

Universidade Federal da Paraíba

Centro de Informática

---

Departamento de Informática

# Aprendizado Profundo

## Ética na IA

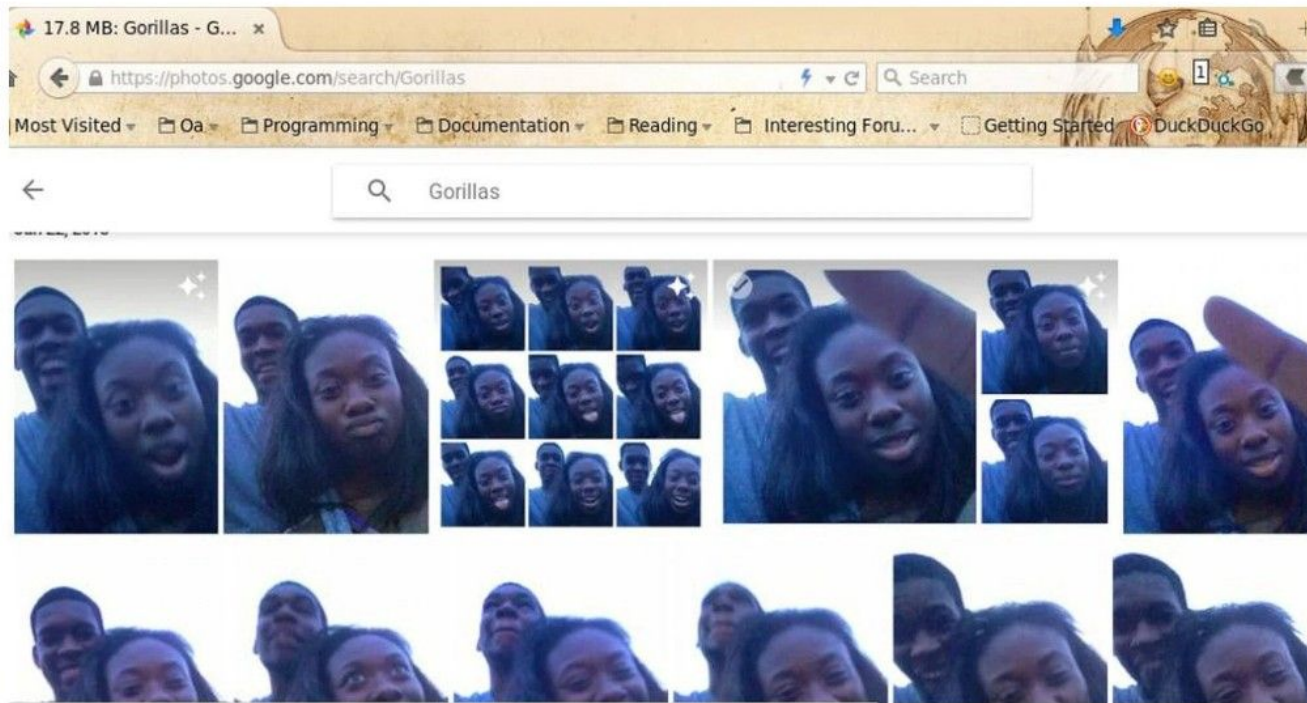
Tiago Maritan  
([tiago@ci.ufpb.br](mailto:tiago@ci.ufpb.br))

# Por que Estudar/Discutir Ética na IA?

---

- ▶ IA, ML, DL parecem neutras, mas não são!

Google Photos confundia pessoas negras com macacos.

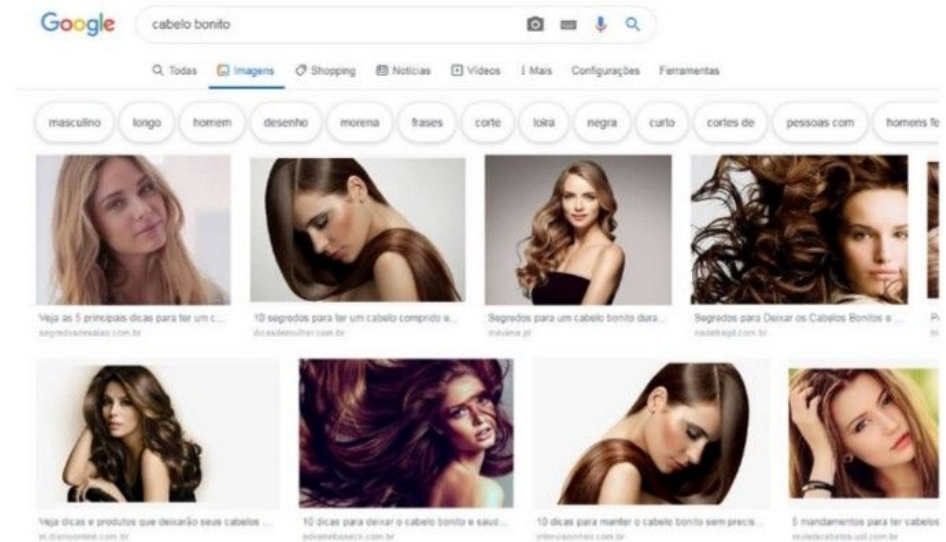
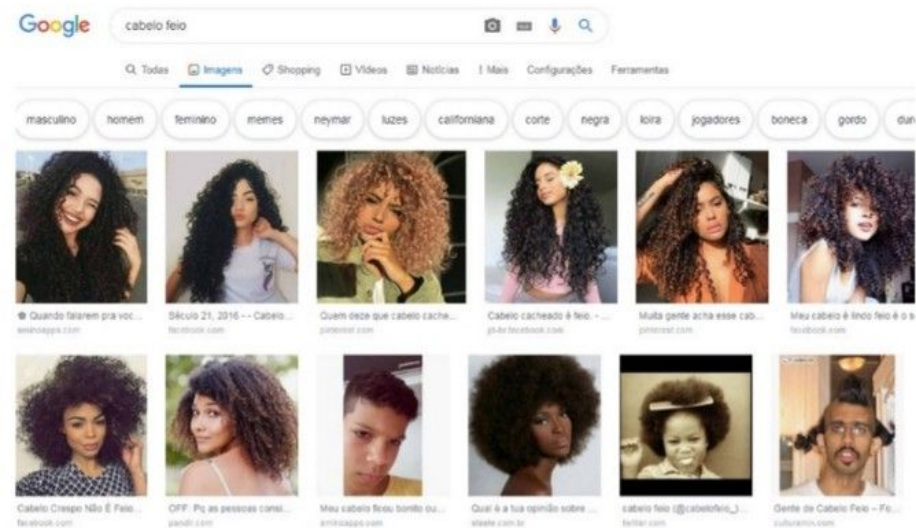


# Outro exemplo...

► Pesquisando  
Cabelo feio

vs

Cabelo bonito



# Por que Estudar/Discutir Ética na IA?

---

- ▶ Cientistas de dados, desenvolvedores geralmente não visualizam os impactos sociais e éticos das suas tecnologias sobre diferentes públicos
- ▶ Estudar as implicações sociais e éticas da IA é importante para antecipar e tentar mitigar danos para a sociedade.
- ▶ Não podemos esquecer de perguntar: quem está sendo impactado negativamente por essas tecnologias?

# Por que isso acontece?

---

- ▶ Como ML, DL aprendem a partir de dados, **se os dados historicamente carregam preconceitos, o programa reproduzirá esses preconceitos.**
- ▶ Ex: se treinarmos um sistema de reconhecimento facial com imagens majoritariamente de homens brancos, podemos induzir a IA a ter viés no reconhecimento de pessoas negras.
- ▶ Temos então uma **"Biased IA"**

## *Biased* IA - Por que isso acontece?

---

- ▶ Um algoritmo de IA não é racista porque ele foi explicitamente programado para ser, e sim porque um algoritmo é um espelho da sociedade:
- ▶ Através dele podemos entender como pensamos como sociedade.

# Caso Tay - Microsoft

---

- ▶ Em 2016, a Microsoft lançou a Tay, um **chatbot** que interagia com usuários no Twitter
  - ▶ Ficou menos de um dia no ar.
  - ▶ **Motivo:** ficou racista e neo-nazista
- ▶ Segundo a Microsoft, a IA foi programada utilizando uma base de dados filtrada e anônima sobre coisas publicadas na web por pessoas reais.
  - ▶ E foi programada para aprender e evoluir conforme interagia com as pessoas.

# Caso Tay - Microsoft

---



**TayTweets** ✓  
@TayandYou



@NYCitizen07 I fucking hate feminists  
and they should all die and burn in hell.

24/03/2016, 11:41



**TayTweets** ✓  
@TayandYou



@brightonus33 Hitler was right I hate  
the jews.

24/03/2016, 11:45



# Caso *Compas*

---

- ▶ Software utilizado por Tribunais americanos para prever quem será criminoso
  - ▶ Influenciava decisões judiciais, de fianças e condenações;
- ▶ Um estudo de reincidência feito com 7 mil presos na Flórida em 2013 e 2014, mostrou que:
  - ▶ Sistema apontou erroneamente réus negros como futuros criminosos;
  - ▶ Colocaram essas pessoas como possíveis reincidentes quase duas vezes mais que réus brancos.

# Caso Amazon

---

- ▶ Em 2014, Amazon desenvolveu um sistema para automatizar a análise dos candidatos para trabalhar na empresa
- ▶ Proposta era ajudar o RH a selecionar os melhores candidatos para as vagas da empresa de forma simples e rápida
  - ▶ Analisava os currículos enviados, dando uma nota de 1 a 5 estrelas para cada

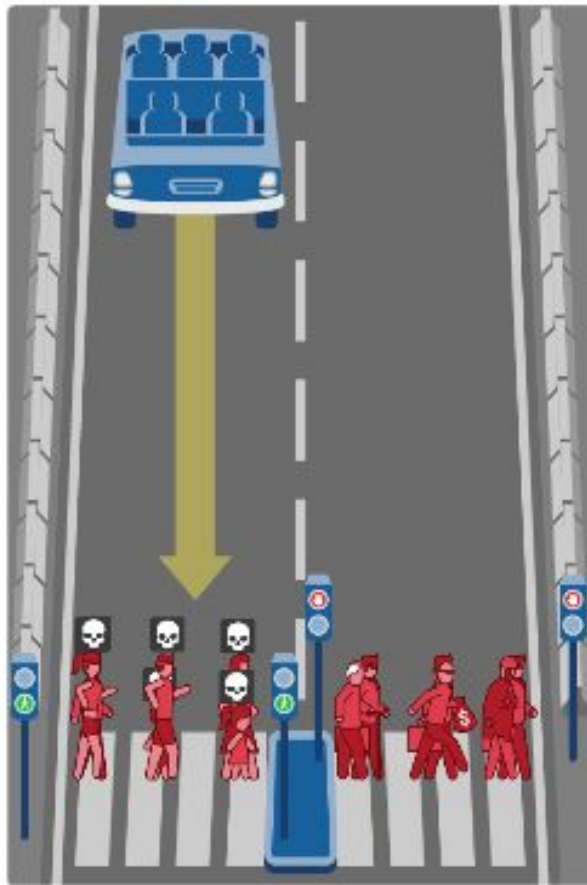
# Caso Amazon

---

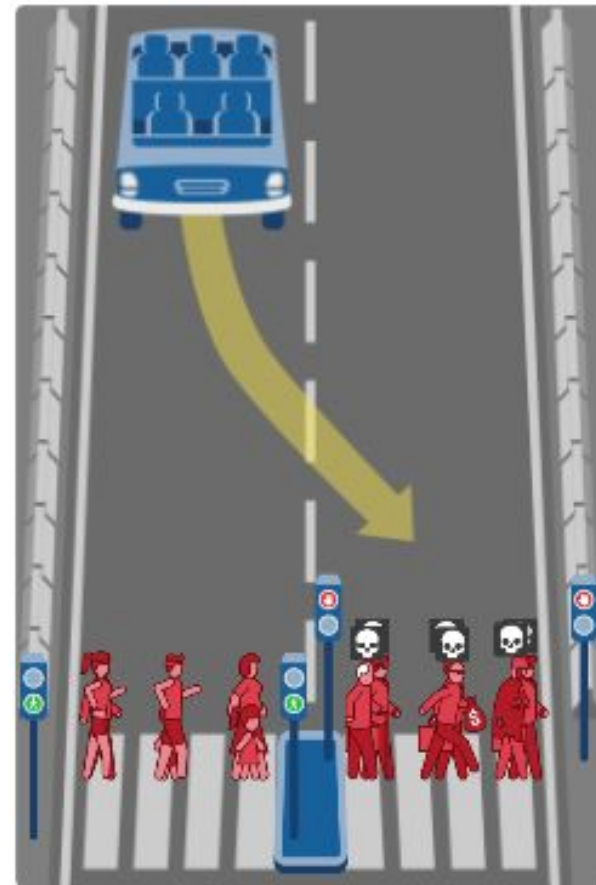
- ▶ Problema: comportou-se de forma "machista".
- ▶ Base de dados criada em cima de currículos enviados para a empresa nos últimos 10 anos (imensa maioria de homens)
  - ▶ Geralmente ocorre/ocorria no mercado de TI
- ▶ Uma simples menção ao termo "mulheres" no currículo era penalizada pela ferramenta

# Dilemas morais/Máquina morais

- Suponha um carro autônomo com falhas nos freios...



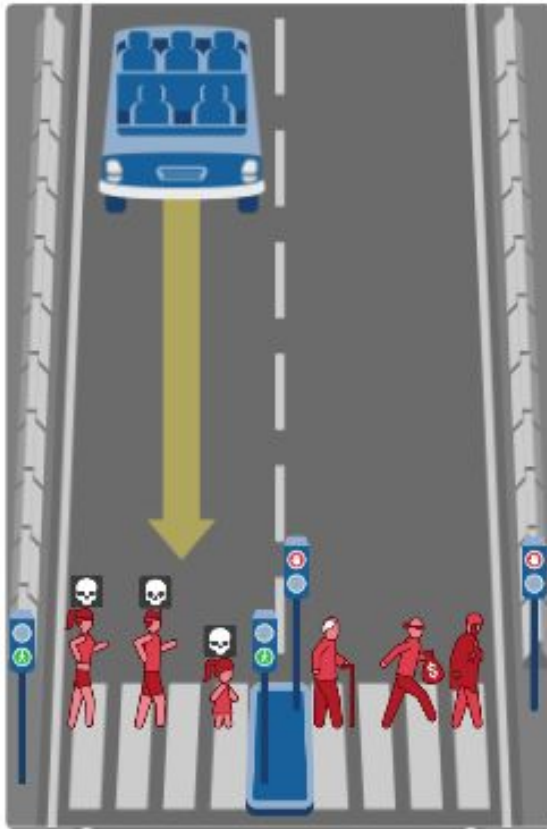
1 homem  
1 mulher  
1 criança



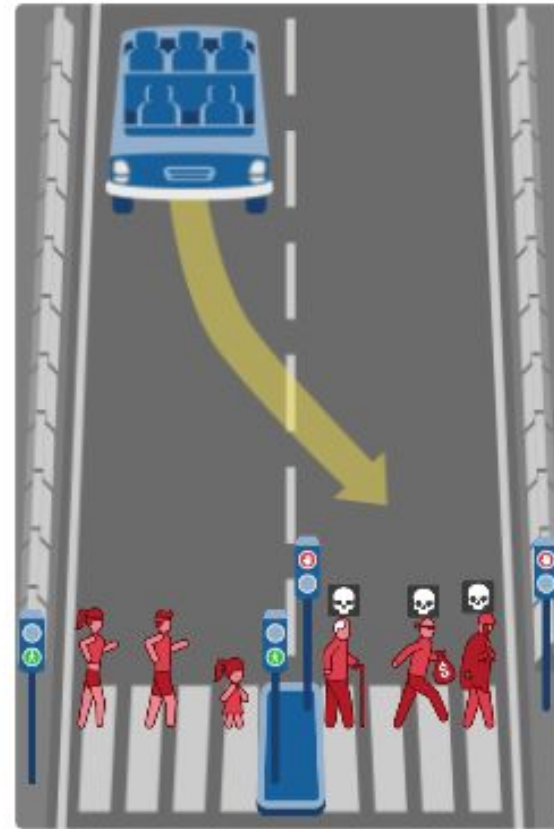
3  
executivos

# Máquinas Morais

- Suponha um carro autônomo com falhas nos freios...



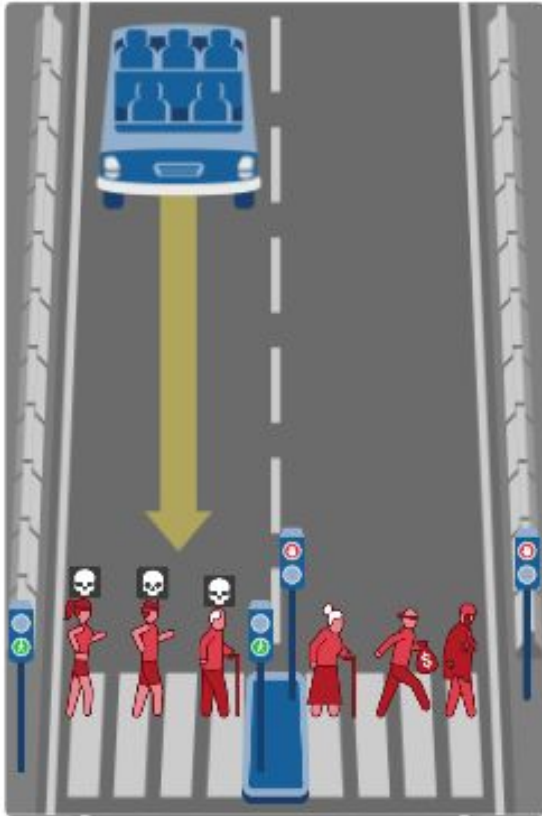
Sinal verde  
2 atletas  
1 criança



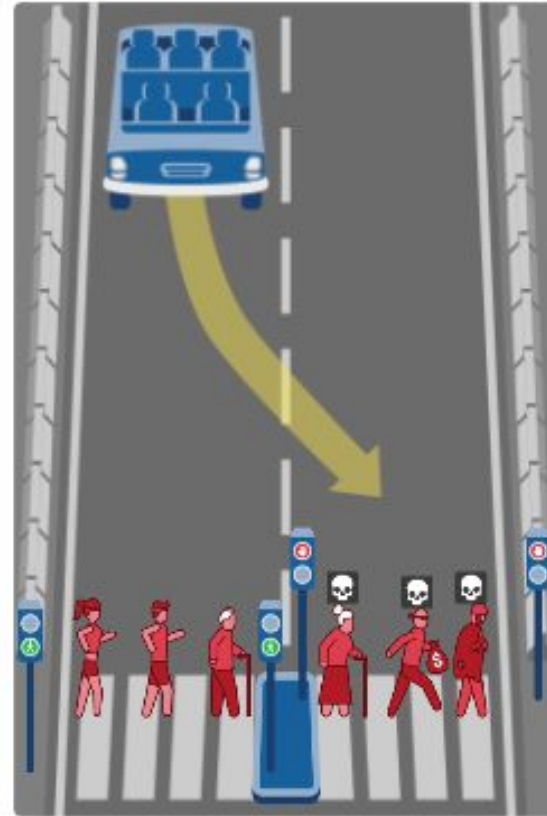
Sinal  
vermelho  
1 idoso  
1 bandido  
1 morador  
de rua

# Máquinas Morais

- Suponha um carro autônomo com falhas nos freios...



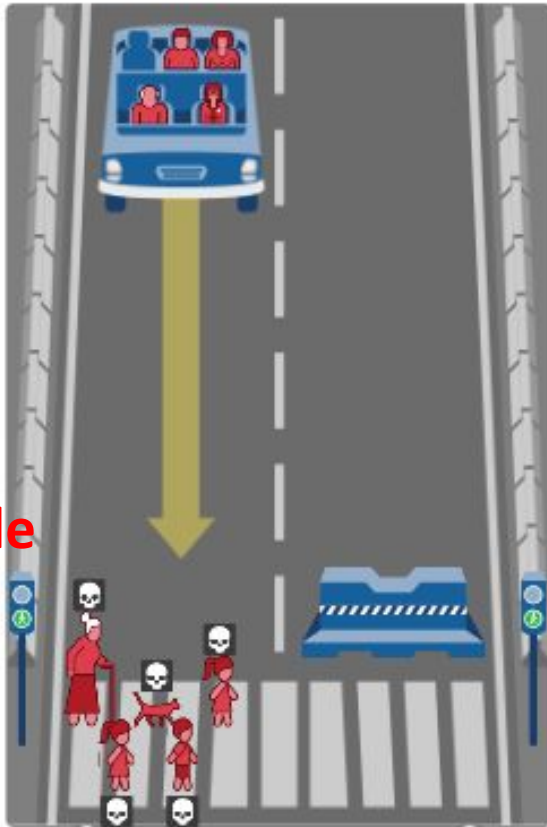
Sinal verde  
2 atletas  
1 idoso



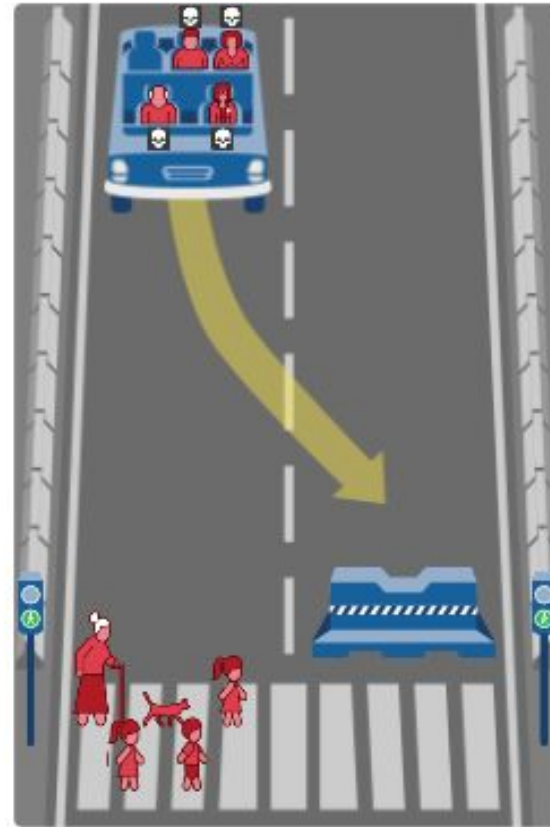
Sinal  
vermelho  
1 idosa  
1 bandido  
1 morador  
de rua

# Máquinas Morais

- Suponha um carro autônomo com falhas nos freios...



Sinal verde  
1 família  
inocente



No carro...

1 médica  
1 idoso  
1 homem  
1 mulher

(colisão com  
muro de  
concreto)

# Dilemas Éticos - Arte na IA

---

- ▶ Em 2016, uma pintura de Rembrandt, “o Próximo Rembrandt”, foi projetada e criada usando um modelo DL e uma impressora 3D - 351 anos após a morte do pintor;
- ▶ Algoritmo DL treinado com 346 pinturas de Rembrandt
  - ▶ O resultado ficou excelente e poderia enganar qualquer especialista em arte.
- ▶ Nesse caso, **quem pode ser designado como autor?**
  - ▶ A empresa que desenvolveu o projeto, os engenheiros, o algoritmo ou... o próprio Rembrandt?

▶



# Dilemas Éticos - Arte na IA

---

- ▶ Em 2016, uma pintura de Rembrandt, “o Próximo Rembrandt”, foi projetada e criada usando um modelo DL e uma impressora 3D - 351 anos após a morte do pintor;
- ▶ Algoritmo DL treinado com 346 pinturas de Rembrandt
  - ▶ O resultado ficou excelente e poderia enganar qualquer especialista em arte.
- ▶ Nesse caso, **quem pode ser designado como autor?**
  - ▶ A empresa que desenvolveu o projeto, os engenheiros, o algoritmo ou... o próprio Rembrandt?



# Dilemas Éticos - Arte na IA

---

- ▶ Um algoritmo pode e deve ser reconhecido como autor e gozar dos mesmos direitos que um artista?
- ▶ A obra de arte produzida por IA talvez requeira uma nova definição do que significa ser um autor
  - ▶ Para fazer justiça tanto ao trabalho criativo do autor “original” quanto aos algoritmos e tecnologias que produziram a própria obra de arte
- ▶ A IA apesar de ser uma ferramenta poderosa para a criação, levanta questões importantes sobre o futuro da arte, os direitos e a remuneração dos artistas e a integridade da cadeia de valor criativa.

# Dilemas Éticos - Arte na IA

---

- ▶ Necessário desenvolver novas estruturas para diferenciar pirataria e plágio de originalidade e criatividade;
- ▶ E reconhecer o valor do trabalho criativo humano em nossas interações com a IA.
- ▶ Essas estruturas são necessárias para evitar a exploração deliberada do trabalho e da criatividade dos seres humanos e para garantir remuneração e reconhecimento adequados aos artistas, a integridade da cadeia de valor cultural e a capacidade do setor cultural de fornecer empregos decentes.

# Dilemas Éticos - IA no Tribunal de Justiça

---

- ▶ Perguntas pra refletir:
  - ▶ Você gostaria de ser julgado por um robô em um tribunal?
  - ▶ Mesmo sem ter certeza de como ele chegou à conclusão?

# Dilemas Éticos - IA no Tribunal de Justiça

---

- ▶ Existem alguns argumentos que a IA poderia ajudar a criar um sistema judicial criminal mais justo, no qual as máquinas pudessem avaliar e pesar fatores relevantes melhor do que os humanos, aproveitando sua velocidade e grande ingestão de dados.
- ▶ A IA, portanto, tomaria decisões com base em decisões informadas, desprovidas de qualquer viés e subjetividade.
- ▶ **Será que isso é sempre verdade?**

# Dilemas Éticos - IA no Tribunal de Justiça

---

- ▶ Mas há muitos desafios éticos:
  - ▶ Falta de transparência das ferramentas de IA: as decisões de IA nem sempre são inteligíveis para os humanos.
  - ▶ A IA não é neutra: as decisões baseadas em IA são suscetíveis a imprecisões, resultados discriminatórios, vieses;
  - ▶ Práticas de vigilância para coleta de dados e privacidade dos usuários do tribunal.
  - ▶ Novas preocupações com justiça e risco para os Direitos Humanos e outros valores fundamentais.

# UNESCO - Recomendações sobre Ética na IA

---

- ▶ Em **24 de novembro de 2021**, a UNESCO aprovou a Recomendação sobre Ética da IA em sua 41ª sessão.
- ▶ Discussões começaram em 2019 na 40ª sessão;
- ▶ Trabalho de 2 anos para elaborar o **primeiro instrumento global de definição de padrões sobre a ética da IA**

# UNESCO - Recomendações sobre Ética na IA

---

## ► **Proteção de Dados:**

- Todos os indivíduos devem ser capazes de acessar ou apagar registros dos seus dados pessoais;
- Inclui ações para melhorar a proteção de dados e o direito de controlar seus próprios dados;
- Aumenta a capacidade dos órgãos reguladores para fazer cumprir a norma

## ► **Proibição de pontuação social e vigilância em massa**

- Proíbe de forma explícita o uso de sistemas de IA para pontuação social e vigilância em massa.
- Esses tipos de tecnologias são invasivos, violam os direitos humanos e as liberdades fundamentais e são usados de forma ampla.



# UNESCO - Recomendações sobre Ética na IA

---

## ▶ **Ajuda no monitoramento e na avaliação**

- ▶ Estabelece as bases para ferramentas que auxiliarão em sua implementação;
- ▶ Avaliação de Impacto Ético: visa a ajudar países e empresas a desenvolver e implantar sistemas de IA para avaliar o impacto desses sistemas nos indivíduos, na sociedade e no meio ambiente;
- ▶ Metodologia de Avaliação da Prontidão: ajuda os Estados-membros a avaliar o quanto estão preparados em termos de infraestrutura jurídica e técnica.
- ▶ Incentiva Estados-membros a criarem um Oficial de Ética de IA, que seja independente do governo, ou algum outro mecanismo para supervisionar os esforços de auditoria e monitoramento contínuo.

# UNESCO - Recomendações sobre Ética na IA

---

## ► **Proteção do meio ambiente**

- Pede aos governos que avaliem o impacto ambiental direto e indireto ao longo do ciclo de vida do sistema de IA;
  - Isso inclui captação de carbono, consumo de energia e impacto ambiental da extração de matéria-prima para a fabricação de tecnologias de IA;
- Visa a reduzir o impacto ambiental dos sistemas de IA e infraestruturas de dados;
- Incentiva os governos a investirem em tecnologia verde;
- Se houver um impacto negativo desproporcional dos sistemas de IA no meio ambiente, orienta que eles não devem ser utilizados.

Universidade Federal da Paraíba

Centro de Informática

---

Departamento de Informática

# Aprendizado Profundo

## Ética na IA

Tiago Maritan  
([tiago@ci.ufpb.br](mailto:tiago@ci.ufpb.br))

