PROCESSAMENTO DE LINGUAGEM NATURAL
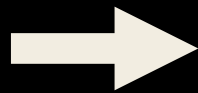
Diogo Cortiz

Ceweb.br

A Breakthrough for A.I. Technology: Passing an 8th-Grade Science Test

By Cade Metz

Sept. 4, 2019

SAN FRANCISCO — Four years ago, more than 700 computer scientists competed in a contest to build artificial intelligence that could pass an eighth-grade science test. There was $80,000 in prize money on the line.

They all flunked. Even the most sophisticated system couldn't do better than 60 percent on the test. A.I. couldn't match the language and logic skills that students are expected to have when they enter high school.

But on Wednesday, the Allen Institute for Artificial Intelligence, a prominent lab in Seattle, unveiled a new system that passed the test with room to spare. It correctly answered more than 90 percent of the questions on an eighth-grade science test and more than 80 percent on a 12th-grade exam.
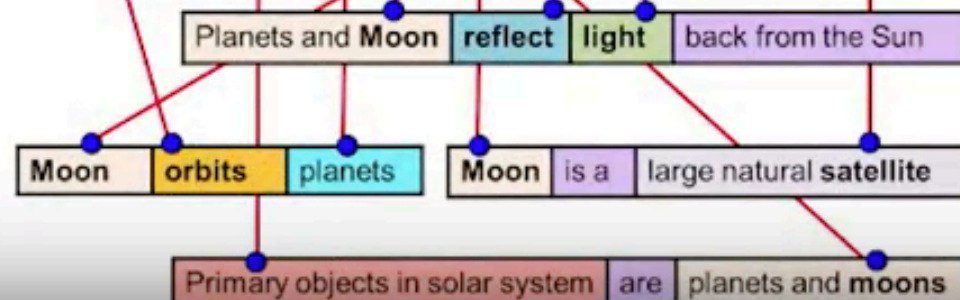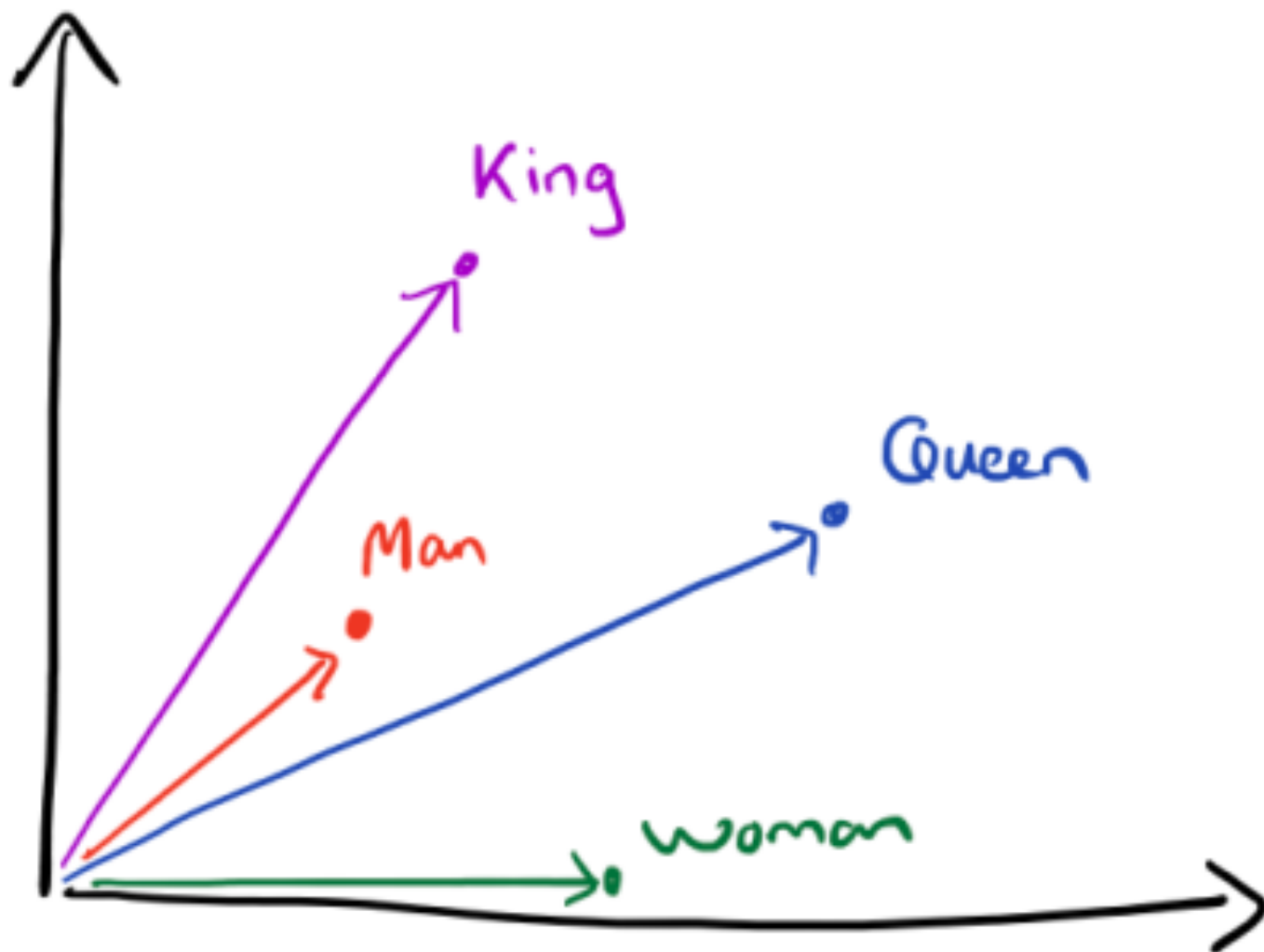
Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings

Tolga Bolukbasi[1], Kai-Wei Chang[2], James Zou[2], Venkatesh Saligrama[1,2], Adam Kalai[2]

[1]Boston University, 8 Saint Mary's Street, Boston, MA

[2]Microsoft Research New England, 1 Memorial Drive, Cambridge, MA

tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

Abstract

The blind application of machine learning runs the risk of amplifying biases present in data. Such a danger is facing us with *word embedding*, a popular framework to represent text data as vectors which has been used in many machine learning and natural language processing tasks. We show that even word embeddings trained on Google News articles exhibit female/male gender stereotypes to a disturbing extent. This raises concerns because their widespread use, as we describe, often tends to amplify these biases. Geometrically, gender bias is first shown to be captured by a direction in the word embedding. Second, gender neutral words are shown to be linearly separable from gender definition words in the word embedding. Using these properties, we provide a methodology for modifying an embedding to remove gender stereotypes, such as the association between between the words *receptionist* and *female*, while maintaining desired associations such as between the words *queen* and *female*. We define metrics to quantify both direct and indirect gender biases in embeddings, and develop algorithms to "debias" the

tote treats subject heavy commit game
browsing sites seconds slow arrival tactical
crafts identity
trimester tanning user parts drop reel firepower
ultrasound busy hoped command
housing caused ill rd scrimmage
modeling beautiful looks builder drafted
sewing dress dance cake victims hay quit brilliant genius
letters nuclear yard
pageant earrings divorce ii firms cocky journeyman
dancers seeking ties guru
salon thighs lust lobby buddy
sassy breasts pearls vases frost vi governor sharply rule
homemaker dancer roses folks friend pal brass buddies burly
feminist babe priest mate beard
she witch witches dads boys cousin chap boyhood he
actresses gals wives lad
queen fiance sons son brothers
sisters girlfriends girlfriend
grandmother wife daddy nephew
ladies
daughters fiancee
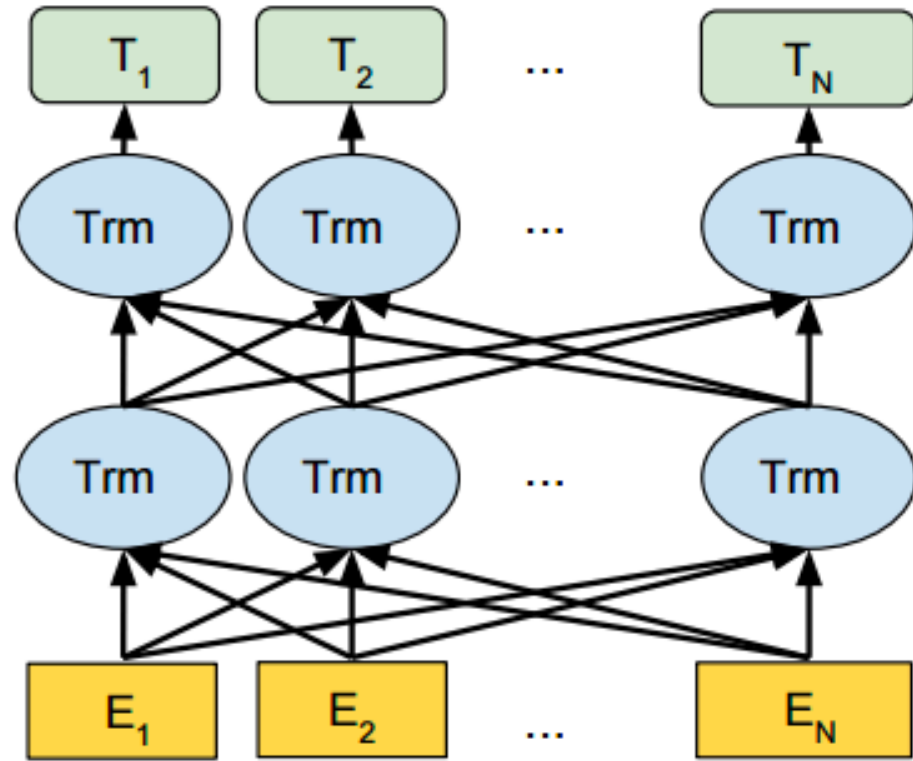
Bidirectional Encoder Representations from Transformers (BERT)
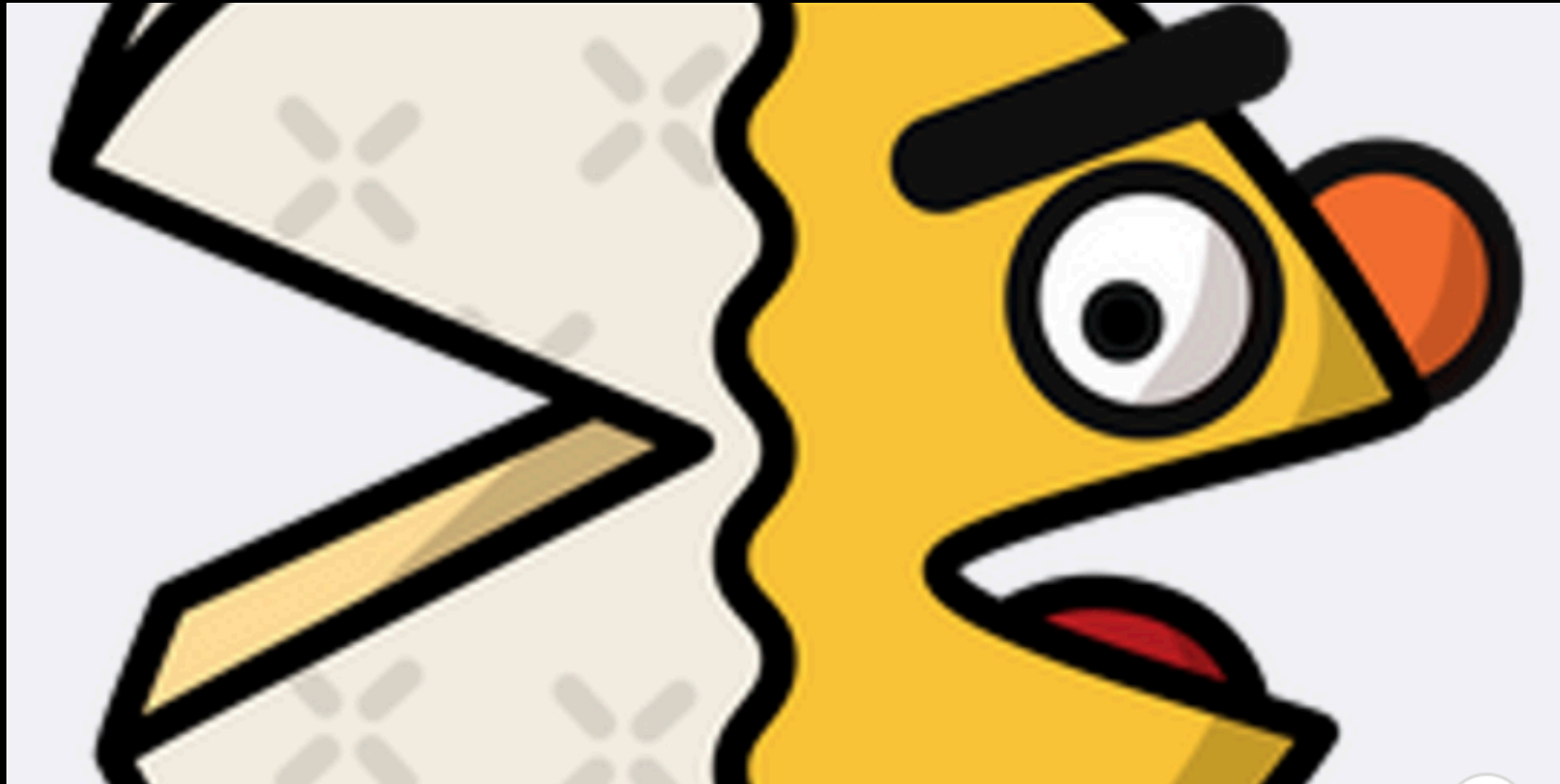
Fair - camembert

SEARCH

Understanding searches better than ever before

Pandu Nayak

Google Fellow and Vice President, Search

Published Oct 25, 2019

If there's one thing I've learned over the 15 years working on Google Search, it's that people's curiosity is endless. We see billions of searches every day, and 15 percent of those queries are ones we haven't seen before--so we've built ways to return results for queries we can't anticipate.

Search query: 2019 brazil traveler to usa need a visa

BEFORE

9:00

google.com

wp Washington Post › 2019/03/21

U.S. citizens can travel to Brazil without the red tape of a visa ...

Mar 21, 2019 · Starting on June 17, you can go to Brazil without a visa and ... Australia, Japan and Canada will no longer need a visa to ... washingtonpost.com; © 1996-2019 The Washington Post ...
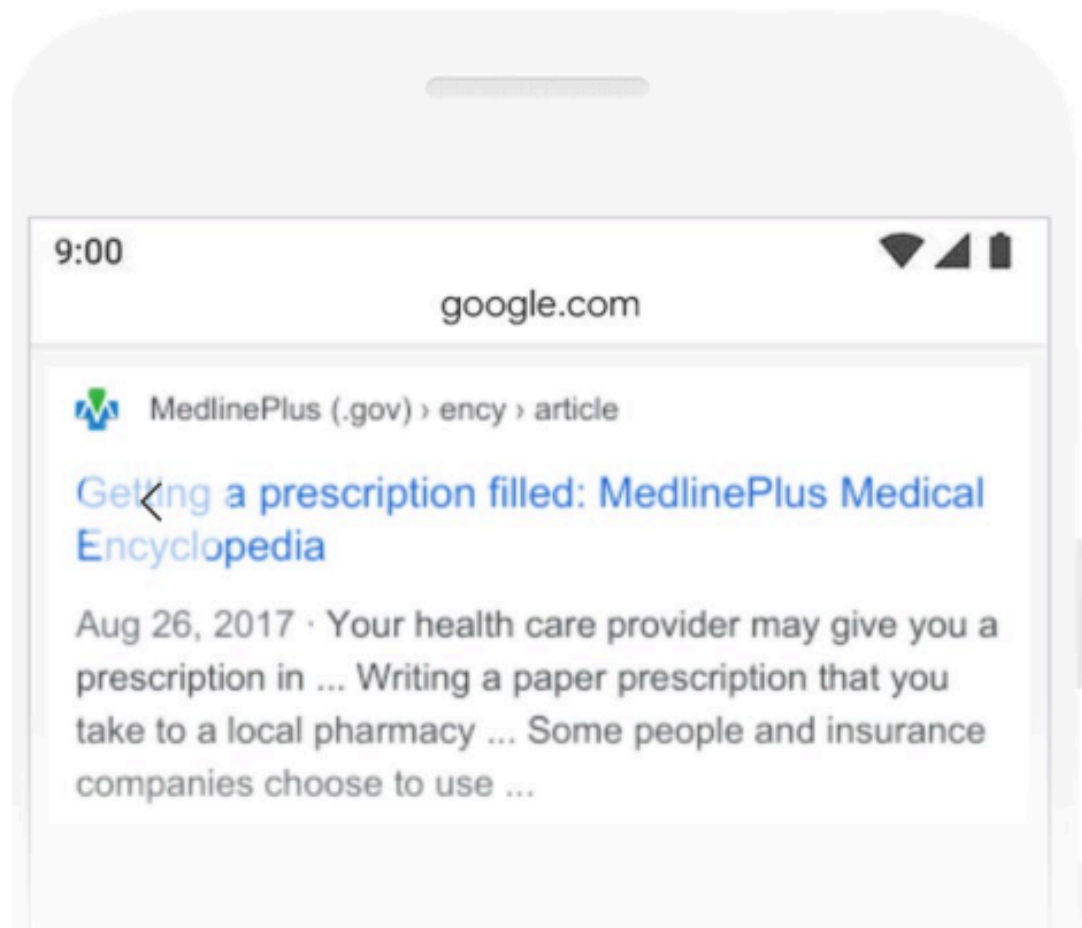
AFTER

9:00

google.com

USEmbassy.gov › br › Visas

Tourism & Visitor | U.S. Embassy & Consulates in Brazil

In general, tourists traveling to the United States require valid B-2 visas. That is unless they are eligible to travel visa ...

Can you get medicine for someone pharmacy

BEFORE

AFTER

9:00

google.com

MedlinePlus (.gov) › ency › article

Getting a prescription filled: MedlinePlus Medical Encyclopedia

Aug 26, 2017 · Your health care provider may give you a prescription in ... Writing a paper prescription that you take to a local pharmacy ... Some people and insurance companies choose to use ...

9:00

google.com

HHS.gov › hipaa › for-professionals

Can a patient have a friend or family member pick up a prescription ...

Dec 19, 2002 · A pharmacist may use professional judgment and experience with common practice to ... the patient's best interest in allowing a person, other that the patient, to pick up a prescription.

Message not received

CITIZEN LAB

The Alan Turing Institute

Citizen participation and machine learning for a better democracy

Using machine learning and natural language processing to overcome information overload in citizen participation platforms

nic.br | ceweb.br

Notícias ▾ Eventos ▾ Publicações ▾ Projetos ▾ Cursos Quem somos ▾ Contato

Notas | 15 JUL 2020

Modelo de Machine Learning que auxilia na detecção de potenciais discursos de ódio nas redes sociais

O Ceweb.br torna público o código de implementação de um modelo de Machine Learning que auxilia na detecção de potenciais discursos de ódio nas redes sociais. Este projeto está sendo realizado em parceria com o laboratório de ciências cognitivas da Queen Mary University of London e implementa as técnicas consideradas estado da arte na área de processamento de linguagem natural. Este código implementa um modelo, da arquitetura BERT, pré-treinado em PT-BR e um treinamento para o ajuste fino (fine-tuning) para a tarefa específica de classificação de sentenças de discurso de ódio. Foi utilizado, como base de treinamento, o corpus publicado por Fortuna, P., Rocha Da Silva, J., Soler-Company, J., Wanner, L., & Nunes, S. (2019). A Hierarchically-Labeled Portuguese Hate Speech Dataset.

Diogo Cortiz