Aula 9

Ética, a nova fronteira da IA

Glauco Arbix
Depto de Sociologia
Instituto de Estudos Avançados - USP

Pós – 2º semestre 2019

Ponto de Partida

1. **Imaginem um sistema que gera enorme impacto na vida das pessoas**
2. **Pensem em um sistema complexo, mas opaco**
3. **A maior parte das pessoas não faz ideia que seus dados são coletados por esse sistema imaginário. Nem que seus dados são relacionados com outros dados**
4. **Um sistema que ajuda a construir um mercado de trabalho desigual, embora seus critérios não sejam claros**
5. **Imaginem agora que nenhum humano ou instituição aceita facilmente assumir a responsabilidade pelas decisões desse sistema. Quem é o responsável? Quais são?**

Quem pensou em um sistema de IA errou.

A descrição acima é a que Franz Kafka fez da burocracia em *O Processo*

Qual é a Fonte do Problema?

1. **Não é coincidência que falta de transparência e *accountability* descritos por Kafka sejam semelhantes aos desafios vividos pela IA**
2. **Tecnologia tem história. Está imersa na sociedade. Envolve gente, poder, disputas. São construções sociais**
3. **IA é mais do que uma nova tecnologia que precisa ser regulamentada**
4. **IA é poderosa força transformadora. E é fundamental que esse poder seja orientado para melhorar a vida das pessoas**

AI é movida a dados. E dados deveriam ser movidos pelo consentimento, privacidade, propriedade e governança

Quem pode e deve discutir ética em IA?

1. **Discussão sobre IA e ética não é só para especialistas**
2. **Princípios éticos não podem ser incorporados por algoritmos. Envolvem valores que mudam com o tempo e muitas vezes são conflitantes**
3. **A universidade têm enorme responsabilidade pelo evolução responsável da IA. Pode pensar o longo prazo e enfatizar a defesa do bem público**
4. **A Usp pode estar na vanguarda do desenvolvimento de uma IA inclusiva, diversificada, ética e voltada para melhorar a vida das pessoas e da sociedade**

Dados + Dados + Dados Desprotegidos

1. **Societies are increasingly delegating complex, risk-intensive decisions to AI systems, such as diagnosing patients, hiring workers, granting parole, managing financial transactions.**
2. **This raises new challenges around liability regarding the limits of current legal frameworks (Bamberger, Mulligan, 2015) in dealing with "disparate impact" (Barocas, 2016), preventing algorithmic harms (Veale, 2018), or social justice issues related to automating law enforcement or social welfare (Eubanks, 2018), or online media consumption (Harambam, 2018).**

Tropeços nas decisões

Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots

By Jacob Snow, Technology & Civil Liberties Attorney, ACLU of Northern California
JULY 26, 2018 | 8:00 AM

TAGS: Face Recognition Technology, Surveillance Technologies, Privacy & Technology

Amazon's face surveillance technology is the target of growing opposition nationwide, and today, there are 28 more causes for concern. In a test the ACLU recently conducted of the facial recognition tool, called "Rekognition," the software incorrectly matched 28 members of Congress, identifying them as other people who have been arrested for a crime.

The members of Congress who were falsely matched with the mugshot database we used in the test include

The false matches were disproportionately of people of color, including six members of the Congressional Black Caucus, among them civil rights legend Rep. John Lewis (D-Ga.).

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini
MIT Media Lab 75 Amherst St. Cambridge, MA 02139
JOYAB@MIT.EDU

Timnit Gebru
rch 641 Avenue of the Americas, New York, NY 10011
TIMNIT.GEBRU@MICROSOFT.COM

A. Friedler and Christo Wilson

Abstract

s demonstrate that machine
ithms can discriminate based
race and gender. In this
nt an approach to evaluate
utomated facial analysis al-
atasets with respect to phe-
ps. Using the dermatolo-
itzpatrick Skin Type clas-
, we characterize the gen-
distribution of two facial
rks, IJB-A and Adience.
datasets are overwhelm-
lighter-skinned subjects
and 86.2% for Adience)
facial analysis dataset
gender and skin type.
nmercial gender clas-
sing our dataset and
inned females are the
oup (with error rates
maximum error rate
ales is 0.8%. The
in the accuracy of
les, lighter female

who is hired, fired, granted a loan, or how long an individual spends in prison, decisions that have traditionally been performed by humans are rapidly made by algorithms (O'Neil, 2017; Citron and Pasquale, 2014). Even AI-based technologies that are not specifically trained to perform high-stakes tasks (such as determining how long someone spends in prison) can be used in a pipeline that performs such tasks. For example, while face recognition software by itself should not be trained to determine the fate of an individual in the criminal justice system, it is very likely that such software is used to identify suspects. Thus, an error in the output of a face recognition algorithm used as input for other tasks can have serious consequences. For example, someone could be wrongfully accused of a crime based on erroneous but confident misidentification of the perpetrator from security video footage analysis.

Many AI syste , e.g. face recognition tools,
arning algorithms that are
d data. It has recently
rithms trained with biased
algorithmic discrimination
6; Caliskan et al., 2017).
showed that the popular
Word2Vec, encodes soci-
authors used Word2Vec
erator that fills in miss-
The analogy man is to
completed with "homemaker", conforming to the
stereotype that programming is associated with
men and homemaking with women. The biases
in Word2Vec are thus likely to be propagated
throughout any system that uses this embedding.

pace of society. From helping determine

* Download our gender and skin type balanced PPB dataset at gendershades.org

Sistemas de Facial Recognition tem Enorme Valor

Mas São Francisco proibiu seu uso pelo setor público

Razões:

A tecnologia não é confiável. A legislação é imprecisa. A sociedade não está pronta

Na esteira de São Francisco, Oakland (CA) e Somerville (MA) também baniram seus sistemas

REKOGNITION

1. **Polícia de Orlando interrompeu uso do sistema da Amazon**
2. **Razão:** *The city didn't have the necessary equipment or bandwidth to get it properly running and never once was able to test it live* **(18 July 2019)**
3. **Denúncias vão mais fundo:** *Rekognition has been plagued by criticism of its contributions to bias policing, unlawful surveillance, and racial profiling, as well as the clandestine way Amazon has gone about selling it to police departments while it's still in active development.*

Letter from Nationwide Coalition to Amazon regarding Rekognition

- “Amazon also encourages the use of Rekognition to monitor “people of interest,” raising the possibility that those labeled suspicious by governments—such as undocumented immigrants or Black activists—will be targeted for Rekognition surveillance. Amazon has even advertised Rekognition for use with officer body cameras, which would fully transform those devices into mobile surveillance cameras aimed at the public.”

- “Amazon Rekognition is primed for abuse in the hands of governments. This product poses a grave threat to communities, including people of color and immigrants, and to the trust and respect Amazon has worked to build. Amazon must act swiftly to stand up for civil rights and civil liberties, including those of its own customers, and take Rekognition off the table for governments.”

June 2018. https://www.aclu.org/letter-nationwide-coalition-amazon-ceo-jeff-bezos-regarding-rekognition

Amazon prepara proposta de regulamentação legal

Impacto disseminado

- **Nos EUA, apenas em 2018, 42% dos adolescentes foram expostos em mídias sociais**
- **A polarização política aumentou e ameaça corroer pilares da democracia**
- **Sistemas de Facial Recognition são utilizados para identificar, isolar e punir minorias e oposicionistas**

Mary Meeker Report-Code Conference 2019

A escala das informações torna mais difícil a eficiência de regras, a delimitação de fronteiras, leis e códigos de conduta capazes de mitigar a corrosão da sociedade

Imaginem um Surveillance State

- **Uma sociedade em que cada cidadão é avaliado regularmente pelo seu desempenho e lealdade ao governo**
- **Para isso, você é avaliado pela sua posição num ranking, construído por sistemas de rating**
- **Uma avaliação alta permite, p.ex., o acesso uma internet mais rápida ou a um *fast track* para tirar visto**
- **Se você expressa sua posição política sem permissão, via um post online que contradiz a narrativa oficial, seu rating diminui**
- **Para calcular seu rating, muitas empresas repassam ao governo os dados coletados nas mídfias sociais, na internet e nos sistemas de comunicação**
- **Dados de suas compras, gosto, pagamento. Dados colhidos por câmeras nas ruas, lojas, escolas, fábricas, acoplados a sistemas de facial recognition**

Parece uma distopia. Mas é o que já está sendo construído na China de hoje

Impacto disseminado

- **Nos EUA, apenas em 2018, 42% dos adolescentes foram expostos em mídias sociais**
- **A polarização política aumentou e ameaça corroer pilares da democracia**
- **Sistemas de Facial Recognition são utilizados para identificar, isolar e punir minorias e oposicionistas**

Mary Meeker Report-Code Conference 2019

A escala das informações torna mais difícil a eficiência de regras, a delimitação de fronteiras, leis e códigos de conduta capazes de mitigar a corrosão da sociedade

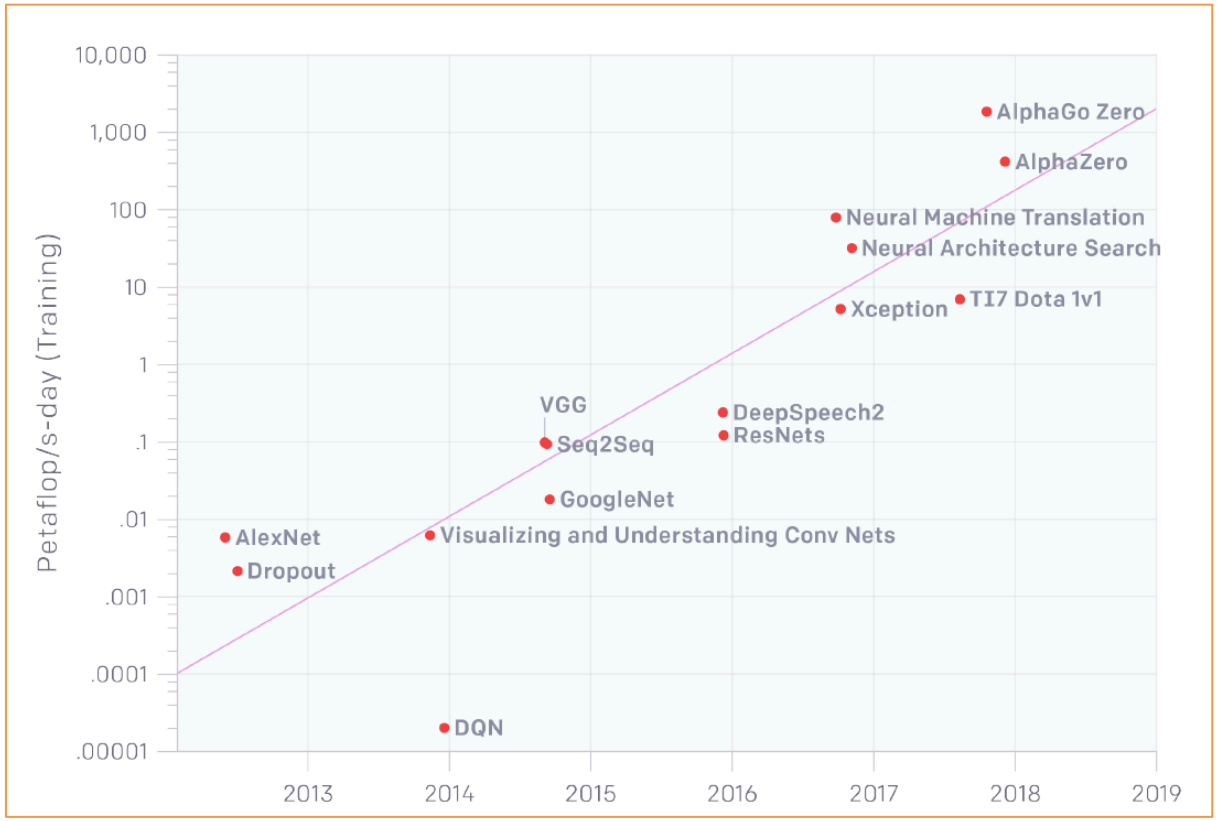
Impacto disseminado

- **Nos EUA, apenas em 2018, 42% dos adolescentes foram expostos em mídias sociais**
- **A polarização política aumentou e ameaça corroer pilares da democracia**
- **Sistemas de Facial Recognition são utilizados para identificar, isolar e punir minorias e oposicionistas**

Mary Meeker Report-Code Conference 2019

A escala das informações torna mais difícil a eficiência de regras, a delimitação de fronteiras, leis e códigos de conduta capazes de mitigar a corrosão da sociedade

Quantidade de dados para treinar algoritmos de Deep Learning cresceu 300.000 vezes em 6 anos



Allen Institute, 2019

- **Não amigável ao meio ambiente**
- **Funciona como barreira de entrada para pesquisadores**

Quando o tema é ética, transparência é palavra-chave

Mas nem sempre é respeitada por robots, chatbots e avatars, que deixam de ser ferramentas para serem apresentados (e decidirem) como uma espécie de agentes inteligentes.

Autônomos

Driverless car da Nvidia

Sistemas opacos são usinas de problemas

1. **É possível conciliar negócios que entram em contradição com princípios éticos e a pesquisa científica?**
2. **Como aprovar uma cirurgia de risco com base no diagnóstico de um sistema que não é explicável?**
3. **É possível dar vida a sistemas autônomos confiáveis para aprovar crédito nos bancos?**
4. **Para contratar funcionários?**
5. **Para decidir quem sai ou entra em prisão domiciliar?**

É possível. Mas a insegurança é grande

Como equacionar os problemas?

1. **IA está em toda parte. Mas seu funcionamento é invisível**
2. **Essa (i)realidade coloca novos problemas éticos**
3. **Como tratar as opacas decisões da IA? As ações são baseadas em inúmeras interações entre muitos agentes, como designers, developers, coders, users, software and hardware.**
4. **As abordagens tradicionais não conseguem lidar com a responsabilidade distribuída (Pagallo, Floridi)**
5. **São recentes as teorias que separam a responsabilidade dos agentes de suas intenções, de modo a repartir a responsabilidade moral entre designers, empresas, reguladores e até mesmo usuários**

Questões de Fundo

- **Algoritmos não têm ética, moral, valores ou ideologia.**
- **Nós temos**

As questões sobre a ética na IA estão ligadas à ética de quem contrata, de quem produz e oferece as tecnologias para o uso na sociedade.

- **As questões éticas são centrais para o futuro da IA. Mesmo do ponto de vista financeiro**

- **Várias pesquisas indicam que os desafios culturais e, em especial, a falta de confiança, são os principais obstáculos que impedem uma adoção mais acelerada da IA**

Qual é a saída?

1. Ao distribuir as responsabilidades entre os agentes, a teoria mitiga o negativo e estimula o positivo.
2. Essa forma compartilhada de estabelecer a responsabilidade é chave para o desenvolvimento da IA.
3. Sem modelos teóricos desse porte, a IA ficará prisioneira da insegurança, da incerteza e do medo.
4. Floridi aponta para uma Translational Ethics

“A translational ethics of AI needs to formulate foresight methodologies to indicate ethical risks and opportunities and prevent unwanted consequences. Impact assessment analyses are an example of this methodology. They provide a step-by-step evaluation of the impact of practices or technologies deployed in a given organization on aspects such as privacy, transparency, or liability.”

1. **A regulação é difícil e mais do que nunca, necessária**
2. **A coordenação de esforços é mais necessária ainda**
3. **E a universidade está no centro desse debate**

Nas questões éticas, os holofotes estão nos cientistas

Para colocar um algoritmo no mundo real é preciso mais do que ciência

- **Quem já se perguntou sobre a formação dos cientistas?**
- **São raros os cursos de ciência da computação que discutem essas questões**
- **A experimentação e a busca de novas linguagens e técnicas, na fronteira do conhecimento, são insuficientes para formar pesquisadores capazes de gerar IA for Good**

Perguntas que todo pesquisador deveria fazer

1. Quais são seus próprios *biases*?
2. Sua equipe de trabalho é diversificada?
3. Você avalia seu trabalho para saber se não está excluindo ou segregando? Ou pede para outros avaliarem?
4. Os modelos que você cria são rastreáveis durante seu ciclo de vida?
5. Quais os direitos fundamentais que você considera imprescindíveis para sua atividade?
6. Você se sente responsável pela qualidade dos dados que você utiliza?
7. Existe alguma fronteira ou linha ética que não pode ser cruzada?

Roteiro para o desenvolvimento de IA

1. **Transparência:** todos têm direito de saber como as decisões oferecidos pelos algoritmos são tomadas
2. **Black Box:** mais do que dados relevantes para transparência IA deve conter informações sobre a base ética na qual o sistema se assenta
3. **Foco nas pessoas:** códigos de ética ajudam a garantir princípios de dignidade, liberdade, privacidade, diversidade e todos os direitos humanos
4. **Feito por pessoas:** IA deve garantir que as máquinas sejam, legalmente, ferramentas e não sejam responsabilizadas por sua atuação
5. **Sem Bias:** é fundamental que os sistemas tenham controles para impedir discriminação de gênero, raça, preferência sexual, idade
6. **Sem aumentar desigualdades:** o desenvolvimento de AI deve se orientar para assegurar a requalificação e o reposicionamento dos atingidos, de modo a que haja uma transição justa para a digitalização da sociedade

Tecnologias de IA (e, principalmente, seus pesquisadores) podem ajudar a diminuir riscos e ameaças

- Emprego e Trabalho
- Ética
- Privacidade
- Diversidade
- Inclusão

- Desemprego
- Bias
- Vigilância
- Desigualdade
- Ódio e Fake News

Estabelecer códigos e diretrizes para a pesquisa é ponto crítico.

- **For design:** The codes of conduct, standards and certification processes that ensure the integrity of developers and users as they research, design, construct, employ and manage AI systems

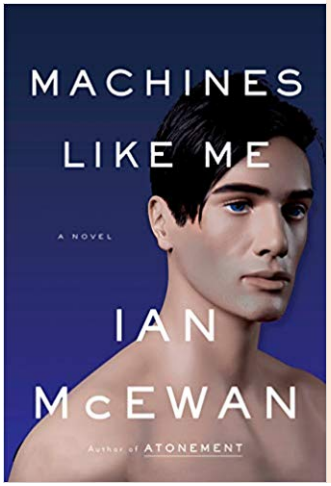
- **By Design:** the technical integration of ethical capabilities as part of the behavior of artificial autonomous system

Direcionadores

- **Trustworthy technologies:** based on clear understanding of AI performance and interaction with real people and social institutions.
- **AI fully explainable:** oriented to increase algorithmic fairness, transparency and accountability, taking into account the clearness of algorithmic decisions remains heavily dependent on the process that explanations can be embedded in AI systems.
- **Ethical auditing:** for complex algorithmic systems, to make accountability viable, and reduce the so called “black-boxes issues”.
- **Human-centered education:** to deal with the most decisive dimension of AI, for the present and future generations.

A universidade – e a USP – pode ajudar a pesquisa em IA se tornar efetivamente *human-centered*

observatório da inovação



When we program morality into robots, are we doomed to disappoint them with our very human ethical inconsistency?

Ian McEwan

observatório da inovação

USP ie]