Dilemas éticos em Inteligência Artificial

DISCIPLINA DIREITO E SOFTWARE
IME-USP

Enrico Roberto
Pesquisador no InternetLab e Lawgorithm. Doutorando na USP.
LL.M. na Universidade de Munique (LMU). Advogado.

AGENDA

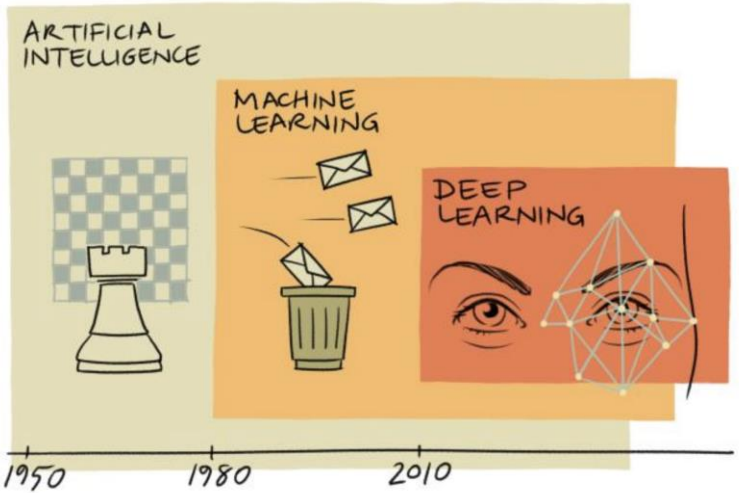
1. Inteligência Artificial e Aprendizado de Máquina

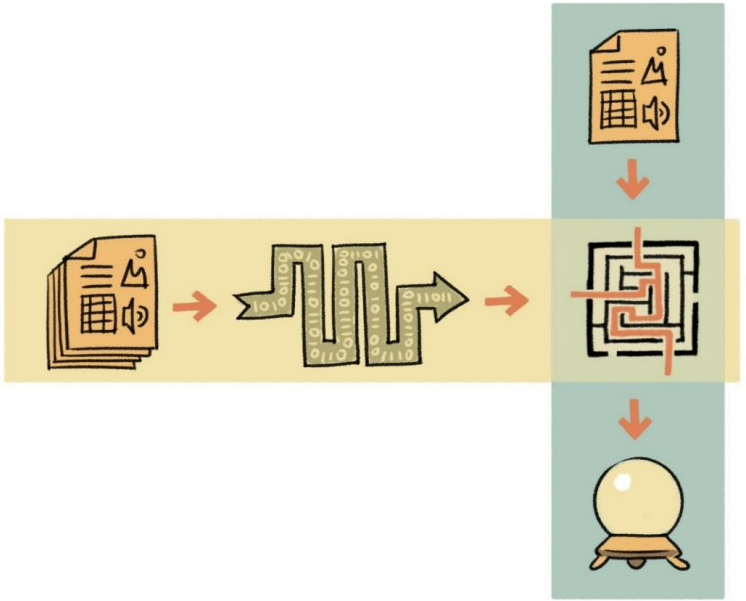
2. Neutralidade, ou IA como processo social

3. Ética e Regulação em IA

Machine Learning

> Ferramenta de processamento de dados





Fonte das imagens: Datatilsynet. "Artificial Intelligence and Privacy". Disponível em: *https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf*

Big Data

Diversas maneiras de captura de informação: sensores, câmeras, *smartphones*, *IoT*, negócios (farmácias, mercados), dados públicos

> *Volume crescente*

> *Geração constante*

> *Variedade de fontes*

> *Capacidade de estruturação e busca*

> *Capacidade de cruzamento de dados*

"DATA IS THE NEW OIL."

Coined in 2006 by Clive Humby, a British data commercialization entrepreneur, this now famous phrase was embraced by the World Economic Forum in a 2011 report, which considered **data to be an economic asset,** like oil.

From the beginning of recorded time until 2003, we created **5 exabytes** (5 billion gigabytes) **of data.**

In 2011 the same amount was created every two days.

By 2013, it's expected that the time will shrink to 10 minutes.

There are nearly as many bits of information in the digital universe as there are stars in our actual universe.

There are 133 million BLOGS on the web.

As of August 2012, there were just over 4 million articles in the English Wikipedia.

Just as a study of activity on Twitter gave residents, family members, and journalists advance warning of details about the devastating earthquake and tsunami in Japan, **high-frequency traders,** with the help of computer algorithms, use Big Data to follow trends and to act quickly on their findings.

These specialized algorithms make split-second decisions to buy or sell a commodity. New cable being laid under the Atlantic will shave **5 milliseconds** from the current 65 milliseconds it takes for trading instructions to travel between New York City and London.

With new fiber-optic cable, the round-trip time between New York and London will be 59.6 milliseconds.

This 5-millisecond saving is worth many millions of dollars to the trading firms who use the cable (and who will pay millions to do so).

How they save 5 milliseconds

The depth of the Atlantic Ocean varies.

The new cable will lie on areas of the ocean floor that are up to 1,000 feet shallower than the current fastest cable. By taking a different route, the new cable is shorter, meaning that the time it takes for messages to travel along it is shortened.

80% of all humans own a mobile phone of some sort. Out of 5 billion mobiles, 1 billion are smartphones. (In Singapore, 54% of citizens are smartphone users.)

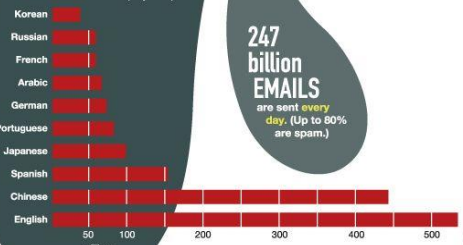
60% of all humans (5.4 billion people) are active texters. In 2010, 193,000 text messages were sent every second.

10% of all photos ever taken were taken in 2011.

247 billion EMAILS are sent every day. (Up to 80% are spam.)

English is the dominant language of the web. But by 2014 it will be **Chinese,** if its current rate of increase continues.

Top languages used on the web (May 2011):



Every hour, we create enough Internet traffic to fill **7 billion DVDs.** Side by side, that's that's seven times the height of Everest.

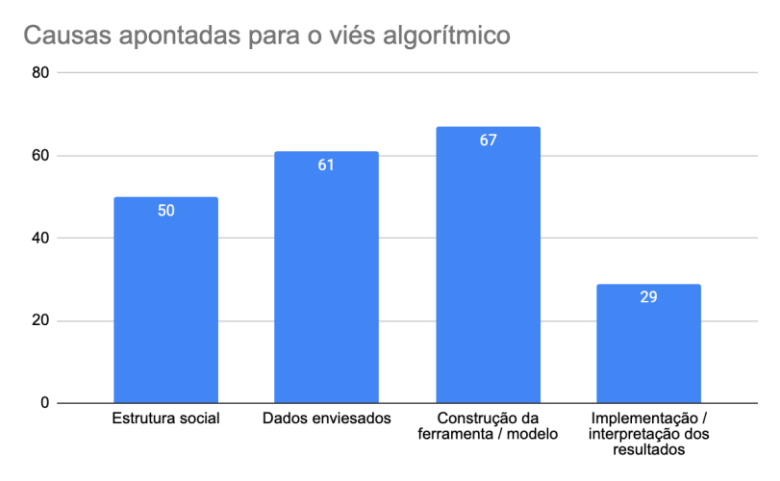
50% of 5-year-old kids in the U.S. are given access to a smartphone.

Information Graphics by Nigel Holmes

Neutralidade?

Viés algorítmico

Literatura "clássica": focada na reprodução de padrões contidos nos dados de treinamento ou na construção do modelo.



AI programs exhibit racial and gender biases, research reveals

Machine learning algorithms are picking up deeply ingrained race and gender prejudices concealed within the patterns of language use, scientists say

The Guardian

Universo amostral: 98 artigos em inglês, português e espanhol. Gráfico de artigo a ser publicado.

Viés Algorítmico - breves exemplos

Police across the US are training crime-predicting AIs on falsified data

> **Caso COMPAS**

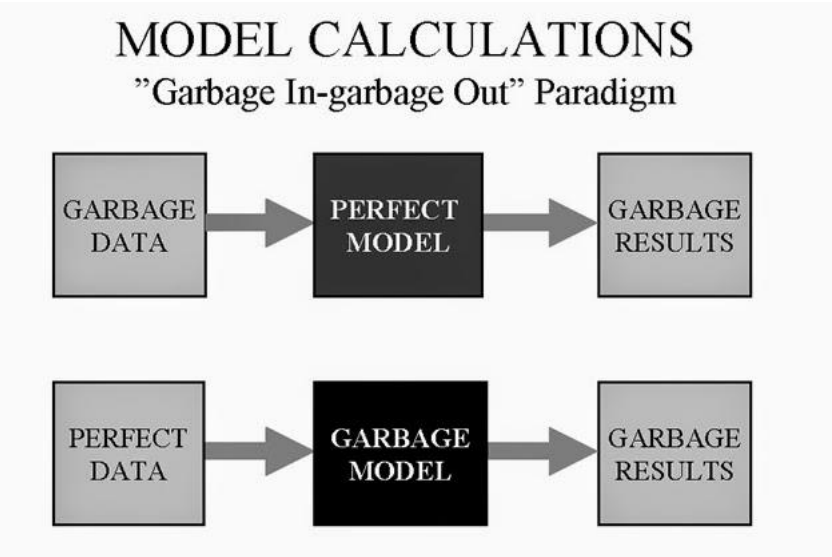
"The classic example in language is that a doctor is male and a nurse is female. If these biases exist in a language then a translation model will learn it and amplify it. **If an occupation is [referred to as male] 60 to 70 percent of the time, for example, then a translation system might learn that and then present it as 100 percent male.**"

Macduff Hughes - Head de Tradução / Google (The Verge)

"Obermeyer et al. find evidence of racial bias in one widely used algorithm, such that Black patients assigned the same level of risk by the algorithm are sicker than White patients (…) **The authors estimated that this racial bias reduces the number of Black patients identified for extra care by more than half. Bias occurs because the algorithm uses health costs as a proxy for health needs.** Less money is spent on Black patients who have the same level of need.

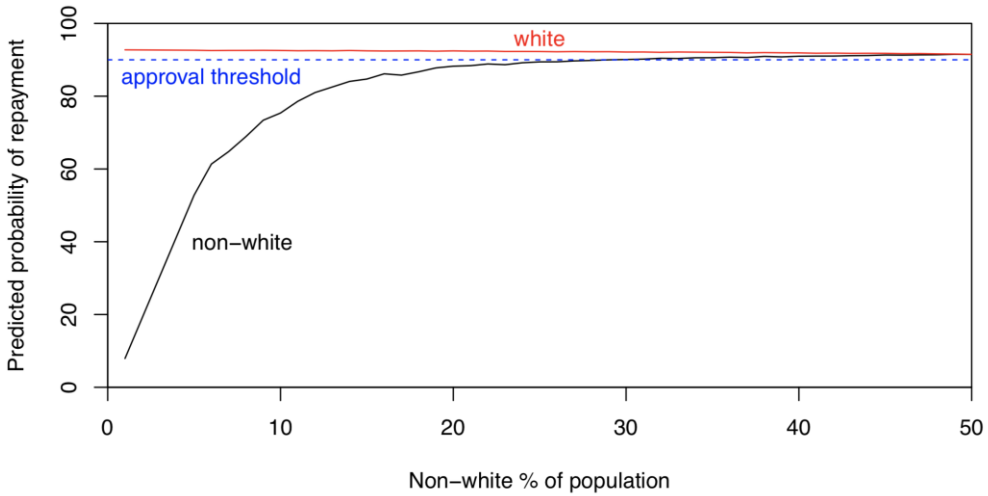
Obermeyer et al. DOI: 10.1126/science.aax2342

Reprodução de Padrões



Thanks to machine-learning algorithms, the robot apocalypse was short-lived.

Viés de Incerteza

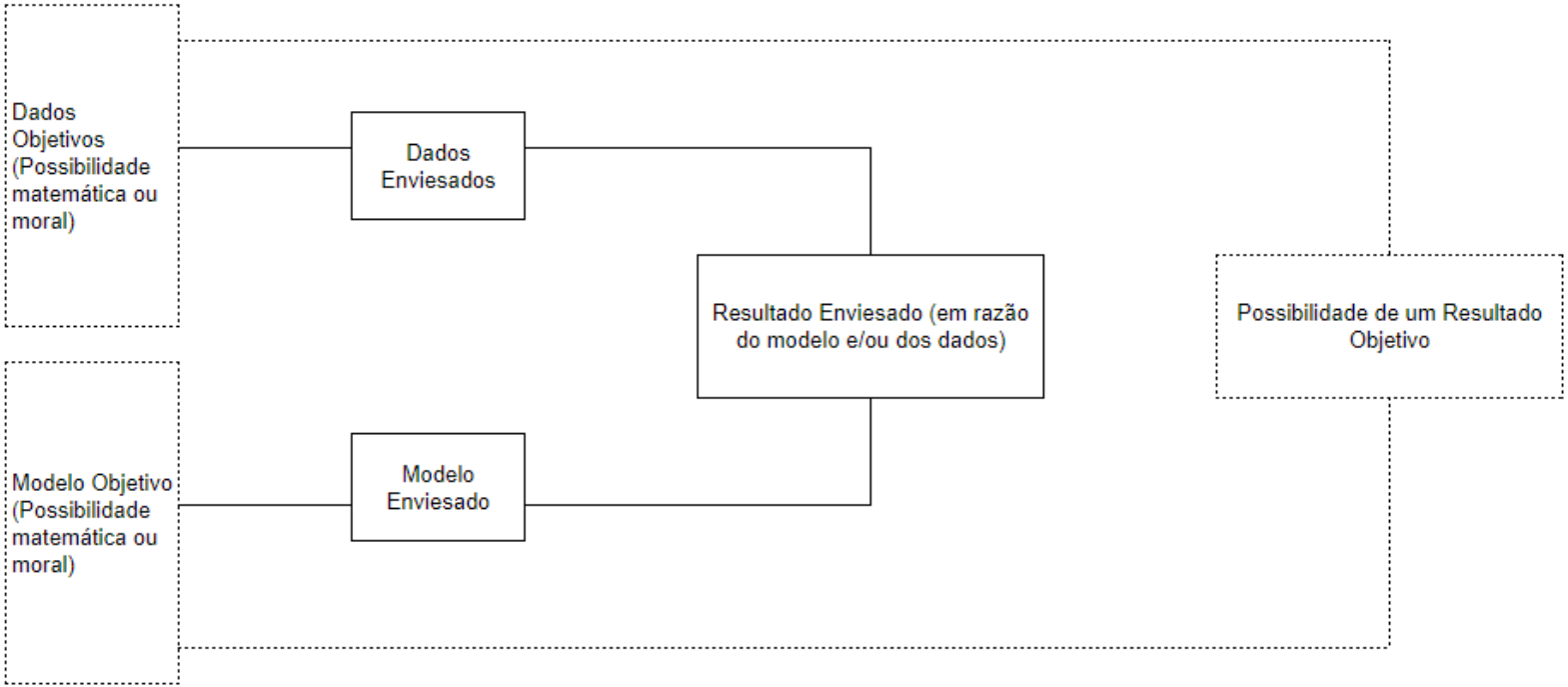


Fato: Mais dados dão maior certeza ao algoritmo

Problema: Grupos sub-representados terão piores chances em algoritmos aversos a risco (como aprovação de crédito)

Gráfico: GOODMAN, B./FLAXMAN, S. *European Union regulations on algorithmic decision-making and a "right to explanation".* 2016. AI Magazine, Vol 38, No 3

Neutralidade: possível?



Dados são neutros?

“Dados são apresentados como se fossem a matéria prima da informação, e algoritmos os agentes neutros para processá-la.”

(Deborah Lupton, Digital Sociology, p. 100)

“Até hoje, a lógica científica baseou-se no raciocínio causal e alcançou explicações quando conseguiu ligar um evento (como efeito) às causas subjacentes - quanto mais preciso e determinado for este encontro das causas, melhor será a explicação. (...) Tudo isso não seria mais necessário, e seria suplantado pelo progresso tecnológico que fornece enormes quantidades de dados e uma alta capacidade de processá-los, até que 'os números falem por si mesmos', sem necessidade de teorias ou modelos. Do processamento de dados derivam-se correlações e padrões que revelam a estrutura e a ordem do universo em jogo. (...) A superabundância de dados disponíveis já contém todas as soluções que podem ser alcançadas, mesmo sem ter definido explicitamente o problema” *(Anderson e Car, 2008)*

Fonte do texto: The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Disponível em *https://www.wired.com/2008/06/pb-theory/*

“Algorithmic Authority”

Existem convenções e **práticas para buscar, gravar e categorizar os dados**: *decisões humanas*

Quais dados serão coletados?
Quais serão analisados e vistos?

> Curtidas, retweets, visualizações, comentários > “engajamento” > economia da atenção / publicidade

> A atenção é um recurso escasso (e que parece ter atingido seu pico)

> O poder retórico do big data (pretensa neutralidade) coloca-o como ***árbitro do que é saudável ou não saudável****, aceitável e inaceitável, normal e anormal, produtivo e improdutivo.*

> Acesso ao conhecimento é mediado por ***incentivos****: incentivos econômicos ou políticos; conhecimento “viralizável” (com certo peso emocional); informações quantificáveis e buscáveis etc.*

Exemplos:

Primeiros resultados do Google. Define qual conhecimento será acessado.

> Indústria da *Search Engine Optimization* (e da compra de *likes*)

Fonte: https://medium.com/@realkevinthompson/icymi-seo-ranking-factors-in-2018-d0d27988a83b

Click Farms

IA e COVID-19

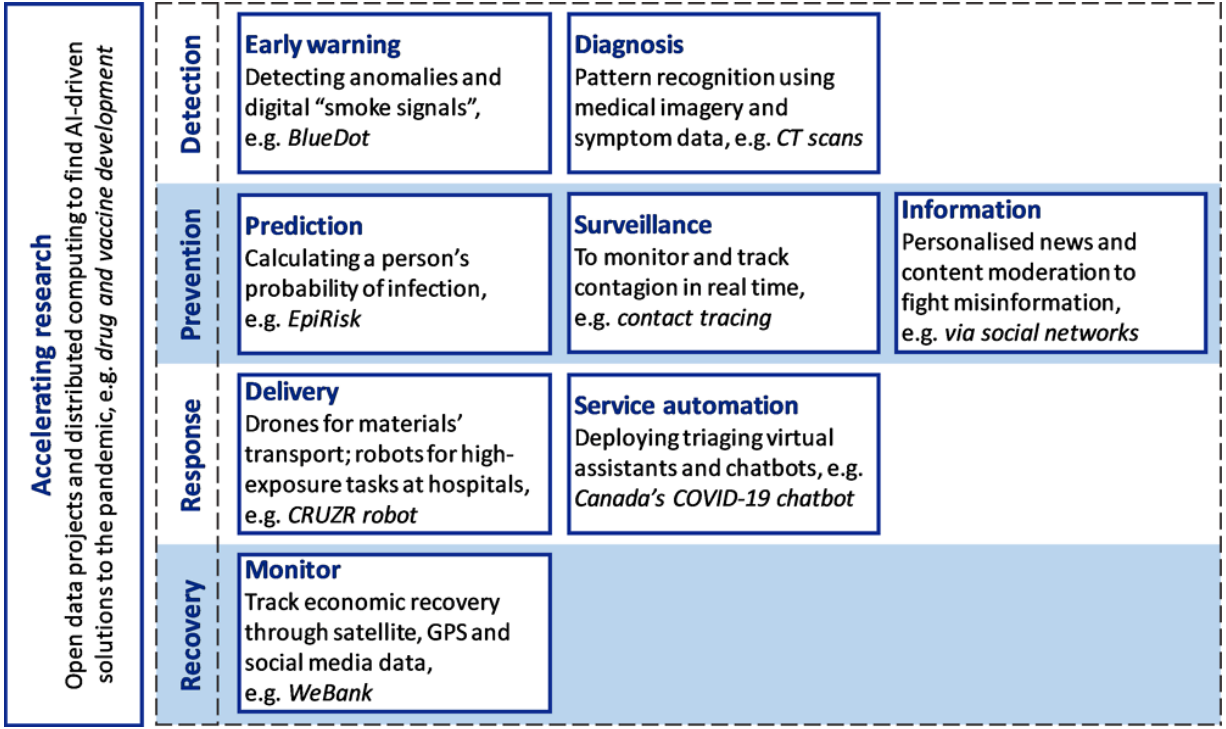
IA sendo utilizada para:

> modelar curva de contaminação (ex.: Websensors Analytics COVID-19, USP);

> diagnosticar doentes (e.g. com base em tomografias, vide Mei et al., https://doi.org/10.1038/s41591-020-0931-3)

> análise de dados de vigilância (*contact tracing* ou similares),

... e outros – *vide tabela*



Fonte: *OECD Policy Responses to Coronavirus (COVID-19): Using artificial intelligence to help combat COVID-19*

IA e COVID-19

Perguntas “práticas”:

> De onde vêm e quais são os dados utilizados para o sistema? Quais populações foram analisadas pelos dados?

> A maneira de construção e disponibilização da base de dados reflete incentivos específicos ou outros vieses?

> Os algoritmos serão utilizados em populações e situações abarcadas pelos dados de treinamento?

> O que exatamente o modelo construído é capaz de prever?

Ética e Regulação em IA

Ética e regulação e IA

*Revisão e ponderação humanas em todos os passos da criação de uma IA – **a IA é tão subjetiva quanto quem a cria e quem a utiliza**.*

> **Questão individual**: *leis de proteção de Dados / transparência para exercício de direitos / direito à revisão humana*

> **Questão sistêmica**: *transparência pública / prestação de contas e responsabilização / normas técnicas*

Regulação e IA

Questão sistêmica:

> Medidas que permitam controle público da tecnologia e responsabilização (accountability) > princípio: Transparência

Transparência não somente como meio de exercício de direitos, mas como princípio que permita o controle coletivo da tecnologia

Ferramentas de transparência coletiva

> **Relatório de Impacto à Proteção de Dados (ou similares):** *sempre que um sistema de inteligência artificial for desenvolvido para finalidades sensíveis ou apresentar risco para direitos e liberdades fundamentais. Deve ser publicamente disponibilizado*

> Informações: o fato de o sistema estar sendo utilizado; previsão de direitos fundamentais afetados pelo sistema (vide *Fundamental Rights Impact Assessment* da UE); quais mecanismos de mitigação de vieses utilizados; origem da base de dados; informações sobre exercício de direitos individuais

Ferramentas de transparência coletiva

> **Medidas posteriores de transparência:** *informações regulares sobre a utilização, publicamente disponibilizadas.*

Informações sobre: O fato de o sistema estar sendo utilizado, finalidade, locais de uso e populações afetadas; pessoas foram atingidas pelo sistema, incluindo porcentagens de pessoas por raça, gênero e etnia, por exemplo; quantas e quais decisões com impacto em direitos e liberdades fundamentais foram tomadas com base em informações fornecidas por tais sistemas; por quais órgãos e com que frequência o sistema está sendo utilizado; informações de contato para o exercício de direitos individuais.

Regulação do ponto de vista coletivo

> **Comitês de controle público** - que cobrem informações e analisem relatórios apresentados;

> **Comitês para elaboração de normas técnicas** - que criem normas de conduta tecnicamente viáveis;

> **Obrigação de testes prévios** que considerem a população potencialmente afetada (no âmbito de *privacy by design*, por exemplo);

> **Criação de órgãos internos** aos órgãos ou empresas (como DPOs) que auditem sistemas, respondam ao público etc.

> **Ações afirmativas** nos times de desenvolvimento, comitês de controle e órgãos internos.

(questões: diferenciar por finalidade. Exemplo: segurança pública ≠ fins comerciais. Diferenciar por setor: privado ≠ público ≠ polícia. Regras e diferenças setoriais necessárias.)

“Avaliação de Impacto Algorítmico”

AI Now Institute **(Universidade de Nova Iorque)**

1. Avaliação dos sistemas automatizados existentes e propostos - impactos potenciais na justiça, proporcionalidade, viés ou similares.
2. Processos de revisão de pesquisadores externos para medir ou rastrear impactos ao longo do tempo;
3. Transparência pública divulgando normas e procedimentos em curso;
4. Audiências públicas para esclarecer preocupações e responder questões; e
5. Os governos devem fornecer mecanismos aprimorados de garantia do devido processo legal para indivíduos ou comunidades contestarem avaliações inadequadas ou injustas.

Direitos individuais

>Medidas de controle do indivíduo sobre os efeitos da IA. Como garantir esse empoderamento individual? ***Exemplos de direitos:*** **Direito à Informação, Direito à Intervenção, Princípios da Transparência, Necessidade, Responsabilização e Prestação de Contas. (LGPD)**

Direito à Informação (Arts. 9º e 20, § 1º)

De que forma prestar informações "claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada"?

Direito à Intervenção (Art. 20, *caput*)

O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses

Privacy by Design

Modelos Transparentes: *Explainable AI.*

Aplicação de métodos para minimização dos dados utilizados (princípio da necessidade):

- ***Generative Adversarial Networks***
- ***Federated Learning***
- ***Matrix capsules***

Aplicação de métodos para maior privacidade*:*

- ***Differential Privacy***
- ***Homomorphic encryption***

Leituras Recomendadas:

- **Datatilsynet. (2017). "Artificial Intelligence and Privacy". Disponível em: *https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf***

- **Thelisson, Eva & Padh, Kirtan & L. Celis, Elisa. (2017). "Regulatory Mechanisms and Algorithms towards Trust in AI/ML." Disponível em: *https://www.researchgate.net/publication/318913104_Regulatory_Mechanisms_and_Algorithms_towards_Trust_in_AIML***

Obrigado!

Enrico Roberto
Pesquisador no InternetLab. Doutorando na USP.
LL.M. na Universidade de Munique (LMU). Advogado.

Twitter: @enricorbt
LinkedIn: Enrico Roberto
https://usp-br.academia.edu/EnricoRoberto