



Appen Limited

Investor Technology Day

19th May 2020



Agenda

- Introductions
- The need for training data, and challenges faced
- Overview of our technology platform
- Break
- How our technology scales our business
- How technology benefits our customers, our crowd and our people
- Questions

Introductions



Mark Brayan
CEO



Wilson Pang
CTO

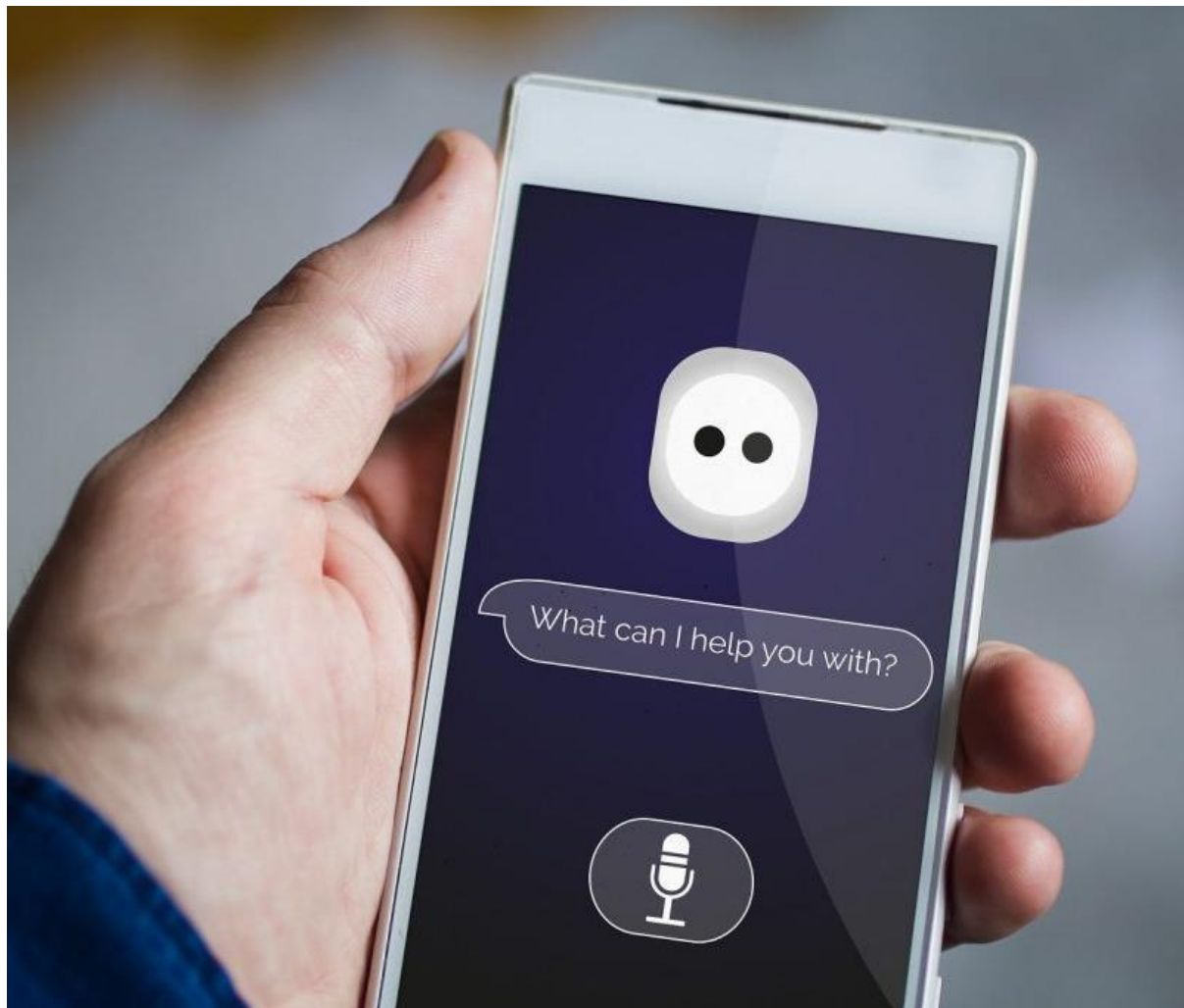


Meeta Dash
VP Product



Ryan Kolln
VP Corporate
Development

AI models are used as the decision engines within applications



Speech-to-text AI: converts the raw audio file into text



Natural Language Understanding (NLU): turns words into an actionable instructions



Computer Vision: Label image attributes to make them searchable beyond product name

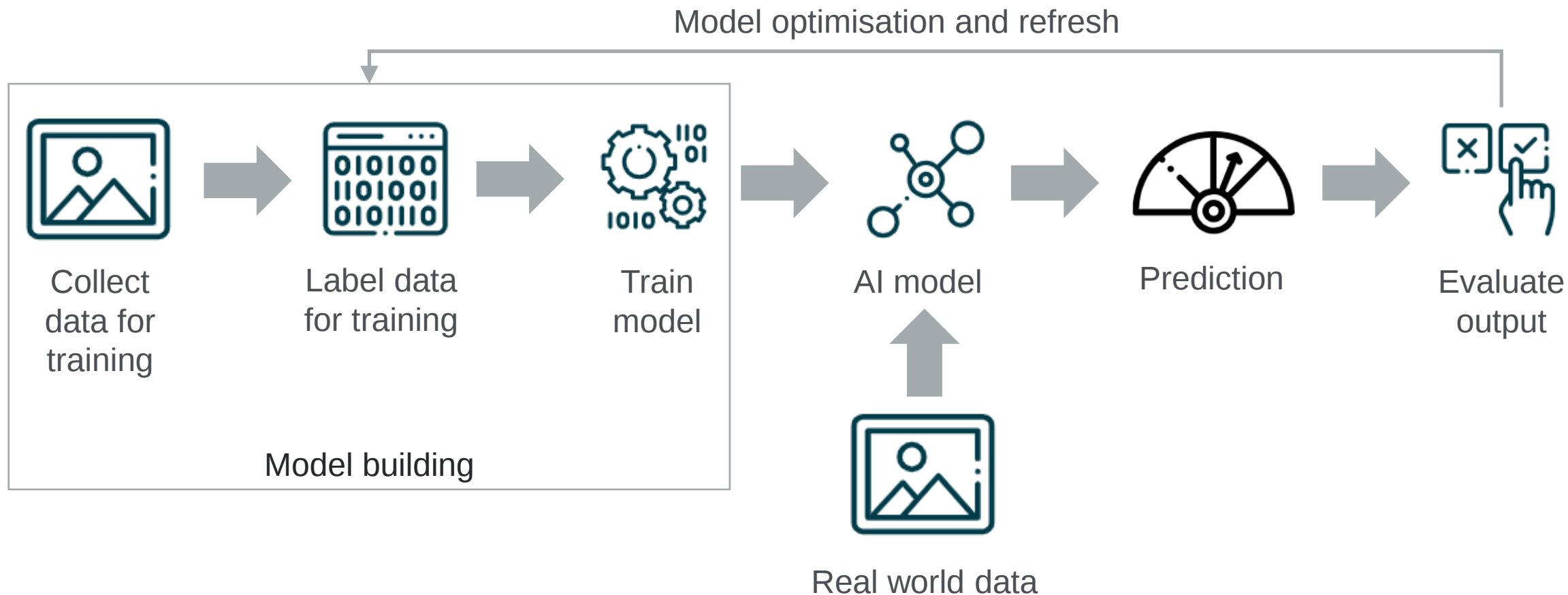


Search Relevance: Return the optimal response for the relevant search query



Product recommendations: Provide a set of alternative product options for the customers

Training data is core to building AI models



Training data is the IP for building artificial intelligence models

- Relevance
- Language
- Image

Relevance training data is typically a subjective judgement

Relevance Judgment

Query:
cowboy boots

Result Title:
Lane Boots Women's 'Scrollie' Cowboy Boot

Result Image:



How well does this Result match the Query? (required)

- ☐ Off Topic
- ☒ Acceptable
- ☐ Good
- ☐ Excellent

Corresponding JSON file

```
{
  "id": 2739608310,
  "data": {
    "title": "Lane Boots Women's 'Scrollie' Cowboy Boot",
    "image_url": "http://ak1.ostkcdn.com/images/products/9754985/P16927168.jpg",
    "query": "cowboy boots"
  },
  "judgments_count": 1,
  "state": "judgable",
  "agreement": null,
  "missed_count": 0,
  "gold_pool": null,
  "created_at": "2020-05-13T10:05:53+00:00",
  "updated_at": "2020-05-13T10:20:08+00:00",
  "job_id": 1580097,
  "results": {
    "judgments": [
      {
        "id": 5597061359,
        "created_at": "2020-05-13T10:19:57+00:00",
        "started_at": "2020-05-13T10:19:27+00:00",
        "missed": null,
        "rejected": null,
        "country": "USA",
        "region": "AZ",
        "city": "Phoenix",
        "unit_id": 2739608310,
        "job_id": 1580097,
        "worker_id": 45837366,
        "trust": 1,
        "worker_trust": 1,
        "unit_state": "judgable",
        "data": {
          "how_well_does_this_result_match_the_query": "acceptable"
        },
        "unit_data": {
          "title": "Lane Boots Women's 'Scrollie' Cowboy Boot",
          "image_url": "http://ak1.ostkcdn.com/images/products/9754985/P16927168.jpg",
          "query": "cowboy boots"
        }
      }
    ]
  },
  "how_well_does_this_result_match_the_query": {
    "agg": "acceptable",
    "confidence": 1
  }
}
```

Detail on the search term and returned item

Time and date

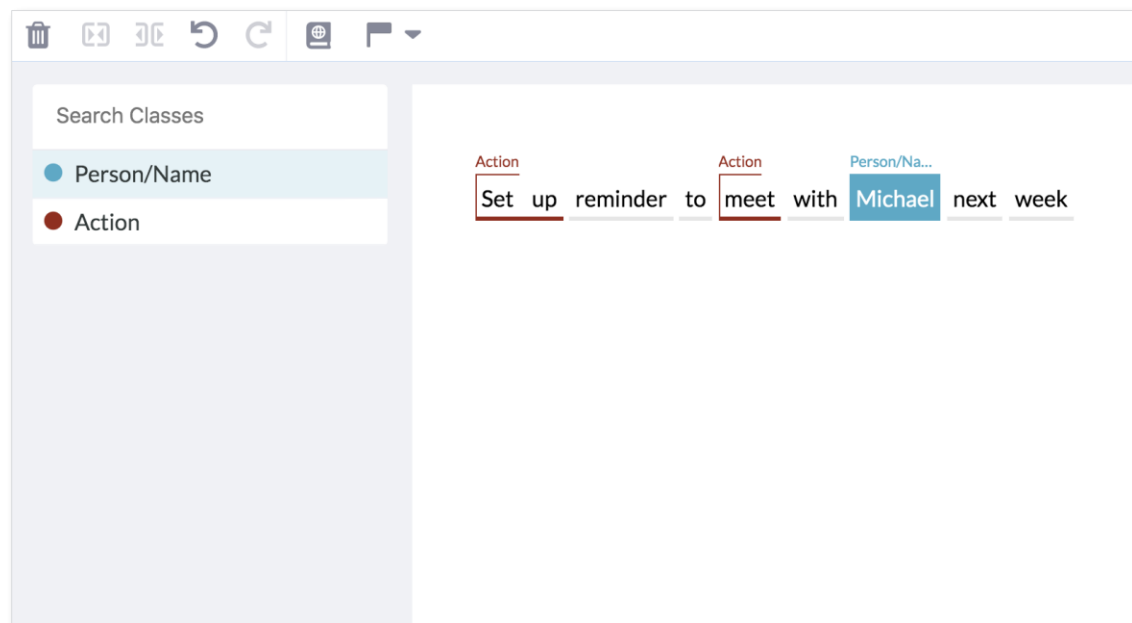
Location

Worker identifier

Result of task

Language training data assigns meaning to written or spoken words

Text Annotation/Tagging



Corresponding JSON file

```
{
  "id": 2740241946,
  "data": {
    "text": "Set up reminder to meet with Michael next week"
  },
  "judgments_count": 1,
  "state": "judgable",
  "created_at": "2020-05-14T06:52:34+00:00",
  "updated_at": "2020-05-14T07:04:07+00:00",
  "job_id": 1580607,
  "results": {
    "judgments": [
      {
        "id": 5598332480,
        "created_at": "2020-05-14T07:03:57+00:00",
        "started_at": "2020-05-14T07:03:21+00:00",
        "acknowledged_at": null,
        "external_type": "cf_internal",
        "country": "USA",
        "region": "AZ",
        "city": "Phoenix",
        "unit_id": 2740241946,
        "job_id": 1580607,
        "worker_id": 45837366,
        "trust": 1,
        "worker_trust": 1,
        "unit_state": "judgable",
        "data": {
          "annotations": "{\"job_id\":\"1580607\",\"annotation_id\":\"b70501bb8d8a52cd10affdcca63981624ee54574cbdc0a14\",\"custom_bucket\":false}"
        },
        "unit_data": {
          "text": "Set up reminder to meet with Michael next week"
        }
      }
    ],
    "annotations": {
      "tagg": "https://f8-text-annotation-live.s3.amazonaws.com/agg_report/2740241946/1141cbeebdd41d4079316814111a8038a64a2a747X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIK6U7WPQWDCESA0Q%2F20200514%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20200514T070501Z&X-Amz-Expires=604800&X-Amz-SignedHeaders=host&X-Amz-Signature=f5f0ee6d8d978994e68251c2f7a2739b6926a7a48970cebf4a57a5238e47ea08",
      "confidence": 1
    }
  }
}
```

Original text

Time and date

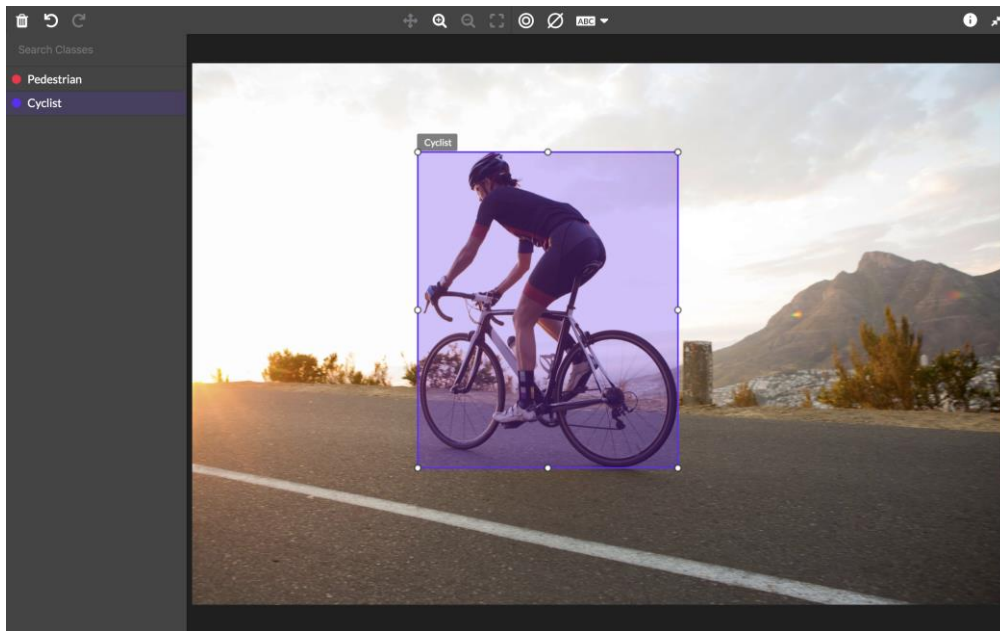
Location

Worker identifier

Annotation

Image and video training data assigns meaning to visual data

Text Annotation/Tagging



Corresponding JSON file

```
{
  "id": 2740230286,
  "data": {
    "image_url": "https://drive.google.com/uc?id=1uPo8X07Db2AD9TeGqzhQbDGoStMXeSEQ"
  },
  "judgments_count": 1,
  "state": "judgable",
  "created_at": "2020-05-14T06:29:04+00:00",
  "updated_at": "2020-05-14T06:48:26+00:00",
  "job_id": 1580600,
  "results": {
    "judgments": [
      {
        "id": 5598319441,
        "created_at": "2020-05-14T06:48:15+00:00",
        "started_at": "2020-05-14T06:47:08+00:00",
        "country": "USA",
        "region": "CA",
        "city": "Lafayette",
        "unit_id": 2740230286,
        "job_id": 1580600,
        "worker_id": 45837366,
        "trust": 1,
        "worker_trust": 1,
        "unit_state": "judgable",
        "data": {
          "annotation": "[{"id": "a78cc97e-855c-4bb0-8bd6-50bb5dd2fb32", "class": "Cyclist", "type": "box", "coordinates": {"x": 586, "y": 227, "w": 681, "h": 827}}]",
          "unit_data": {
            "image_url": "https://drive.google.com/uc?id=1uPo8X07Db2AD9TeGqzhQbDGoStMXeSEQ"
          }
        }
      }
    ],
    "nothing_to_box": {
      "agg": "",
      "confidence": 1
    }
  },
  "annotation": []
}
```

Location of image file

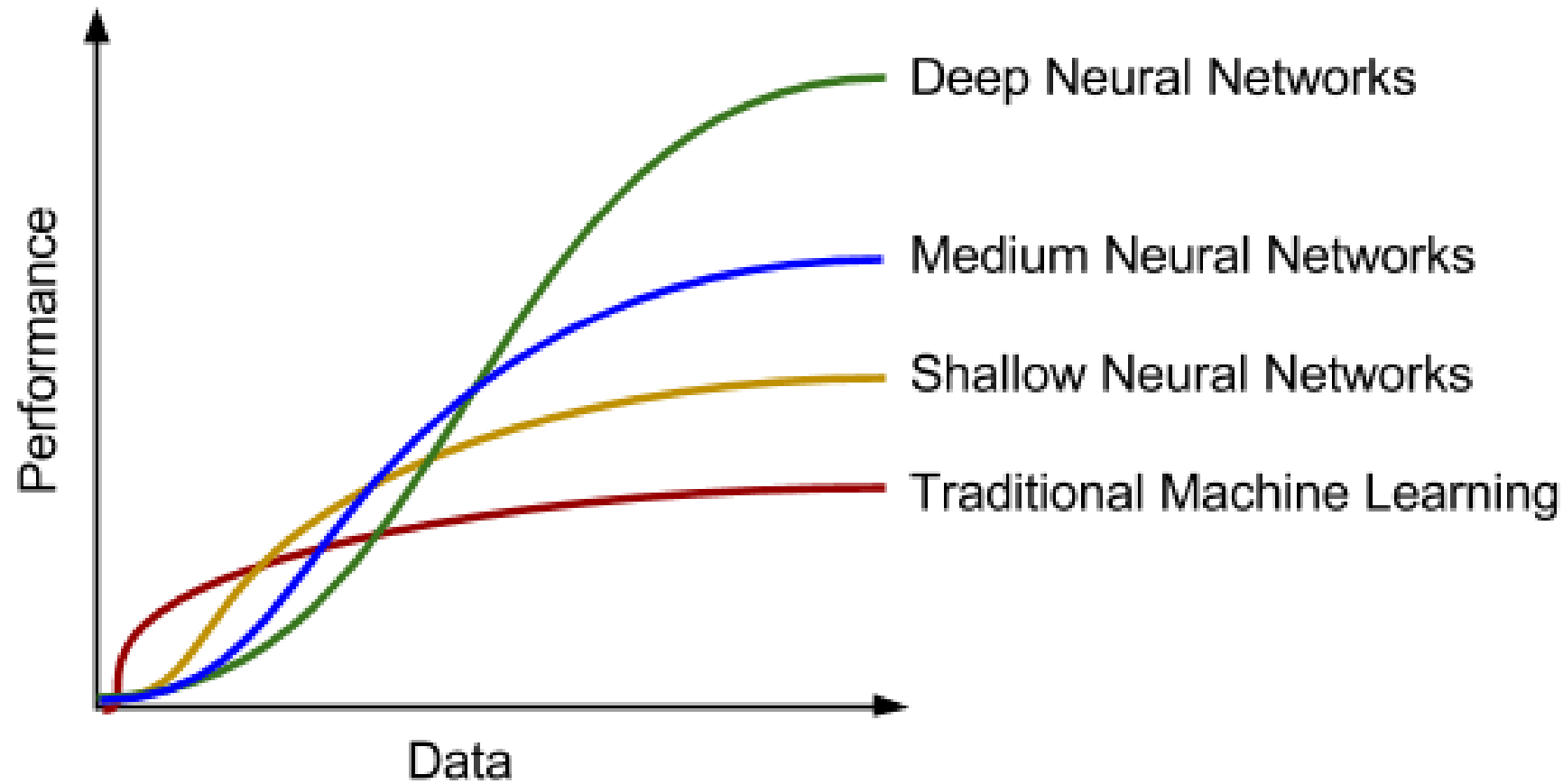
Time and date

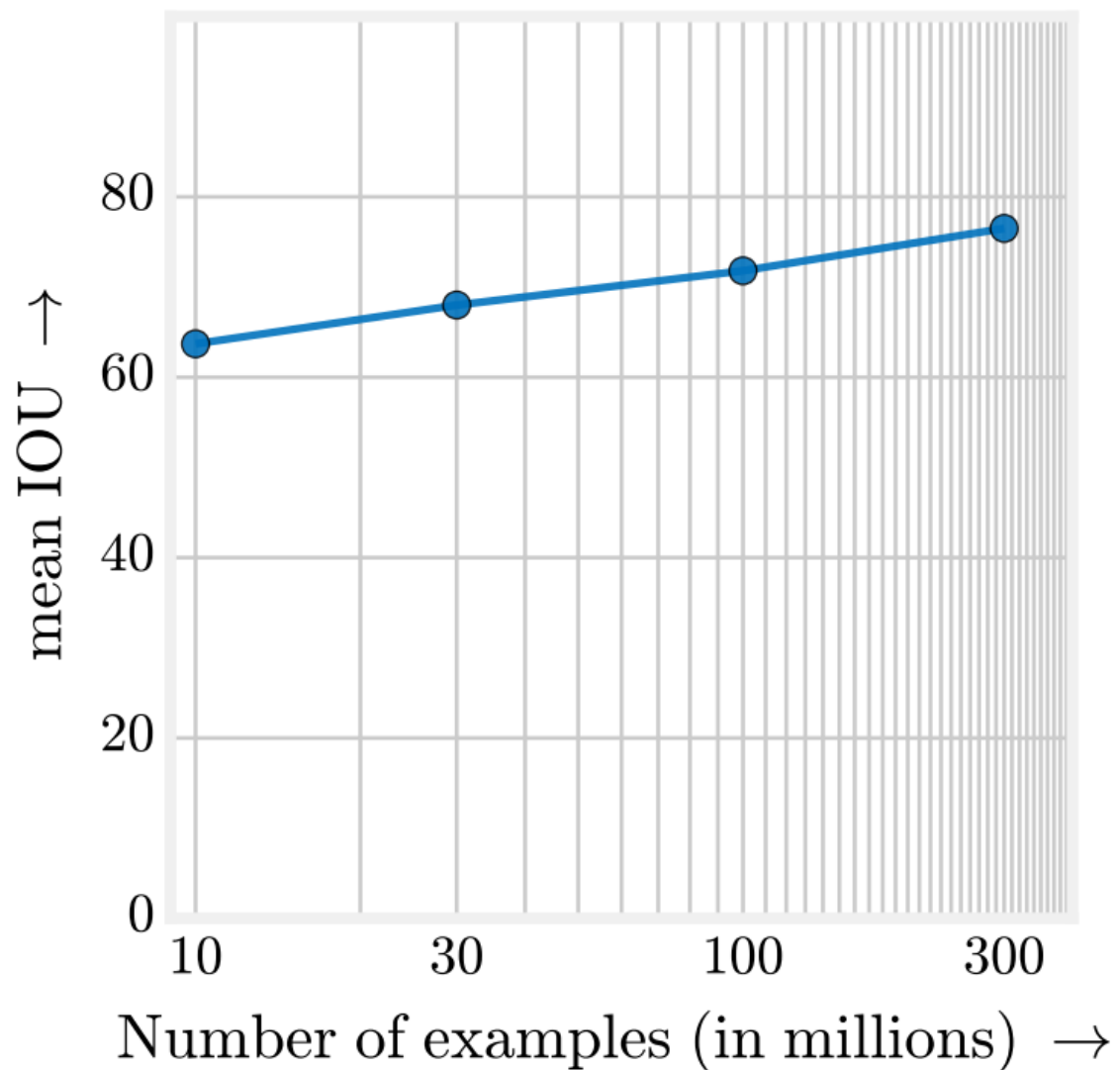
Location

Worker identifier

Coordinates of box and class "Cyclist"

AI performance is correlated with the volume of data used for training



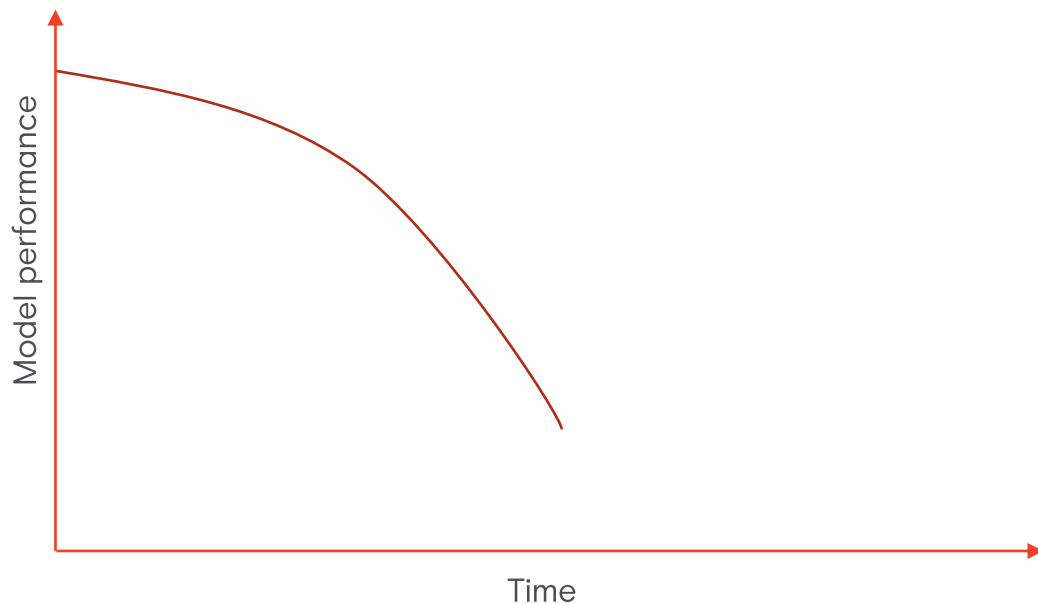


Source: Sun, C. et al., Revisiting Unreasonable Effectiveness of Data in Deep Learning Era, <https://arxiv.org/abs/1707.02968>, Aug. 2017.

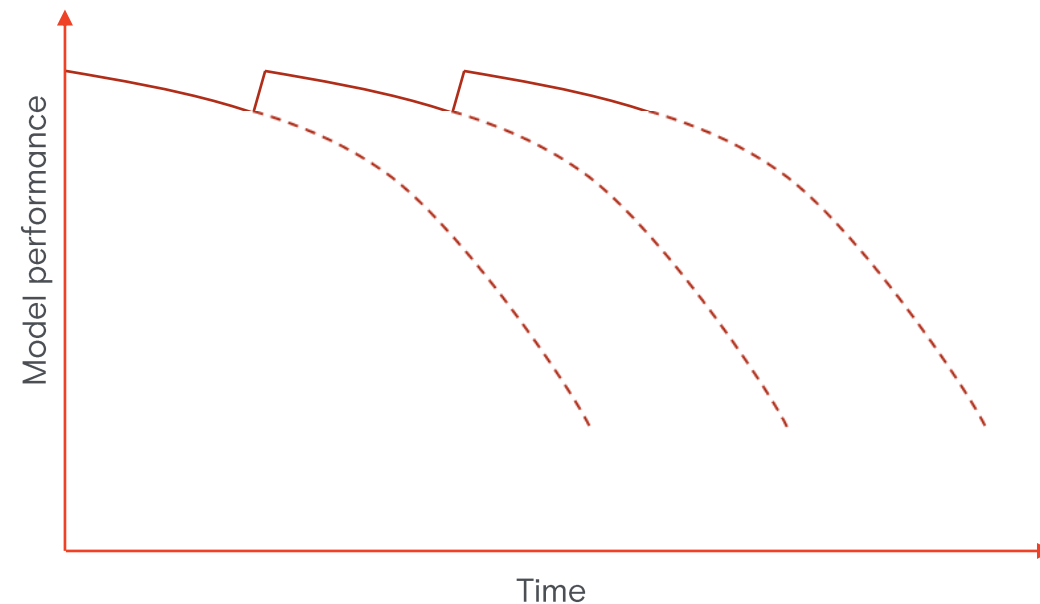
Studies show that performance increases logarithmically based on volume of training data

Volume of data is critical, but it also needs to be current

“Model drift” is a common in AI models



Refreshing training data ensures optimal model performance



- ~34% of models need to be refreshed monthly¹
- Large portion of our revenue is to support models from recurring project to maintain data recency

1. McKinsey

Even the best AI models are far from perfect



Color



Texture



Erratic

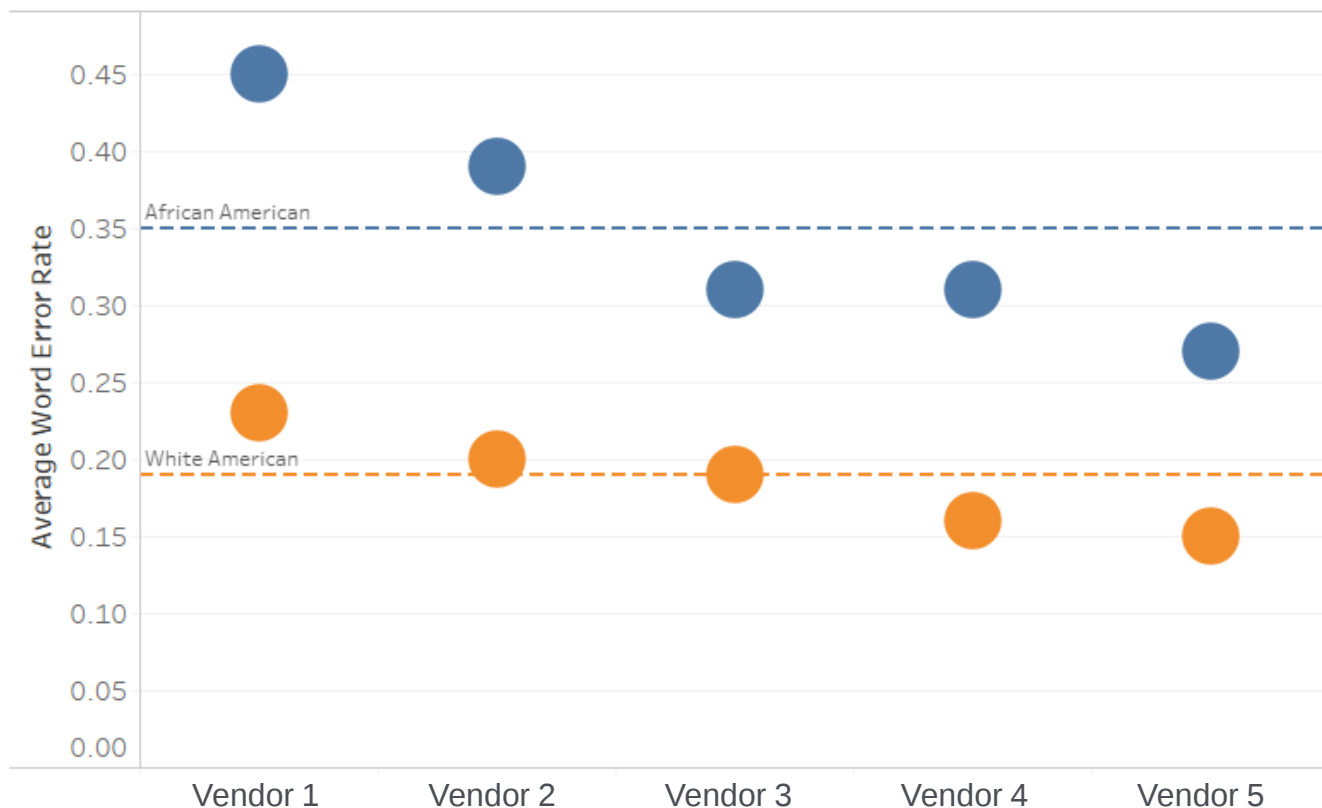


Overgeneralization

Source: <https://arxiv.org/pdf/1907.07174.pdf>

Performance issues often arise from insufficient training data

Average word error rates for machine transcriptions of African Americans and White Americans



*“..using **more diverse training datasets** that include African American Vernacular English will reduce performance differences and ensure speech recognition technology is inclusive”*

Proceedings of the National Academy of Sciences in the United States of America

Source: Proceedings of the National Academy of Sciences in the United States of America – Racial disparities in automated speech recognition (March 2020)

Our mission is to make
AI work in the real world
by creating large
volumes of high-quality
training data faster.

We combine our industry leading technology with a crowd of over 1m and deep internal expertise



The Appen Platform



**Crowd
management**



**Client
Workspace**



**Annotation
Tools**



Crowd Management provides a frictionless experience for crowd workers and project managers



Registration



Training



Project
Qualification



Ticketing



Payments



Social

Project Worker Smart Match



Campaign
Mgmt



Report



Ticketing



Quality
Assurance



Recruiting



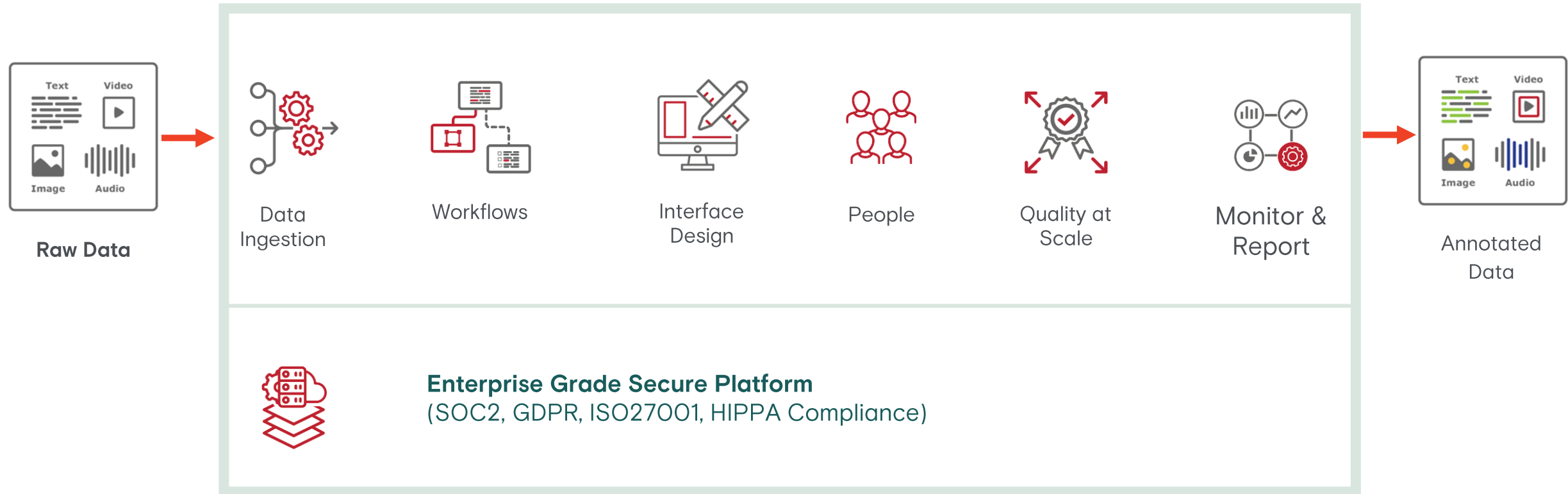
Project
Setup



Project
Managers



Client Workspace helps data scientists manage training data preparation efficiently





Annotation tools support a wide variety of data labelling use cases



Content Relevance

Search Relevance

Sentiment Analysis

Data Categorization

Data Validation



Speech & Language

Voice Transcription

Translation



Natural Language Understanding

Name Entity Extraction

Text Relationship



Computer Vision

Image Transcription

Image Annotation

Pixel Level Semantic Segmentation

Video Annotation

LiDAR Annotation

LiDAR Semantic Annotation



Data Collection

Mobile Recorder

Telephony

Studio Recorder

In-Car Recorder



Job Designer / CML (Custom Markup Language)



**Crowd
management**



**Client
Workspace**



**Annotation
Tools**

Our platform provides the building blocks to provide large volumes of high-quality training data faster

Building on core tech to accelerate scalability

Build the platform



**Crowd
management**



**Client
Workspace**



**Annotation
Tools**

Accelerate scalability

1. Product breadth

Support a wider variety of use cases and customers to increase market share

2. Crowd productivity

Utilise AI to automate and increase crowd productivity

3. Internal efficiency

Apply AI to our internal operations to deliver large-scale high-quality data faster

Increasing our product breadth will enable us to capture greater market share

1. Product breadth

2. Crowd productivity

3. Internal efficiency

Examples of how we are increasing product breadth

- ❖ Support a wider variety of data annotation types, including vertical specific capabilities
- ❖ Improve our low-touch service models to increase our addressable market
- ❖ Enhance our secure products to meet privacy requirements
- ❖ Create tighter integration with our customers' model building pipelines

Utilise AI and automation to improve the throughput of our crowd



1. Product breadth

2. Crowd productivity

3. Internal efficiency

Examples of how we are increasing crowd throughput

- ❖ Incorporate AI into our data labelling process to radically increase the output quality and volume of our crowd

Apply AI to our internal operations to deliver large-scale high-quality data



1. Product breadth

2. Crowd
productivity

3. Internal
efficiency

Examples of how we are increasing internal efficiency

- ❖ Leveraging AI to improve the crowd recruitment process
- ❖ Improve data quality by automatically matching crowd workers with tasks better suited to their capabilities

Technology plays a critical role for our customers, crowd workers, and employees



Customers



Crowd workforce



Employees



How technology benefits our Customers

- ❖ Customisable platform to support annotation tasks across a wide variety of data types
- ❖ Labelling automation improves speed and quality data annotation
- ❖ Workflows simplifies complex multi-step annotation tasks
- ❖ Integration with customers real-time data pipelines
- ❖ Greater security of customers' data



Benefits to the business

Broader product set enables us to capture a wider share of the market

Greater integration with our customers' operations drives loyalty

Improved unit economics accelerates scalability



How technology benefits our Crowd

- ❖ Predictive matching of crowd workers to tasks improves worker experience and output
- ❖ Automated resume screening accelerates onboarding
- ❖ Annotation automation reduces repetition, creating a more rewarding crowd experience
- ❖ Improved user-interface of annotation tools reduces friction in the labelling process



Benefits to the business

Accelerated onboarding improves our ability to fulfill project demands

More tenured crowd workers ramp faster and reduce rework, resulting in better margins

A large, diverse and engaged crowd enables us to serve more high-volume annotation requirements



How technology benefits our Employees

- ❖ Automated quality management through AI validation
- ❖ Faster and more flexible project setup (for managed service customers using our platform)
- ❖ Significant increases in the speed and volume of applicant processing



Benefits to the business

Internal workforce productivity lowers project overhead costs

Improved quality delivers better customer outcomes and results in more loyal customers

Technology is essential to the delivery of large volumes of high-quality training data faster

Build the platform



**Crowd
management**



**Client
Workspace**



**Annotation
Tools**

Accelerate scalability

1. Product breadth

**2. Crowd
productivity**

**3. Internal
efficiency**

Revenue growth and quality

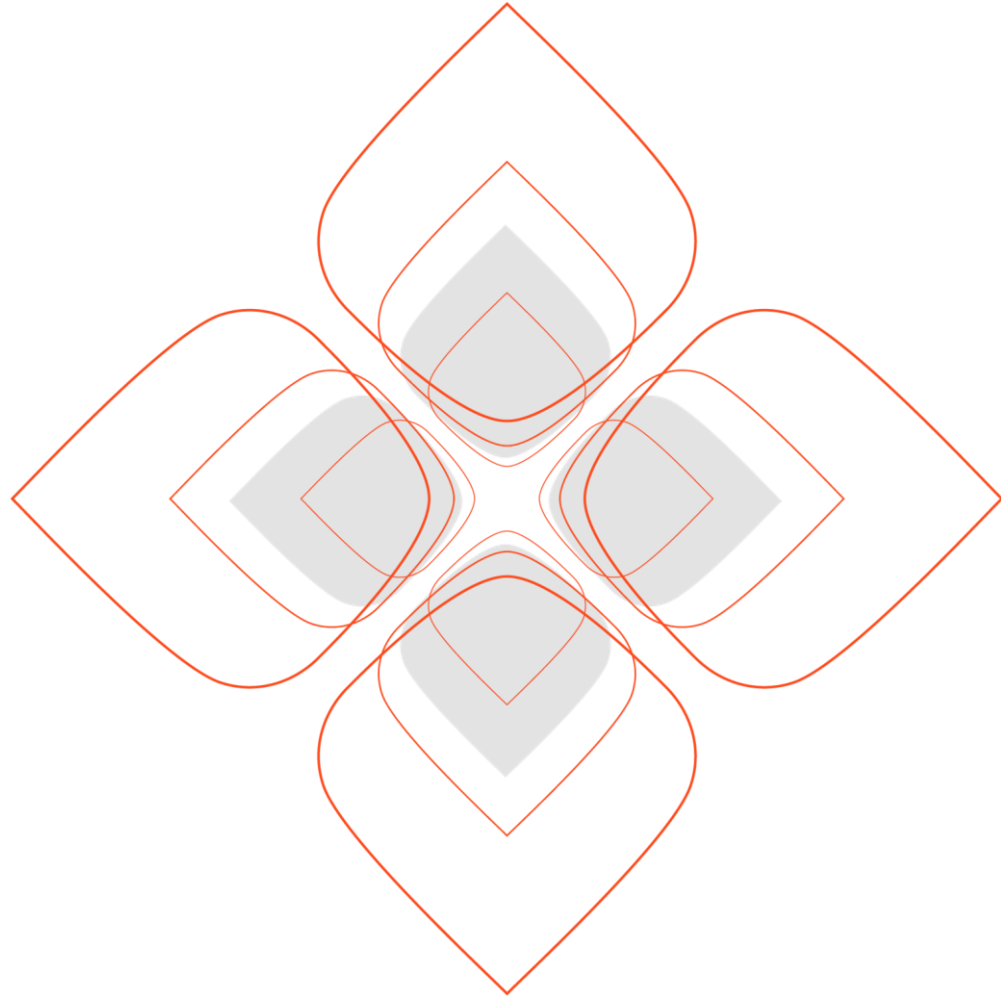
- Increased market share by supporting more use cases and customers
- Diversified customer base and greater stickiness

Improved margins

- Lower cost to serve due to crowd and internal productivity gains

Competitive advantage

- Technology and scale creates competitive moat



Thank you