Artificial Intelligence for Social Media Marketing: Data, Methods, and Insights

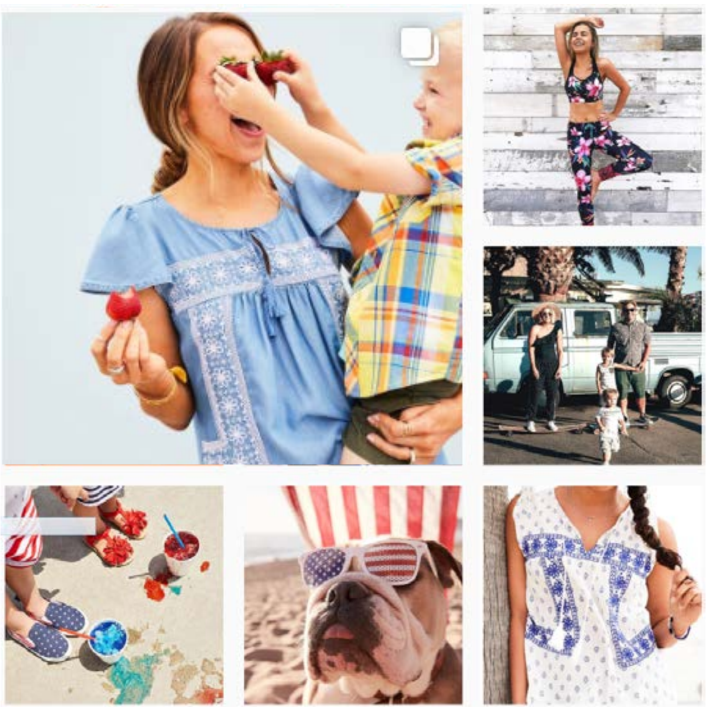
Liu Liu

University of Colorado Boulder – Leeds School of Business

MSI Accelerator, September 15th 2022

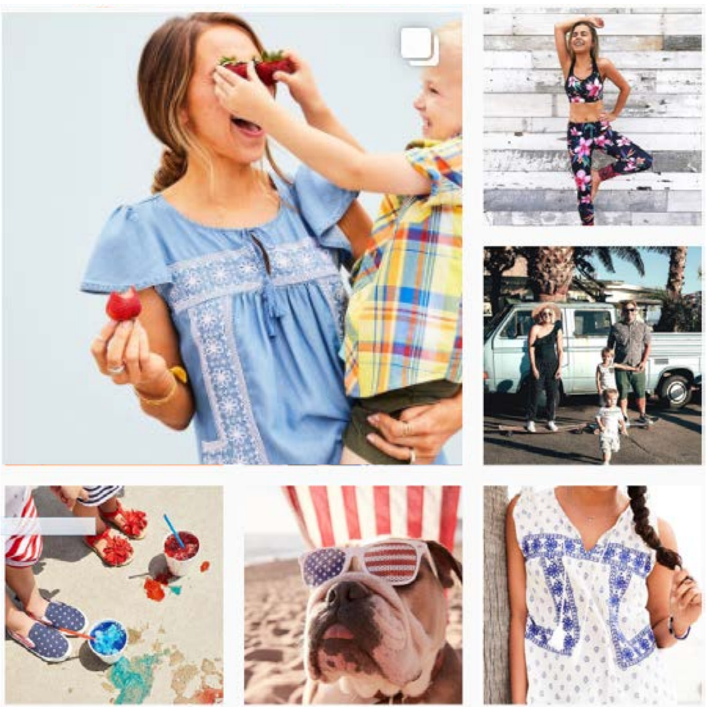
Brand Manager

@oldnavy



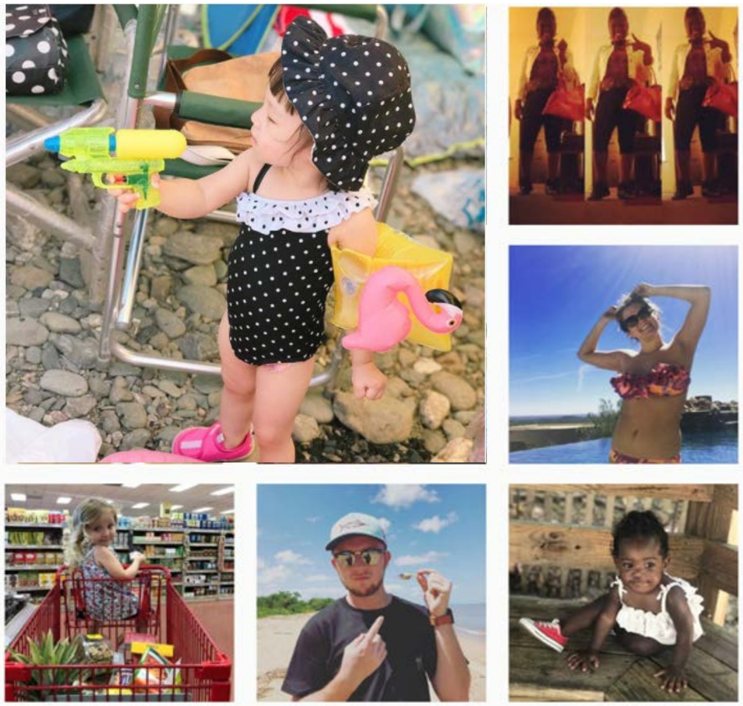
Brand Manager

@oldnavy



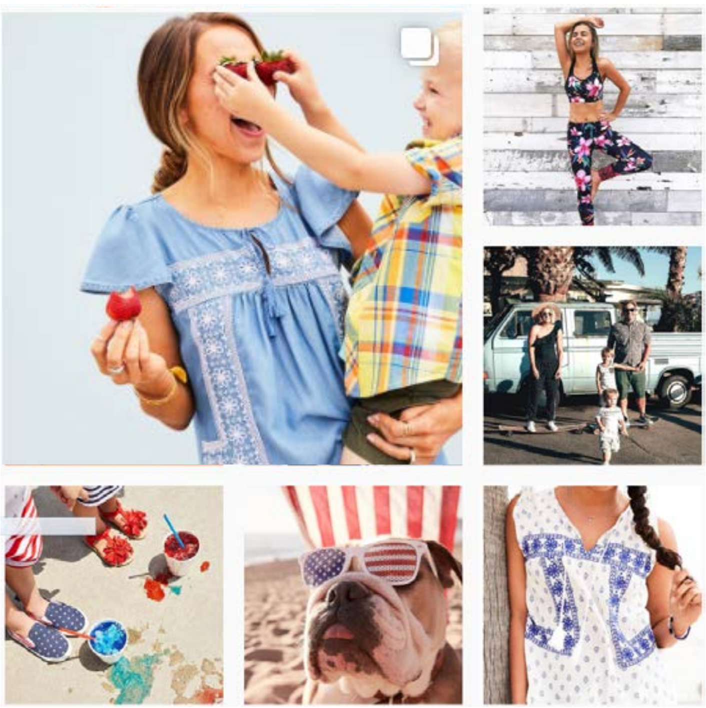
Consumer

#oldnavy



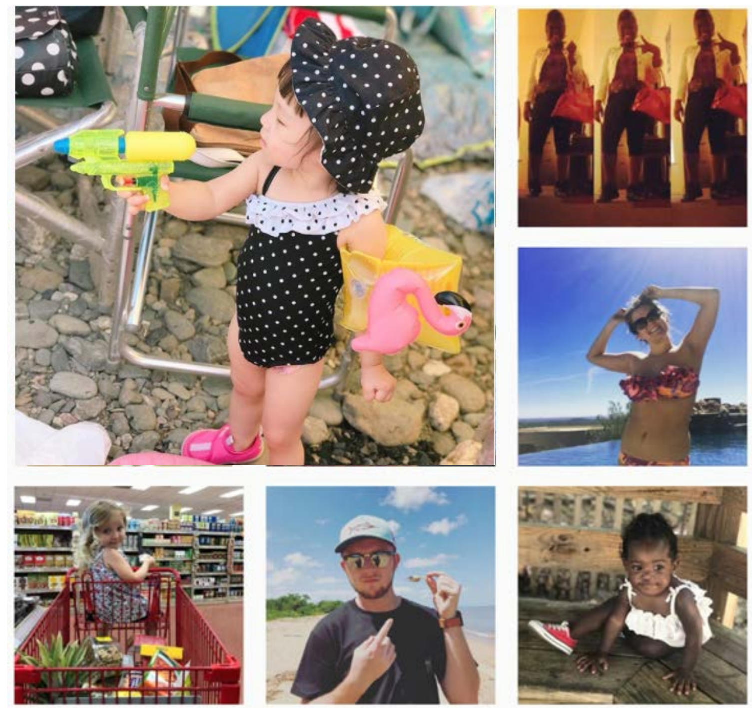
Brand Manager

@oldnavy



Consumer

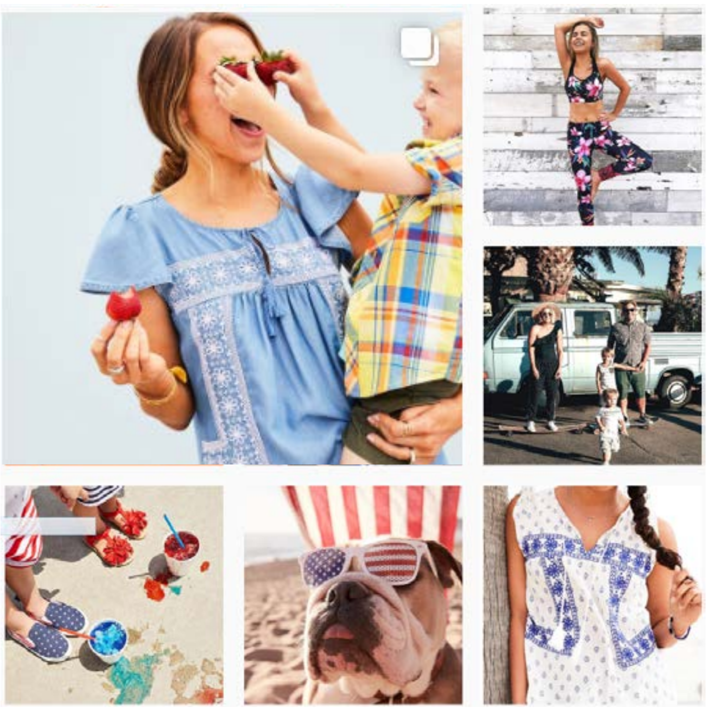
#oldnavy



How is my brand portrayed in **consumer photos**?

Brand Manager

@oldnavy



Consumer

#oldnavy

How is my brand portrayed in **consumer photos**?

Increasing Amount of Unstructured Social Conversations

Text

Image

#eddiebauer

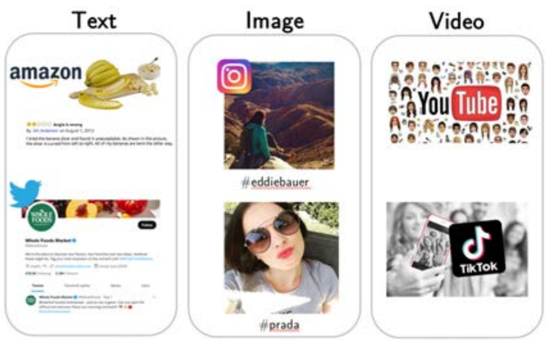
#prada

Video

How can we extract insights and get value?

Insights/Problems

Data



Product Design

- Understand customers needs

Branding

- Measure brand perceptions

Advertising and social media

- Generate social media posts and ads

Influencer marketing

- What makes a good influencer video ads?

...

Agenda

Artificial Intelligence (AI): A brief intro and recent breakthroughs with Deep learning

AI for Social Media Marketing: Cases studies

- Visual Listening In: Extracting Brand Image Portrayed on Social Media (Liu et al. 2020)
- Identify Customer Needs from User-Generated Content (Timoshenko and Hauser 2019)
- Supporting Content Marketing with Natural Language Generation (Reisenbichler et al. 2022)

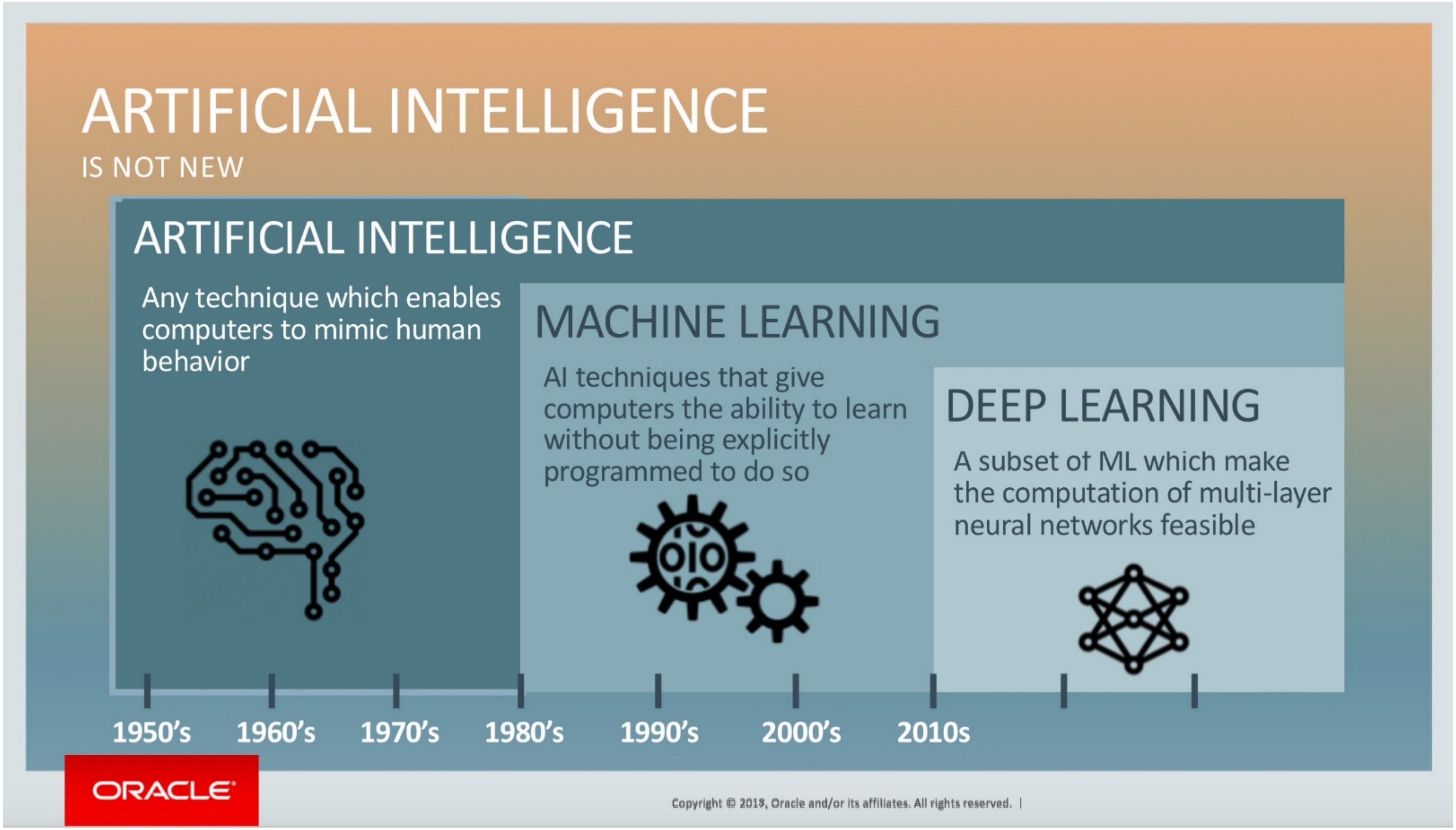
Brainstorming and group work

Sharing group work and reflections (e.g., pitfalls)

AI and Recent Breakthroughs with Deep Learning

AI History Outline

1. - 1956 Prehistory of Artificial Intelligence
2. 1956 – 1974 First Artificial Intelligence Spring
3. Break for Activities
4. 1980 - 1987 Second Artificial Intelligence Spring
5. Break for Activities
6. **2011 - Present Third Artificial Intelligence Spring**
7. What is next?


ARTIFICIAL INTELLIGENCE
IS NOT NEW
ARTIFICIAL INTELLIGENCE
Any technique which enables computers to mimic human behavior
MACHINE LEARNING
AI techniques that give computers the ability to learn without being explicitly programmed to do so
DEEP LEARNING
A subset of ML which make the computation of multi-layer neural networks feasible
1950's
1960's
1970's
1980's
1990's
2000's
2010s
ORACLE
Copyright © 2018, Oracle and/or its affiliates. All rights reserved.

Deep learning are taking over

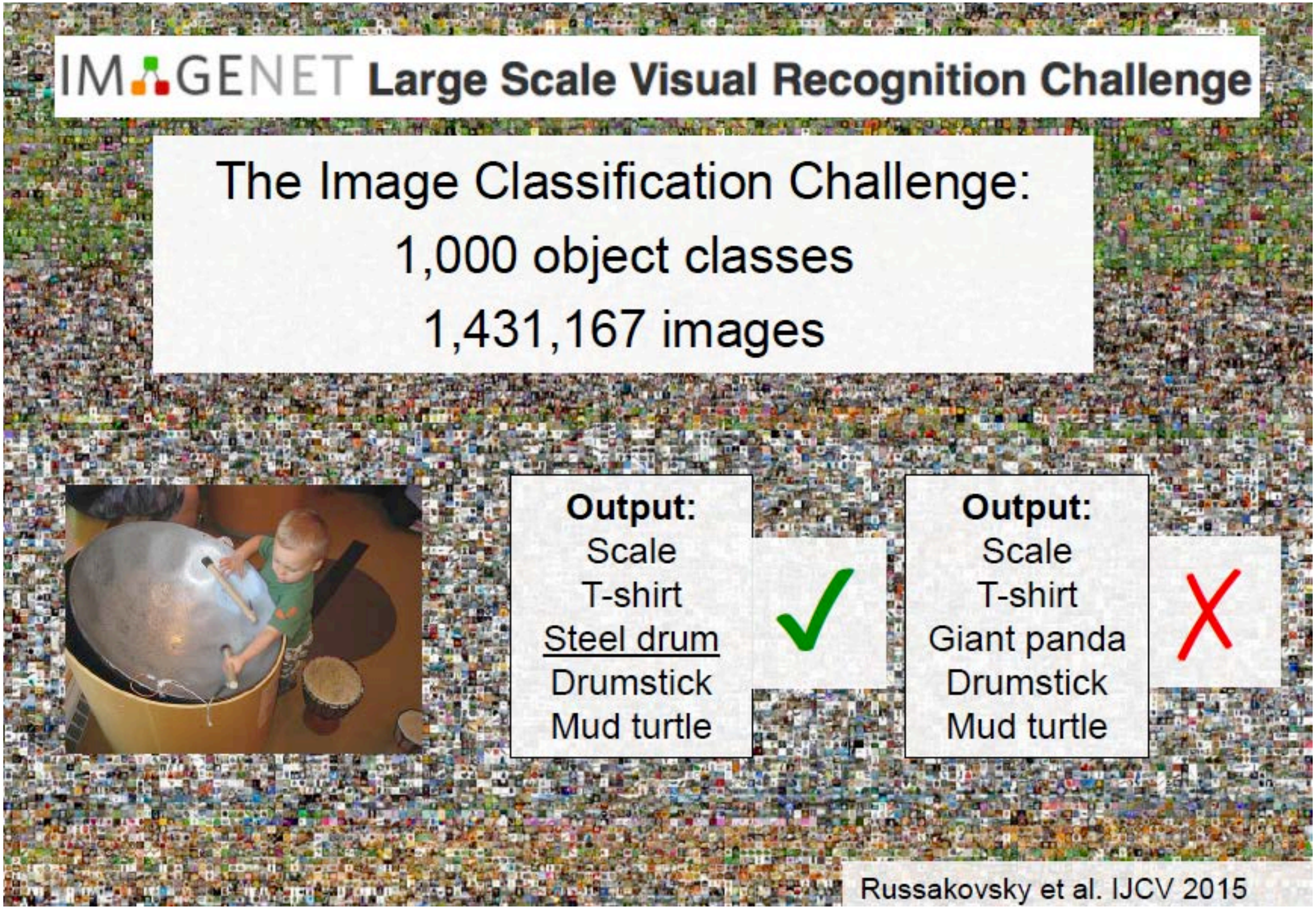
Deep learning have become one of the main approaches to AI

They have been successfully applied to various fields

They have established the state of the art

- Often exceeding previous benchmarks by large margin
- Sometimes solving problems you couldn't solve using earlier ML methods

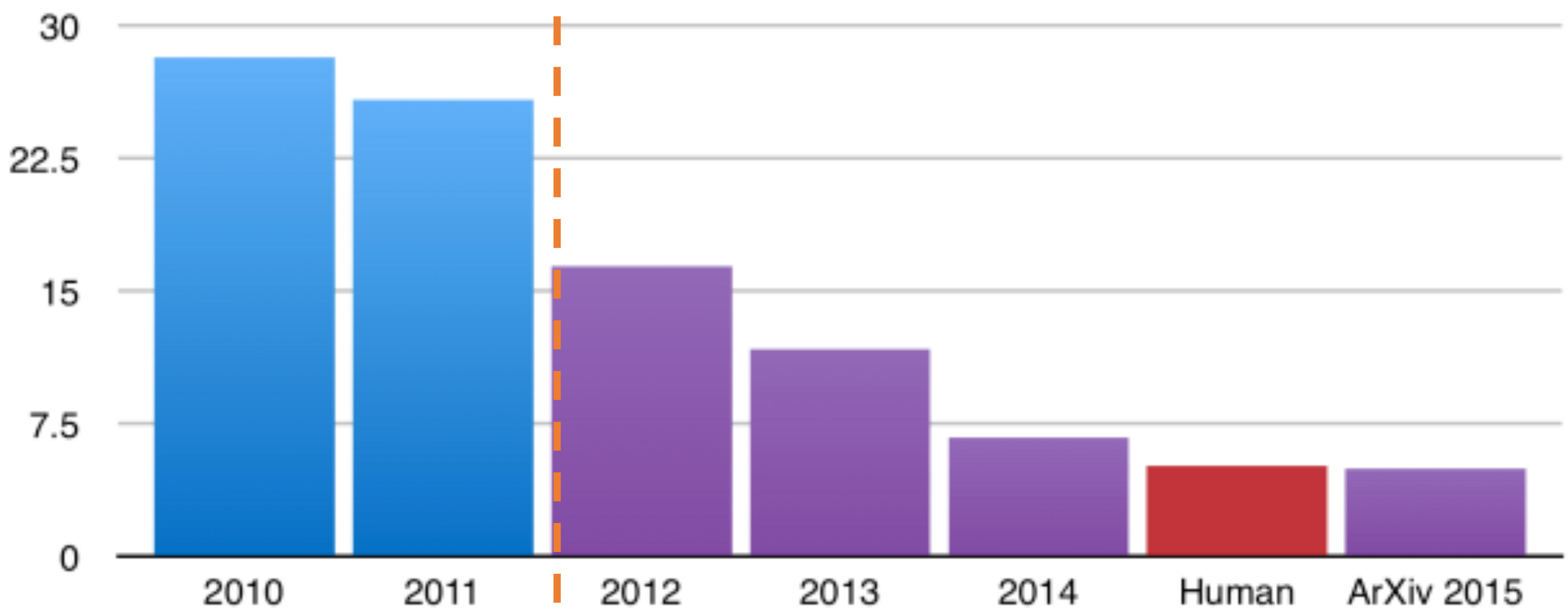

IMAGENET
22,000 categories
15,000,000 images
J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li & L. Fei-Fei. CVPR, 2009.

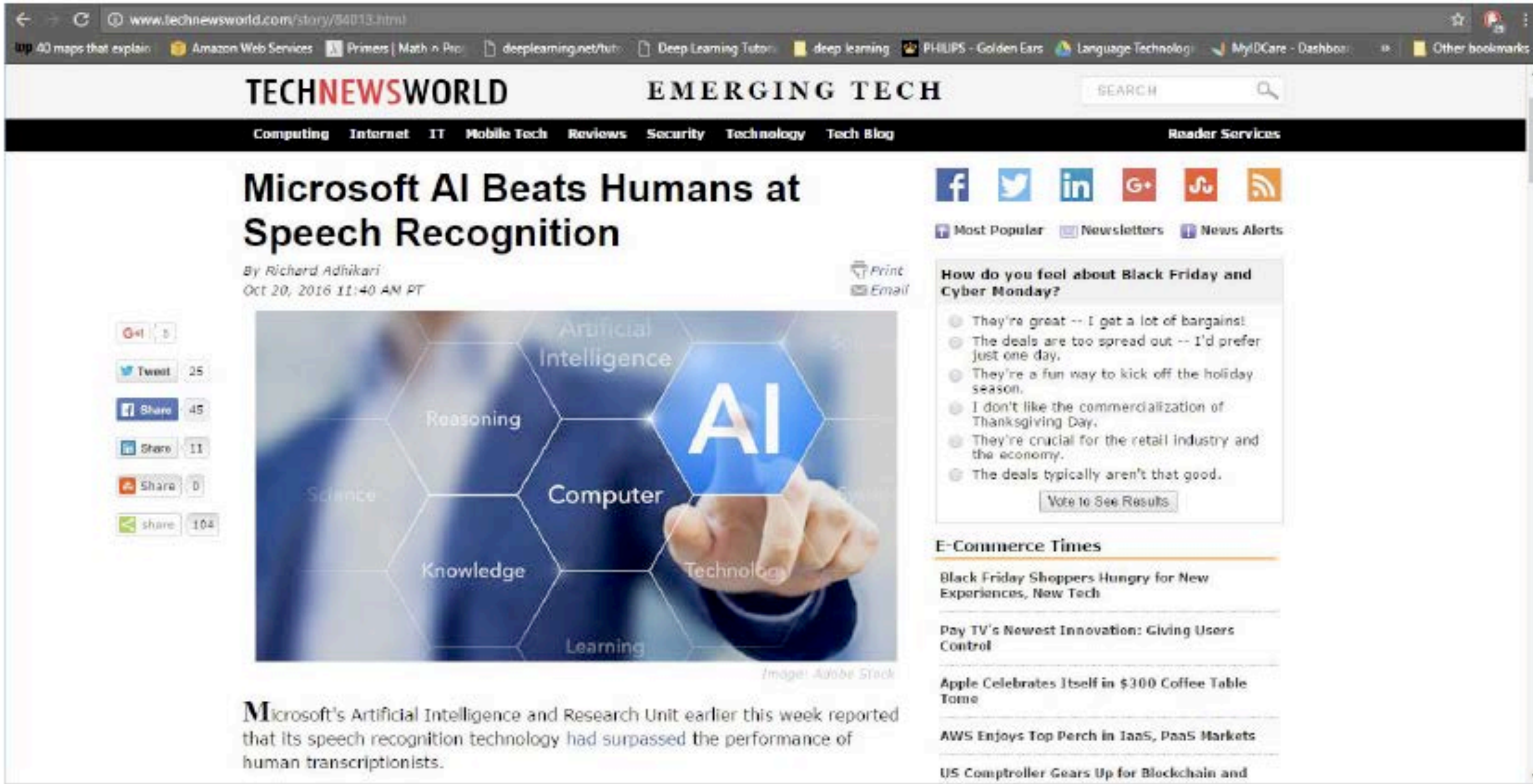

IMAGENET Large Scale Visual Recognition Challenge
The Image Classification Challenge:
1,000 object classes
1,431,167 images
Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle
Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle
Russakovsky et al. IJCV 2015

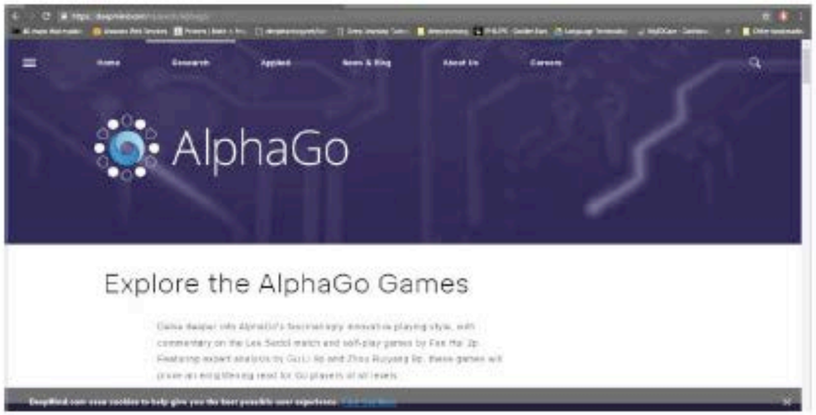
Pre-deep learning

Deep learning,
e.g., Convolutional Neural Network

Top-5 error rate on ImageNet




www.technewsworld.com/story/84013.html
TECHNEWSWORLD
EMERGING TECH
SEARCH
Computing Internet IT Mobile Tech Reviews Security Technology Tech Blog
Reader Services
Microsoft AI Beats Humans at Speech Recognition
By Richard Adhikari
Oct 20, 2016 11:40 AM PT
Print
Email
Artificial Intelligence
Reasoning
AI
Computer
Knowledge
Learning
Microsoft's Artificial Intelligence and Research Unit earlier this week reported that its speech recognition technology had surpassed the performance of human transcriptionists.
Most Popular
Newsletters
News Alerts
How do you feel about Black Friday and Cyber Monday?
They're great -- I get a lot of bargains!
The deals are too spread out -- I'd prefer just one day.
They're a fun way to kick off the holiday season.
I don't like the commercialization of Thanksgiving Day.
They're crucial for the retail industry and the economy.
The deals typically aren't that good.
Vote to See Results
E-Commerce Times
Black Friday Shoppers Hungry for New Experiences, New Tech
Pay TV's Newest Innovation: Giving Users Control
Apple Celebrates Itself in $300 Coffee Table Tome
AWS Enjoys Top Perch in IaaS, PaaS Markets
US Comptroller Gears Up for Blockchain and



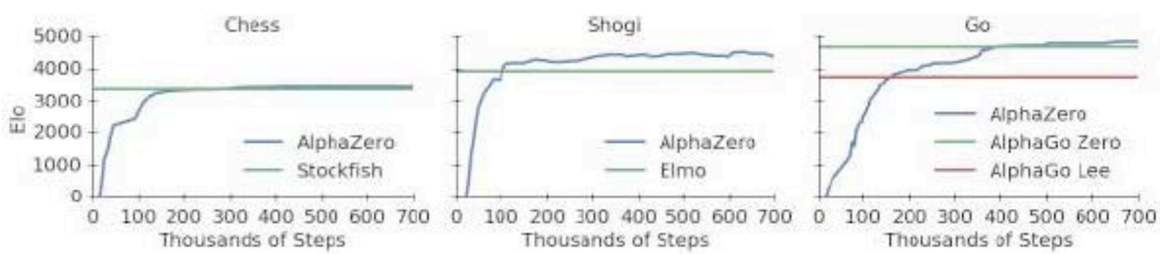
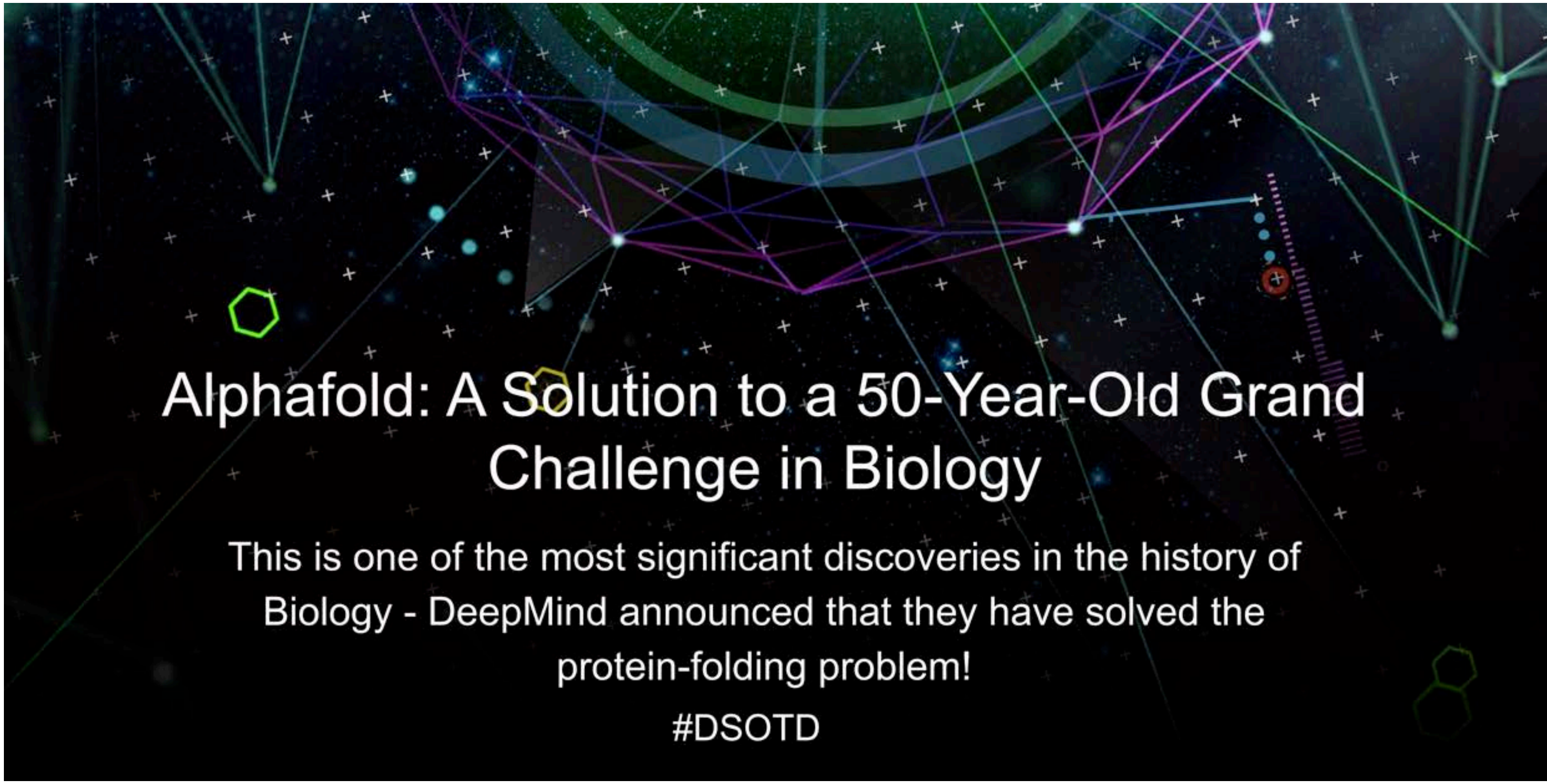
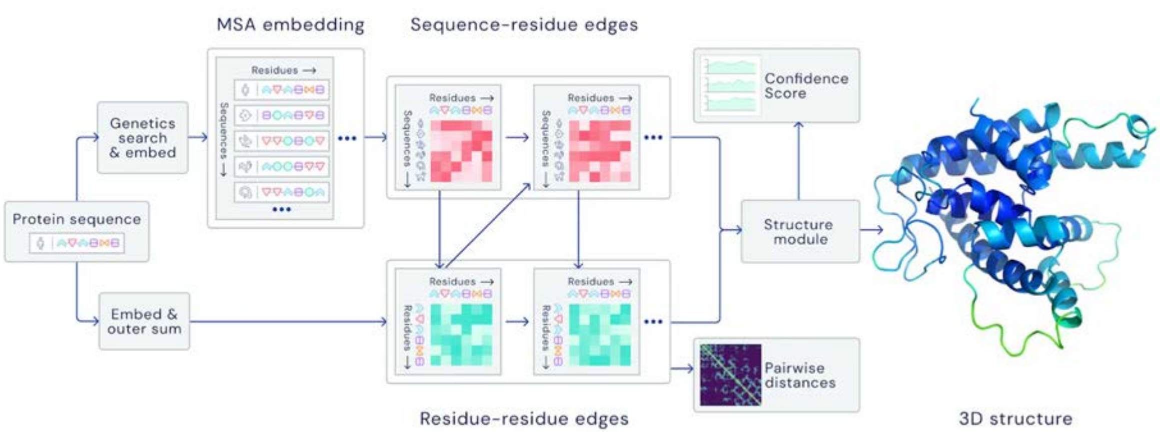


Figure 1: Training *AlphaZero* for 700,000 steps. Elo ratings were computed from evaluation games between different players when given one second per move. **a** Performance of *AlphaZero* in chess, compared to 2016 TCEC world-champion program *Stockfish*. **b** Performance of *AlphaZero* in shogi, compared to 2017 CSA world-champion program *Elmo*. **c** Performance of *AlphaZero* in Go, compared to *AlphaGo Lee* and *AlphaGo Zero* (20 block / 3 day) (*29*).





https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

ThisPersonDoesNotExist.com uses AI to generate endless fake faces

Hit refresh to lock eyes with another imaginary stranger

By James Vincent | Feb 15, 2019, 7:38am EST

SHARE

A few sample faces — all completely fake — created by ThisPersonDoesNotExist.com

https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan

MIT Technology Review

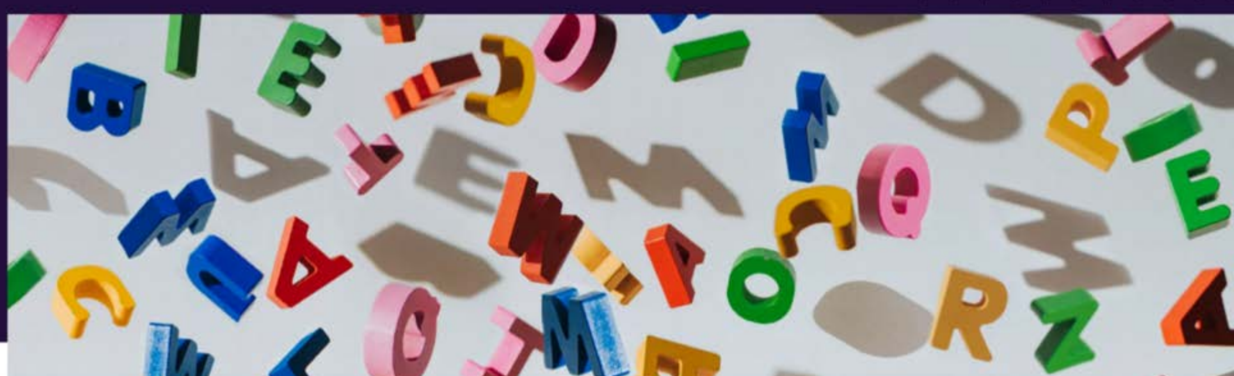
Featured Topics Newsletters Events Podcasts Sign in Subscribe

ARTIFICIAL INTELLIGENCE

OpenAI’s new language generator GPT-3 is shockingly good—and completely mindless

The AI is the largest language model ever created and can generate amazing human-like text on demand but won't bring us closer to true intelligence.

By Will Douglas Heaven July 20, 2020



https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/

GPT-3 and Text Generation

The New York Times

Meet GPT-3. It Has Learned to Code (and Blog and Argue).

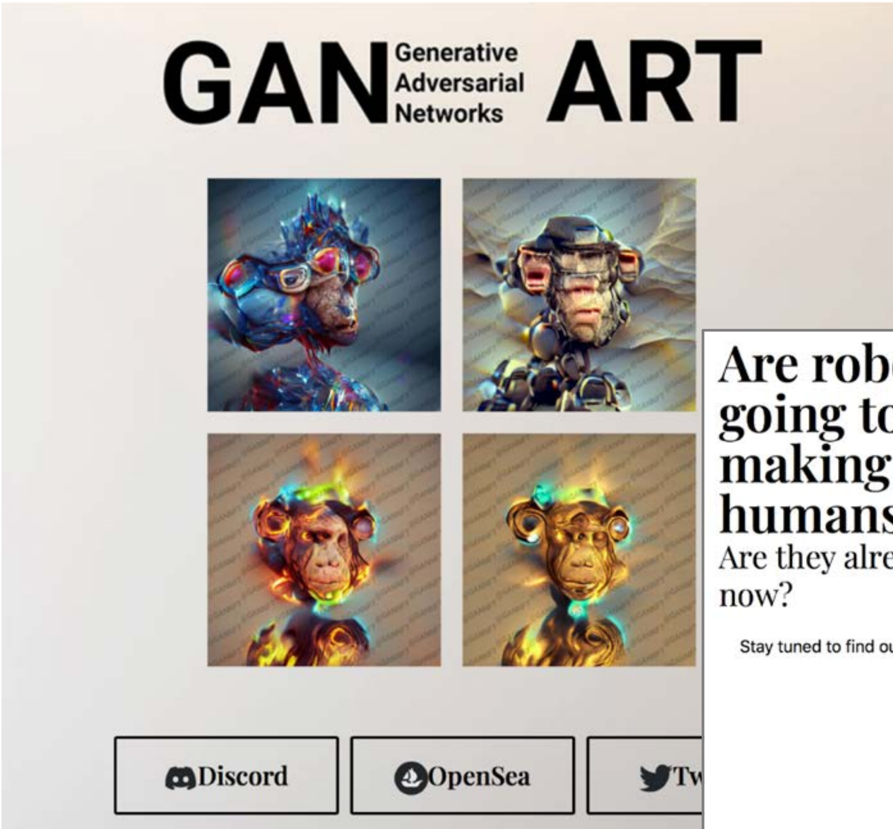
The latest natural-language system generates tweets, pens poetry, summarizes emails, answers trivia questions, translates languages and even writes its own computer programs.

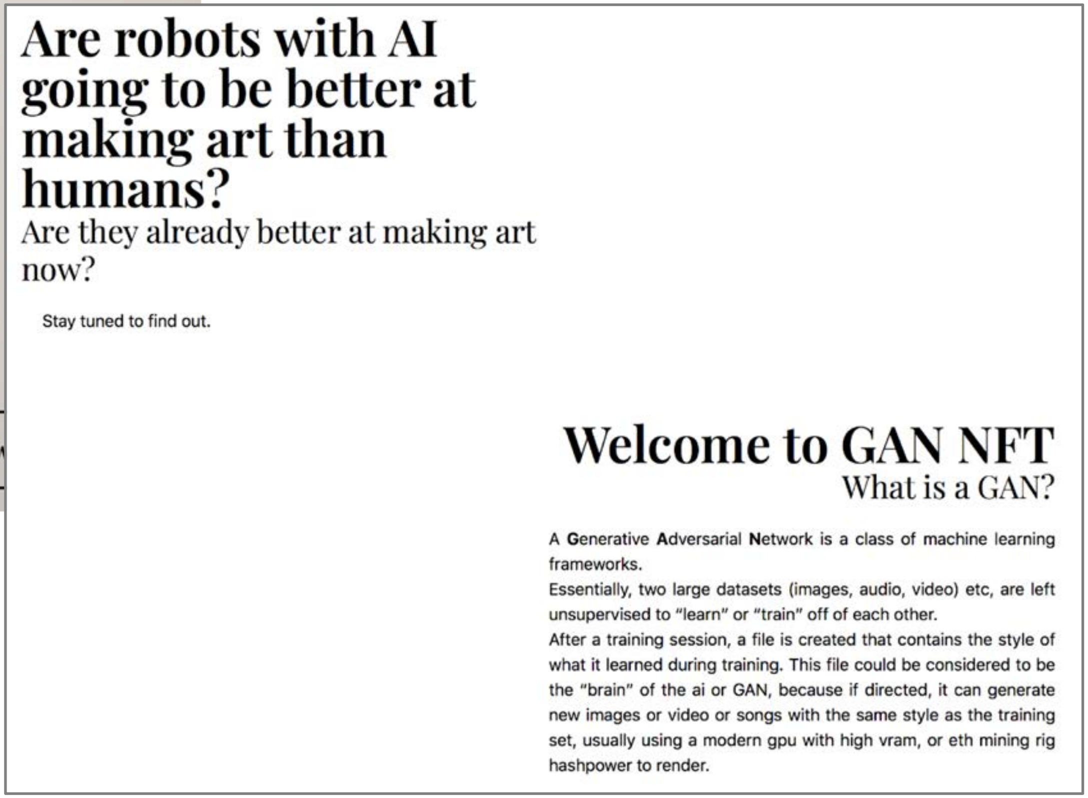
https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html?smid=url-share

One of his experiments involved a pop psychologist, Scott Barry Kaufman. The system took in Mr. Kaufman's name and a topic for discussion: creativity. Then, when asked "How do we become more creative?" GPT-3 responded instantly:

> I think creative expression is a natural byproduct of growing up in a diverse world. The more diverse the world is, the more you get exposed to different people, to different opportunities, to different places and to different challenges. And the more diverse that is, the more likely you'll be to be able to put the dots together to form something new. And in many ways, I think if you want to be creative, you have to go for it. If you want to be a writer, you have to write, if you want to be a musician, you have to create music, if you want to be a comedian, you have to create comedy, and the more you create, the more likely it is that you'll stumble onto some interesting stuff.

Later, when Mr. Wrigley posted the paragraph on Twitter, somebody looped in the real Scott Barry Kaufman. He was stunned. "It definitely sounds like something I would say," the real Mr. Kaufman tweeted, later adding, "Crazy accurate A.I."





Are robots with AI going to be better at making art than humans?

Are they already better at making art now?

Stay tuned to find out.

Welcome to GAN NFT

What is a GAN?

A **G**enerative **A**dversarial **N**etwork is a class of machine learning frameworks.

Essentially, two large datasets (images, audio, video) etc, are left unsupervised to "learn" or "train" off of each other.

After a training session, a file is created that contains the style of what it learned during training. This file could be considered to be the "brain" of the ai or GAN, because if directed, it can generate new images or video or songs with the same style as the training set, usually using a modern gpu with high vram, or eth mining rig hashpower to render.

An AI-Generated Artwork Won First Place at a State Fair Fine Arts Competition, and Artists Are Pissed

Jason Allen's AI-generated work "Théâtre D'opéra Spatial" took first place in the digital category at the Colorado State Fair.

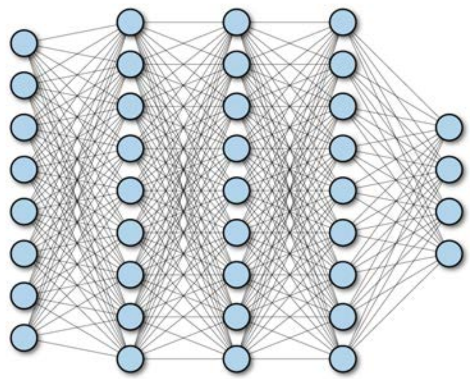
SCREENGRAB: DISCORD.

A man came in first at the Colorado State Fair's fine art competition using an AI generated artwork on Monday. "I won first place," a user going by Sincarnate said in a Discord post above photos of the AI-generated canvases hanging at the fair.

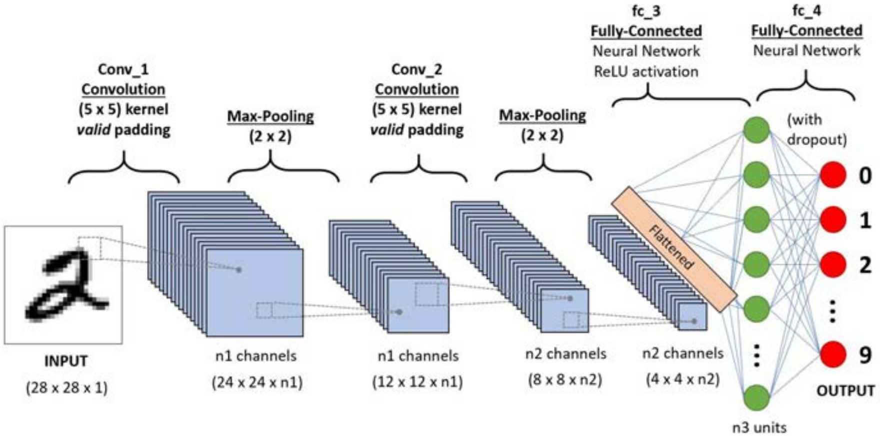
https://www.vice.com/en/article/bvmvqm/an-ai-generated-artwork-won-first-place-at-a-state-fair-fine-arts-competition-and-artists-are-pissed

https://openai.com/dall-e-2/

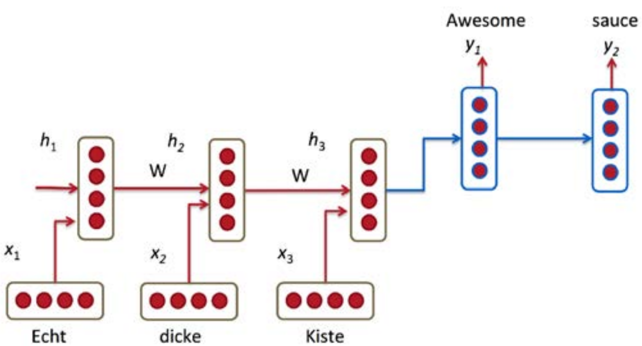
What are deep neural networks?



Fully-Connected Network



Convolutional Neural Network (CNN)



Recurrent Neural Network (RNN)

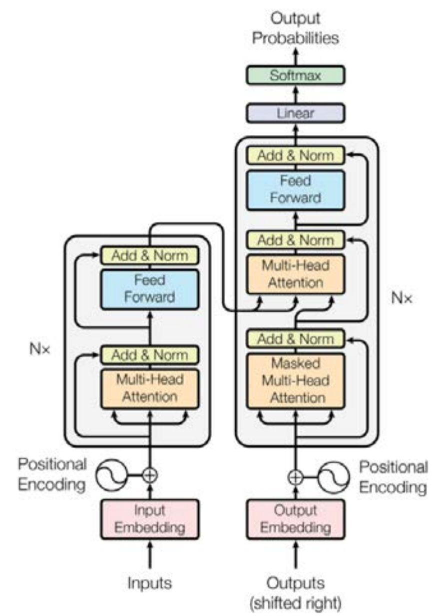
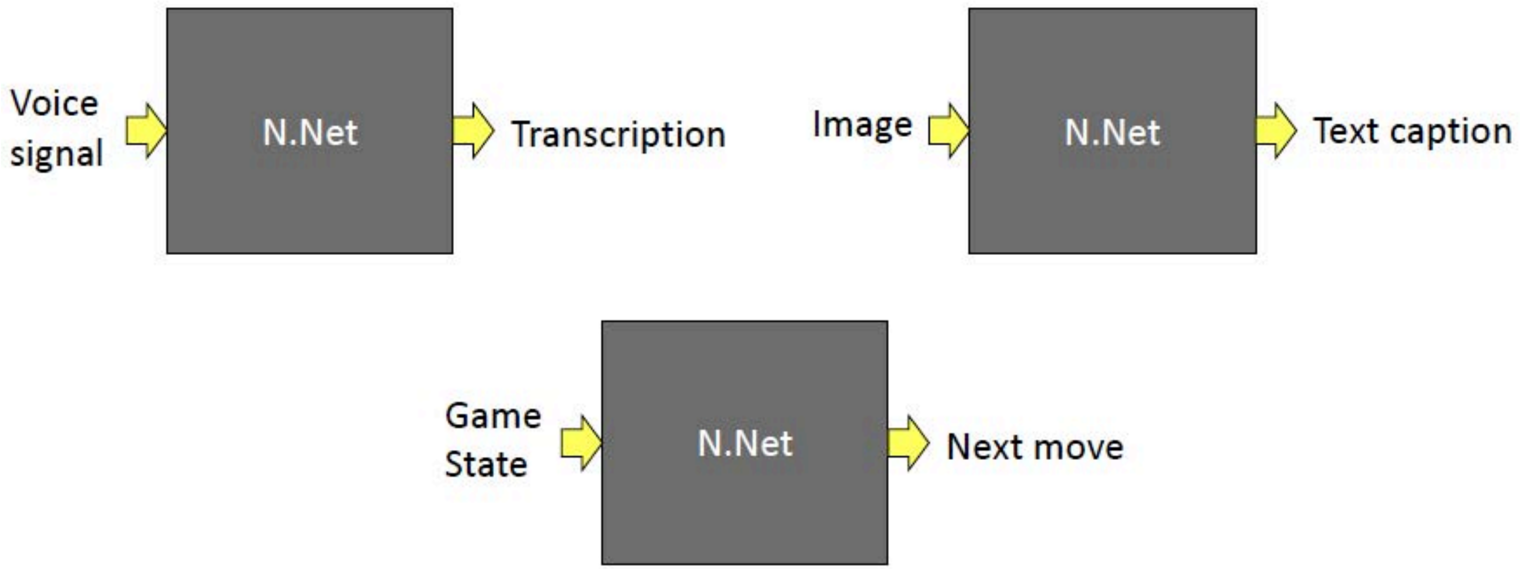


Figure 1: The Transformer - model architecture.

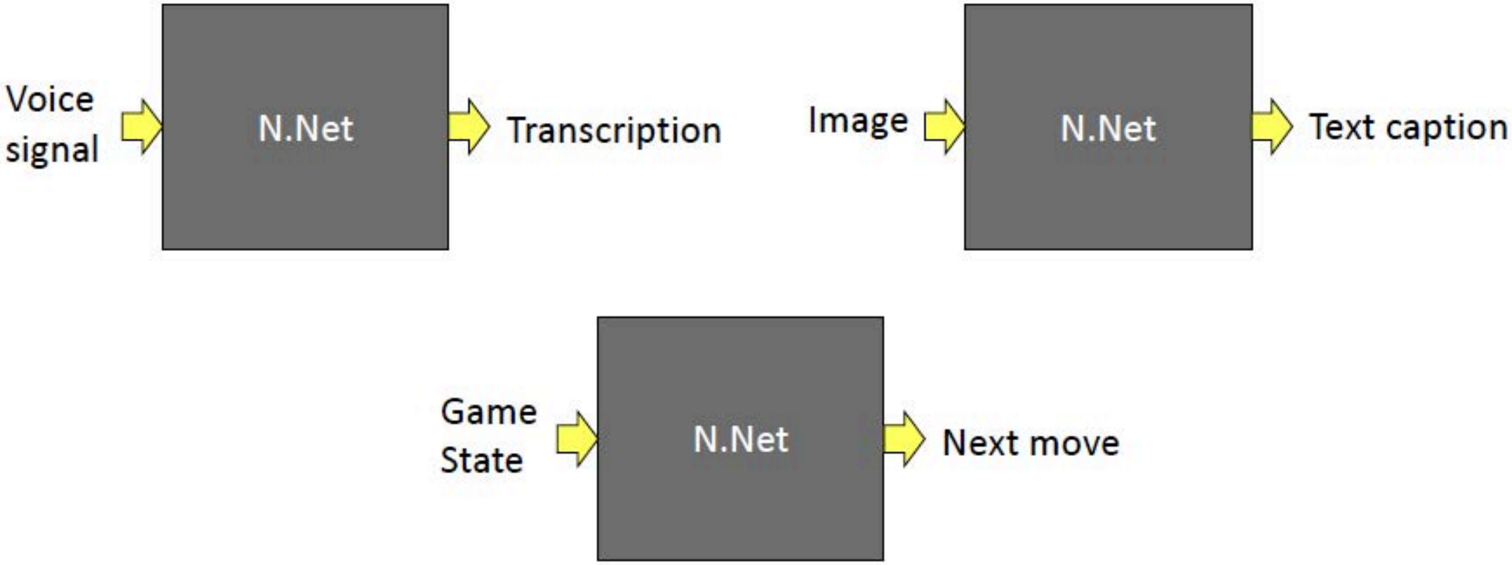
Transformer based neural network, e.g., Bert, GPT

What are deep neural networks?



What are these boxes?

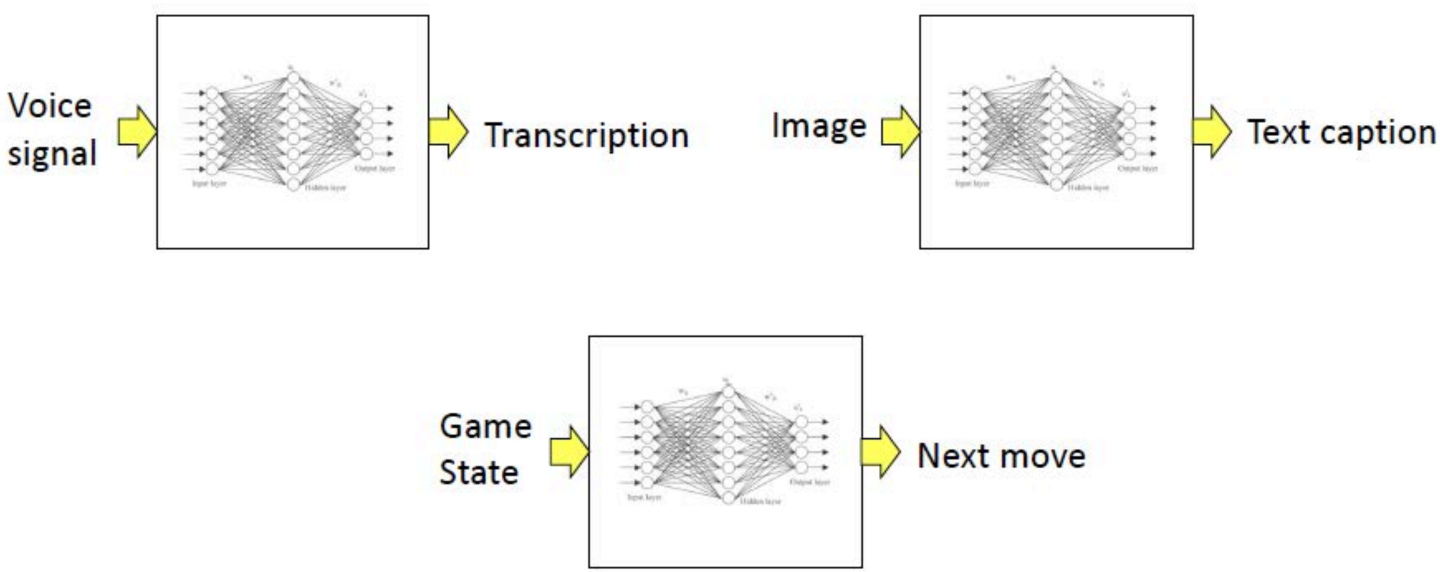
What are deep neural networks?



What are these boxes?

Each of these boxes is actually a function

What are deep neural networks?



What are these boxes?

Each of these boxes is actually a function

It can be approximated by a neural network

Agenda

Artificial Intelligence (AI): A brief intro and recent breakthroughs with Deep learning

AI for Social Media Marketing: Cases studies

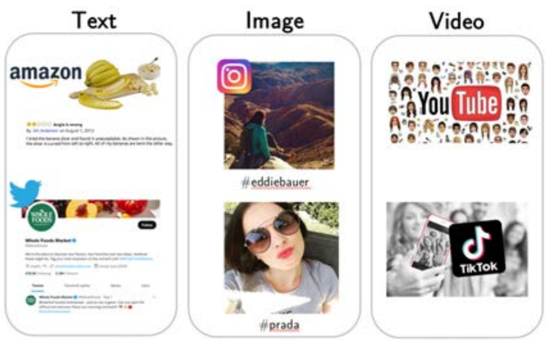
- Visual Listening In: Extracting Brand Image Portrayed on Social Media (Liu et al. 2020)
- Identify Customer Needs from User-Generated Content (Timoshenko and Hauser 2019)
- Supporting Content Marketing with Natural Language Generation (Reisenbichler et al. 2022)

Brainstorming and group work

Sharing group work and conclude

Insights/Problems

Data



Product Design

- Understand customers needs

Branding

- Measure brand perceptions

Advertising and social media

- Generate social media posts and ads

Influencer marketing

- What makes a good influencer video ads?

...

Problem

Identify important problems

- opportunities to use data and analytics to create value for consumers or firms

Data

What is the ideal data to solve this problem?

- e.g., internal data, external data

Training data vs. Production/Application data

Method

What is the right method?

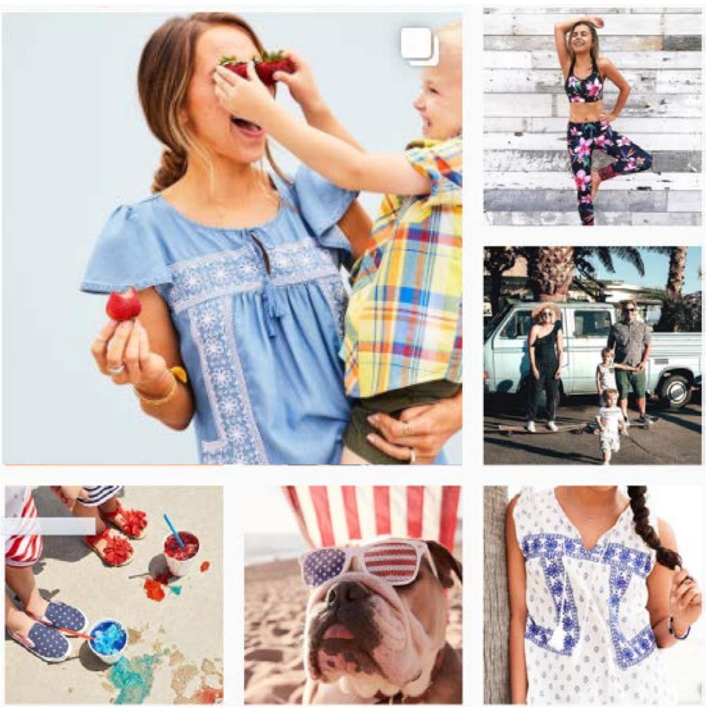
How to evaluate the method and interpret results?

Visual Listening In:
Extracting Brand Image Portrayed on Social Media

Liu Liu, Daria Dzyabura, Natalie Mizik

Brand Manager

@oldnavy



Consumer

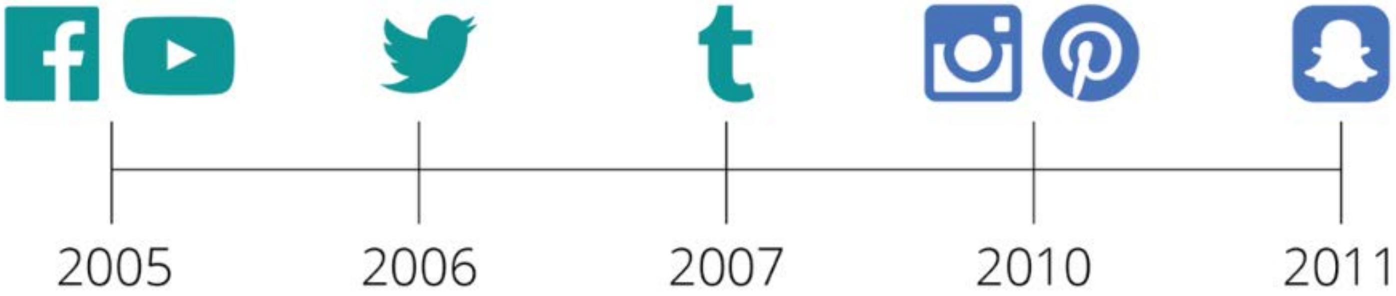
#oldnavy

How is my brand portrayed in **consumer photos**?

Photos Are the New Social Conversation

Image-based social media platforms are on the rise

- E.g., Instagram has 700 million monthly active users
- 95 million photos/videos uploaded daily[1]



1. https://www.instagram.com/press/
Photo credit: Tom McGrath@Crimson Hexagon

Consumers Associate Brands with Contexts

Consumers hashtag brands and depict interactions with brands

- E.g., 53 million posts on Instagram with #nike

Link brands with usage context and experiences

#eddiebauer

#prada

Social Media Marketing

"Listen in" on consumer conversations

- E.g., Archak, Ghose, & Ipeirotis, 2011; Lee & Bradlow, 2011; Netzer et al., 2012; Tirunillai & Tellis, 2012, 2014; Liu, Lee, & Srinivasan, 2019; Timoshenko and Hauser, 2019

→ Text mining

→ Functional attributes of products

This Research

Measure brand image portrayed on consumer photos

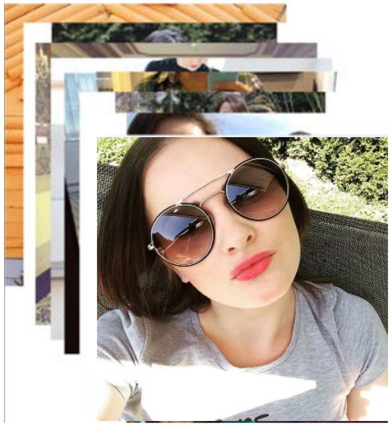
This Research

Measure brand image portrayed on consumer photos

"How are brands portrayed along *intangible brand attributes?*"

$\xrightarrow{?}$ ***rugged***

#eddiebauer



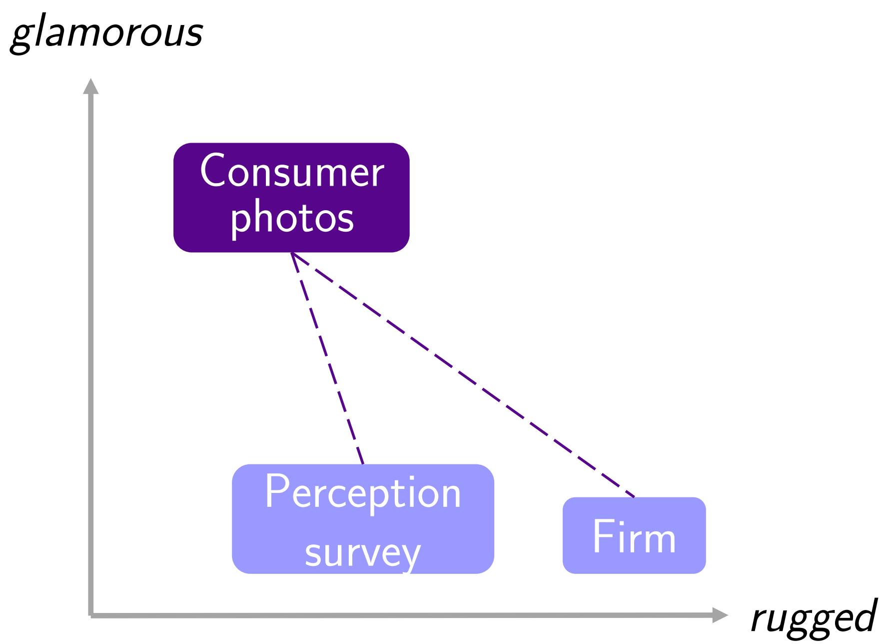
$\xrightarrow{?}$ ***glamorous***

#prada

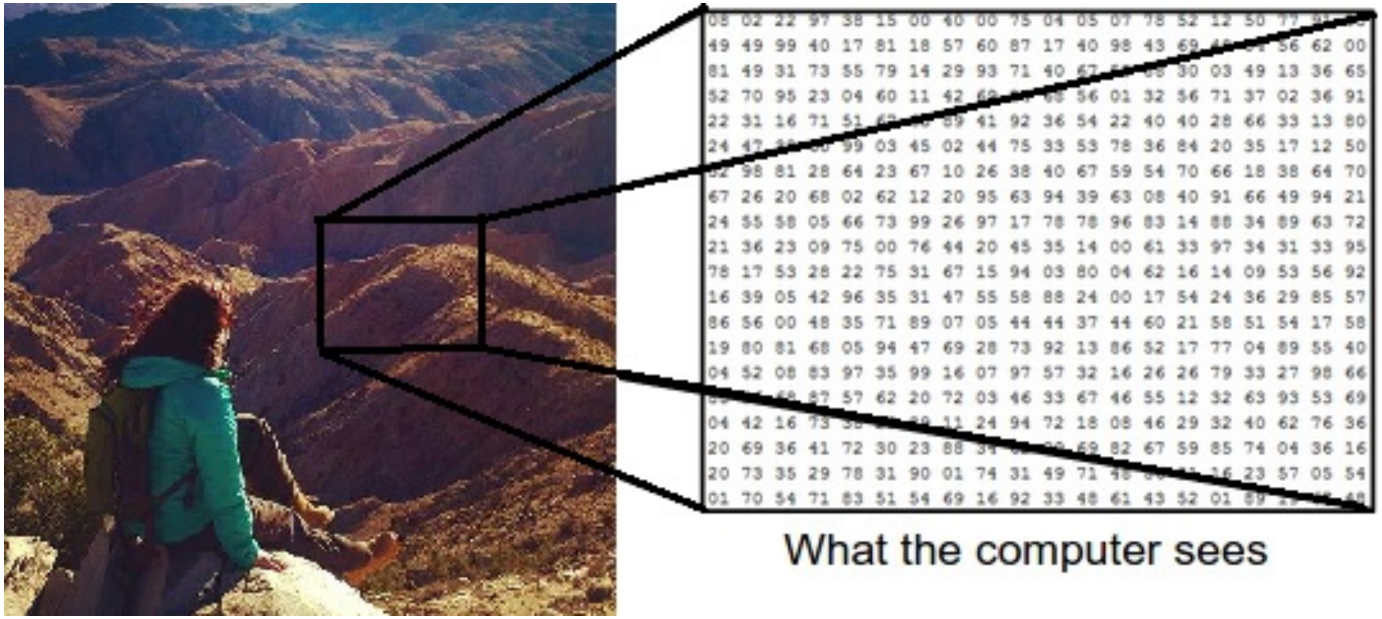
Why Do We Care?

Consumers are co-creating brand image on social media

Allow firms to correct, or leverage, or identify new opportunities for brand positioning and differentiation



How to Measure Brand Attributes from Photos?



Challenging : Data is very unstructured.

Require new methods!

Contributions

Do consumers' photos contain brand image/perception info?

Yes - Create a new photo-based metric about consumers' brand perception

⇩

How to measure?

Measure brand attributes from photos using Deep learning

- Convolutional neural networks for brand attributes

⇩

What insights are generated from social media?

Application to Instagram brand photos

Consumer photos vs. Firm photos vs. Brand Perception survey

- Brand image on social media reflects perception survey
- Identify gaps in positioning strategy

Steps

Measure brand attributes from photos

- Data: Photos labeled with brand attributes
  - *glamorous, rugged, healthy, fun*
- Algorithms: multi-label image classification

Application to Instagram brand photos

- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  1. Product category level consistency
  2. Brand maps
  3. Gaps in positioning strategy

Steps

Measure brand attributes from photos
- Data: Photos labeled with brand attributes
  - *glamorous, rugged, healthy, fun*
- Algorithms: multi-label image classification

$$\boldsymbol{X}, \boldsymbol{Y} \longrightarrow \boldsymbol{F}(\boldsymbol{X})$$

Application to Instagram brand photos
- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  - i. Product category level consistency
  - ii. Brand maps
  - iii. Gaps in positioning strategy

Steps

Measure brand attributes from photos

- Data: Photos labeled with brand attributes
  - *glamorous, rugged, healthy, fun*
- Algorithms: multi-label image classification

$$\boldsymbol{X}, \boldsymbol{Y}$$

$$\downarrow$$

$$\boldsymbol{F}(\boldsymbol{X})$$

$$\downarrow$$

$$\widehat{\boldsymbol{Y}}$$

Application to Instagram brand photos

- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  - i. Product category level consistency
  - ii. Brand maps
  - iii. Gaps in positioning strategy

Steps

Measure brand attributes from photos
- Data: Photos labeled with brand attributes
  - *glamorous, rugged, healthy, fun*
- Algorithms: multi-label image classification

$$\boldsymbol{X}, \boldsymbol{Y} \longrightarrow \boldsymbol{F}(\boldsymbol{X}) \longrightarrow \hat{Y}$$

Application to Instagram brand photos
- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  i. Product category level consistency
  ii. Brand maps
  iii. Gaps in positioning strategy

Multi-label Image Classification on Brand Attributes

?

Does it convey ruggedness?
Does it convey glamour?
Does it convey healthiness?
Does it convey fun?

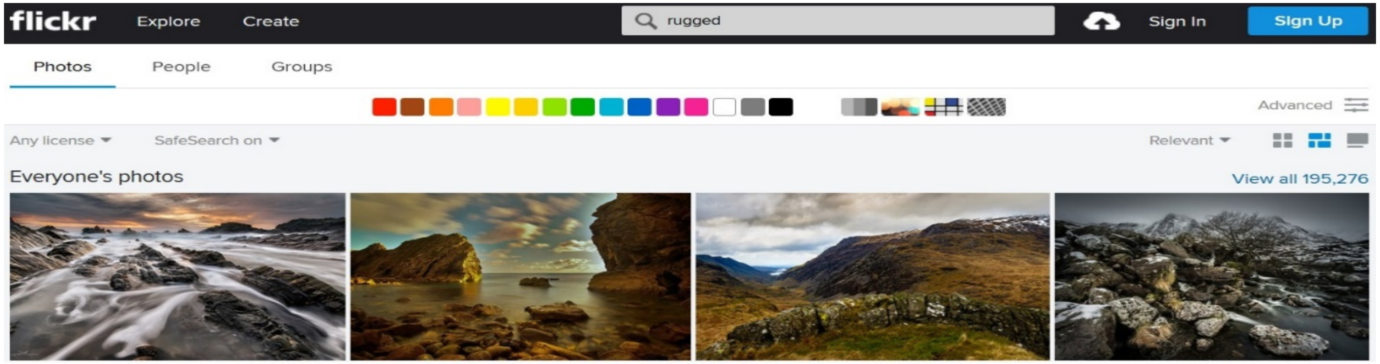
1. Images labeled with brand attributes
2. Algorithms that learn mapping between images and attributes

Collect Images Labeled with Brand Attributes

Need positive and negative instances for each brand attribute

Collect data using Flickr search engine

- Query attributes (e.g., healthy) and antonyms (e.g., unhealthy)



Example Images

glamorous

drab

rugged

gentle

healthy

unhealthy

fun

dull

16,368 images in total

Example Images

glamorous

drab

rugged

gentle

healthy

unhealthy

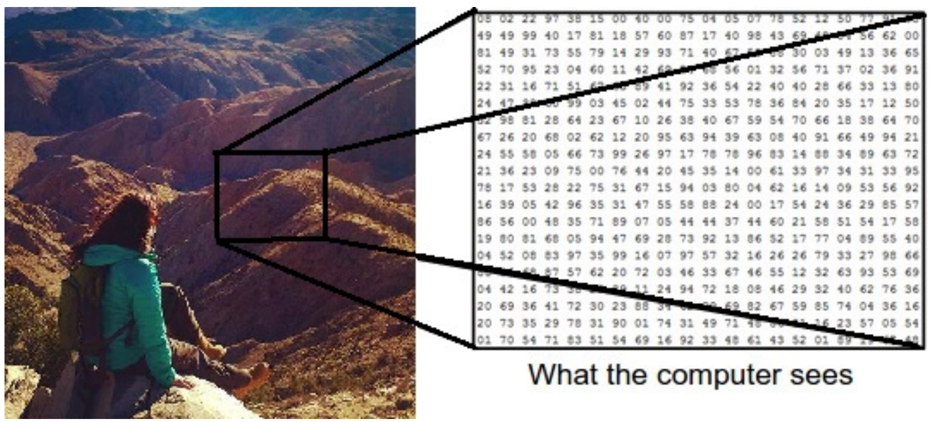
fun

dull

16,368 images in total

Algorithms: Mapping Between Images and Attributes

For example, for "rugged"



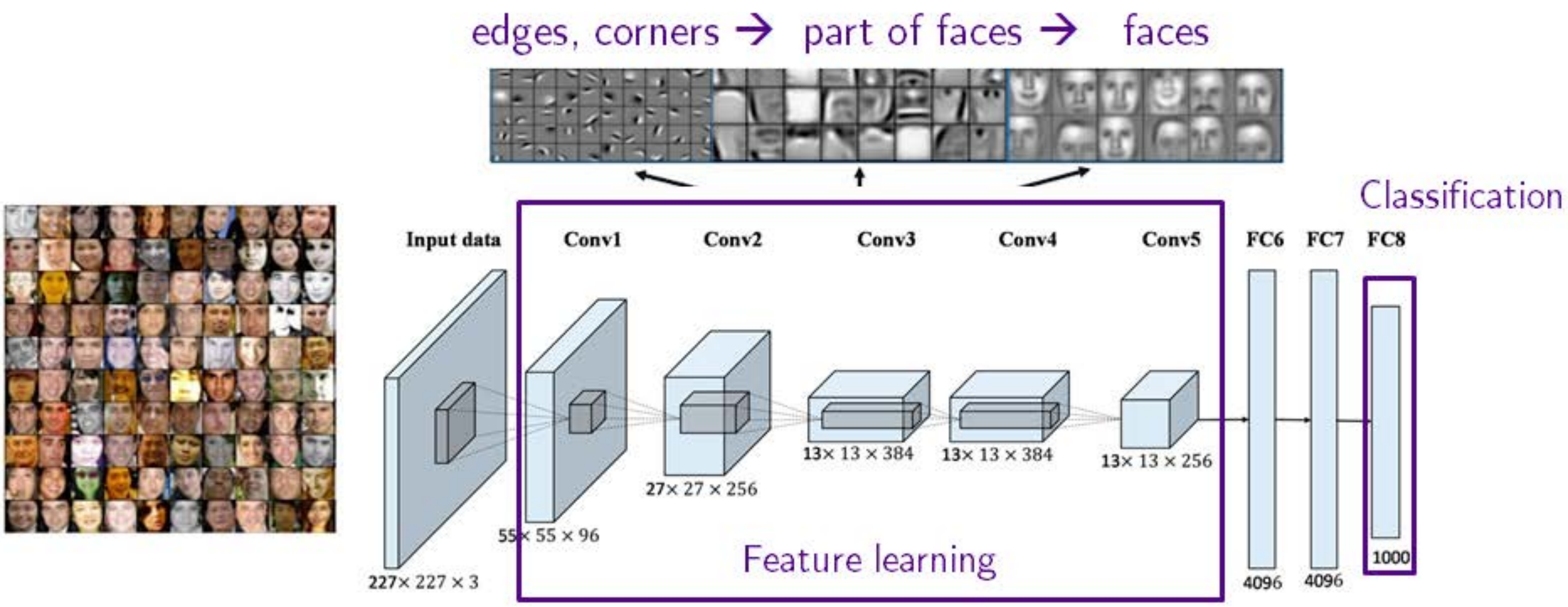
F(x) → rugged

Very unstructured

Deep Learning

Automatic feature extraction

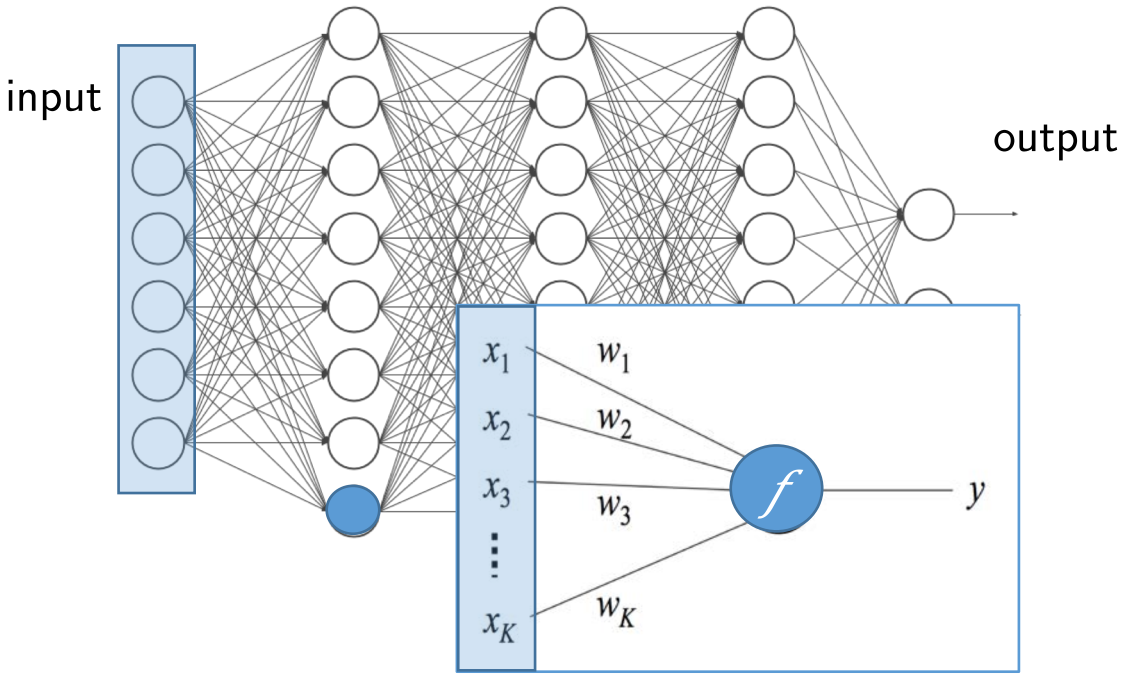
Represent hierarchy of concepts (Bengio et al. 2015)



Deep Learning: How Does It Work?

Model complex non-linear relationship

via many simple non-linear transformations one after another



$$y = f(\sum_{j=1}^{K} w_j x_j)$$

$$F(\mathbf{x}, \mathbf{W}) = f(w^K, f(w^{K-1}, f(...f(w^0, x)...)$$

Estimation: back propagation (gradient descent)

A Multi-label Convolutional Neural Network (ConvNet)

Brand attribute prediction

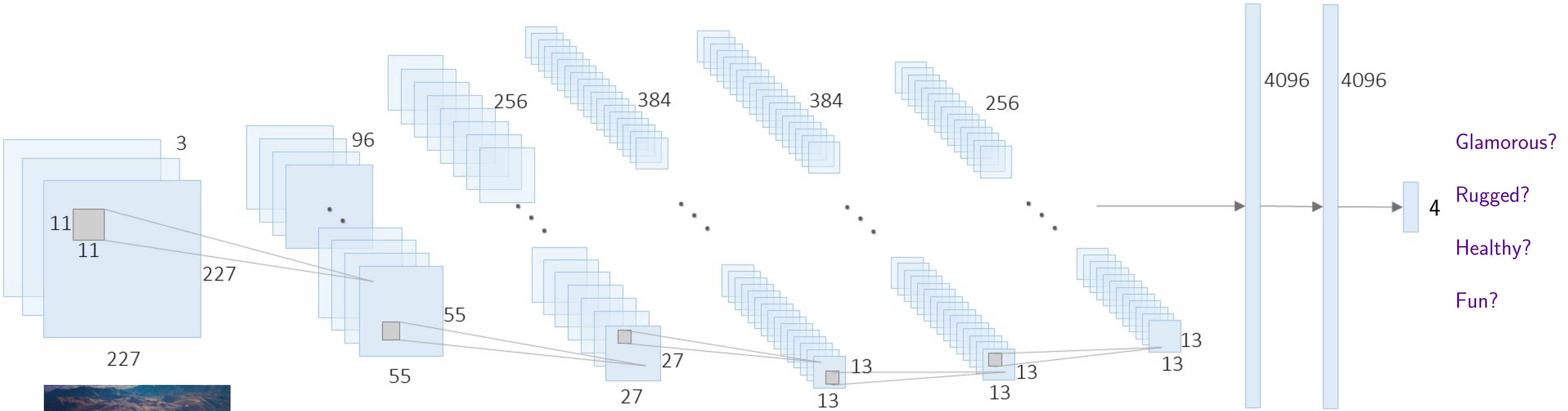


Image in pixels:
227 x 227 x 3

Train ConvNets with Transfer Learning

- Require large amount of data
- Transfer learning
  - Transfer parameters from one domain to another
  - Initialize with parameters from pre-trained models
    - ImageNet model (Krizhevsky et al., 2012): object classification
    - Flickrstyle model (Karayev et al., 2013) : style recognition
  - Fine-tune the model on our training data

Models Trained on GPU

Code and train model using Caffe deep learning framework

5000 iterations in a K80 GPU on a university high performance cluster

80% training data, 10% validation, 10% hold-out sample

- Pick the iteration with the lowest loss in a validation set

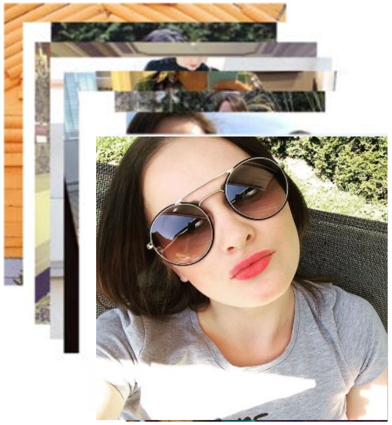
Out-of-Sample Predictive Performance

	Multi-label ConvNet	
	Accuracy	AUC
Glamorous	88.6%	0.846
Rugged	91.3%	0.853
Healthy	89.9%	0.859
Fun	89.4%	0.827
Mean	89.8%	0.846

Note: 80% training, 10% validation, 10% hold-out sample

$\xrightarrow{?}$ ***rugged***

#eddiebauer



$\xrightarrow{?}$ ***glamorous***

#prada

Steps

Measure brand attributes from photos

- Data: Photos labeled with brand attributes
  - *glamorous, rugged, healthy, fun*
- Algorithms

$$Y \downarrow F(x) \downarrow \hat{Y}$$

Application to Instagram brand photos

- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  1. Product category level consistency
  2. Brand maps
  3. Gaps in positioning strategy

Brand Photos on Instagram

56 brands from Apparel and Beverage categories

Consumer: photos on Instagram (#brand)
- About 2,000 per brand
- 114,367 total

Firms: photos on official Instagram accounts
- 72,089 total

Model Performance on Instagram Data Sample

Data Sample (600 images in total, 150 images per attribute)

Group 1	Group 2	Group 3
Attribute is present 50 images	*Ambiguous* 50 images	*Attribute is not present* 50 images

Human judges

- Level-3 US judges on Figure Eight platform[1]
- 20 judges for each attribute and each image

1. Previously Crowdflower

Model vs. Human Judges: Overall Agreement

Table. Aggregate Model Performance According to Human-Based Image Labels

	AUC: Model vs. the Majority Vote of Human Judges	Agreement: Model vs. the Majority Vote of Human Judges	Agreement: A Single Human Judge vs. the Majority Vote of Human Judges, Average
glamorous	0.93	83%	85%
rugged	0.96	85%	83%
healthy	0.91	78%	80%
fun	0.94	84%	80%
Average:	0.94	82.5%	82%

Note: 600 images (150 images for each attribute), 12,000 total judgments (20 judgments for each image and attribute). The model-based label is equal to 1 if the model-based probability estimate for attribute presence is greater than 50%, and zero otherwise. The human-based label for an image is equal to 1 if the majority of the judges indicate attribute presence, and zero otherwise. Agreement is the percentage of images for which the majority of human judges evaluating an image assign this image the same label as our model. The total cost of data collection is $288.96, with an average cost per judgment of $0.024.

Image-Based Brand Image (IBBI) Metric

Compute the average probability of brand *j* images that express the brand attribute *p*:

$$IBBI_{ba} = \frac{\sum_{i=1}^{N^{(b)}} \Pr\left(y_n^{(b)}(a) = 1 \middle| \boldsymbol{X}_n^{(b)}\right)}{N^{(b)}}.$$

For example

$$IBBI_{\text{Prada, rugged}} = 0.07$$

$$IBBI_{\text{Eddie Bauer, rugged}} = 0.17$$

Three Different but Related Brand Metrics

Consumer-created brand photos	- Brand image conveyed from consumer photos - Context, usage, consumption experience
Firm-created brand photos	- Part of firms' marketing effort to create brand identity
Brand perception survey	- Young and Rubicam's Brand Asset Valuator (BAV) - How do consumers perceive brands?

Empirical Studies and Insights

1. Product category level
   - Check consistency between different brand metrics

2. Brand maps
   - Compare brand maps created from different brand metrics

3. Identify Gaps in positioning strategy
   - A case study of underwear brands

Empirical Studies and Insights

1. Product category level
   - Check consistency between different brand metrics

2. Brand maps
   - Compare brand maps created from different brand metrics

3. Identify Gaps in positioning strategy
   - A case study of underwear brands

Consistency Between Different Brand Metrics

Table. Correlation Analyses of Model Predictions for Consumer and Firm-Created Images on Instagram and the BAV Survey-Based Measures of Brand Perceptions

Apparel	Consumer Images vs. Firm Images	Consumer Images vs. BAV	Firm Images vs. BAV
glamorous	0.7838***	0.5519***	0.6100***
rugged	0.9122***	0.5467**	0.5035**
healthy	0.4680**	0.1794	0.3225*
fun	0.6061***	0.3583*	0.2883
Beverages	**Consumer Images vs. Firm Images**	**Consumer Images vs. BAV**	**Firm Images vs. BAV**
glamorous	0.5518**	0.4568**	0.6582***
rugged	0.8259***	0.3596*	0.4708*
healthy	0.7370***	0.6976***	0.4766**
fun	0.3775*	0.1791	0.2584

Note: We see further increases in correlations between consumer and firm-image based data and survey-based measures of brand perceptions when we substitute BAV measures with our survey data from Instagram users. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Survey from Instagram Users

Table. Correlations between Model Predictions for Consumer and Firm-Created Images on Instagram and Survey-Based Measures of Brand Perceptions from Instagram Users

Apparel	BAV vs. Survey	Consumer Images vs. Survey	Firm Images vs. Survey
glamorous	0.9503***	0.5824***	0.6325***
rugged	0.9338***	0.6831***	0.6630***
healthy	0.8600***	0.0842	0.1941
fun	0.6486***	0.5672***	0.4914**
Beverages	**BAV vs. Survey**	**Consumer Images vs. Survey**	**Firm Images vs. Survey**
glamorous	0.9238***	0.5001**	0.5743**
rugged	0.5485***	0.7899***	0.6645***
healthy	0.9482***	0.7127***	0.5350**
fun	0.8714***	0.2130	0.3648*

Note: the shaded cells represent instances of improvement over the correlations between consumer- and firm-image based data and BAV reported in Table 2. Average number of respondents per brand is 62. Cost of the survey data collection is $346.10. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

→ Overall convergence

→ Brand image portrayed on social media, i.e., our IBBI measure, reflect consumers' brand perceptions

→ What I "say" vs. What I "post"

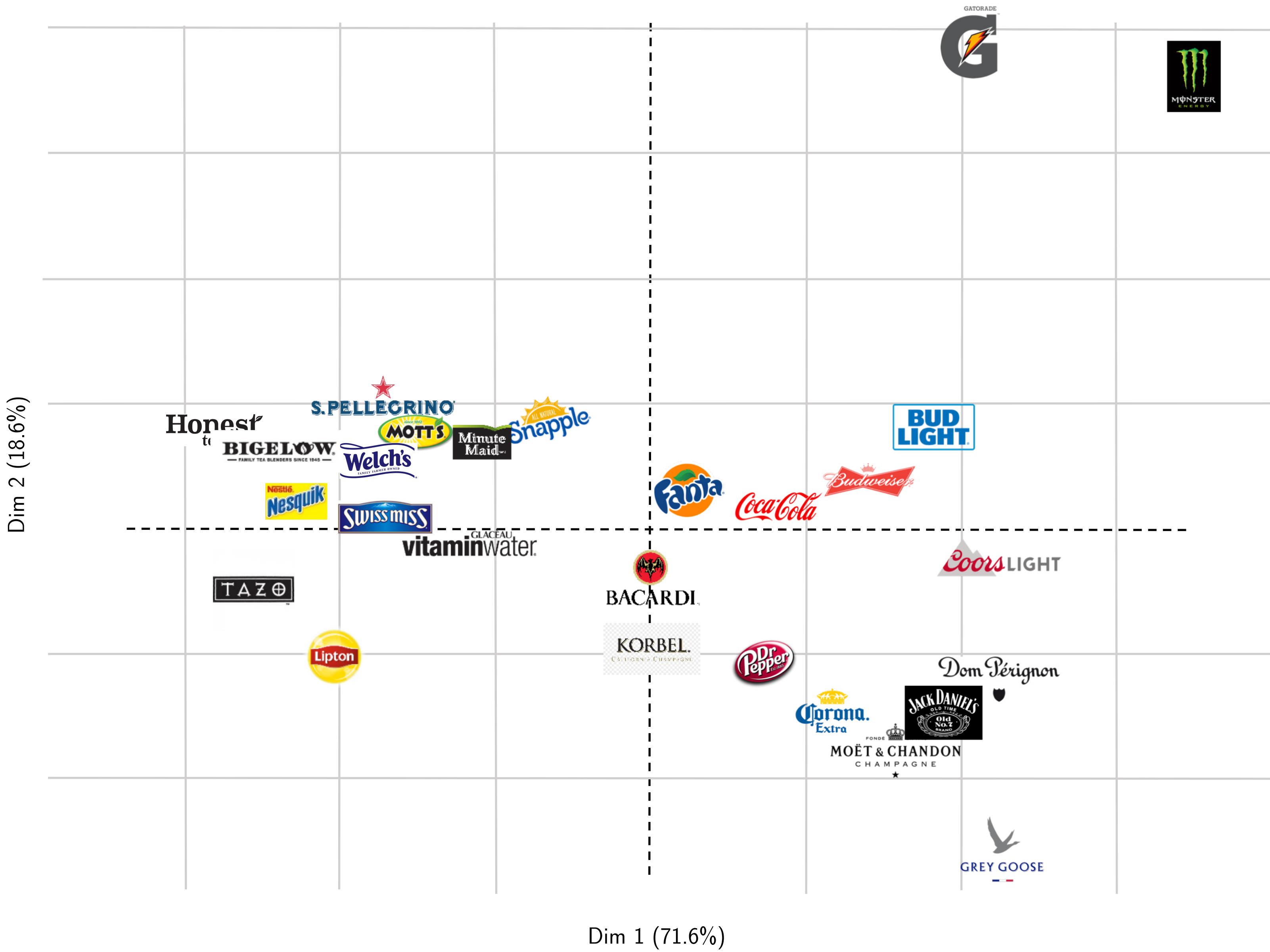
Empirical Studies and Insights

1. Product category level
   - Check consistency between different brand metrics

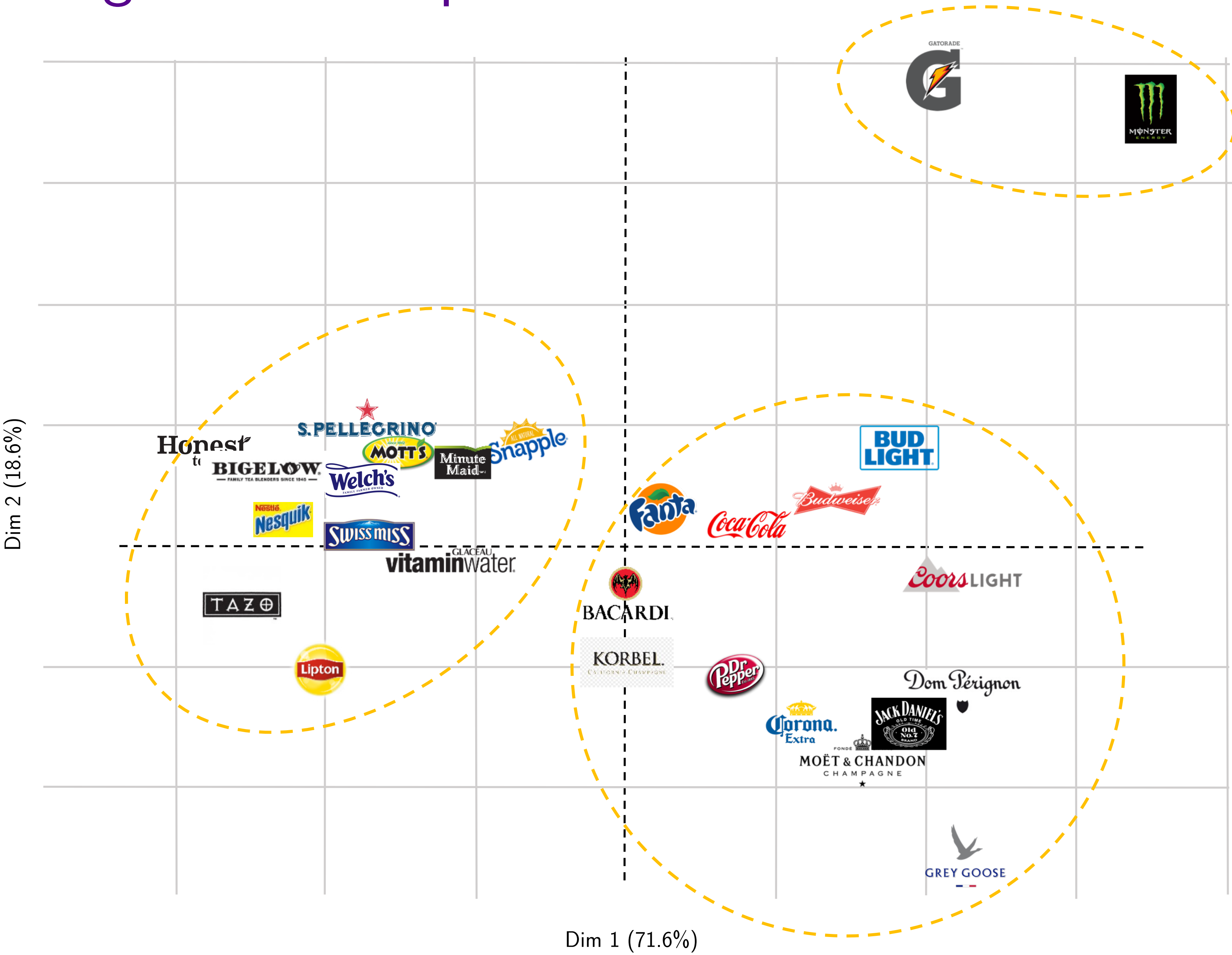
2. Brand maps
   - Compare brand maps created from different brand metrics

3. Identify Gaps in positioning strategy
   - A case study of underwear brands

Beverage Brand Map from Consumer Photos



Beverage Brand Map from Consumer Photos



Empirical Studies and Insights

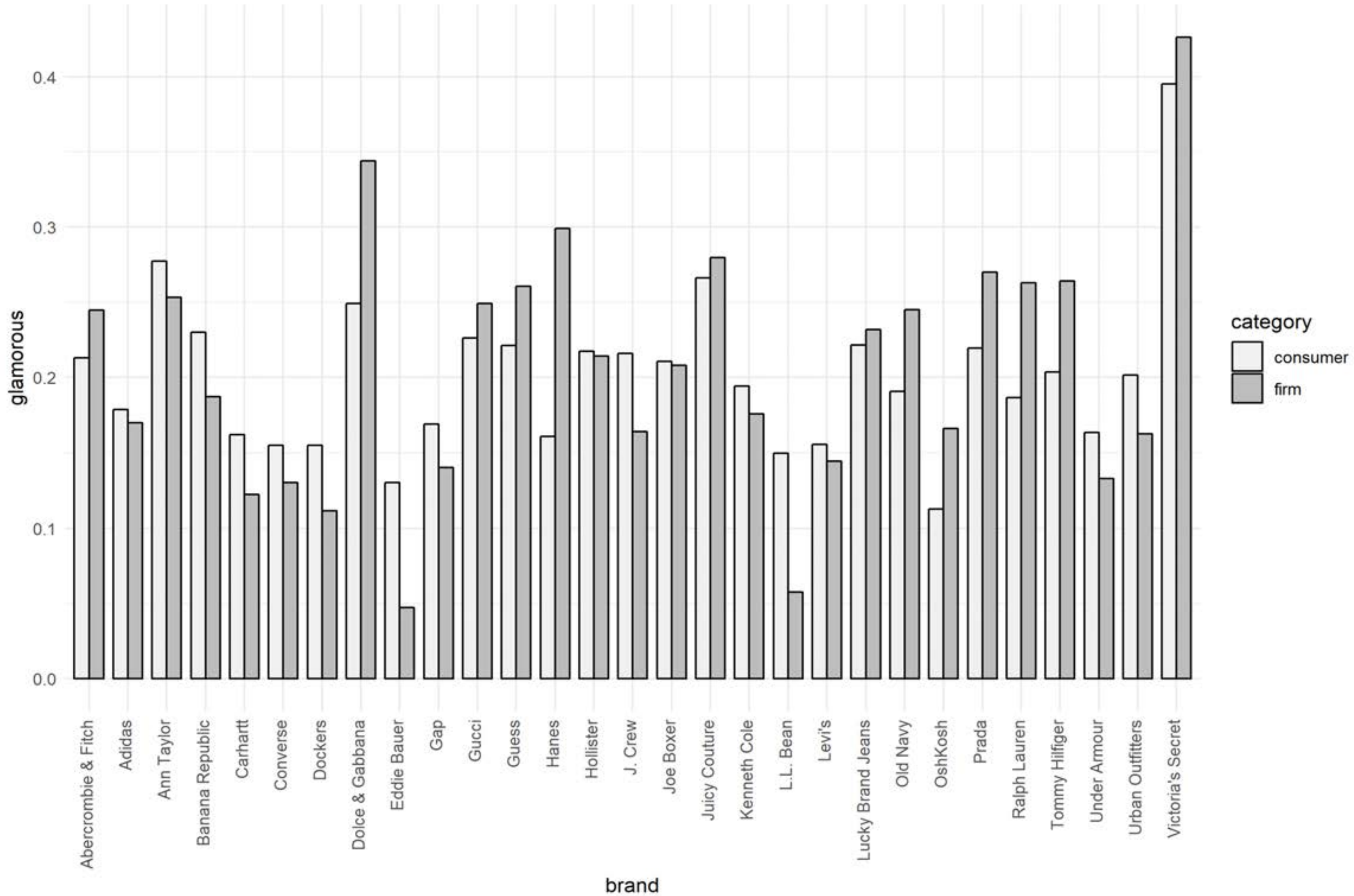
1. Product category level
   - Check consistency between different brand metrics

2. Brand maps
   - Compare brand maps created from different brand metrics

3. Identify Gaps in positioning strategy
   - A case study of underwear brands

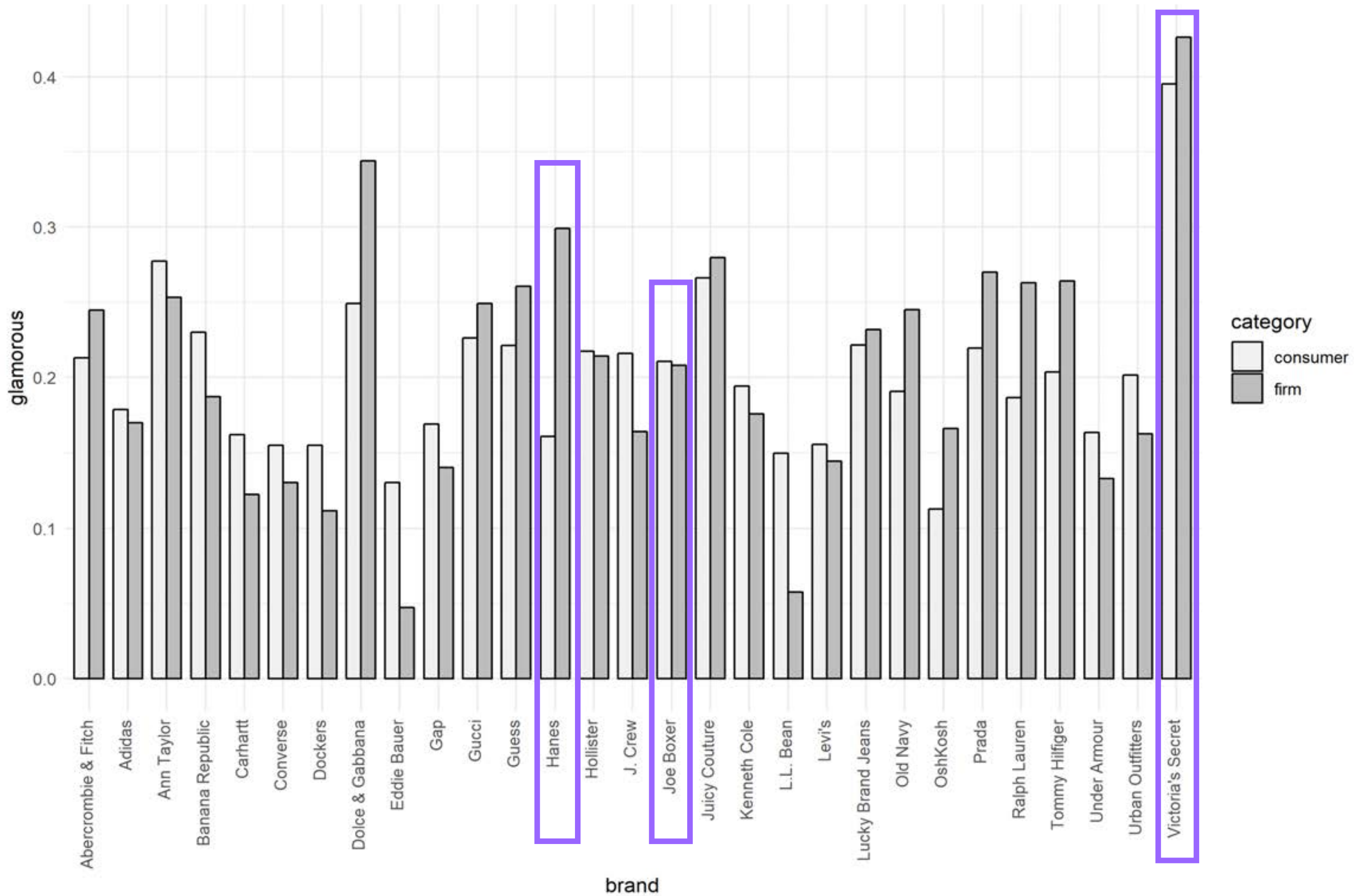
Identify Gaps in Positioning

Consumer vs. Firm: e.g., glamorous in Apparel



Identify Gaps in Positioning

Consumer vs. Firm: e.g., glamorous in Apparel



Validate Differences using Human Judges

Victoria's secret, Joe Boxer, and Hanes

- 500 random consumer images and 500 random firm images
- 10 human judgments per image

(a) **IBBI Scores: BrandImageNet Model and Human Judges**

		BrandImageNet[a]	Human Judges[b]
Victoria's Secret	Consumer images (N=500)	0.39	0.47
	Firm images (N=500)	0.43	0.63
Joe Boxer	Consumer images (N=500)	0.20	0.11
	Firm images (N=134)	0.21	0.21
Hanes	Consumer images (N=500)	0.16	0.12
	Firm images (N=163)	0.30	0.36

Model Predictions and Human Judgements are Consistent

Consistent difference between consumer and firm portrayal

Consistent difference between brands

(b) IBBI Score Differentials between Firm and Consumer Images: Difference (T-stat, p-value)

	BrandImageNet	Human Judges
Victoria's Secret: Firm vs. Consumer images	0.04 (2.25, 0.02)	0.16 (7.92, <.001)
Joe Boxer: Firm vs. Consumer images	0.01 (0.24, 0.81)	0.10 (5.86, <.001)
Hanes: Firm vs. Consumer images	0.14 (7.61, <.001)	0.24 (13.98, <.001)

(c) IBBI Score Differentials between Brands: Difference (T-stat, p-value)

		BrandImageNet	Human Judges
Victoria's Secret vs. Joe Boxer	Consumer images	0.19 (12.14, <.001)	0.39 (22.67, <.001)
	Firm images	0.22 (8.11, <.001)	0.42 (15.75, <.001)
Victoria's Secret vs. Hanes	Consumer images	0.23 (15.01, <.001)	0.35 (21.58, <.001)
	Firm images	0.13 (4.98, 0.03)	0.27 (10.43, <.001)
Joe Boxer vs. Hanes	Consumer images	0.04 (3.71, <.001)	-0.01 (-1.68, 0.09)
	Firm images	-0.09 (-3.26, 0.001)	-0.15 (-5.53, <.001)

Problem

Measure brand perceptions from consumer-created brand photos

- Perceptual map
- Identify gaps between firm and consumers
- Track consumers' brand perceptions in real time

Data

Training data: "labeled" Flicker data

Production/Application data: Instagram data (consumer vs. firm)

Method

Convolutional neural network

Evaluation:

- Out-of-sample prediction performance
- Compare with traditional measures

Identifying Customer Needs from User-Generated Content

Artem Timoshenko, John Hauser

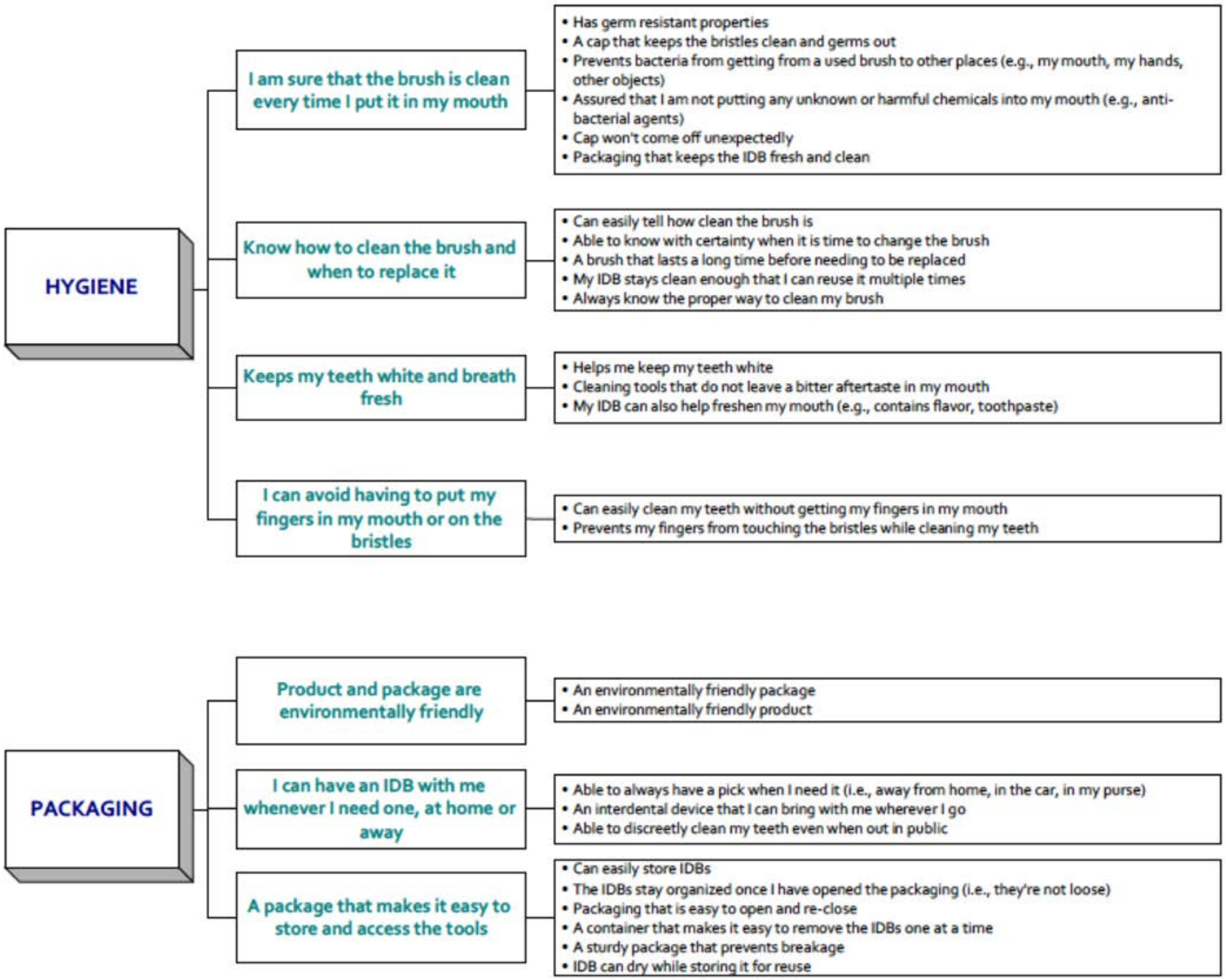
Problem: Identify Customer Needs

Examples of customer needs for **oral care products** from interviews and focus groups:

- Able to make my teeth look whiter
- Oral care items I carry around are easy to keep clean
- Oral care items that match my bathroom décor

Slides credit: Artem Timoshenko

Customer Needs for Oral Care



Slides credit: Artem Timoshenko

Illustrative Example

Amazon Review:

"I replaced an old brush with a new one BUT the description doesn't say that this model no longer has a 30 second timer. The brush shuts off after 2 minutes but the 30 second timer is missing. I would not have purchased this product if I had known".

Slides credit: Artem Timoshenko

Illustrative Example

Amazon Review:

"I replaced an old brush with a new one BUT the description doesn't say that this model no longer has a 30 second timer. The brush shuts off after 2 minutes but the 30 second timer is missing. I would not have purchased this product if I had known".

I know the right amount of time to spend
on each step of my oral care routine

Slides credit: Artem Timoshenko

Illustrative Example

Amazon Review:

"I replaced an old brush with a new one BUT the description doesn't say that this model no longer has a 30 second timer. The brush shuts off after 2 minutes but the 30 second timer is missing. ~~I would not have purchased this product if I had known".~~

I know the right amount of time to spend
on each step of my oral care routine

Slides credit: Artem Timoshenko

Customer Needs from UGC

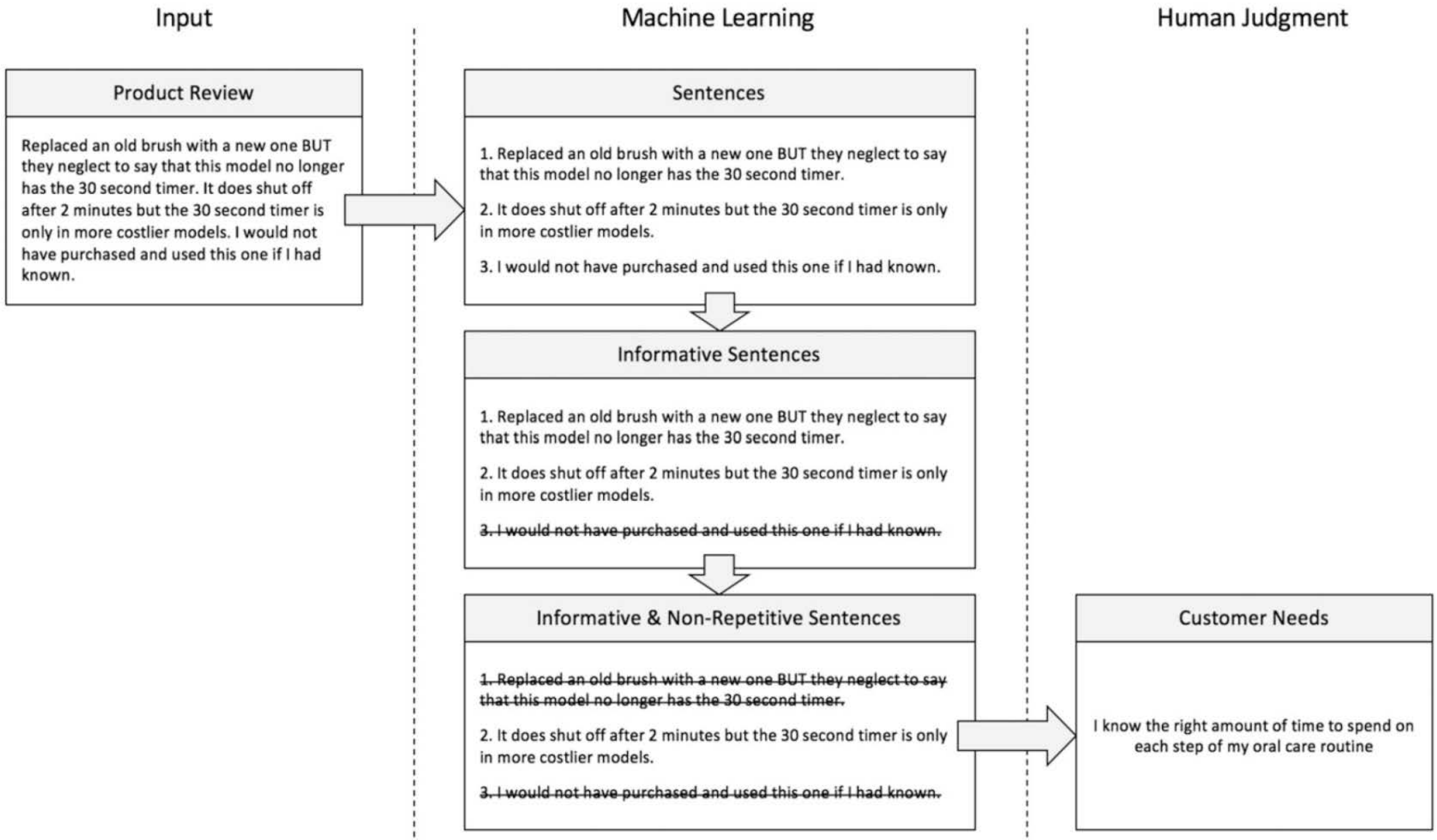
Amazon Review:

~~"I replaced an old brush with a new one BUT the description doesn't say that this model no longer has a 30 second timer.~~ The brush shuts off after 2 minutes but the 30 second timer is missing. ~~I would not have purchased this product if I had known".~~

I know the right amount of time to spend
on each step of my oral care routine

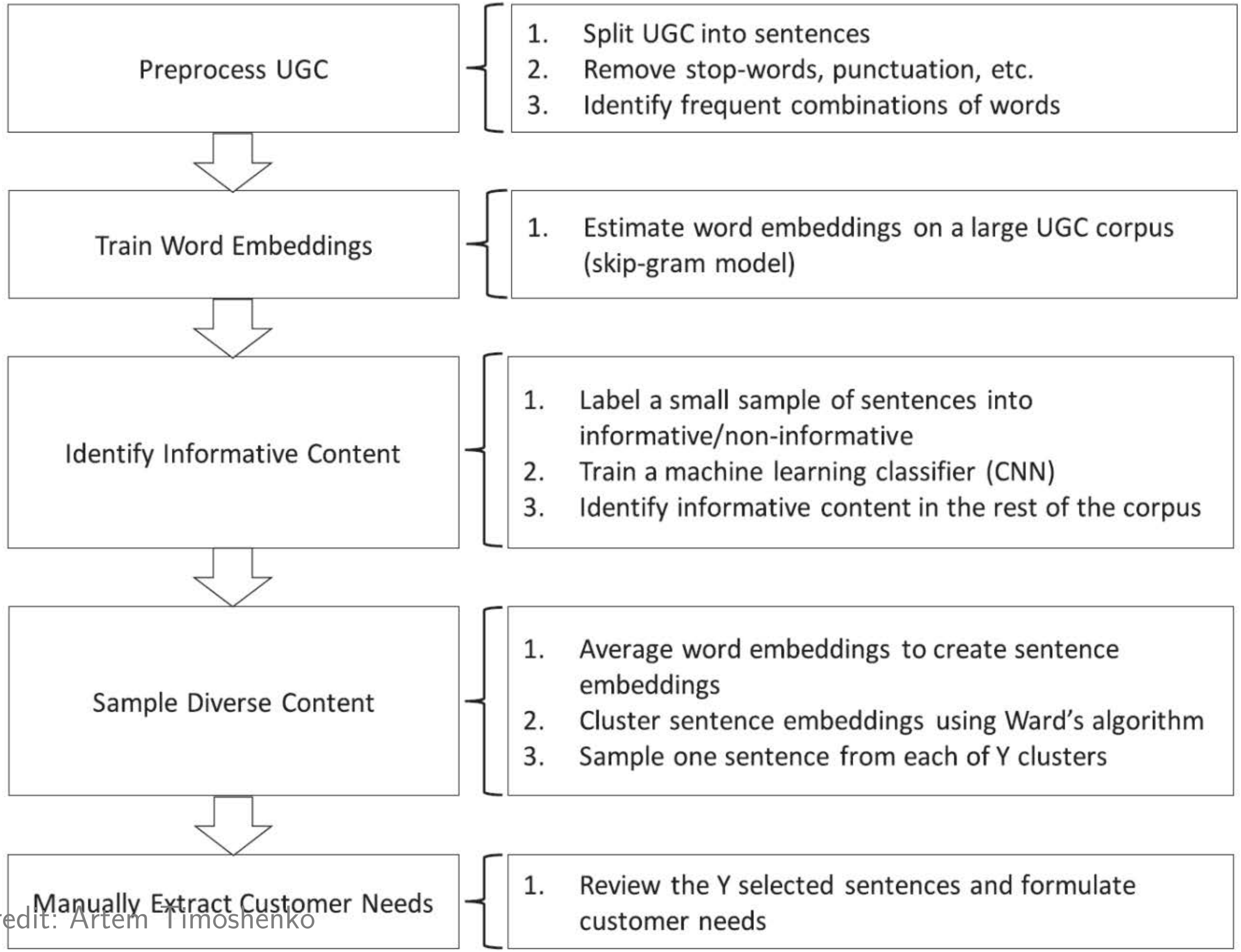
Slides credit: Artem Timoshenko

Figure A.1. Demonstration of the Application of the Proposed Machine-Learning Hybrid Approach to an Amazon Review



Slides credit: Artem Timoshenko

Figure 1. System Architecture for Identifying Customer Needs from UGC



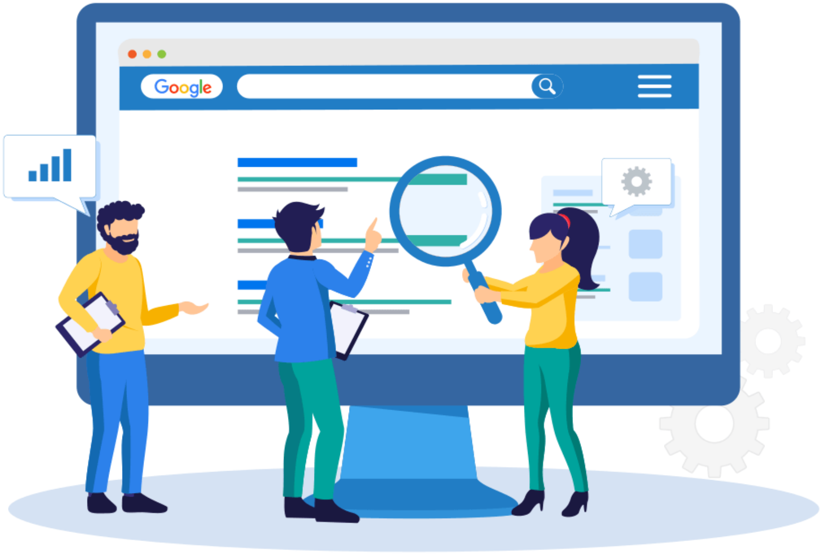
Slides credit: Artem Timoshenko

Support Content Marketing with Natural Language Generation

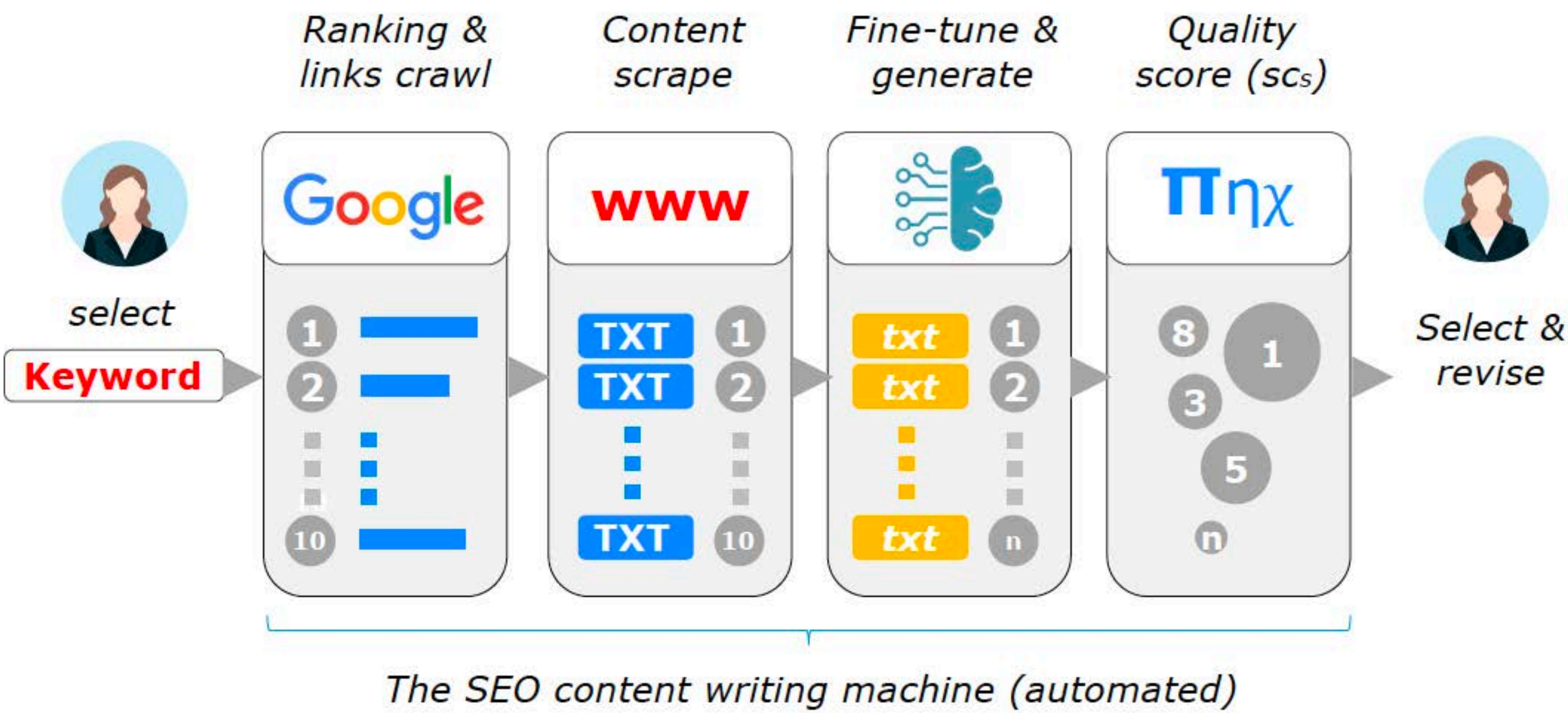
Martin Reisenbichler, Thomas Reutterer, David A. Schweidel, and Daniel Dan

Problem: How to write search engine (SE) optimized content?

1. Identifying current top performing content
2. Machine learning to capture linguistic patterns (e.g., keyword density, readability)
3. Creation of unique content that mirrors the linguistic patterns of top-performing content
4. Human editing of content before publication



Slide credit: David A. Schweidel



Slide credit: David A. Schweidel

Agenda

Artificial Intelligence (AI): A brief intro and recent breakthroughs with Deep learning

AI for Social Media Marketing: Cases studies

- Visual Listening In: Extracting Brand Image Portrayed on Social Media (Liu et al. 2020)
- Identify Customer Needs from User-Generated Content (Timoshenko and Hauser 2019)
- Supporting Content Marketing with Natural Language Generation (Reisenbichler et al. 2022)

Brainstorming and group work

Sharing group work and conclude

Problem

Identify one important problem

- opportunities to use data and analytics to create value for consumers or firms

Data

What is the ideal data to solve this problem?

- e.g., internal data, external data

Training data vs. Production/Application data

Method

What is the right method?

How to evaluate the method and interpret results?

Agenda

Artificial Intelligence (AI): A brief intro and recent breakthroughs with Deep learning

AI for Social Media Marketing: Cases studies

- Visual Listening In: Extracting Brand Image Portrayed on Social Media (Liu et al. 2020)
- Identify Customer Needs from User-Generated Content (Timoshenko and Hauser 2019)
- Supporting Content Marketing with Natural Language Generation (Reisenbichler et al. 2022)

Brainstorming and group work

Sharing group work and reflections (e.g., pitfalls)

Causation (vs. Correlation)

Shifting from Model-Centric to Data-Centric AI

Conventional model-centric approach:

AI System = Code + Data

(model/algorithm)

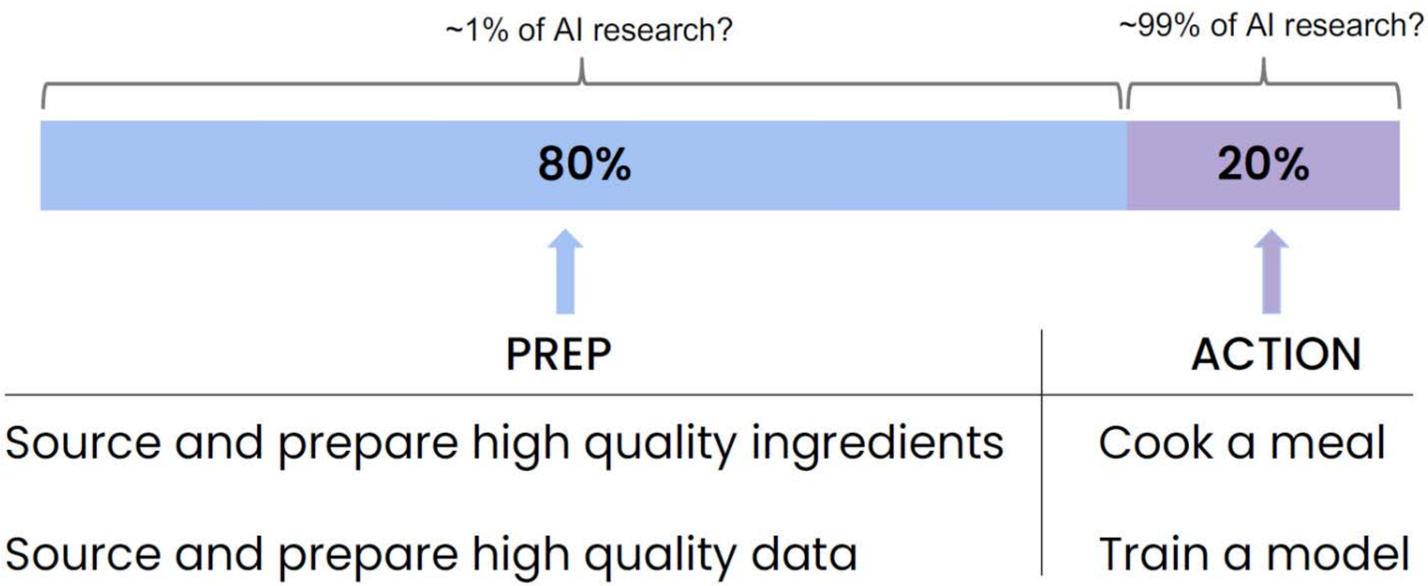
Data-centric approach:

AI System = Code + Data

Slide credit: Andrew Ng, “MLOps: From Model-centric to Data-centric AI”

Data Quality is Important

Data is food to AI



Slide credit: Andrew Ng, “MLOps: From Model-centric to Data-centric AI”

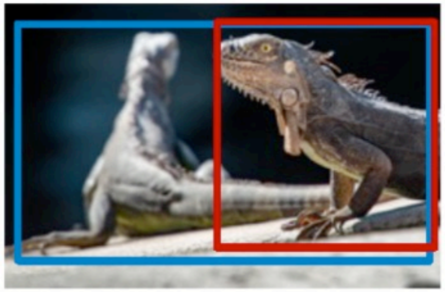
Iguana Detection Example

Labeling instruction:

Use bounding boxes to indicate the position of iguanas

Labeler 1

Labeler 2



Labeler 3

Slide credit: Andrew Ng, “MLOps: From Model-centric to Data-centric AI”

From Big Data to Good Data

Defined consistently (definition of labels y is unambiguous)

Cover of important cases (good coverage of inputs x)

Has timely feedback from production data (distribution covers data drift and concept drift)

Good governance (reasonably free from bias; satisfies privacy; data provenance/lineage, regulatory requirements)

Slide credit: Andrew Ng, "MLOps: From Model-centric to Data-centric AI"

Data-centric AI

Model-centric AI

How can you change the model (code) to improve performance?

Data-centric AI

How can you systematically change your data (inputs x or labels y) to improve performance?

Slide credit: Andrew Ng, "MLOps: From Model-centric to Data-centric AI"

Data-centric AI

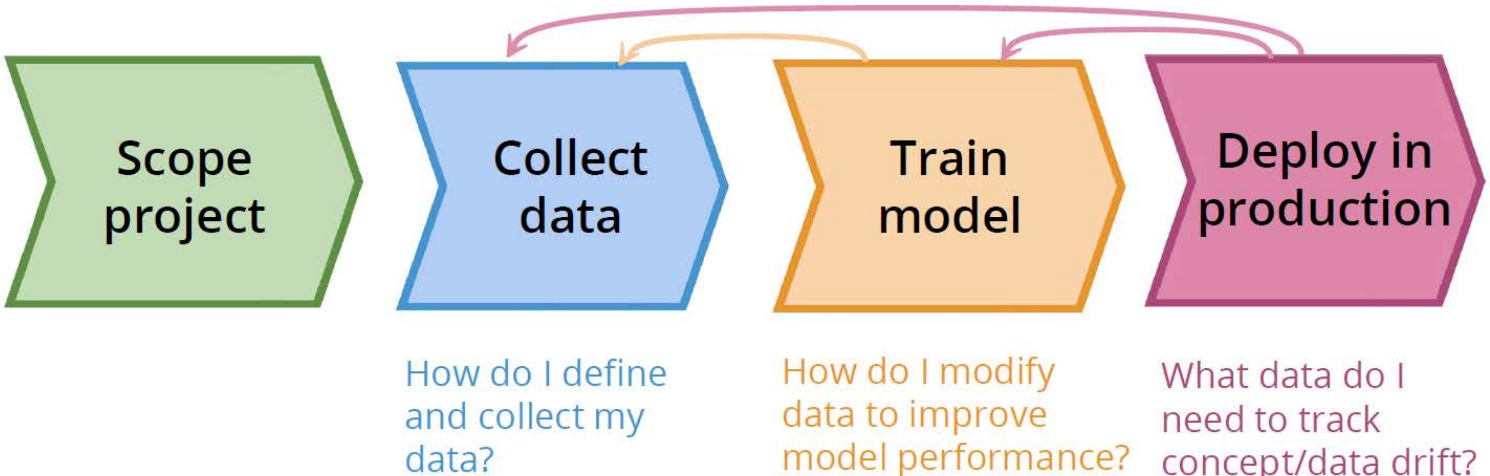
Model-centric AI

How can you change the model (code) to improve performance?

Data-centric AI

How can you systematically change your data (inputs x or labels y) to improve performance?

Make high quality data available through all stages of the ML project lifecycle



Slide credit: Andrew Ng, "MLOps: From Model-centric to Data-centric AI"

Tips for Data-centric AI

1. Make the labels y consistent
2. Use multiple labelers to spot inconsistencies
3. Clarify labeling instructions by tracking down ambiguous examples
4. Toss out noisy examples. More data is not always better
5. Use error analysis to focus on subset of data to improve

Slide credit: Andrew Ng, "MLOps: From Model-centric to Data-centric AI"

Thank you!

Liu Liu (liu.liu-1@colorado.edu)

https://www.colorado.edu/faculty/liu-liu/

References

Liu, Liu, Daria Dzyabura, and Natalie Mizik. "Visual listening in: Extracting brand image portrayed on social media." *Marketing Science* 39, no. 4 (2020): 669-686.

Timoshenko, Artem, and John R. Hauser. "Identifying customer needs from user-generated content." *Marketing Science* 38, no. 1 (2019): 1-20.

Reisenbichler, Martin, Thomas Reutterer, David A. Schweidel, and Daniel Dan. "Frontiers: Supporting Content Marketing with Natural Language Generation." *Marketing Science* 41, no. 3 (2022): 441-452.