

Aprendizado de Máquina: Aspectos Éticos

1. Prelúdio

- Temos discutido técnicas que permitem à máquina resolver problemas que, no passado, só podiam ser resolvidos pelo ser humano. Isso gera um fascínio que, não raro, aproxima a tecnologia do domínio da magia, mas também levanta questões graves.
- Tendo em vista que o cientista de dados atua junto à sociedade, é imperativo que reflitamos a respeito. Faremos isso sem uma sequência rígida.

1.1. IA e Desemprego

Almost half of all US workers are at risk of losing their jobs to robots, according to a new report



Heather Stewart, The Guardian Nov. 5, 2015, 11:12 AM



- Os avanços da IA têm levado à substituição de trabalho humano em tarefas cada vez mais complexas. Isso pode levar a um mundo em que cada pessoa teria mais tempo para se dedicar ao que podemos chamar, genericamente, de ócio (sem sentido pejorativo algum). Mas há a possibilidade de que ocorra uma grave crise social, especialmente nos países menos desenvolvidos.
- Erik Brynjolfsson (MIT): *“It’s one of the dirty secrets of economics: technology progress does grow the economy and create wealth, but there is no economic law that*

says everyone will benefit." [Rotman, 2013]. Brynjolfsson e McAfee falam em "grande desacoplamento", como ilustrado na Fig. 1 (extraída de [Rotman, 2013]).

- Como exemplo, empresas como Google, Tesla, Uber, Ford, GM e BMW trabalham fortemente na tecnologia de carros autônomos. Esses carros, em certo sentido, superam os motoristas humanos pela fusão de dados de diferentes sensores [Su, 2018].

The screenshot shows the top navigation bar of The Guardian website. It includes the 'Support The Guardian' banner with 'Contribute' and 'Subscribe' buttons, the 'The Guardian' logo, and links for 'Search jobs', 'Sign in', 'Search', and 'International edition'. Below the navigation bar is a horizontal menu with categories: News, Opinion, Sport, Culture, Lifestyle (highlighted), and More. A secondary menu lists various topics: Fashion, Food, Recipes, Travel, Health & fitness, Women, Men, Love & sex, Beauty, Home & garden, Money, and Cars. The main content area features a news article by Uber, titled 'Uber's self-driving car unit valued at \$7.3bn as it gears up for IPO'. A yellow banner above the title states 'This article is more than 1 month old'. To the right of the article is an advertisement box with the text 'Ad closed by Google' and a button 'Stop seeing this ad', along with a link 'Why this ad? ①'.

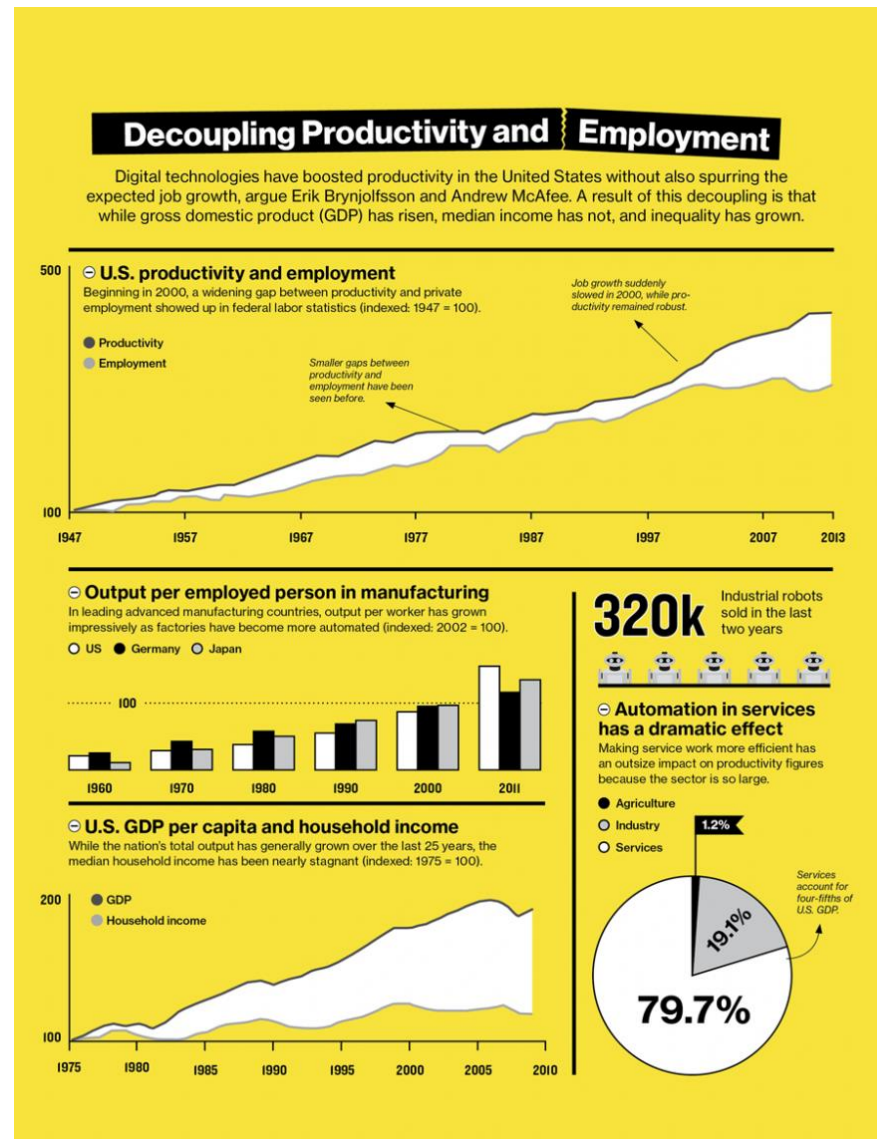


Figura 1 – O Grande Desacoplamento (de [Rotman, 2013]).

- Ideias como “renda mínima” e “impostos sobre robôs” surgem como paliativos, pelo menos nos países mais desenvolvidos. O que restará, na divisão internacional do trabalho, para países pobres como o Brasil?

LIFE WITH A.I.

Elon Musk: Robots will take your jobs, government will have to pay your wage

Published Fri, Nov 4 2016 • 2:19 PM EDT • Updated Mon, Jan 29 2018 • 4:45 PM EST



Catherine Clifford
@CATCLIFFORD

Share [f](#) [t](#) [in](#) [✉](#)

- No Brasil, essa discussão tão importante passa longe do debate político concreto. O desmanche dos mecanismos de proteção / bem-estar social que se desenha no horizonte pode levar a uma tragédia social ainda maior que aquela que temos hoje.

1.2. Fair Machine Learning

- Num livro muito interessante, “Weapons of Math Destruction” [O’Neil, 2016], Cathy O’Neil discute alguns perigos trazidos pela era da ciência de dados. Questionamentos como os que ela apresenta ilustram uma área de investigação que pode ser de grande valia para a sociedade em geral: a de *fair machine learning*.
- Um ponto central da área é analisar a questão do *viés (bias)* que pode estar presente, mesmo que furtivamente, em algoritmos projetados por técnicos competentes. Esses vieses podem levar a laços de feedback potencialmente nocivos, prejudicando inúmeras pessoas e/ou instituições.
- Em 2015, foram discutidos possíveis exemplos ligados ao motor de busca do Google [Yapo e Weiss, 2018]. Um deles foi que a busca por “CEO” apresentava,

como resultados principais, exclusivamente fotos de homens brancos. Pior: um estudo realizado pela Carnegie Mellon University mostrou que resultados ligados a empregos executivos de alto salário eram obtidos com menor frequência quando o motor de busca supunha que o usuário era do sexo feminino. Essa conjunção é uma realimentação negativa, pois contribui para uma piora da desigualdade no futuro.

- Algoritmos são onipresentes, mas geralmente não fazemos ideia de seu modo de funcionamento – estamos numa grande caixa-preta, e muitas vidas são diretamente influenciadas. Também não podemos saber, geralmente, quão balanceado / bem construído é o conjunto de dados usado para treinar um método com o qual interagimos.

Google Image Search Has A Gender Bias Problem

Emily Cohn
The Huffington Post

- Em 2015, a ferramenta de reconhecimento de imagens Flickr foi acusada de produzir resultados racistas, associando a imagem de pessoas negras a tags como “animal” ou “macaco”. O software da Nikon, por sua vez, identificava pessoas asiáticas como estando de olhos fechados [Yapo e Weiss, 2018].

Support The Guardian

Available for everyone, funded by readers

[Contribute →](#)

[Subscribe →](#)

[Search jobs](#)

[Sign in](#)

[Search](#)

[International edition](#)

The Guardian

News

Opinion

Sport

Culture

Lifestyle

More

[World](#) [UK](#) [Science](#) [Cities](#) [Global development](#) [Football](#) **Tech** [Business](#) [Environment](#) [Obituaries](#)

Technology

This article is more than 4 years old

Flickr faces complaints over 'offensive' auto-tagging for photos

Auto-tagging system slaps 'animal' and 'ape' labels on images of black people, and tags concentration camps with 'jungle gym' and 'sport'

Advertisement

Ad closed by Google

[Stop seeing this ad](#)

[Why this ad?](#)

The New York Times

Opinion

OPINION

Artificial Intelligence's White Guy Problem

By Kate Crawford

June 25, 2016



- A análise de risco de cometimento de crimes por indivíduos também é problemática. Embora os algoritmos usados não usem informações diretas sobre raça / etnia, o viés surge, por exemplo, na taxa de previsões erradas de reincidência, como mostra a Fig. 2.

	White	African American
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

"Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes."

Figura 2 – Dados de Análise de Reincidência (de [Yapo e Weiss, 2018]).

- É necessário buscar formas de combater esse problema, seja na forma de diretrizes saudáveis de projeto, seja por um ativo processo de verificação / análise. Julgamos que levar em conta o risco de viés é dever do projetista e da instituição à qual ele se vincula.

1.3. A Democracia Sobreviverá à IA?

- Um dos grandes problemas na política atual é o efeito que a ciência de dados tem em processos democráticos. Através do uso de algoritmos de IA e de “robôs”, é possível, por exemplo, monitorar o perfil ou estado de espírito de determinado público e gerar material para manipulação / desinformação em tempo real. Vivemos a era das “fake news” que se espalham rapidamente pelas redes sociais e das milícias cibernéticas.

- O escândalo da Cambridge Analytica chamou a atenção das pessoas para o fato de que os dados que elas produzem constantemente e voluntariamente podem ser utilizados para finalidades que elas nem concebem. É lugar-comum dizer que dados são o bem mais valioso de nossa era – no caso, esses dados foram usados para alavancar campanhas políticas. Perfis de “likes” e de uso de outros recursos de redes sociais levavam a uma análise de perfil do usuário (sem sua permissão). A partir dessa análise, torna-se simples “customizar” conteúdos para atingir determinadas finalidades “publicitárias”.
- Costuma-se dizer que “uma mentira contada muitas vezes se torna verdade” – com a ciência de dados e as redes sociais, essa frase toma proporções assustadoras. As pessoas são sugestionáveis (pelo menos até certo ponto), e podem ser levadas a construir imagens falsas de diversos fatos e ideias.

- Caso não se proponham contramedidas adequadas, pode acontecer de a democracia como a conhecemos se tornar um simples fantoche nas mãos de quem produzir os algoritmos e conteúdos mais eficientes.
- A própria IA pode ser usada para combater esse processo – para isso, é importante que haja centros de pesquisa e universidades com independência suficiente para não dependerem do beneplácito de governos e grandes corporações.

1.4. A Singularidade Está Próxima?

- Um conceito que tem impacto no imaginário de muitas pessoas é o de “singularidade tecnológica” [Kurzweil, 2005]. Basicamente, a ideia diz respeito à possibilidade de que haja uma espécie de “reação em cadeia” junto a uma

inteligência artificial de propósito geral – essa inteligência poderia projetar máquinas mais inteligentes, que projetariam máquinas mais inteligentes, numa “explosão de inteligência” [Wikipedia, 2019].

- Pode ser, então, que a máquina atinja níveis muito superiores de desempenho em tarefas que normalmente associamos à capacidade cognitiva humana. Nesse caso, cumprir-se-ia a visão de Claude Shannon, de que “seremos para as máquinas o que os cães são para nós”.
- Também se associa o conceito de singularidade a desenvolvimentos que levem o ser humano a “se fundir com a máquina”, “aumentando” diversas habilidades e mesmo superando as doenças e até mesmo a morte.
- Naturalmente, um dos fatores que são evocados pelos que creem na singularidade é uma visão da lei de Moore como algo que impacta na computação como um todo. A Fig. 3 traz uma ilustração representativa.

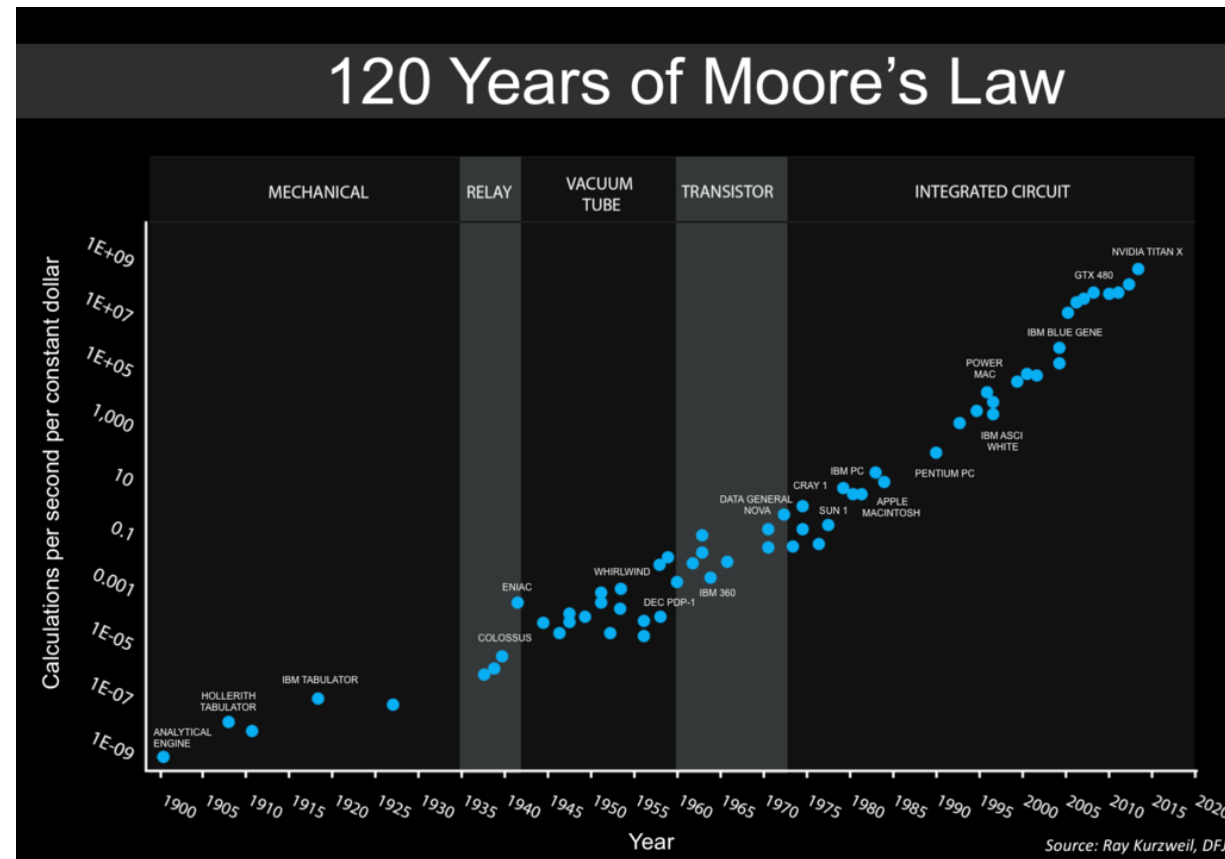


Figura 3 – Gráfico da Lei de Moore (de [Wikipedia]).

2. Referências bibliográficas

KURZWEIL, R., *The Singularity is Near*, Penguin, 2005.

O'NEIL, C., *Weapons of Math Destruction*, Broadway Books, 2016.

ROTMAN, D., "How Technology is Destroying Jobs", MIT Technology Review Magazine, July / August, 2013.

SU, G., "Unemployment in the AI Age", AI Matters, Vol. 3, No. 4, 2018.

WIKIPEDIA Artigos Diversos, 2019.

YAPO, A., WEISS, J., "Ethical Implications of Bias in Machine Learning", Proceedings of the 51st Hawaii International Conference on System Sciences, 2018.