SINGULARITY
TECH DAY

#STechDay2020

Fighting Phishing with Deep Learning and Image Recognition

SINGULARITY
TECH DAY

#STechDay2020

ORGANIZATION

plain concepts

SPONSORS

Microsoft

SUPPORT

World Central Kitchen

THANK YOU!

SINGULARITY
TECH DAY

#STechDay2020

Ignasi Paredes

Lead Data Scientist

ignasi.paredes@es.nestle.com

Francisco José Pérez

Machine Learning Engineer

fplopez@plainconcepts.com

SINGULARITY
TECH DAY

#STechDay2020

Agenda

- Context
- Solution Overview
- Architecture
- Initial attempts
- Final Solution
- Challenges ahead

SINGULARITY
TECH DAY

#STechDay2020

Agenda

- Context
- Solution Overview
- Architecture
- Initial attempts
- Final Solution
- Challenges ahead

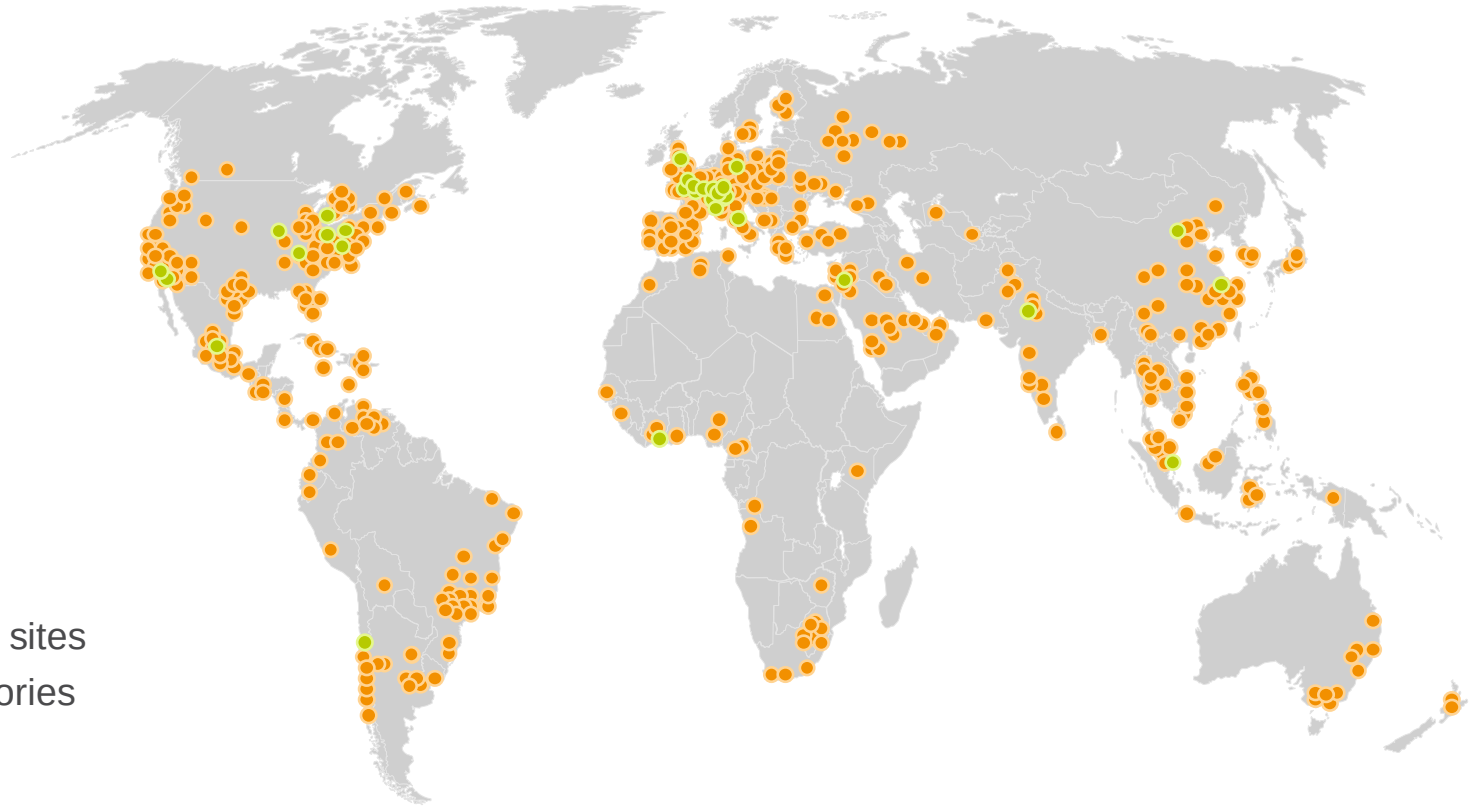
Context: Nestlé

- CHF 91.4 billion in sales in 2018.
- 328.000 employees in over 190 countries.
- 413 factories in 85 countries.
- Over 2000 brands.
- 1 billion Nestlé products sold every day.

SINGULARITY
TECH DAY

#STechDay2020

Context: Nestlé

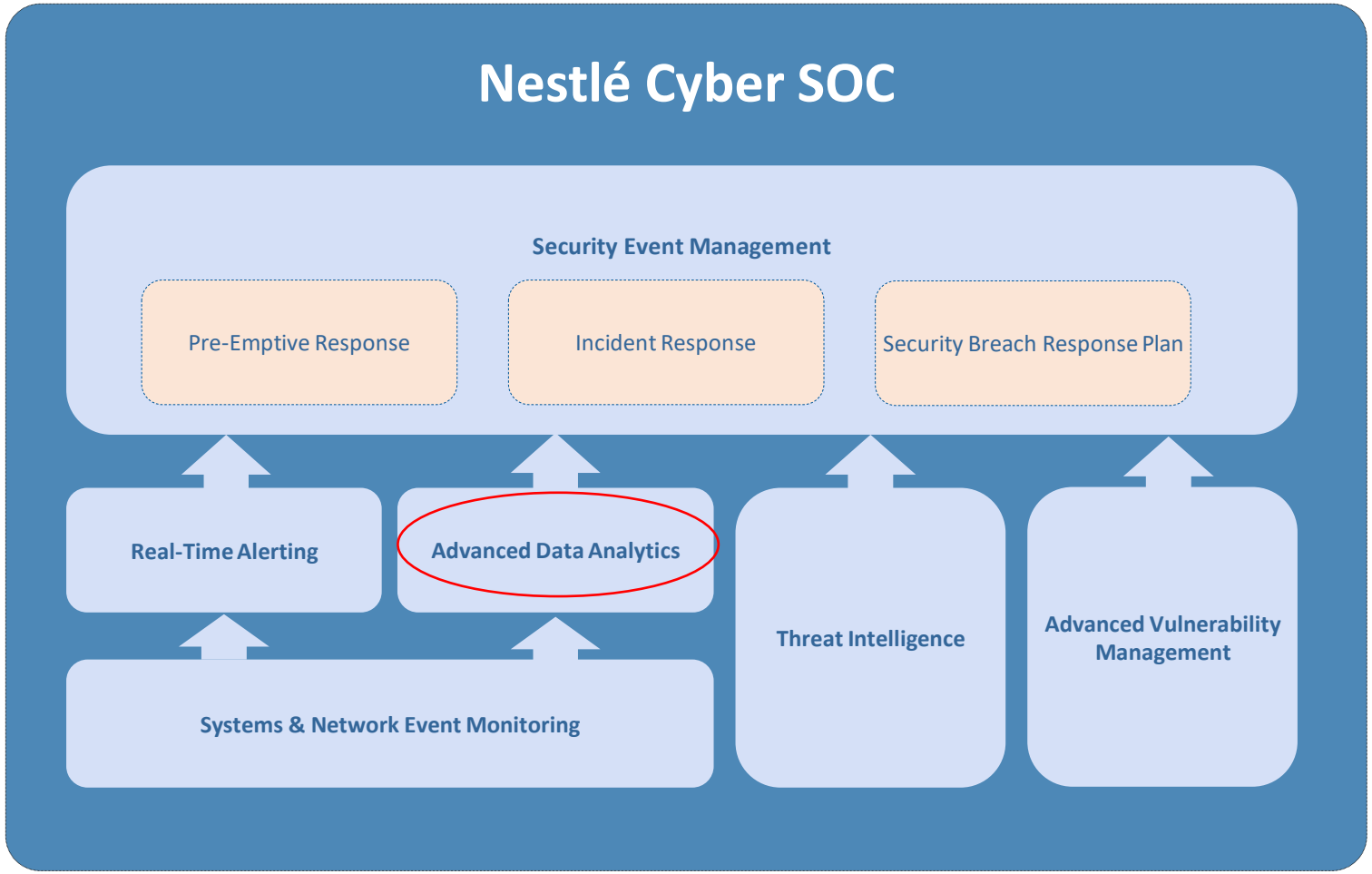


Context: Nestlé

SINGULARITY
TECH DAY

#STechDay2020

Context: The Cyber Security Operations Center



Context: Cybersecurity – should I care?

130
Average number of security breaches in 2017

145
Average number of security breaches in 2018

+11%
Increase in the last year

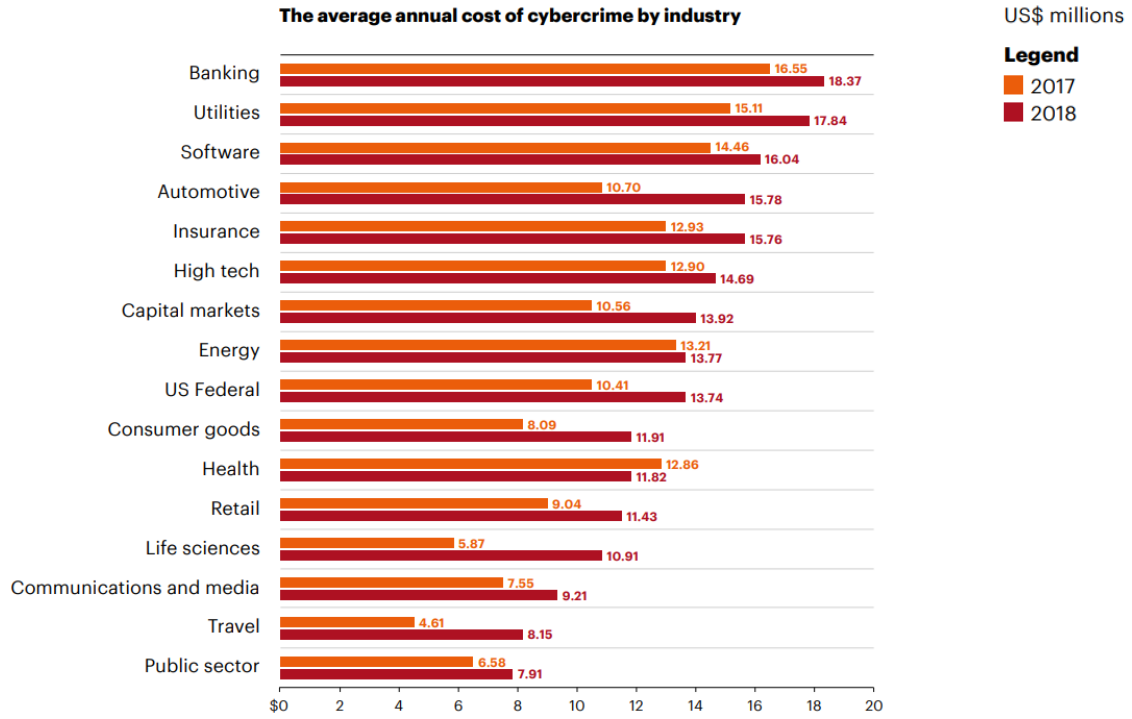
=67%
Increase in the last 5 years

$11.7m
Average cost of cybercrime in 2017

$13.0m
Average cost of cybercrime in 2018

+12%
Increase in the last year

=72%
Increase in the last 5 years



https://www.accenture.com/_acnmedia/pdf-96/accenture-2019-cost-of-cybercrime-study-final.pdf

SINGULARITY
TECH DAY

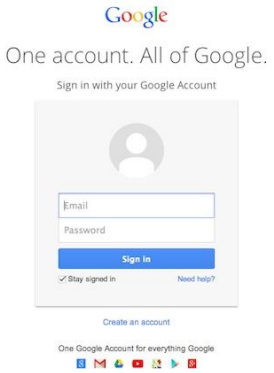
#STechDay2020

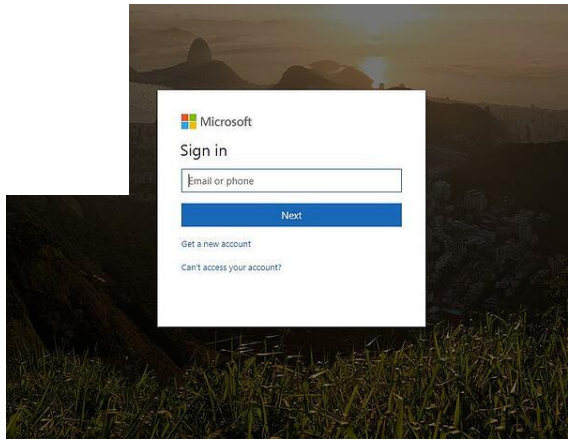
Agenda

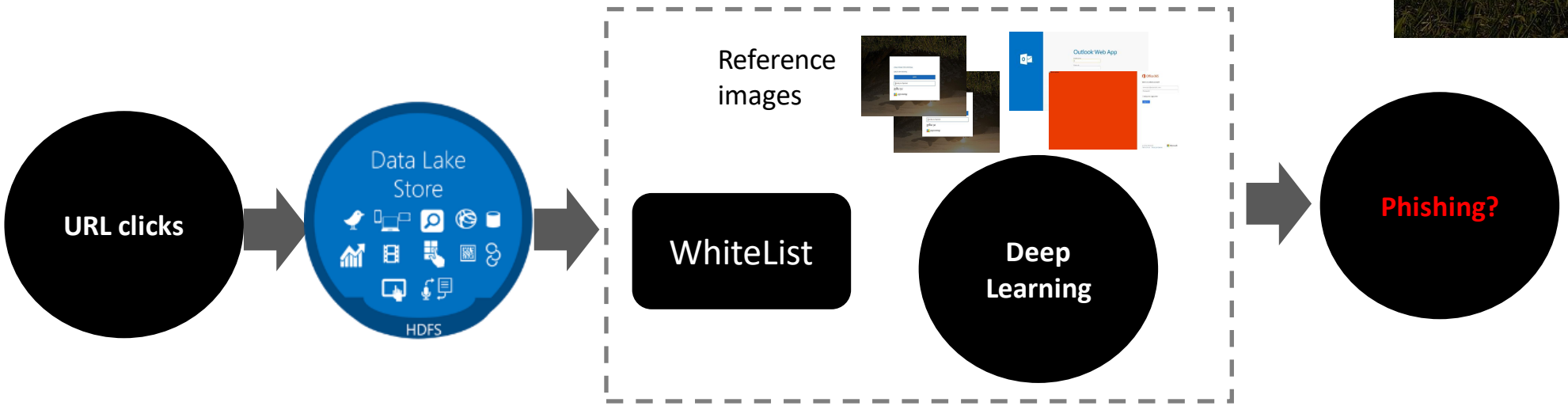
- Context
- Solution Overview
- Architecture
- Initial attempts
- Final Solution
- Challenges ahead

Solution Overview

- Monitor dangerous urls from emails
- Flag login pages trying to impersonate legitimate sites (phishing)
- Identify landing pages that are virtually identical to well-known services (Office365, G Suite)







SINGULARITY
TECH DAY

#STechDay2020

Agenda

- Context
- Solution Overview
- Architecture
- Initial attempts
- Final Solution
- Challenges ahead

Architecture

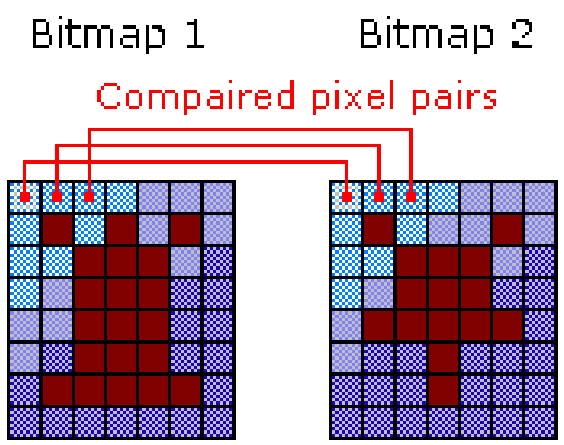
SINGULARITY
TECH DAY

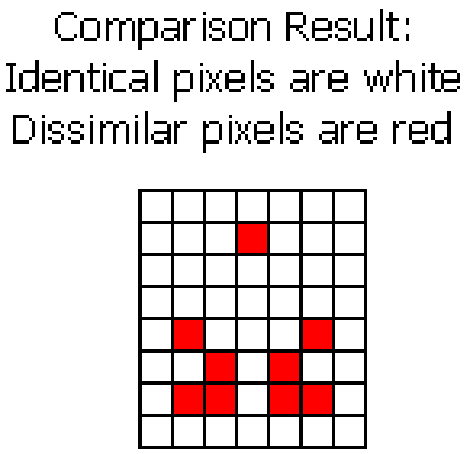
#STechDay2020

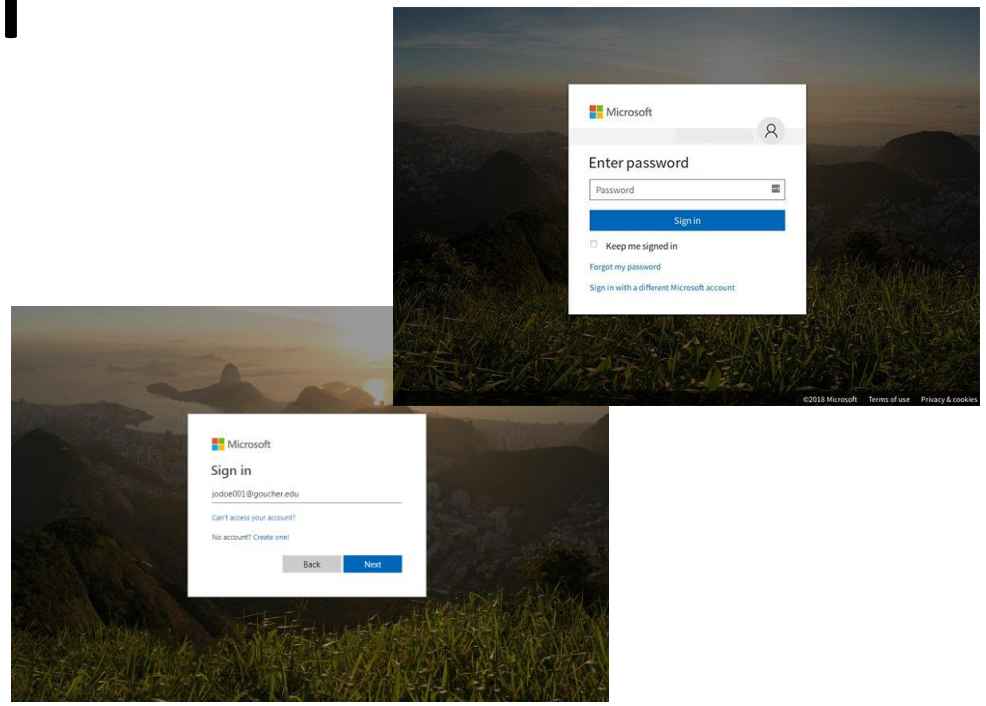
Agenda

- Context
- Solution Overview
- Architecture
- Initial attempts
- Final Solution
- Challenges ahead

Attempt #1: naive, pixel to pixel







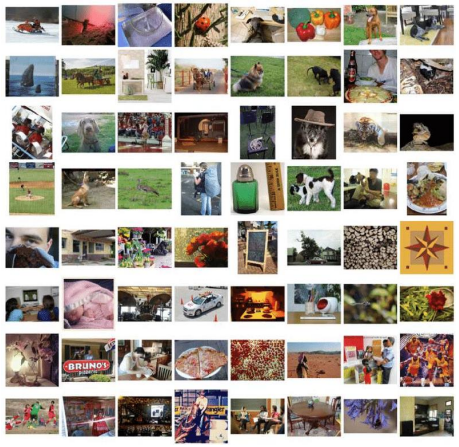
Multiple metrics:

- **MAE, MSE, PAE, AE…**

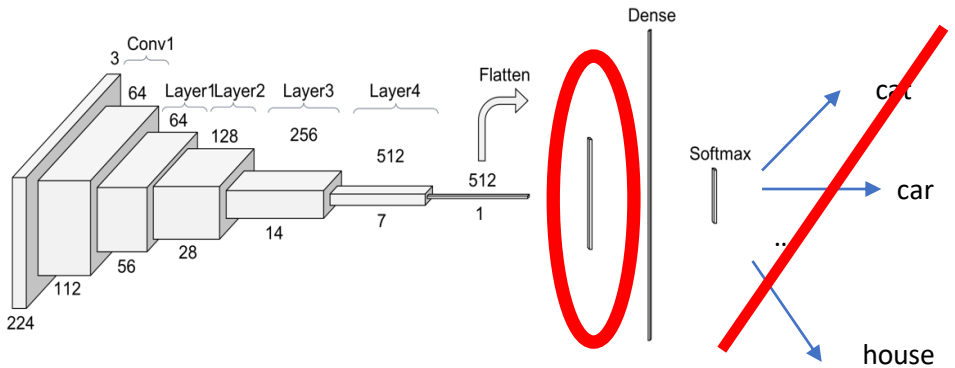
Too strict for small differences!

Attempt #2: pre-trained Neural Network

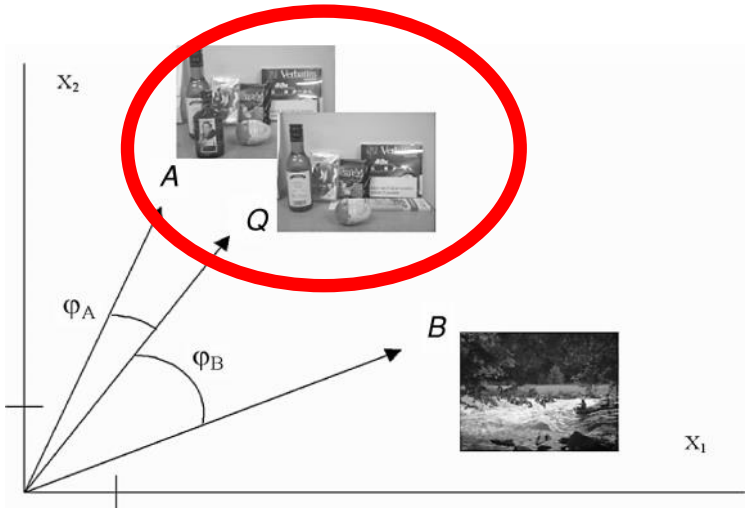
1. Transform images to a lower-level representation (Resnet-101)
   - NN last convolution layer flattened (1D)
2. Compare that last layer (image embedding) with the equivalent representation for reference images (cosine similarity)



ImageNet sample

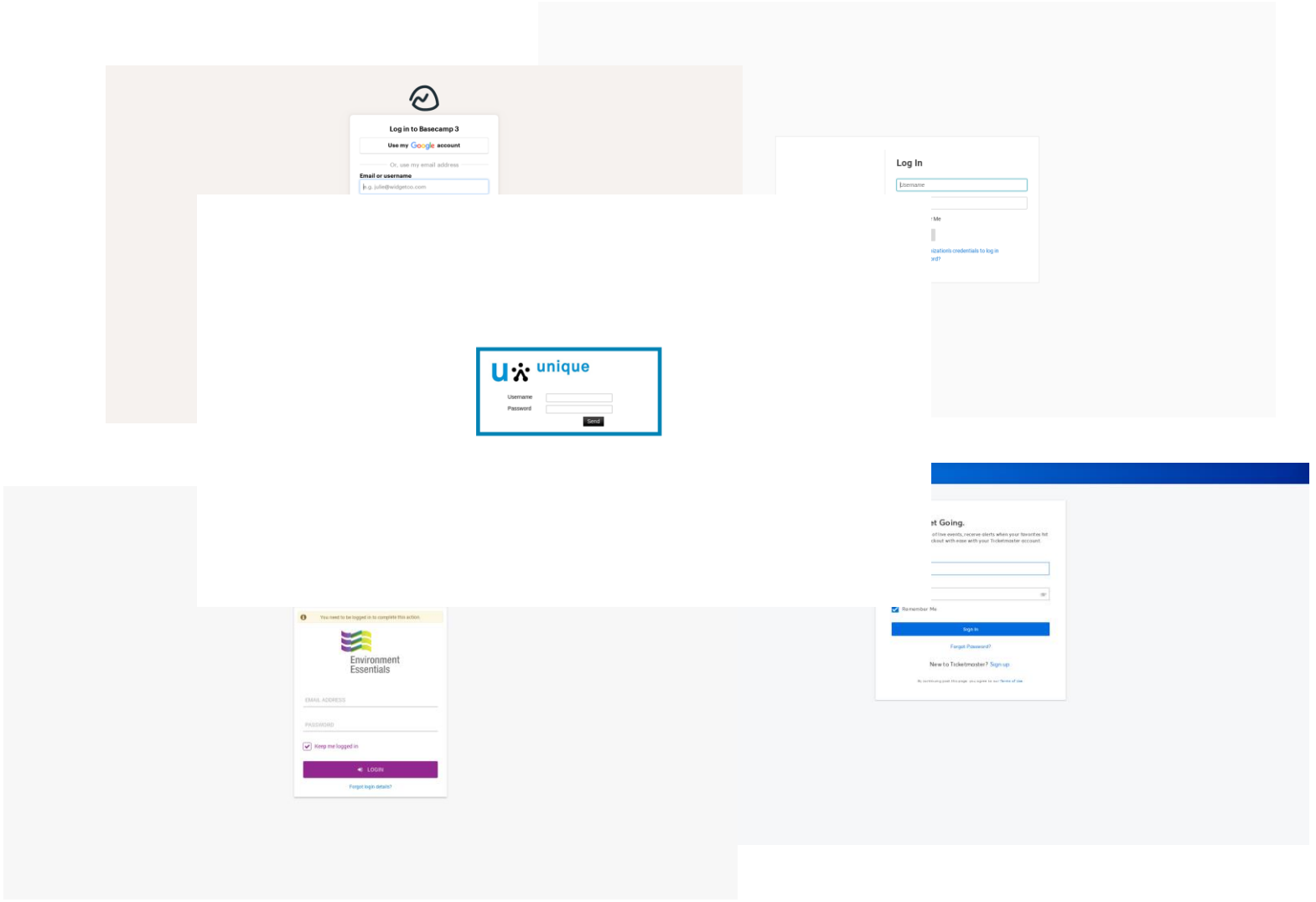


Resnet-34



Attempt #2: issues

- Login pages are not really a representation of real-world objects (ImageNet is full of cotidian things)
- Not able to extract logins of interest only!



SINGULARITY
TECH DAY

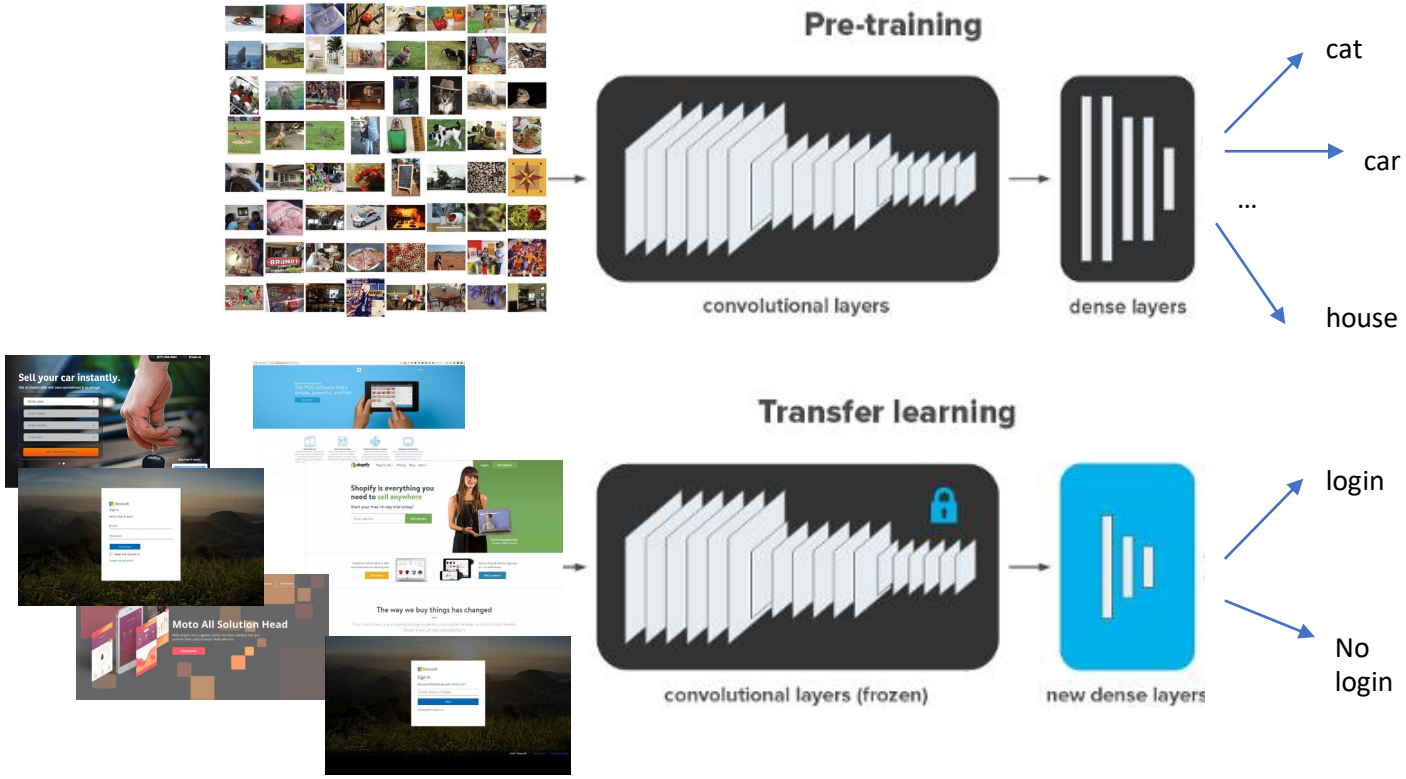
#STechDay2020

Agenda

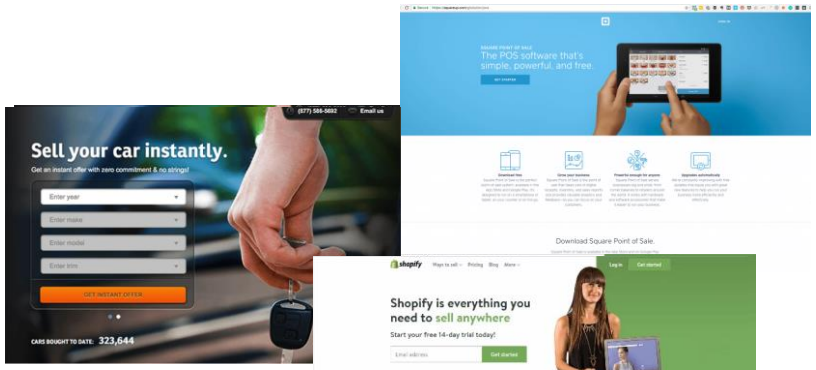
- Context
- Solution Overview
- Architecture
- Initial attempts
- Final Solution
- Challenges ahead

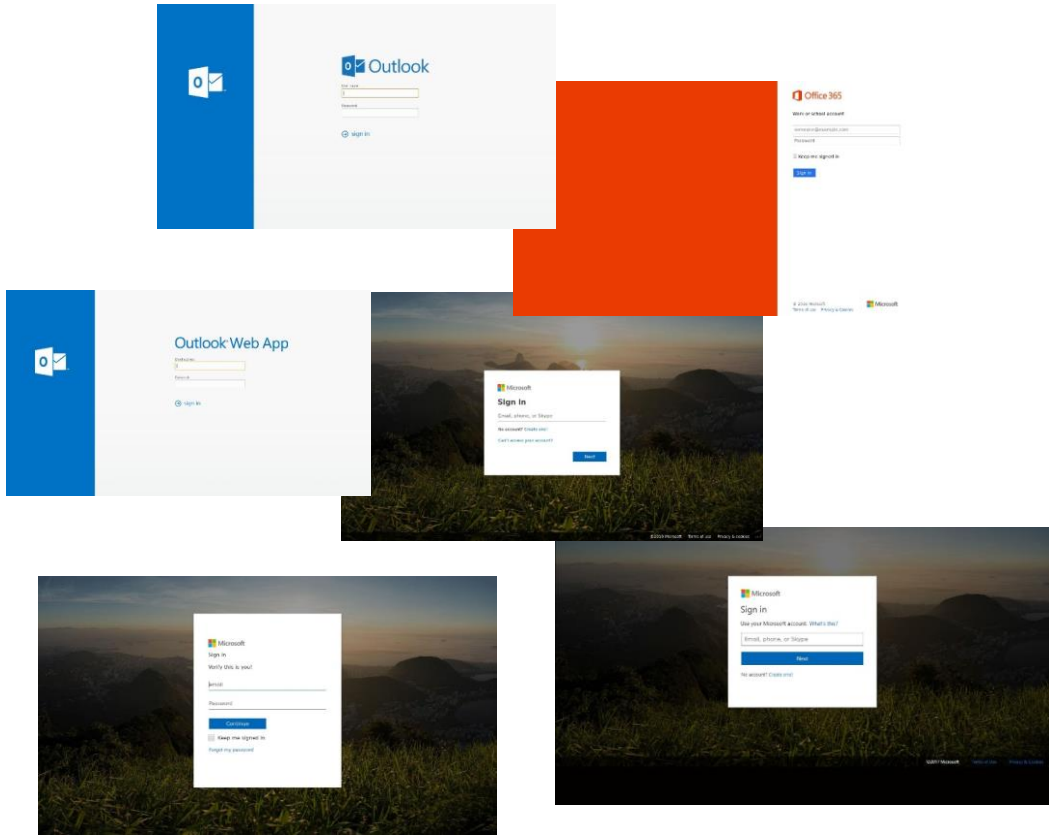
Final solution: transfer learning

- Customize embeddings with our specific classification problem: transfer learning on login/no-login
- Use the new image embeddings to compute similarities



Final solution: real examples





SINGULARITY
TECH DAY

#STechDay2020

Agenda

- Context
- Solution Overview
- Architecture
- Initial attempts
- Final Solution
- Challenges ahead

Challenges ahead

Deep Learning can be really useful to detect phishing but…

- Model drift
  - Cyber Security threats are constantly changing and proper monitoring and automatic retraining mechanisms for ML are a must
- Adversarial Machine Learning
  - Deep Learning (and ML in general) are quite vulnerable to manipulated inputs (designed specifically to trick them)
- Explainable AI
  - The more complex the models we use, the harder it becomes to understand their decisions

SINGULARITY
TECH DAY

#STechDay2020

THANKS AND...

SEE YOU SOON!

ORGANIZATION

plain concepts

SPONSORS

Microsoft

SUPPORT

World Central Kitchen