AI and Deep Learning applications in Industry

Kathy Copic

VP, Growth

The Insight Fellows Program

750 ⁺

DATA SCIENTISTS AND
DATA ENGINEERS

INSIGHT

Insight Fellows are **Data Scientists** and **Data Engineers** at:

Silicon Valley ● New York ● Boston ● Seattle
+ many others…

Data team segmentation

| Analytics & Decision Science | Machine Learning Engineering | AI & Deep Learning | Data Engineering |

Machine Learning Engineering

Products where Data is at the Core

**AI &
Deep
Learning**

When the
usual tools just
won't do

Data Engineering
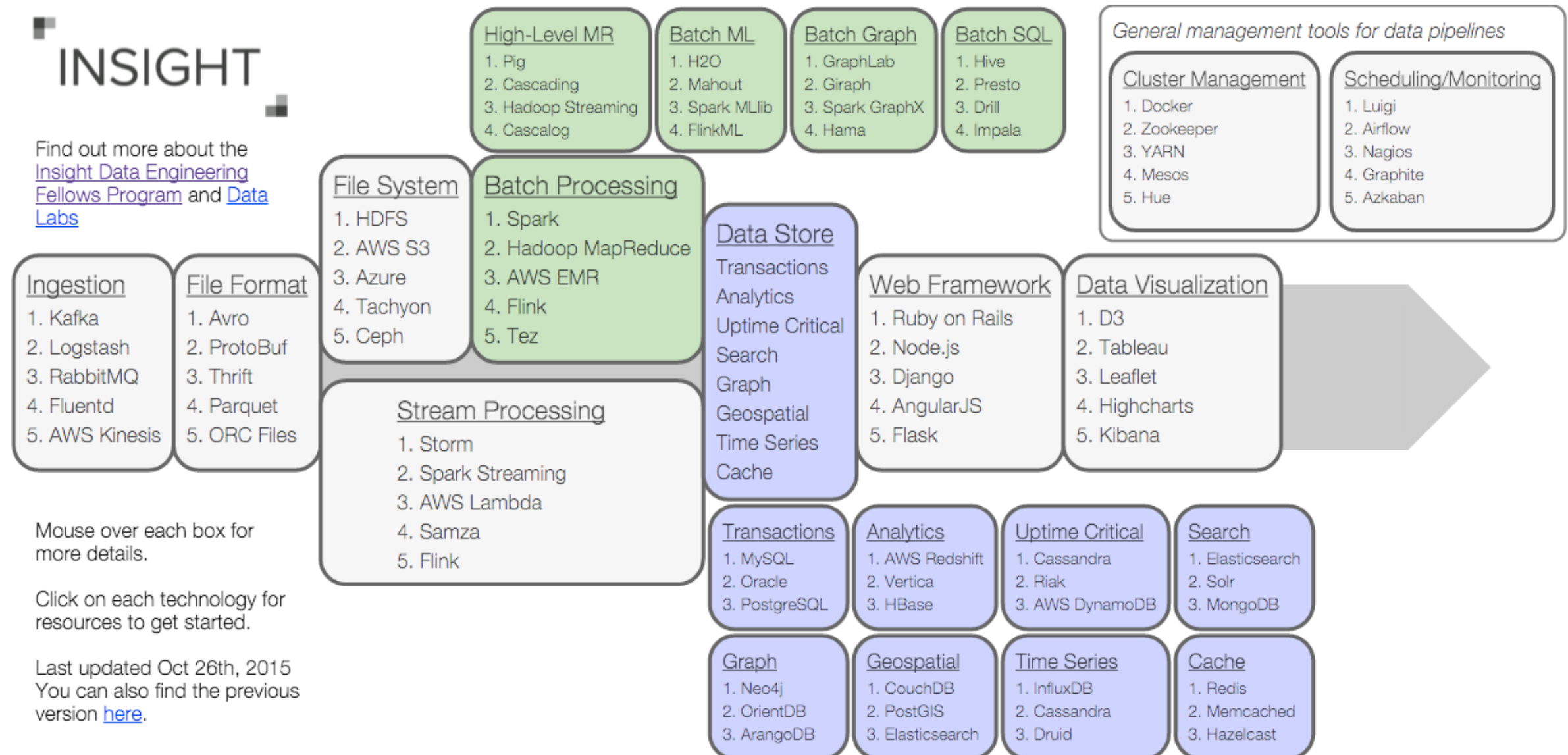
Volume
Velocity
Variety

INSIGHT

Find out more about the Insight Data Engineering Fellows Program and Data Labs

Mouse over each box for more details.

Click on each technology for resources to get started.

Last updated Oct 26th, 2015
You can also find the previous version here.

High-Level MR
1. Pig
2. Cascading
3. Hadoop Streaming
4. Cascalog

Batch ML
1. H2O
2. Mahout
3. Spark MLlib
4. FlinkML

Batch Graph
1. GraphLab
2. Giraph
3. Spark GraphX
4. Hama

Batch SQL
1. Hive
2. Presto
3. Drill
4. Impala

General management tools for data pipelines

Cluster Management
1. Docker
2. Zookeeper
3. YARN
4. Mesos
5. Hue

Scheduling/Monitoring
1. Luigi
2. Airflow
3. Nagios
4. Graphite
5. Azkaban

File System
1. HDFS
2. AWS S3
3. Azure
4. Tachyon
5. Ceph

Batch Processing
1. Spark
2. Hadoop MapReduce
3. AWS EMR
4. Flink
5. Tez

Data Store
Transactions
Analytics
Uptime Critical
Search
Graph
Geospatial
Time Series
Cache

Web Framework
1. Ruby on Rails
2. Node.js
3. Django
4. AngularJS
5. Flask

Data Visualization
1. D3
2. Tableau
3. Leaflet
4. Highcharts
5. Kibana

Ingestion
1. Kafka
2. Logstash
3. RabbitMQ
4. Fluentd
5. AWS Kinesis

File Format
1. Avro
2. ProtoBuf
3. Thrift
4. Parquet
5. ORC Files

Stream Processing
1. Storm
2. Spark Streaming
3. AWS Lambda
4. Samza
5. Flink

Transactions
1. MySQL
2. Oracle
3. PostgreSQL

Analytics
1. AWS Redshift
2. Vertica
3. HBase

Uptime Critical
1. Cassandra
2. Riak
3. AWS DynamoDB

Search
1. Elasticsearch
2. Solr
3. MongoDB

Graph
1. Neo4j
2. OrientDB
3. ArangoDB

Geospatial
1. CouchDB
2. PostGIS
3. Elasticsearch

Time Series
1. InfluxDB
2. Cassandra
3. Druid

Cache
1. Redis
2. Memcached
3. Hazelcast

Insight Alumni in AI roles

Justin Gilmer
Deep Learning Resident, Google Brain
Insight Fellow Winter 2015

Melissa Runfeldt
Machine Learning Engineer, Salesforce Einstein
Insight Fellow Summer 2016

TJ Torres
Deep Learning Data Scientist, Stitch Fix R&D
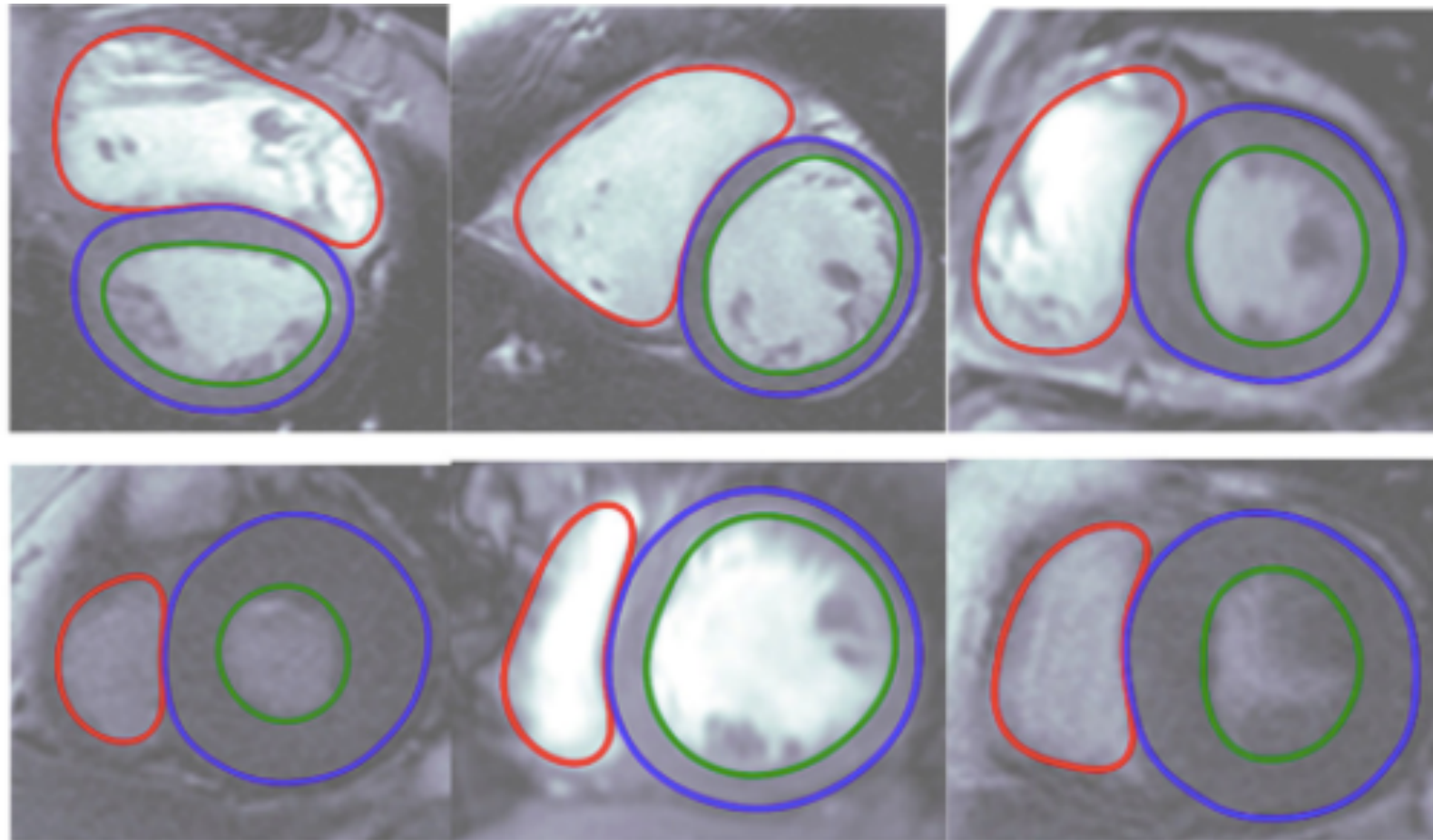Insight Fellow Summer 2015

Alex Gude
Deep Learning Data Scientist, Lab41, an IQT Lab
Insight Fellow Summer 2015

ARTERYS

Arterys Receives FDA Clearance For The First Zero-Footprint Medical Imaging Analytics Cloud Software With Deep Learning For Cardiac MRI

Font Finder - a deep learning based font search engine
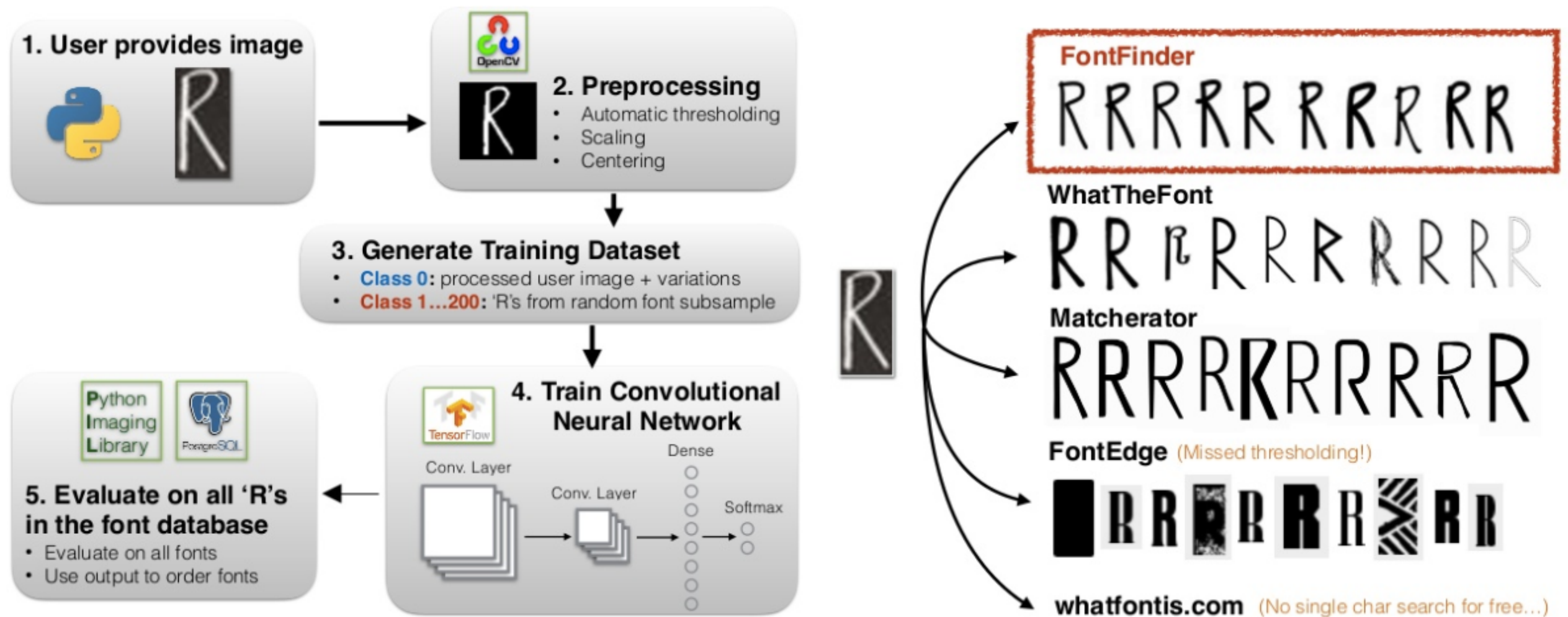
Michel Trottier-McDonald, ATLAS, now at Opendoor

iSee Method

Face **With** Glasses

Face, **No** Glasses

Input Image

Encode

Decode

Reconstructed Image

z

Align your face.

NEXT
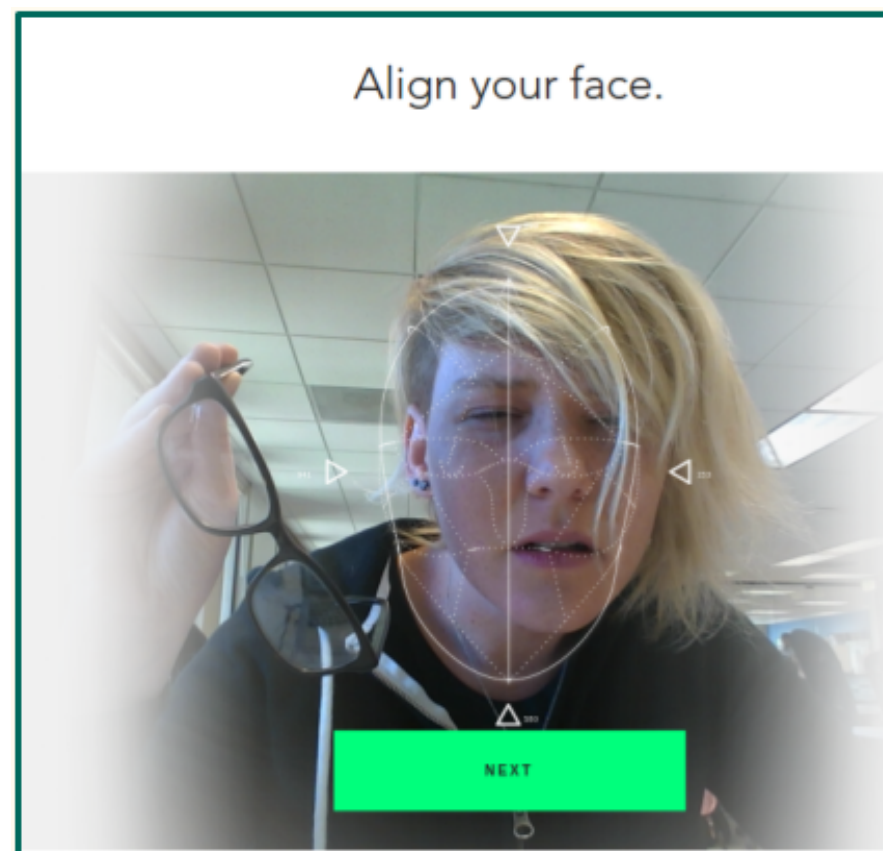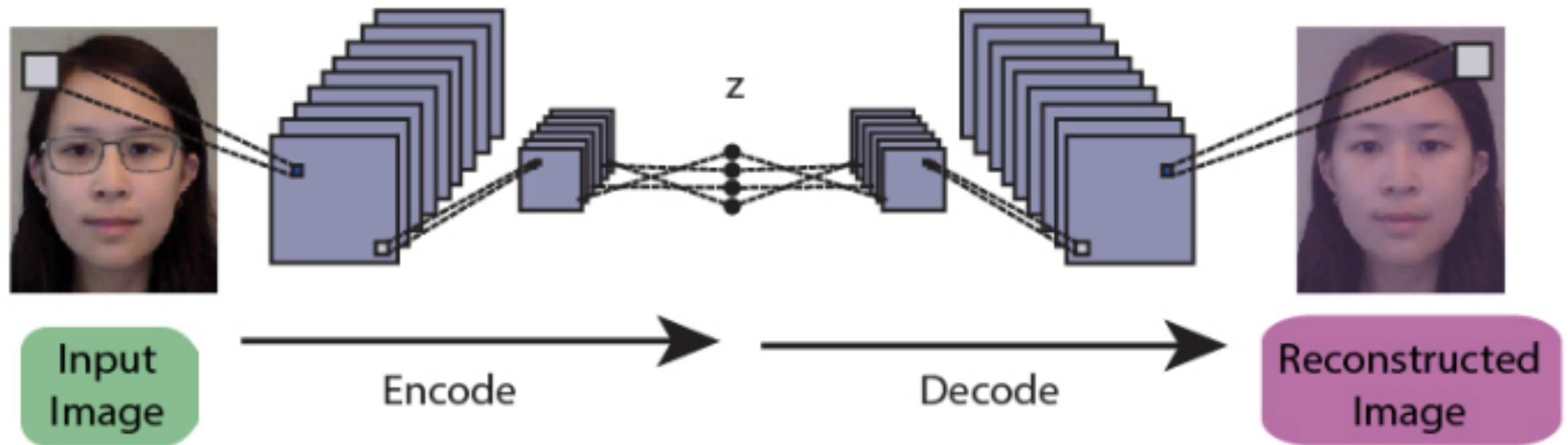
Separating Overlapping Chromosomes with Deep Learning
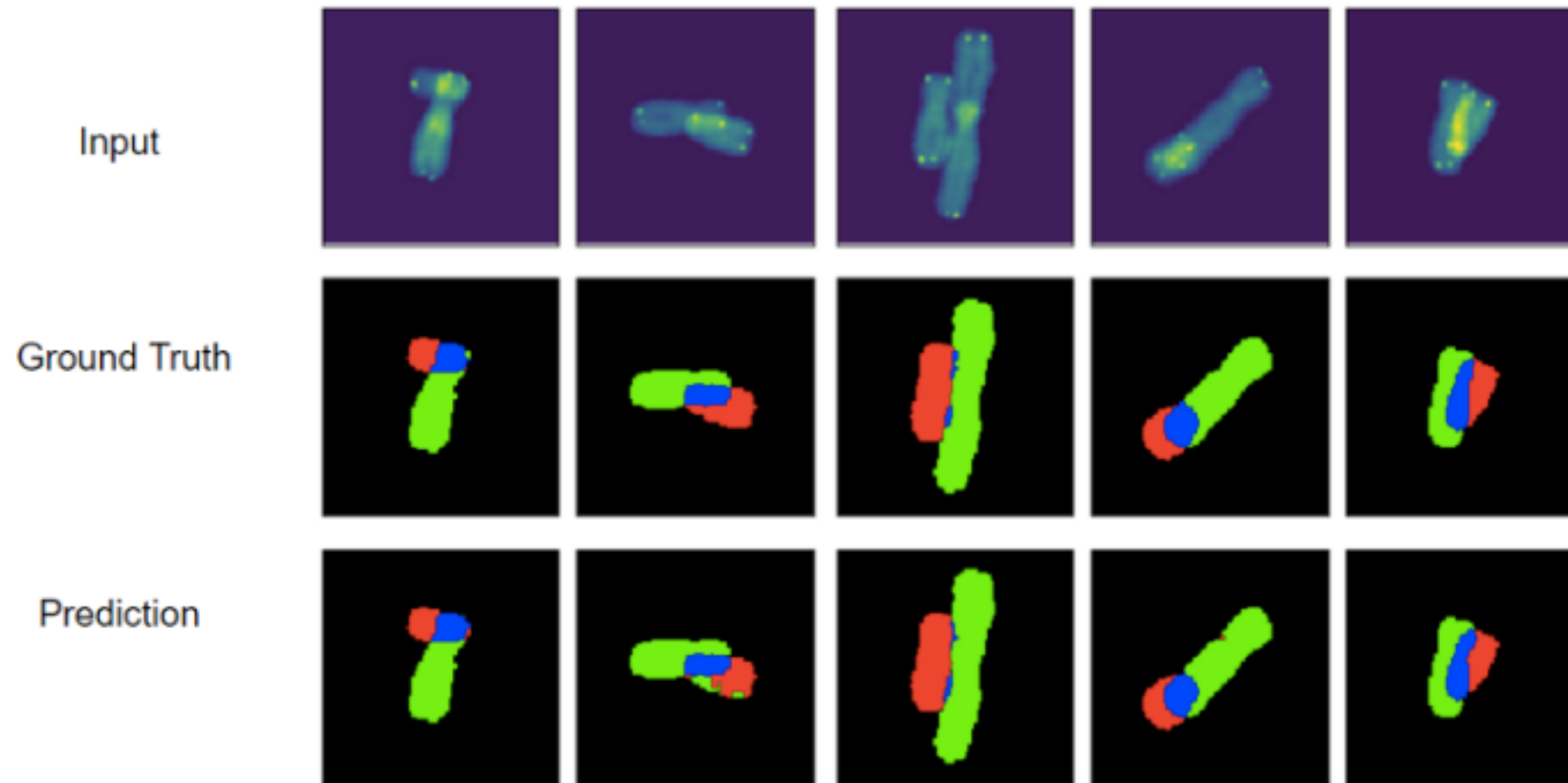
AI•ON
Artificial Intelligence
Open Network

Who are AI Professionals?

AI Researcher ← → AI Engineer

Baidu
1000 person team
ML experts from academia, prototype models
ML engineers from industry, put models into production

Facebook AI Research (FAIR)
Research Scientist: PhD, postdoc, many publications
Research Engineer: Master's, exposure to ML/AI

Ian Goodfellow @goodfellow_ian · Dec 11

#rocketai definitely has the most popular Jacobian-Optimized Kernel Expansion of NIPS 2016

↩ 6   ⇄ 42   ♥ 214   •••

Metrics for the Rocket AI launch party

Email RSVPs to party: 316

People who emailed in their resume: 46

Large name brand funds who contacted us about investing: 5

Media: Twitter, Facebook, HackerNews, Reddit, Quora, Medium etc
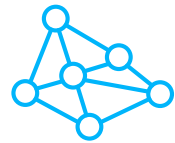
Time Planning: < 8 hours

Money Spent: $79 on the domain, $417 on alcohol and snacks + (police fine)

For reference, NIPS sponsorship starts at $10k.

Estimated value of Rocket AI: *in the tens of millions*.

Questions?

kathy@insightdatascience.com
Kathy Copic: LinkedIn
@KathyScientist

Academia to Industry

Challenges academics face when making the transition

Tools

Results published O(months) → **Pace** → **Results needed O(weekly), MVP**

Analysis for scientific community → **Product** → **Results need to impact bottom line**

Technical/domain-specific jargon → **Vocabulary** → **Business terms**