

THIS AYN RANDOM NUMBER GENERATOR YOU WROTE *CLAIMS* TO BE FAIR, BUT THE OUTPUT IS BIASED TOWARD CERTAIN NUMBERS.

*WELL, MAYBE THOSE NUMBERS ARE JUST INTRINSICALLY BETTER!*



# Bias

Eric Wong

9/1/2022

# Announcements

- Ed Discussion is up
- Sign up for paper presentations
- Materials from first lecture are posted

# Paper presentations

- Readings posted by the end of Friday for next week (suggestions welcome)
- Presenters: prepare notes to lead seminar
- Everyone else: post in Ed Discussion before class

# What is bias?

“A tendency for or against something”

## Contents [\[hide\]](#)

### 1 Etymology

### 2 Types of bias

#### 2.1 Cognitive biases

##### 2.1.1 Anchoring

##### 2.1.2 Apophenia

##### 2.1.3 Attribution bias

##### 2.1.4 Confirmation bias

##### 2.1.5 Framing

##### 2.1.6 Halo effect and horn effect

##### 2.1.7 Self-serving bias

##### 2.1.8 Status quo bias

#### 2.2 Conflicts of interest

##### 2.2.1 Bribery

##### 2.2.2 Favoritism

##### 2.2.3 Lobbying

##### 2.2.4 Regulatory issues

##### 2.2.5 Shilling

#### 2.3 Statistical biases

##### 2.3.1 Forecast bias

##### 2.3.2 Observer-expectancy effect

##### 2.3.3 Reporting bias & social desirability bias

##### 2.3.4 Selection bias

#### 2.4 Prejudices

##### 2.4.1 Ageism

##### 2.4.2 Classism

##### 2.4.3 Lookism

##### 2.4.4 Racism

##### 2.4.5 Sexism

### 3 Contextual biases

#### 3.1 Biases in academia

##### 3.1.1 Academic bias

##### 3.1.2 Experimenter bias

##### 3.1.3 Funding bias

##### 3.1.4 Full text on net bias

##### 3.1.5 Publication bias

#### 3.2 Biases in law enforcement

##### 3.2.1 Driving while black

##### 3.2.2 Racial profiling

##### 3.2.3 Victim blaming

#### 3.3 Biases in media

##### 3.3.1 Agenda setting

##### 3.3.2 Gatekeeping

##### 3.3.3 Sensationalism

#### 3.4 Other contexts

##### 3.4.1 Educational bias

##### 3.4.2 Inductive bias

##### 3.4.3 Insider trading

##### 3.4.4 Match fixing

#### 3.5 Implicit bias

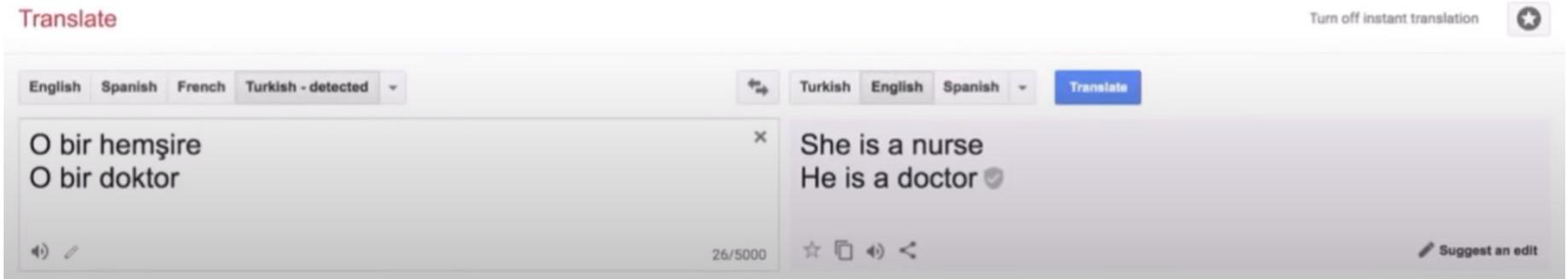
# Types of bias

- Statistical bias (*board*)
- Learning bias (*board*)
- Data bias (*these slides*)

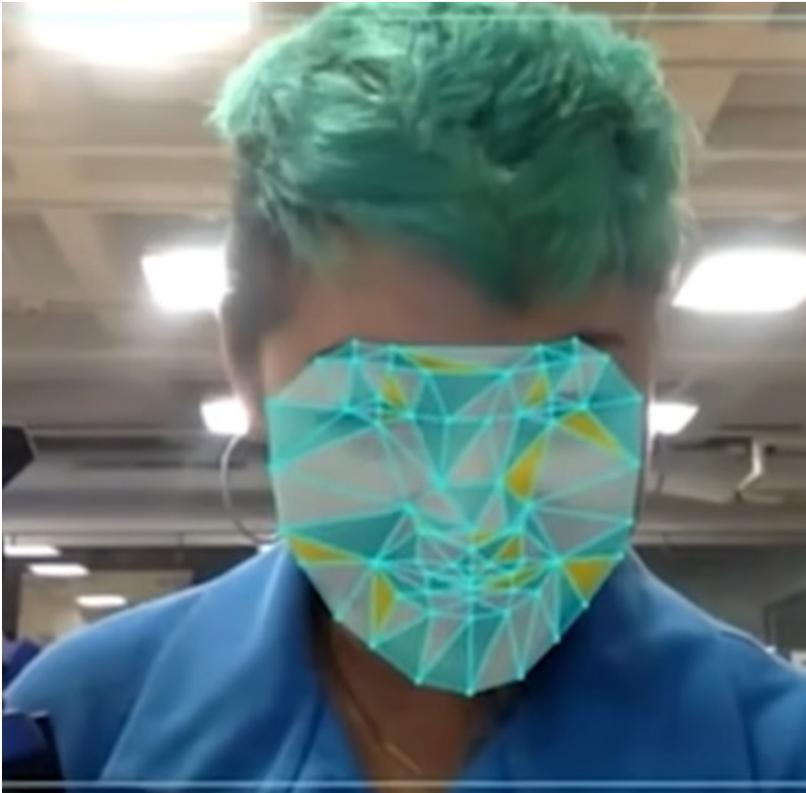
# Harms of representation

1. Stereotyping
2. Recognition
3. Denigration
4. Under representation

# 1. Stereotyping

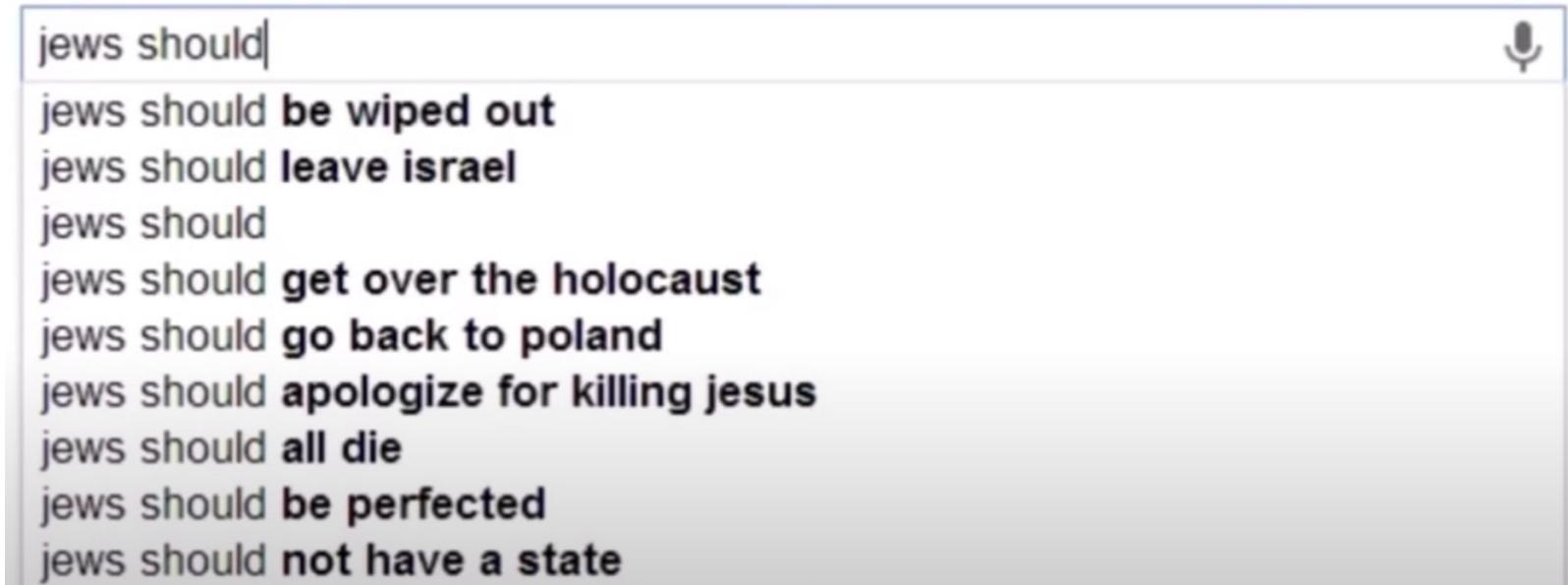


## 2. Recognition



Buolamwini & Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification"

# 3. Denigration



# 4. Under representation



Kay et al. 2015, "Unequal Representation and Gender Stereotypes in Image Search Results for Occupations"

PC: Kate Crawford, "The Trouble with Bias" NeurIPS 2017 keynote

The screenshot shows the Google Translate interface. On the left, the source text in Turkish is "O bir hemşire" and "O bir doktor". On the right, the target text in English is "She is a nurse" and "She is a doctor" with a gender icon. A note above the English text states "Gender-specific translations are limited. [Learn more](#)". The interface includes language selection tabs for Turkish, English, and Spanish, a character count of 26 / 5,000, and icons for voice input, output, copy, and share.

[Send feedback](#)

## Understand gender-specific translation order

Gender-specific translations show in alphabetical order by gender label. For example, in English, “feminine” comes alphabetically before “masculine.”

In languages without alphabets, gender-specific translations show in each language’s standard indexing order by gender label.

Q jews

Q jews **harp**

 **Jews In Space**  
Song by Mel Brooks and Ralph Burns

Q jews **in india**

Q jews **definition**

Q jews **and gentiles**

Q jews **of the wild west**

Q jews **in baseball**

Q jews **in new york**

Q jews **in ireland**

Q jews **in iran**

Google Search I'm Feeling Lucky

*Report inappropriate predictions*

Q jews should

Google Search I'm Feeling Lucky

Q asians should

Google Search I'm Feeling Lucky

Q ducks should

Q ducks **what** should **we da**

Q ducks should **not eat bread**

Q ducks shouldn't eat bread

Q **why** shouldn't ducks eat bread

Q ducks shoulder **tattoo**

Q **how many** ducks should i **get**

Q **what gender** ducks should i **get**

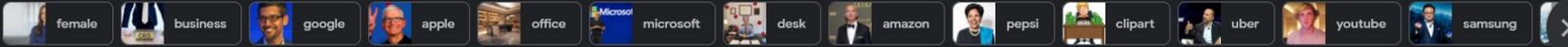
Q **what** ducks should **not eat**

Q **how many** ducks should i **keep**

Q **what kind of** ducks should i **get**

Google Search I'm Feeling Lucky

*Report inappropriate predictions*



CEO vs. Owner: Key Differences | Ohio ...  
onlinemasters.ohio.edu



Chief Executive Officer (CEO): What ...  
investopedia.com



Satya Nadella - Wikipe...  
en.wikipedia.org



Chief Executive Officer (CEO) Job ...  
thebalancecareers.com



CEO Job Description  
betterteam.com



World's most powerful CEOs come from ...  
cntraveller.in



Harvard study: What CEOs do all day  
cnbc.com



Famous CEOs in International Business ...  
internationalbusinessguide.org



Is your CEO a credible spokesperson ...  
cuttingedgepr.com



You Are The CEO of Your Life - Personal...  
personalexcellence.co



CEO vs. President: What's the ...  
smartasset.com



What Are the Duties of a CEO?  
entrepreneur.com



Becoming a CEO: Unexpected traits top ...  
execed.economist.com



Do women make better CEOs than men ...  
lighthouse.mq.edu.au



330,536 Ceo Images, Stock Photos ...  
shutterstock.com



What do CEOs do? A CEO Job Description ...  
steverobbins.com



as CEO to Help Your Business Grow ...  
inc.com



7 Personality Traits Every CEO Should Have  
forbes.com



Video & CEO TV | The CEO Magazine  
theceomagazine.com



Female CEO: &#8216;Women Will Never Be ...  
inc.com



What it takes to be a great CEO ...  
welcome-center-germany.com



CEO vs. President  
findlaw.com



CEO vs. CFO  
betterteam.com



CEO job description template | Workable  
resources.workable.com

Related searches

- female ceo
- ceo logo
- business ceo

# Bias part 2

Eric Wong

9/6/2022

# Today

- Notebook on calculating bias
- Project proposal guidelines posted
  - Form teams and submit by Oct 4th
- Specific biases
- Paper presentation

# Reminders

- Sign up for paper presentations
- Post/reply on Ed discussion
- Start thinking about projects and form groups

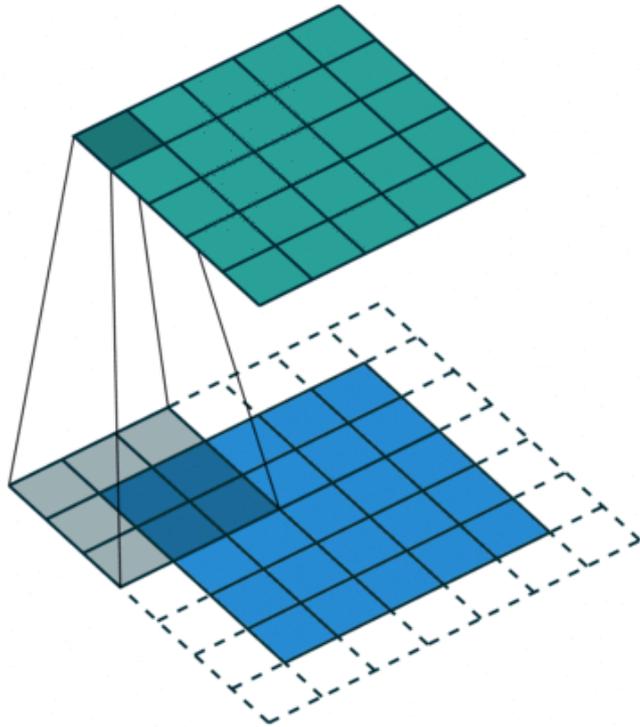
# Feedback

- Temperature on DRL
- Lights in CHEM

# Missingness bias

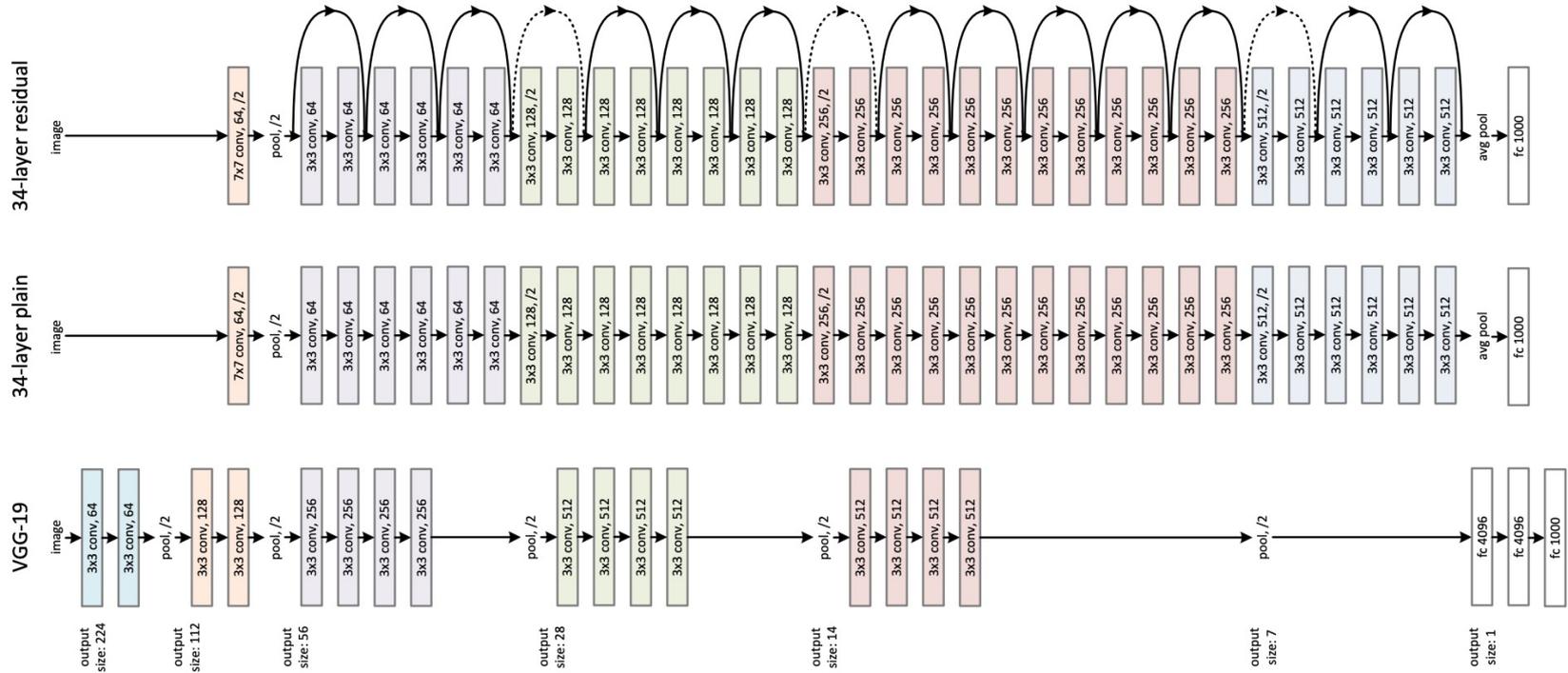
A bias that occurs *when data is missing*

# Convolution layer



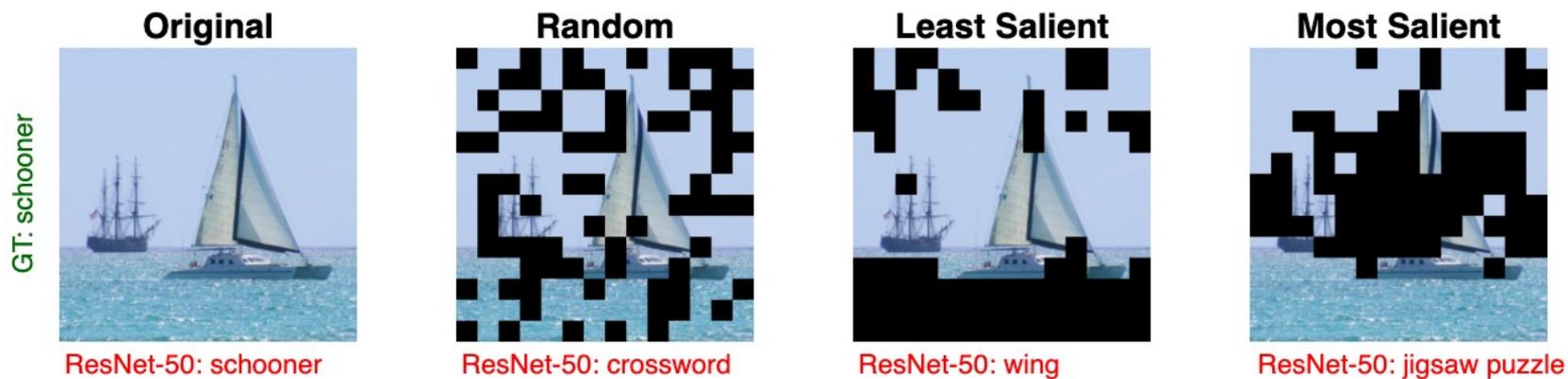
$$f(x)_{m,n} = \sum_i \sum_j w_{ij} x_{m-i, n-j}$$

# Case Study: ResNets

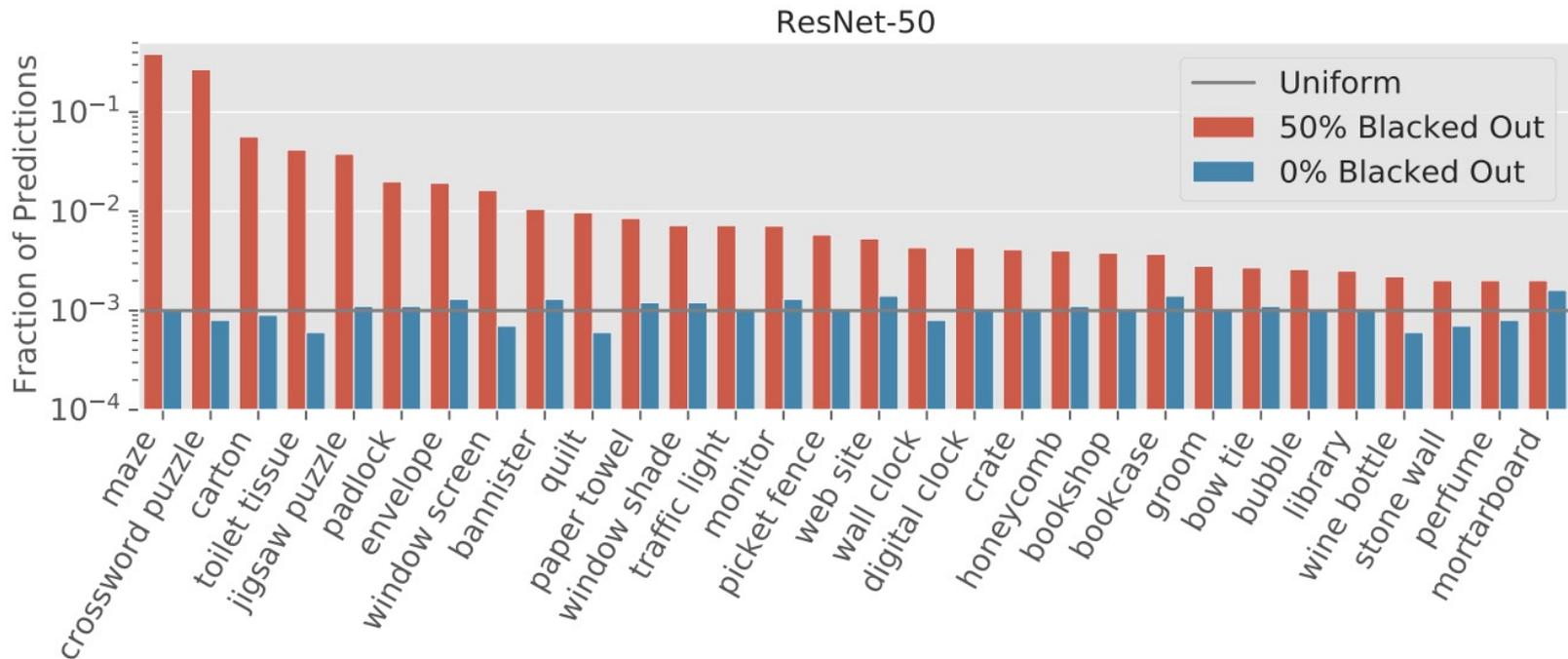


He, et al. 2016 "Deep residual learning for image recognition."

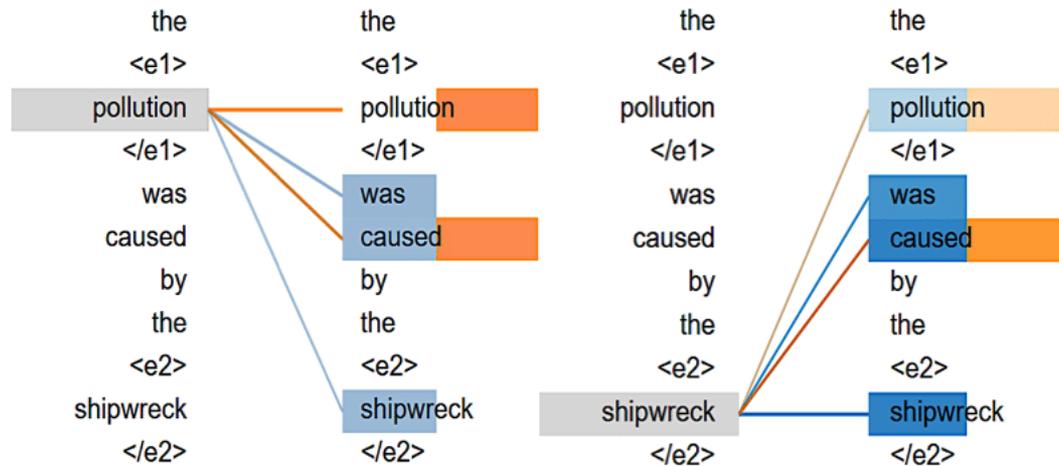
# Blacking out pixels



# Skewed class distribution



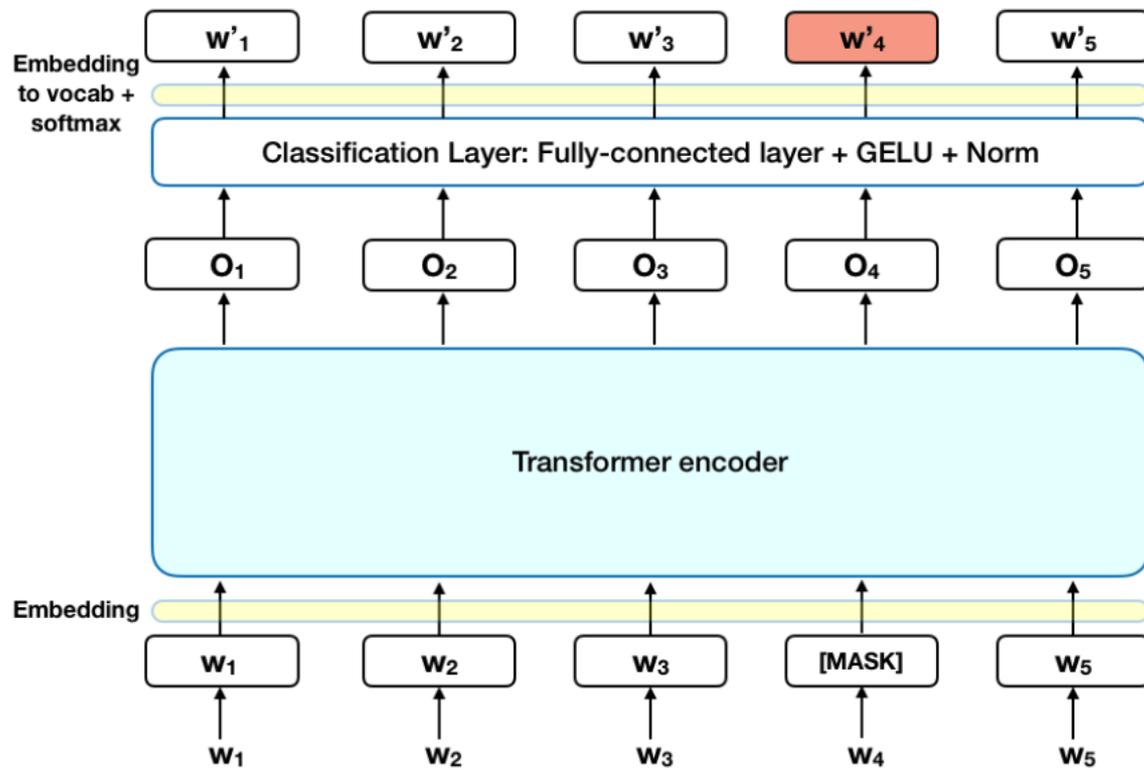
# Attention layer



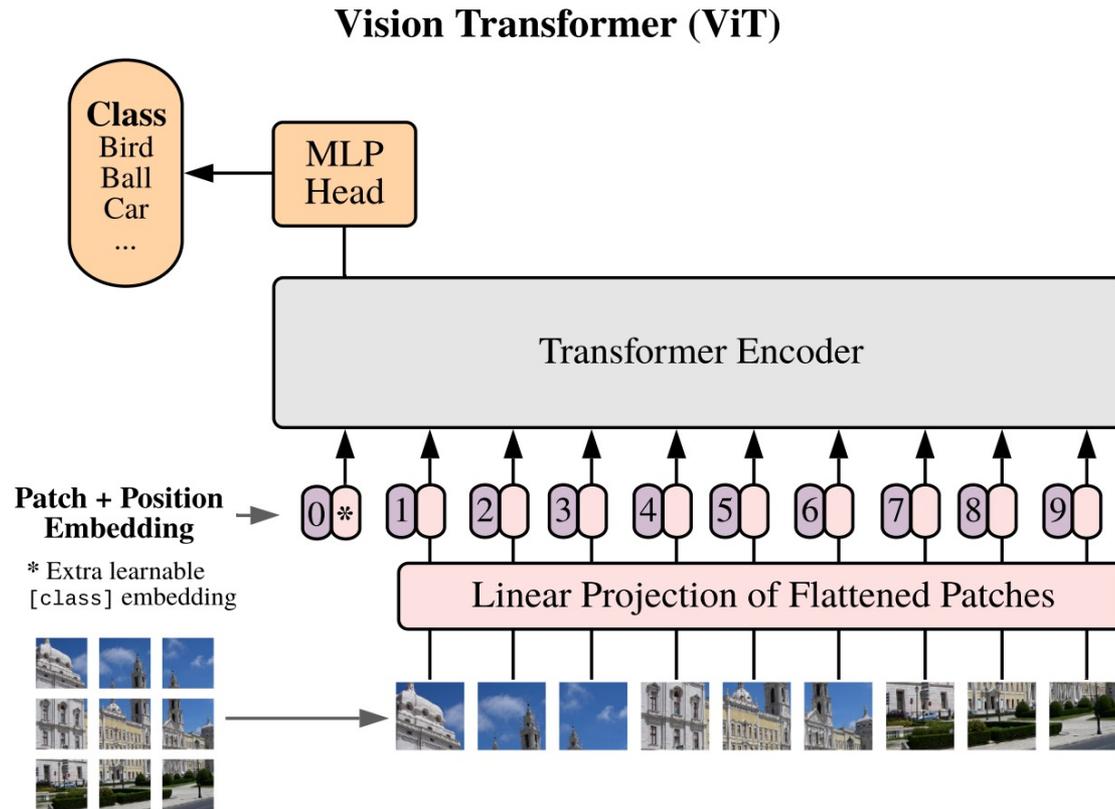
$$f(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V$$

Self attention:  $Q, K, V = W_Q x, W_K x, W_V x$

# Transformer

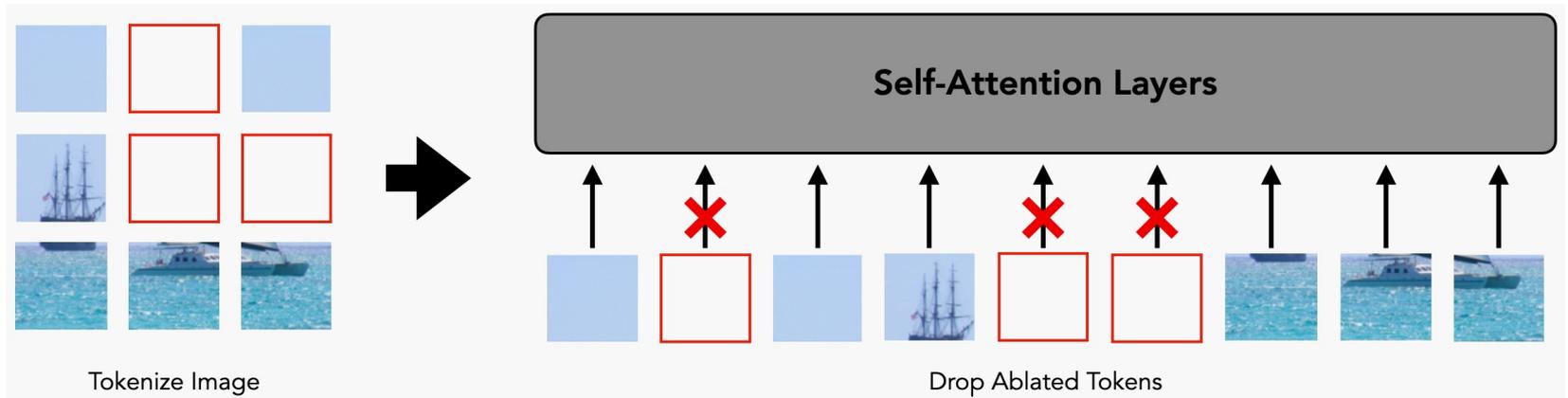


# Case Study: Vision transformer



Dosovitskiy et al. 2021 "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale"

# “True” missingness



# Mitigates learning bias

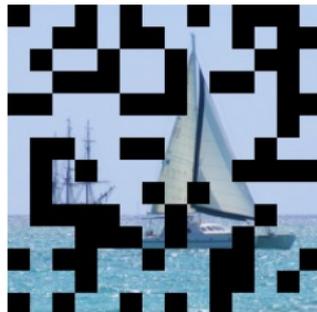
GT: schooner

**Original**



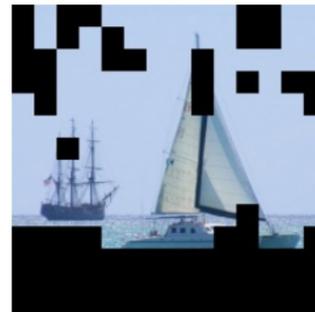
ResNet-50: schooner  
ViT-S: catamaran

**Random**



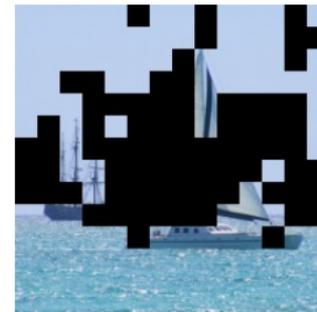
ResNet-50: crossword  
ViT-S: catamaran

**Least Salient**



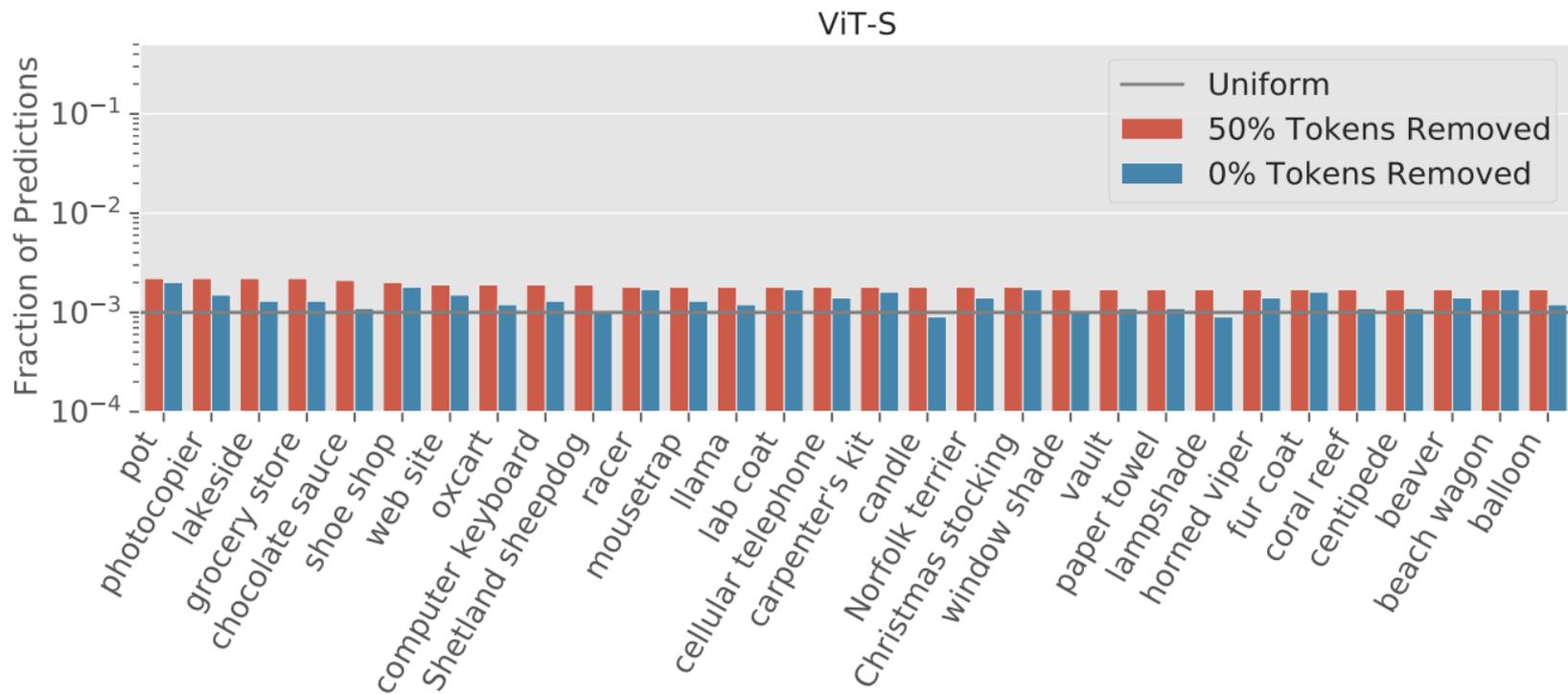
ResNet-50: wing  
ViT-S: catamaran

**Most Salient**



ResNet-50: jigsaw puzzle  
ViT-S: catamaran

# Uniform class distribution



# Missing words in text

Occupation templates:

“[MASK] is a doctor/nurse/dancer”

Compare probability scores for  
different pronouns

# Measuring data bias in word embeddings

- Analogies
- Implicit Association Test
- Distances or alignment to group subspace

# Analogies

Prediction of stereotypes as a  
measure of bias

*doctor : man :: woman : nurse*

# Implicit Association

Cosine similarity between word embeddings of concepts and attributes

Examples:

- European vs American names (concepts)
- Pleasant vs Unpleasant (attributes)