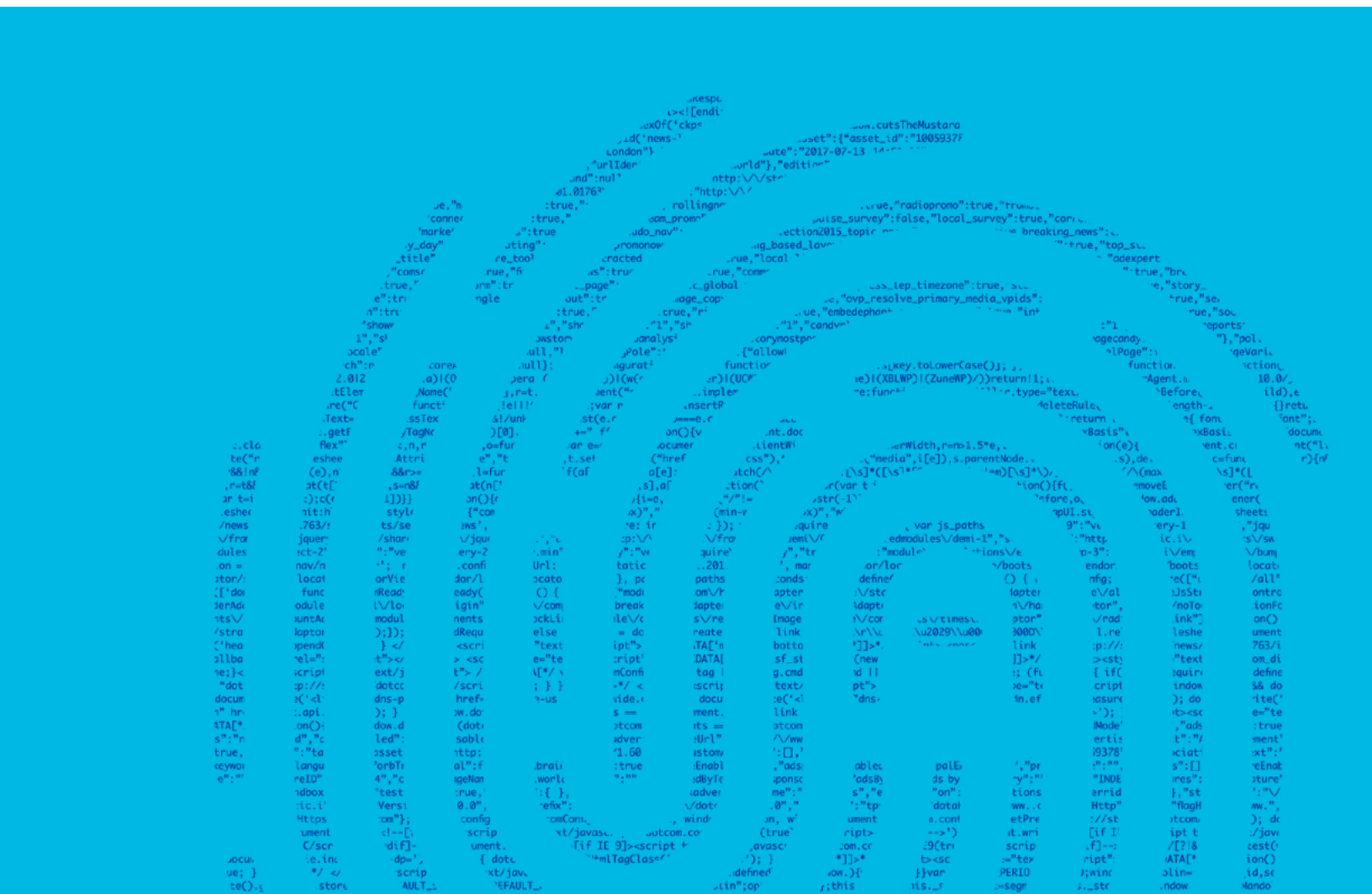


< >

Declaração de Montreal
pela IA responsável_

</ >

DECLARAÇÃO DE MONTREAL PELO DESENVOLVIMENTO RESPONSÁVEL DA INTELIGÊNCIA ARTIFICIAL 2018



Este documento é uma parte do RELATÓRIO DA DECLARAÇÃO DE MONTREAL PELO DESENVOLVIMENTO RESPONSÁVEL DA INTELIGÊNCIA ARTIFICIAL 2018.

Você encontrará o relatório completo [AQUI](#).

ÍNDICE

LER A DECLARAÇÃO DE MONTREAL PELA IA RESPONSÁVEL	5
PREÂMBULO	7

OS PRINCÍPIOS

1. PRINCÍPIOS DO BEM-ESTAR	8
2. PRINCÍPIOS DO RESPEITO À AUTONOMIA	9
3. PRINCÍPIOS DE PROTEÇÃO DA INTIMIDADE E DA VIDA PRIVADA	10
4. PRINCÍPIOS DA SOLIDARIEDADE	11
5. PRINCÍPIOS DA PARTICIPAÇÃO DEMOCRÁTICA	12
6. PRINCÍPIOS DA EQUIDADE	13
7. PRINCÍPIOS DA INCLUSÃO DA DIVERSIDADE	14
8. PRINCÍPIOS DA PRUDÊNCIA	15
9. PRINCÍPIOS DA RESPONSABILIDADE	16
10. PRINCÍPIOS DO DESENVOLVIMENTO SUSTENTÁVEL	17

DEFINIÇÃO DOS TERMOS USADOS NESTE DOCUMENTO	18
---	----

CRÉDITOS	I
PARCEIROS	II

LER A DECLARAÇÃO DE MONTREAL PELA IA RESPONSÁVEL

UMA DECLARAÇÃO PARA QUÊ?

A Declaração de Montreal pelo Desenvolvimento Responsável da Inteligência Artificial persegue 3 objetivos:

1. Elaborar um quadro ético para o desenvolvimento e a implementação da IA;
2. Orientar a transição digital para que todos possam se beneficiar dessa revolução tecnológica;
3. Abrir um espaço para o diálogo nacional e internacional de modo a obter sucesso coletivo no desenvolvimento inclusivo, equitativo e ecologicamente sustentável da IA.

UMA DECLARAÇÃO DE QUÊ?

PRINCÍPIOS

O objetivo principal desta Declaração é identificar princípios e valores éticos que promovam os interesses fundamentais de indivíduos e grupos. Estes princípios aplicados aos domínios digital e da inteligência artificial permanecem gerais e abstratos. Para lê-los adequadamente, os seguintes pontos devem ser mantidos em mente:

- > Embora estejam apresentados em forma de lista, eles não estão hierarquizados. O último princípio não é menos importante que o primeiro. Mas é possível, conforme as circunstâncias, atribuir mais peso a um princípio do que a outro, ou considerar que um princípio é mais pertinente que outro.
- > Embora sejam diversos, devem ser interpretados de forma coerente, a fim de evitar qualquer conflito que impeça sua aplicação. Em geral, os limites da aplicação de um princípio são traçados pelo domínio de aplicação de outro princípio.
- > Embora reflitam a cultura moral e política da sociedade em que foram elaborados, fornecem uma base para o diálogo intercultural e internacional.
- > Embora possam ser interpretados de maneiras diferentes, eles não podem ser interpretados de qualquer maneira. É imperativo que as interpretações sejam coerentes.
- > Embora estes sejam princípios éticos, eles podem ser traduzidos em linguagem política e interpretados de maneira jurídica.

A partir desses princípios foram elaboradas recomendações cujo objetivo é fornecer diretrizes para realizar a transição digital dentro do marco ético da Declaração. Elas cobrem alguns temas intersetoriais fundamentais para pensar sobre a transição para uma sociedade na qual a IA promova o bem comum: governança algorítmica, alfabetização digital, inclusão digital da diversidade e sustentabilidade ecológica.

UMA DECLARAÇÃO PARA QUEM?

A Declaração de Montreal é dirigida a qualquer pessoa, organização da sociedade civil ou empresa que pretenda participar do desenvolvimento da inteligência artificial de forma responsável, seja para contribuir científica e tecnologicamente, para desenvolver projetos sociais, para elaborar regras (regulamentos, códigos) que se apliquem a ela, para contestar orientações erradas ou imprudentes, ou ainda para poder emitir alertas à opinião pública quando necessário.

É dirigida também aos responsáveis políticos, eleitos ou nomeados, de quem os cidadãos esperam que tomem decisões e adotem medidas voltadas para as mudanças sociais em gestação, que coloquem rapidamente em prática ações para a transição digital voltada para o bem de todos e também que antecipem os sérios riscos apresentados pelo desenvolvimento da IA.

UMA DECLARAÇÃO QUE ADOTA QUAL MÉTODO?

A Declaração é o resultado de um processo deliberativo inclusivo que envolve o diálogo entre cidadãos, especialistas, responsáveis públicos, partes interessadas do setor industrial, organizações da sociedade civil e associações profissionais. O interesse deste encaminhamento é triplo:

1. **Arbitrar coletivamente as controvérsias éticas e societais sobre a IA;**
2. **Melhorar a qualidade da reflexão sobre a IA responsável;**
3. **Fortalecer a legitimidade das propostas para uma IA responsável.**

A elaboração dos princípios e das recomendações foi um trabalho de construção cooperativa, que envolveu diversos participantes em locais públicos, em salas de reuniões de organizações profissionais, em mesas redondas com especialistas internacionais, em laboratórios de pesquisa, em salas de aula e em encontros online via Internet, sempre com o mesmo rigor.

E DEPOIS DA DECLARAÇÃO?

Como a Declaração se concentra em uma tecnologia que vem sendo constantemente aprimorada desde a década de 1950, e cujo ritmo de inovações importantes está se acelerando exponencialmente, é essencial conceber a Declaração como um documento com diretrizes abertas, adaptáveis e passíveis de revisão, em vista da evolução dos conhecimentos e das técnicas, assim como do feedback sobre o uso da IA na sociedade. Ao final do processo de elaboração da Declaração, chegamos ao ponto de partida para uma conversa aberta e inclusiva sobre o futuro da humanidade atendida pelas tecnologias da Inteligência Artificial.

PREÂMBULO

Pela primeira vez na história da humanidade é possível criar sistemas autônomos capazes de realizar tarefas complexas que se pensava serem reservadas para a inteligência natural: processar grandes quantidades de informação, calcular e prever, aprender e adaptar suas respostas a situações em mudança, além de reconhecer e classificar os objetos.

Devido à natureza imaterial das tarefas que desempenham, e por analogia com a inteligência humana, designa-se esses sistemas muito diversos pelo termo geral de “Inteligência Artificial”. A inteligência artificial é um grande avanço científico e tecnológico, que pode gerar benefícios sociais consideráveis, melhorando as condições de vida, de saúde e de acesso à justiça, criando riqueza, fortalecendo a segurança pública ou controlando o impacto das atividades humanas sobre o meio ambiente e o clima. Máquinas inteligentes não são apenas capazes de computações superiores às de seres humanos, elas podem fazer companhia, cuidar e interagir com seres sencientes.

O desenvolvimento da inteligência artificial, no entanto, apresenta grandes desafios éticos e riscos sociais. De fato, máquinas inteligentes podem restringir as escolhas de indivíduos e grupos, diminuir a qualidade de vida, perturbar a organização do trabalho e do mercado de trabalho, influenciar a vida política, entrar em conflito com os direitos fundamentais, exacerbar desigualdades econômicas e sociais e afetar os ecossistemas, o meio ambiente e o clima. Embora não haja progresso científico nem vida social sem riscos, cabe aos cidadãos determinar as finalidades morais e políticas que dão sentido aos riscos em um mundo incerto.

Os benefícios da inteligência artificial serão maiores se os riscos associados à sua implementação forem pequenos. No entanto, o principal perigo apresentado pelo desenvolvimento da inteligência

artificial consiste em dar a ilusão de que o futuro é controlado pela computação. Reduzir a sociedade a números e governá-la por meio de procedimentos algorítmicos é um sonho antigo que ainda alimenta as ambições humanas. Mas nos assuntos humanos, o amanhã raramente se parece com os dias de hoje, e os números não podem determinar o que tem valor moral, nem o que é socialmente desejável.

Os princípios desta Declaração são os rumos de uma bússola ética que orienta o desenvolvimento da inteligência artificial em direção a fins moral e socialmente desejáveis. Eles também fornecem uma estrutura ética para promover os direitos humanos reconhecidos internacionalmente nas áreas de implementação de inteligência artificial. Tomados como um todo, os princípios formulados estabelecem as bases para a confiança social em sistemas artificialmente inteligentes.

Os princípios desta Declaração baseiam-se no entendimento comum de que os seres humanos procuram florescer como seres sencientes, dotados de emoções e pensamentos, e se esforçam para realizar o seu potencial, exercendo livremente suas capacidades afetivas, morais e intelectuais. É responsabilidade de diferentes atores públicos e privados e de tomadores de decisão nos níveis local, nacional e internacional assegurar que o desenvolvimento e a implementação da inteligência artificial sejam compatíveis com a proteção e o desabrochar das capacidades humanas mais fundamentais. É de acordo com este objetivo que os princípios propostos devem ser interpretados de maneira coerente, levando em conta os contextos sociais, culturais, políticos e jurídicos específicos de sua aplicação.

1

PRINCÍPIOS DO BEM-ESTAR

O desenvolvimento e o uso de Sistemas de Inteligência Artificial (SIAs) devem permitir aumentar o bem-estar de todos os seres sencientes.

1. Os SIAs devem permitir que os indivíduos melhorem suas condições de vida, de saúde e de trabalho.
2. Os SIAs devem permitir que os indivíduos satisfaçam suas preferências, dentro dos limites do que não cause danos a outro ser senciente.
3. Os SIAs devem permitir que os indivíduos exerçam suas capacidades físicas e intelectuais.
4. Os SIAs não devem ser fonte de desconforto, a menos que este último possa gerar um maior bem-estar que não possamos alcançar de outra forma.
5. O uso de SIAs não deve contribuir para o aumento do estresse, da ansiedade e de sentimentos de assédio ligados ao ambiente digital.



2

PRINCÍPIOS DO RESPEITO À AUTONOMIA

Os SIAs devem ser desenvolvidos e usados respeitando-se a autonomia das pessoas e com o objetivo de aumentar o controle, pelos indivíduos, de sua vida e de seu meio ambiente.

1. Os SIAs devem capacitar os indivíduos a realizar seus próprios objetivos morais e sua concepção de uma vida digna de ser vivida.
2. Os SIAs não devem ser desenvolvidos ou usados para prescrever aos indivíduos um modo particular de vida, direta ou indiretamente, implementando mecanismos restritivos de monitoramento, avaliação ou incitação.
3. As instituições públicas não devem usar os SIAs para promover ou para desvalorizar uma concepção do que seja uma boa vida.
4. É essencial capacitar os cidadãos em relação às tecnologias digitais, garantindo o acesso a diferentes tipos de conhecimento relevantes, ao desenvolvimento de competências estruturantes (alfabetização digital e midiática) e a formação do pensamento crítico.
5. Os SIAs não devem ser desenvolvidos para propagar informações não confiáveis, mentiras e propaganda, e devem ser projetados com o propósito de reduzir tal propagação.
6. O desenvolvimento dos SIAs deve evitar criar dependências por meio de técnicas de captação da atenção e de imitação da aparência humana, que possam induzir a uma confusão entre SIAs e seres humanos.

3

A privacidade e a intimidade devem ser protegidas de intrusão por SIAs e por Sistemas de Aquisição e Arquivamento de Dados Pessoais (SAADs).

PRINCÍPIOS DE PROTEÇÃO DA INTIMIDADE E DA VIDA PRIVADA

1. Espaços de intimidade em que as pessoas não são sujeitas a vigilância ou avaliação digital devem ser protegidos da intrusão de SIAs ou de SAADs – Sistemas de Aquisição e Arquivamento de Dados Pessoais.
2. A intimidade do pensamento e das emoções deve ser estritamente protegida contra o uso de SIAs e SAADs que poderiam lhe causar danos, em particular quando usados para julgar moralmente pessoas ou suas escolhas de vida.
3. As pessoas devem sempre ter a opção de desconexão digital em relação à sua vida privada, e os SIAs devem oferecer explicitamente a possibilidade de escolha da desconexão a intervalos regulares, sem incitar o indivíduo a permanecer conectado.
4. As pessoas devem ter amplo controle sobre as informações relativas às suas preferências. Os SIAs não devem criar perfis de preferências individuais que visem influenciar o comportamento dos envolvidos sem seu consentimento livre e informado.
5. Os SAADs devem garantir a confidencialidade dos dados e o anonimato de perfis pessoais.
6. Toda pessoa deve ser capaz de manter um amplo controle sobre seus dados pessoais, em particular no que diz respeito à sua coleta, uso e disseminação. O uso de SIAs e de serviços digitais não pode estar condicionado à renúncia aos direitos de propriedade sobre seus dados pessoais.
7. Qualquer pessoa pode doar seus dados pessoais a organizações de pesquisa para contribuir para o progresso do conhecimento.
8. A integridade da identidade individual deve ser garantida. Os SIAs não devem ser usados para imitar ou modificar a aparência física, voz e outras características individuais com o propósito de prejudicar a reputação de uma pessoa ou manipular outras pessoas.

4

PRINCÍPIOS DA SOLIDARIEDADE

O desenvolvimento de SIAs deve ser compatível com a manutenção de relações solidárias entre pessoas e gerações.

1. Os SIAs não devem prejudicar a preservação de relações afetivas e morais que floresçam entre as pessoas, e devem ser desenvolvidos com o objetivo de fomentá-las de modo a reduzir a vulnerabilidade e o isolamento das pessoas.
2. Os SIAs devem ser desenvolvidos para colaborar com os seres humanos em tarefas complexas e devem fomentar o trabalho colaborativo entre humanos.
3. Os SIAs não devem ser implementados para substituir pessoas em tarefas que exigem relacionamento humano de qualidade, mas devem ser desenvolvidos para facilitar essas relações.
4. Os sistemas de saúde que usam SIAs devem levar em consideração a importância, para os pacientes, das relações com equipe médica e a família.
5. O desenvolvimento de SIAs não deve estimular comportamentos cruéis com robôs que se apresentam como seres humanos ou animais e que pareçam agir como eles.
6. Os SIAs devem ajudar a melhorar o gerenciamento de riscos e criar as condições para uma sociedade mais eficaz de compartilhamento de riscos individuais e coletivos.

5

PRINCÍPIOS DA PARTICIPAÇÃO DEMOCRÁTICA

Os SIAs devem atender aos critérios de inteligibilidade, justificabilidade e acessibilidade, e devem estar sujeitos a escrutínios, debates e controles democráticos.

1. O funcionamento dos SIAs que tomam decisões que afetam a vida, a qualidade de vida ou a reputação dos indivíduos deve ser inteligível para seus desenvolvedores.
2. As decisões dos SIAs que afetam a vida, a qualidade de vida ou a reputação dos indivíduos devem sempre ser justificadas em linguagem compreensível para aqueles que os utilizam ou que sofrem as consequências de seu uso. As justificativas consistem em explicar os fatores e parâmetros mais importantes para a tomada de uma decisão, e devem ser semelhantes às justificativas que seriam exigidas de um ser humano que tomasse o mesmo tipo de decisão.
3. O código dos algoritmos, públicos ou privados, deve estar sempre acessível às autoridades públicas competentes e às partes interessadas para fins de verificação e controle.
4. A descoberta de erros operacionais nos SIAs, de seus efeitos imprevistos ou indesejados, de violações de segurança e vazamentos de dados, deve ser obrigatoriamente relatada às autoridades públicas, às partes interessadas e às pessoas afetadas pela situação.
5. De acordo com a exigência de transparência nas decisões públicas, o código de algoritmos de decisão utilizado pelas autoridades públicas deve ser acessível a todos, com exceção dos algoritmos que apresentam, em caso de uso indevido, alta probabilidade de perigo sério.
6. Para que os SIAs públicos tenham um impacto significativo na vida dos cidadãos, estes devem ter a oportunidade e a competência para deliberar sobre os seus parâmetros sociais, objetivos e limites de uso.
7. Deve ser assegurado em todos os momentos que os SIAs façam aquilo para o qual foram programados e para o qual devem ser utilizados.
8. Todo usuário de um serviço deve saber se uma decisão que lhe diz respeito ou que o afete foi tomada por um SIA.
9. Todo usuário de um serviço que use agentes de conversação deve ser capaz de identificar facilmente se está interagindo com um SIA ou com uma pessoa.
10. A pesquisa no campo da inteligência artificial deve permanecer aberta e acessível a todos.

6

PRINCÍPIOS DA EQUIDADE

O desenvolvimento e o uso dos SIAs devem contribuir para a obtenção de uma sociedade justa e equitativa.

1. Os SIAs devem ser projetados e treinados de forma a não criar, reforçar ou reproduzir discriminação baseada em diferenças sociais, sexuais, étnicas, culturais e religiosas, entre outras.
2. O desenvolvimento de SIAs deve contribuir para eliminar as relações de dominação entre pessoas e grupos com base na diferença de poder, riqueza ou conhecimento.
3. O desenvolvimento de SIAs deve beneficiar econômica e socialmente a todos, de modo a reduzir a precariedade e as desigualdades sociais.
4. O desenvolvimento industrial de SIAs deve ser compatível com condições de trabalho dignas, em todas as fases do seu ciclo de vida, desde a extração dos recursos naturais até sua reciclagem, passando pelo processamento de dados.
5. A atividade digital de usuários de serviços digitais e SIAs deve ser reconhecida como um trabalho que contribui para o funcionamento de algoritmos e agrega valor.
6. O acesso aos recursos, conhecimentos e ferramentas digitais de base deve ser garantido a todos.
7. O desenvolvimento de “comuns algorítmicos” – a concepção e gestão de ferramentas digitais pelos usuários-cidadãos – assim como o de dados livres, com o objetivo de treiná-los e operá-los, é uma meta socialmente equitativa que deve ser apoiada.

7

PRINCÍPIOS DA INCLUSÃO DA DIVERSIDADE

O desenvolvimento e o uso de SIAs devem ser compatíveis com a manutenção da diversidade social e cultural e não devem restringir o leque de escolhas de vida e experiências pessoais.

1. O desenvolvimento e o uso de SIAs não devem levar a uma padronização da sociedade através da normalização de comportamentos e opiniões.
2. O desenvolvimento e a implantação de SIAs devem levar em conta as múltiplas expressões de diversidade social e cultural, e isso deve ocorrer desde a concepção dos algoritmos.
3. Os ambientes de pesquisa em IA, tanto na investigação científica quanto na indústria, devem ser inclusivos e refletir a diversidade de indivíduos e grupos na sociedade.
4. Os SIAs devem evitar o confinamento de indivíduos em um perfil de usuário ou em uma “bolha filtradora”, evitar definir identidades pessoais por meio do processamento de dados obtidos a partir de suas atividades anteriores, e também evitar a redução de suas opções de desenvolvimento pessoal, especialmente nas áreas da educação, da justiça e das práticas empresariais.
5. Os SIAs não devem ser usados ou desenvolvidos para limitar a liberdade de expressar ideias e de comunicar opiniões, cuja diversidade é a condição da vida democrática.
6. Para cada categoria de serviço, a oferta dos SIAs deve ser diversificada para que os monopólios de fato não se constituam e não prejudiquem as liberdades individuais.

8

PRINCÍPIOS DA PRUDÊNCIA

Todas as pessoas envolvidas no desenvolvimento de SIAs devem ser cautelosas, antecipando, tanto quanto possível, as consequências adversas do uso dos SIAs, e tomando as medidas apropriadas para evitá-las.

1. É necessário desenvolver mecanismos que levem em conta o potencial de uso duplo (benéfico e prejudicial) da pesquisa em IA (tanto pública quanto privada) e também a possibilidade de desenvolvimento de SIAs para limitar seu uso prejudicial.
2. Quando o uso inapropriado de um SIA puder representar uma séria ameaça à segurança ou saúde pública, com alta probabilidade, é prudente restringir a disseminação pública ou o acesso aberto ao seu algoritmo.
3. Antes de serem colocados no mercado, sejam eles pagos ou gratuitos, os SIAs devem atender a critérios rigorosos de confiabilidade, segurança e integridade, e estar sujeitos a testes que não ponham em risco a vida das pessoas, não afetem sua qualidade de vida, nem prejudiquem sua reputação ou integridade psicológica. Estes testes devem estar abertos às autoridades públicas competentes e às partes interessadas.
4. O desenvolvimento de SIAs deve prevenir os riscos de uso nocivo dos dados do usuário e proteger a integridade e a confidencialidade dos dados pessoais.
5. Os erros e vulnerabilidades descobertos nos SIAs e SAADs devem ser compartilhados publicamente e em escala mundial por instituições públicas e empresas, em setores que representem uma ameaça significativa à integridade pessoal e à organização social.

9

PRINCÍPIOS DA RESPONSABILIDADE

O desenvolvimento e o uso de SIAs não devem contribuir para a desresponsabilização dos seres humanos quando uma decisão vier a ser tomada.

1. Somente os seres humanos podem ser responsabilizados por decisões decorrentes de recomendações feitas por SIAs e pelas ações decorrentes delas.
2. Em todas as áreas onde deva ser tomada uma decisão que afeta a vida, a qualidade de vida ou a reputação de uma pessoa, além de a decisão final dever recair sobre ser humano, deve ser livre e informada.
3. A decisão de matar deve sempre ser tomada por seres humanos e a responsabilidade por esta decisão não pode ser transferida para um SIA.
4. Pessoas que autorizem um SIA a cometer um crime ou delito, ou que sejam negligentes ao permiti-los, são responsáveis por eles.
5. No caso de um erro ter sido infligido por um SIA, e o SIA se provar confiável e tiver sido usado de maneira normal, não é razoável culpar as pessoas envolvidas em seu desenvolvimento ou uso.

10

O desenvolvimento e o uso de SIAs devem ser realizados de forma a garantir uma forte sustentabilidade ecológica do planeta.

PRINCÍPIOS DO DESENVOLVIMENTO SUSTENTÁVEL

1. Os equipamentos que usam SIAs, sua infraestrutura digital e os objetos conectados a eles e dos quais eles dependem, como centros de dados, devem buscar a mais alta eficiência energética e minimizar as emissões de gases de efeito estufa (GEE) ao longo de todo o seu ciclo de vida.
2. Os equipamentos de SIAs, suas infraestruturas digitais e os objetos conectados sobre as quais se apoiam, devem ter como objetivo gerar um mínimo de resíduos elétricos e eletrônicos, e prever procedimentos de manutenção, de reparo e reciclagem com economia de custos dentro de uma lógica de economia circular.
3. Os equipamentos de SIAs, suas infraestruturas digitais e os objetos conectados sobre as quais se apoiam, devem minimizar os impactos nos ecossistemas e na biodiversidade em todas as fases de seu ciclo de vida, particularmente durante a extração de recursos naturais e nas etapas do fim da vida útil.
4. Os atores públicos e privados devem apoiar o desenvolvimento de SIAs ambientalmente responsáveis, a fim de combater o desperdício de recursos naturais e bens produzidos, de estabelecer cadeias de fornecimento e comércio sustentáveis e de reduzir a poluição ambiental em escala planetária.

DEFINIÇÃO DOS TERMOS USADOS NESTE DOCUMENTO

AGENTE CONVERSACIONAL (chatbot)

Um agente de conversação é um sistema de IA que pode interagir com seu usuário em linguagem natural.

ALGORITMO

Um algoritmo é um método de resolução de problemas por meio de uma sequência finita e inequívoca de operações. Mais precisamente no campo da inteligência artificial, diz respeito a submeter os dados de entrada a um conjunto de operações para chegar a um resultado desejado.

ALFABETIZAÇÃO DIGITAL

A alfabetização digital de um indivíduo se refere à capacidade de acessar, gerenciar, entender, integrar, comunicar, avaliar e criar informações de forma segura e apropriada por meio de ferramentas digitais e de tecnologias em redes, de forma a permitir a sua participação na vida econômica e social.

APRENDIZADO DE MÁQUINA (learning machine)

É o ramo da inteligência artificial que consiste em programar um algoritmo para que este aprenda por si mesmo. Entre a multiplicidade de técnicas disponíveis, existem três tipos principais de aprendizado de máquina:

- > No aprendizado supervisionado, o sistema de inteligência artificial (SIA) aprende a prever um valor a partir de um dado de entrada. Isso requer que os pares de valores de entrada e de saída sejam anotados durante o treinamento. Por exemplo, um sistema pode aprender a reconhecer o objeto em uma foto;

- > No aprendizado não supervisionado, o SIA aprende a encontrar semelhanças entre os dados que não foram anotados, por exemplo, para separá-los em diferentes populações homogêneas. Assim, um sistema pode, por exemplo, reconhecer comunidades de usuários de redes sociais;
- > No aprendizado por reforço, o SIA aprende a agir em seu ambiente, de modo a maximizar a obtenção de uma recompensa que lhe é dada durante o treinamento. Essa é a técnica com a qual os SIAs puderam vencer seres humanos no jogo do Go ou no videogame Dota2.

APRENDIZADO PROFUNDO (deep learning)

Aprendizado profundo é o ramo do aprendizado de máquina que usa redes neurais artificiais de múltiplos níveis. Esta é a tecnologia por trás dos últimos avanços em IA.

ATIVIDADE DIGITAL

Refere-se a todas as ações tomadas por um indivíduo em um ambiente digital, seja em um computador, telefone ou qualquer outro objeto a eles conectado.

BENS DIGITAIS COMUNS (digital commons)

Bens digitais comuns são aplicativos ou dados produzidos por uma comunidade. Ao contrário dos ativos físicos, eles são facilmente compartilháveis e não se deterioram quando usados. Assim, ao contrário do software privado, o software de código aberto – que é frequentemente o resultado da colaboração entre programadores – é um bem digital comum porque seu código-fonte é aberto, isto é, acessível a todos.

BOLHA DE FILTRO (filter bubble)

A expressão "bolha de filtro" (ou bolha filtrante) refere-se à informação "filtrada" que atinge um indivíduo quando este está na Internet. De fato, vários serviços, como redes sociais ou mecanismos de busca, oferecem resultados personalizados para seus usuários. Isso pode ter o efeito de isolar indivíduos (em "bolhas"), uma vez que eles não têm mais acesso a informações comuns a todos os usuários.

CONFIABILIDADE

Um sistema de IA é confiável quando ele executa a tarefa para a qual ele foi concebido, e isto da maneira esperada. A confiabilidade é a probabilidade de sucesso que varia entre 51% e 100%, isto é, que é estritamente superior ao acaso. Quanto mais um sistema é confiável, mais seu comportamento é previsível.

DADOS ABERTOS (open data)

Dados abertos são os dados digitais que os usuários podem acessar livremente. Este é, por exemplo, o caso da maioria dos resultados de pesquisa publicados em IA.

DADOS PESSOAIS

Dados pessoais são aqueles que permitem a identificação de um indivíduo, direta ou indiretamente.

DEPENDÊNCIA DE CAMINHOS

Um mecanismo social pelo qual as decisões tecnológicas, organizacionais ou institucionais, consideradas racionais no passado, mas ineficazes no presente, continuam a influenciar a tomada de decisões. Um mecanismo mantido por causa de um viés cognitivo ou porque sua mudança levaria a um custo ou esforço muito alto. Este é, por exemplo, o caso das infraestruturas rodoviárias urbanas, quando são alvo de programas de otimização

da circulação, em vez de sofrerem mudanças destinadas a organizar uma mobilidade urbana associada a emissões menores de carbono. Esse mecanismo deve ser notado quando se usa IA em projetos sociais, pois dados de treinamento supervisionados podem, às vezes, reforçar paradigmas organizacionais antigos, cuja relevância é questionada nos dias de hoje.

DESCONEXÃO DIGITAL

Entende-se por desconexão digital o desligamento temporário ou permanente, por um indivíduo, de sua atividade digital.

DESENVOLVIMENTO SUSTENTÁVEL

Desenvolvimento sustentável designa o desenvolvimento de sociedades humanas de forma compatível com a capacidade dos sistemas naturais de fornecer os recursos e serviços necessários a essas sociedades. É um desenvolvimento econômico e social que atende às necessidades das pessoas de hoje sem comprometer a vida das gerações futuras.

EFEITO REBOTE

O efeito rebote é o mecanismo pelo qual uma maior eficiência energética ou um melhor desempenho ambiental de bens, equipamentos e serviços leva a um aumento desproporcionalmente maior em seu consumo. Por exemplo, o aumento no tamanho das telas, o crescimento na quantidade de dispositivos eletrônicos nas residências e o aumento das distâncias percorridas de carro ou avião. Globalmente isso resulta em maior pressão sobre os recursos e o meio ambiente.

GAN

Acrônimo de Generative Adversarial Network. Em um GAN, duas redes antagônicas competem entre si para gerar uma imagem. Elas podem ser, por exemplo, usadas para criar uma imagem, uma gravação ou um vídeo que pareçam quase reais para um ser humano.

INTELIGÊNCIA ARTIFICIAL (IA)

Inteligência Artificial designa todas as técnicas que permitem que uma máquina simule a inteligência humana, especialmente para aprender, prever, tomar decisões e perceber o mundo ao redor. No caso de um sistema de computador, o termo inteligência artificial é aplicado aos dados digitais.

INTELIGIBILIDADE

Um sistema de IA é inteligível quando um ser humano com o conhecimento necessário pode entender seu funcionamento, ou seja, seu modelo matemático e os processos que o determinam.

JUSTIFICABILIDADE DE UMA DECISÃO

A decisão de um sistema de IA é justificada quando existem razões não-triviais que a motivem e quando essas razões são transmissíveis em linguagem natural.

SISTEMA DE AQUISIÇÃO E DE ARQUIVAMENTO DE DADOS (SAAD)

Um SAAD é qualquer sistema de computador que possa coletar e salvar dados. Esses dados poderão ser usados para o desenvolvimento de um SIA ou como parâmetros para uma tomada de decisão.

SISTEMA DE INTELIGÊNCIA ARTIFICIAL (SIA)

Um sistema de IA designa qualquer sistema de computador usando algoritmos de inteligência artificial, seja um software, um objeto conectado ou um robô.

SUSTENTABILIDADE ECOLÓGICA FORTE

A noção de sustentabilidade ecológica forte remete à ideia de que, para ser sustentável, a taxa de consumo de recursos naturais e as emissões de poluentes devem ser compatíveis com os limites ambientais planetários, com o ritmo de renovação de recursos e ecossistemas, bem como com a estabilidade climática. Ao contrário da sustentabilidade fraca, menos exigente, a sustentabilidade forte não admite a substituição das perdas de recursos naturais por capital artificial.

TREINAMENTO

Treinamento é o processo de aprendizado de máquina durante o qual o SIA constrói um modelo a partir dos dados fornecidos. As performances do SIA vão depender da qualidade do modelo, que por sua vez depende da quantidade e da qualidade dos dados utilizados durante o treinamento.

CRÉDITOS

A redação da Declaração de Montreal pelo desenvolvimento responsável da inteligência artificial é o fruto do trabalho de uma equipe científica multidisciplinar e interuniversitária, que se baseou em um processo cidadão de consulta a especialistas e partes interessadas no desenvolvimento da IA.

Christophe Abrassart, professor associado da École de design e co-diretor do Lab Ville Prospective na Faculdade de Planejamento da Université de Montréal, membro do Centro de pesquisa em ética (CRÉ)

Yoshua Bengio, professor titular do Departamento de informática e pesquisa operacional (DIRO) da Université de Montréal, diretor científico do Mila e do IVADO

Guillaume Chicoisne, diretor dos programas científicos, IVADO

Nathalie de Marcellis-Warin, presidente diretora geral do Centro interuniversitário de pesquisa em análise das organizações (CIRANO), professora titular da Politécnica de Montreal

Marc-Antoine Dilhac, professor associado do Departamento de Filosofia da Université de Montréal; diretor do eixo Ética e Política, Centro de pesquisa em ética; diretor do Instituto de Filosofia Cidadania e Juventude; Cátedra de pesquisa do Canadá em Ética pública e teoria política

Sébastien Gambs, professor de informática na UQAM, Cátedra de pesquisa do Canadá em análise respeitosa da vida privada e ética dos dados massivos

Vincent Gauthrais, professor titular na Faculdade de Direito da Université de Montréal; diretor do Centro de pesquisa em direito público (CRDP)

Martin Gibert, conselheiro em ética para IVADO e pesquisador do Centro de pesquisa em ética

Lyse Langlois, professora titular e vice-diretora de pesquisa, Faculdade de Ciências Sociais, Departamento de relações industriais, Université Laval; diretora do Instituto de ética aplicada (IDÉA), pesquisadora do Centro interuniversitário sobre a globalização do trabalho (GRIMT)

François Laviolette, professor titular do Departamento de informática e engenharia de programas da Université Laval, diretor do Centro de pesquisa de dados massivos (CRDM)

Pascale Lehoux, professor titular da Escola de Saúde Pública da Université de Montréal (ESPUM); Cátedra da Université de Montréal sobre a inovação responsável em saúde

Jocelyn Maclure, professor titular da Faculdade de Filosofia da Université Laval; presidente da Comissão de ética em ciência e tecnologia (CEST)

Marie Martel, professora adjunta na Escola de biblioteconomia e ciências da informação, Université de Montréal

Joëlle Pineau, professora associada na School of Computer Science da Université McGill, diretora do Laboratório de IA do Facebook, co-diretora do Laboratório Reasoning and Learning

Peter Railton, Gregory S. Kavka Distinguished University Professor; John Stephenson Perrin Professor; Arthur F. Thurnau Professor, Departamento de filosofia da Universidade de Michigan e membro da Academia americana de artes e ciências

Catherine Régis, professora associada da Faculdade de Direito da Université de Montréal; Cátedra de pesquisa do Canadá sobre a cultura colaborativa em direito e políticas da saúde; pesquisadora regular do Centro de pesquisa em direito público (CRDP)

Christine Tappolet, professora titular do Departamento de Filosofia da Université de Montréal, diretora do Centro de pesquisa em ética (CRÉ), responsável pelo grupo interuniversitário sobre a normalidade (GRIN)

Nathalie Voarino, coordenadora científica da Declaração, candidata ao doutorado em Ciências biomédicas, opção Bioética, Université de Montréal

PARCEIROS

Université 
de Montréal



CENTRE DE RECHERCHE EN ETHIQUE



ICRA
Programme
IA et
société



Québec 
Fonds de recherche – Nature et technologies
Fonds de recherche – Santé
Fonds de recherche – Société et culture





< / >

montrealdeclaration-responsibleai.com