



Appen Limited
Level 6, 9 Help Street
Chatswood NSW 2067
Tel: 02 9468 6300
www.appen.com
ACN 138 878 298

ASX ANNOUNCEMENT

20 May 2021

INVESTOR TECHNOLOGY DAY PRESENTATION

Appen Limited (ASX: APX) is holding an online Investor Technology Day event commencing at 11.00am AEST today. To join the webcast, please pre-register at <https://kapara.rdbk.com.au/landers/12ee87.html>. Registration will close at 10.00am AEST today.

The presentation materials for the event are attached.

Authorised for release by the CEO.

Please contact for more information:

Linda Carroll
Investor Relations
+61 2 9468 6300
investorrelations@appen.com
www.appen.com/investors

About Appen Limited

Appen is a global leader in the development of high-quality, human annotated datasets for machine learning and artificial intelligence. Appen brings 25 years of experience collecting and enriching a wide variety of data types including speech, text, image and video. With expertise in 235 languages and access to a global crowd of over 1,000,000 skilled contractors, Appen partners with leading technology, automotive and ecommerce companies - as well as governments worldwide - to help them develop, enhance and use products that rely on natural languages and machine learning. www.appen.com



Appen Limited

Investor Technology Day

20 May 2021

Agenda

11:00am CEO's opening remarks and introductions
Mark Brayan

11:15am AI market update
Ryan Kolln, VP Corporate Development

12:00pm Break

12:20pm How our technology creates value for customers
Wilson Pang, CTO

1:20pm Q&A

1:55pm CEO's closing remarks
Mark Brayan

2:00pm End

Introductions



Mark Brayan
CEO



Wilson Pang
CTO



Ryan Kolln
VP Corporate
Development

Appen is accelerating
its transformation into
an AI powered provider
of AI data and solutions

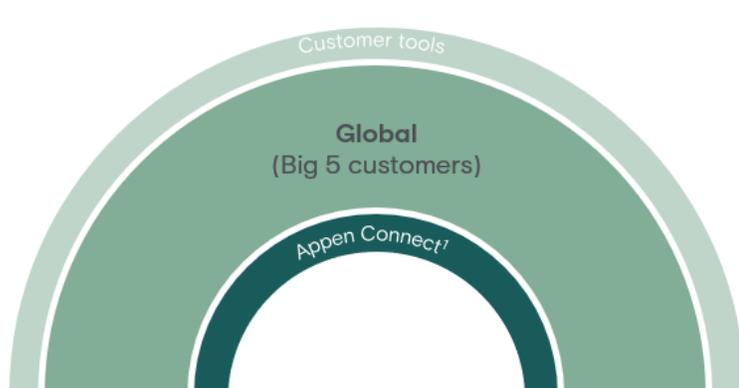
Appen continues to evolve

	From	To
Data type	Language data	AI data
Delivery model	Service led	Product led
Revenue	Project based	Committed
Customer	Concentration	Diversification
Org structure	Functional alignment	Customer alignment
Reporting	Data modality, AUD	Strategy led, USD

Appen's evolution

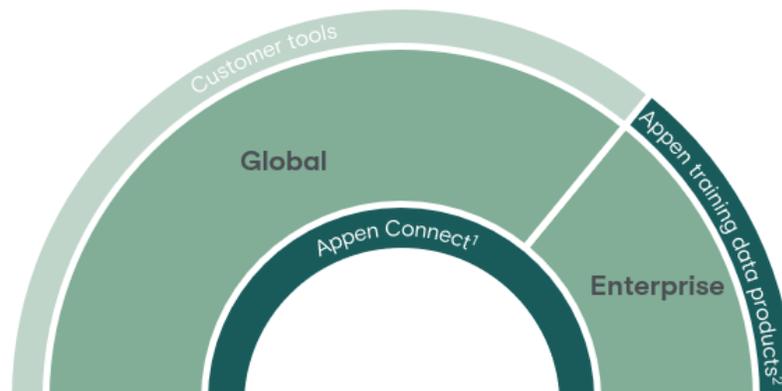
Phase 1: Dominate the AI data annotation services market

Acquired Butler Hill 2010
Acquired Leapforce 2017



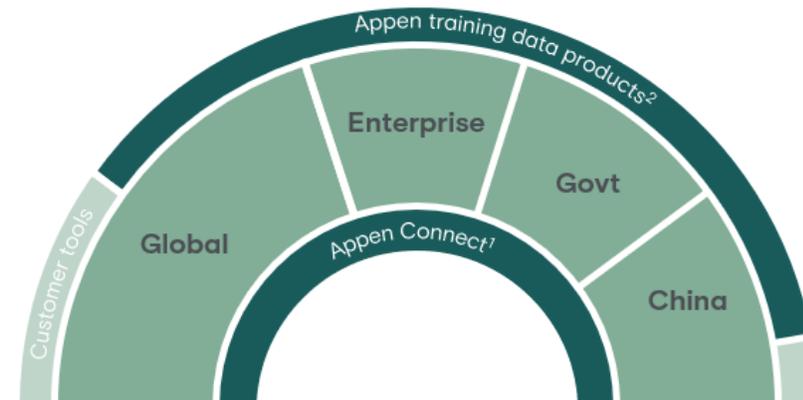
Phase 2: Build leading data annotation products

Acquired Figure Eight 2019
Ongoing product investment
New Enterprise customer base



Phase 3: Product-led expansion

Increased sales and marketing
China and Govt BUs established
Global customers on Appen products



Our areas of focus

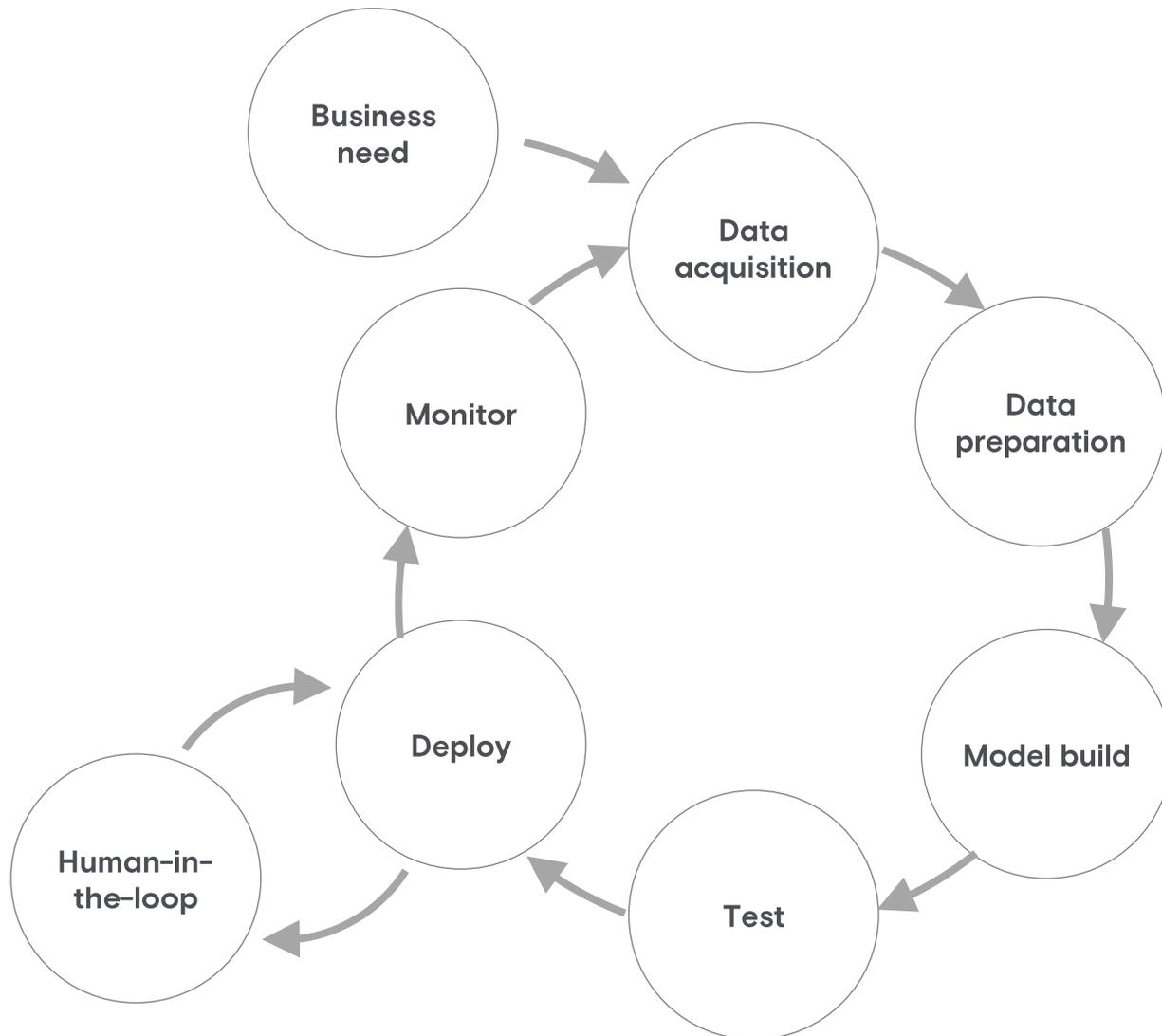
Product Led

Our future is product led, enabling us to deliver high-quality training data, faster, at larger scale, with improved unit economics, and is a foundation for future capabilities

Customer Centric

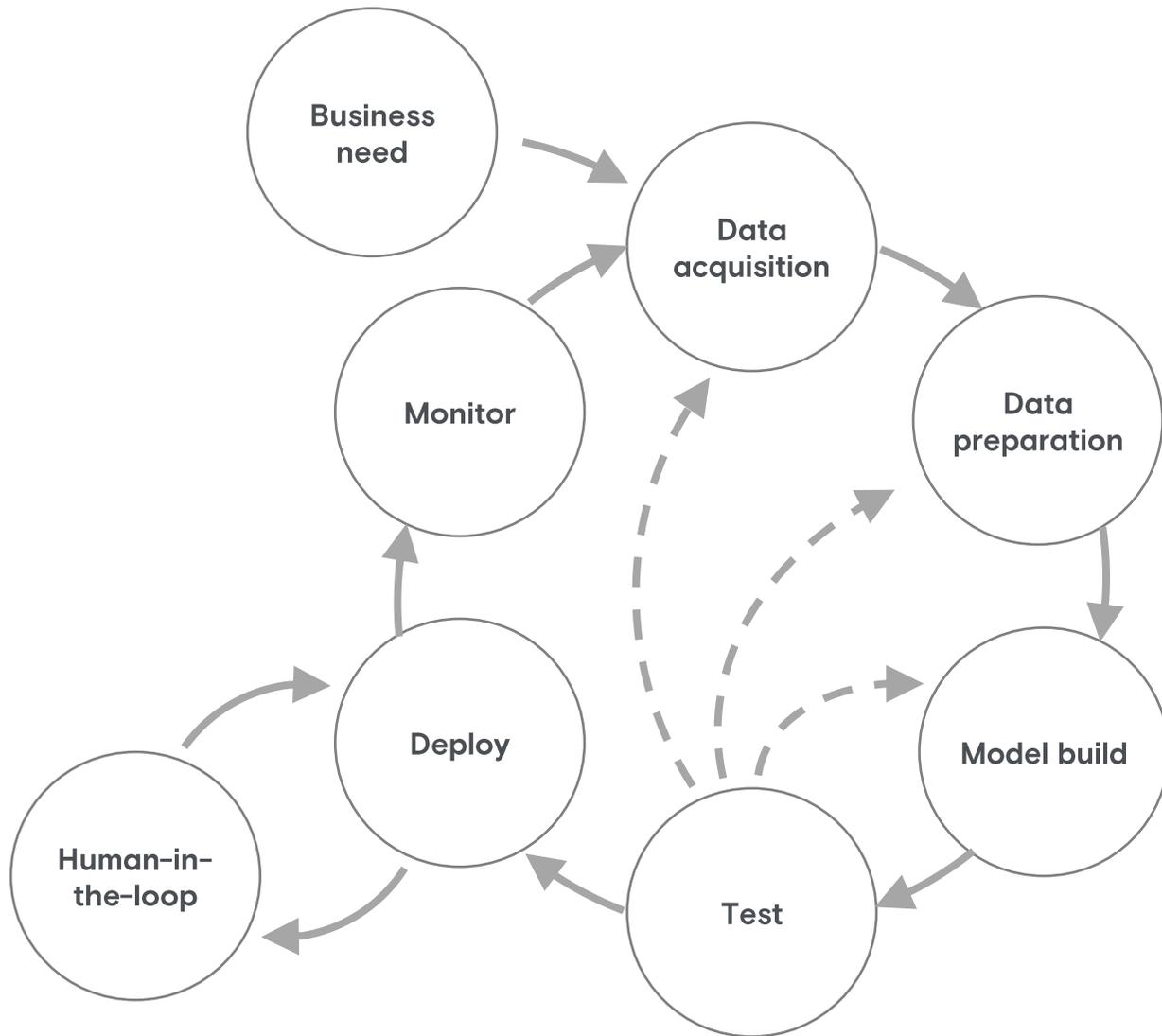
We are aligning our operations to better support the needs of our target customer cohorts

Appen plays an
important role in the AI
application lifecycle



AI lifecycle

An ongoing, iterative approach



AI lifecycle

An ongoing, iterative approach ... with multiple iteration loops

AI model = Model instructions + Training Data



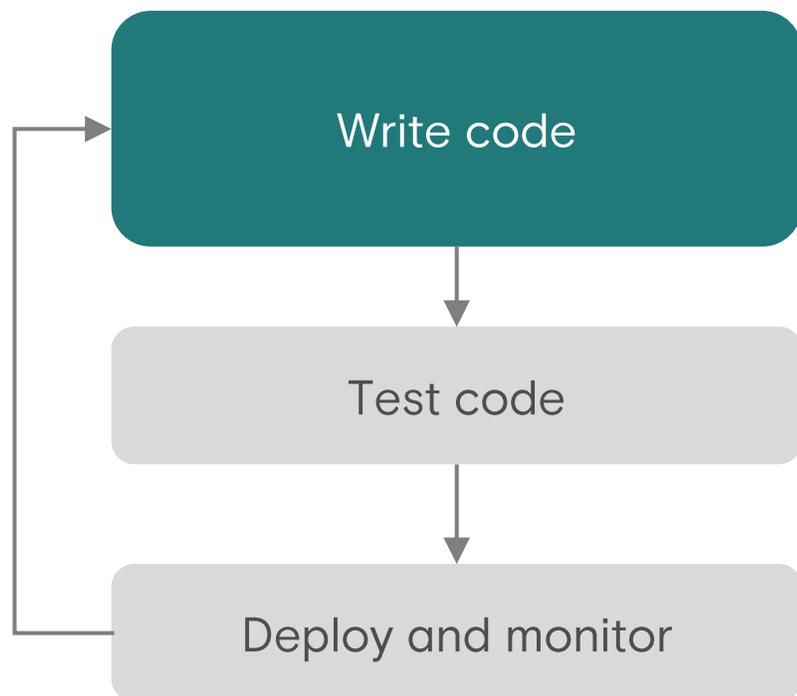
AI model = Model instructions + Training Data

Architecture for the model
to learn – can be as little
as 10 lines of code

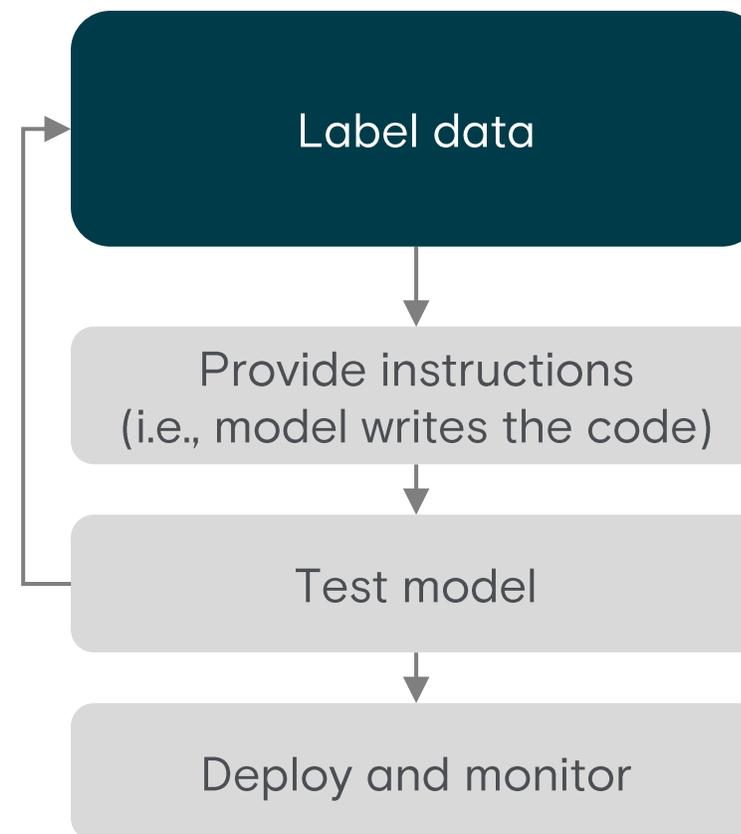
Examples used in the
model training process,
typically more the better

In AI, labelling data is *the most* important step

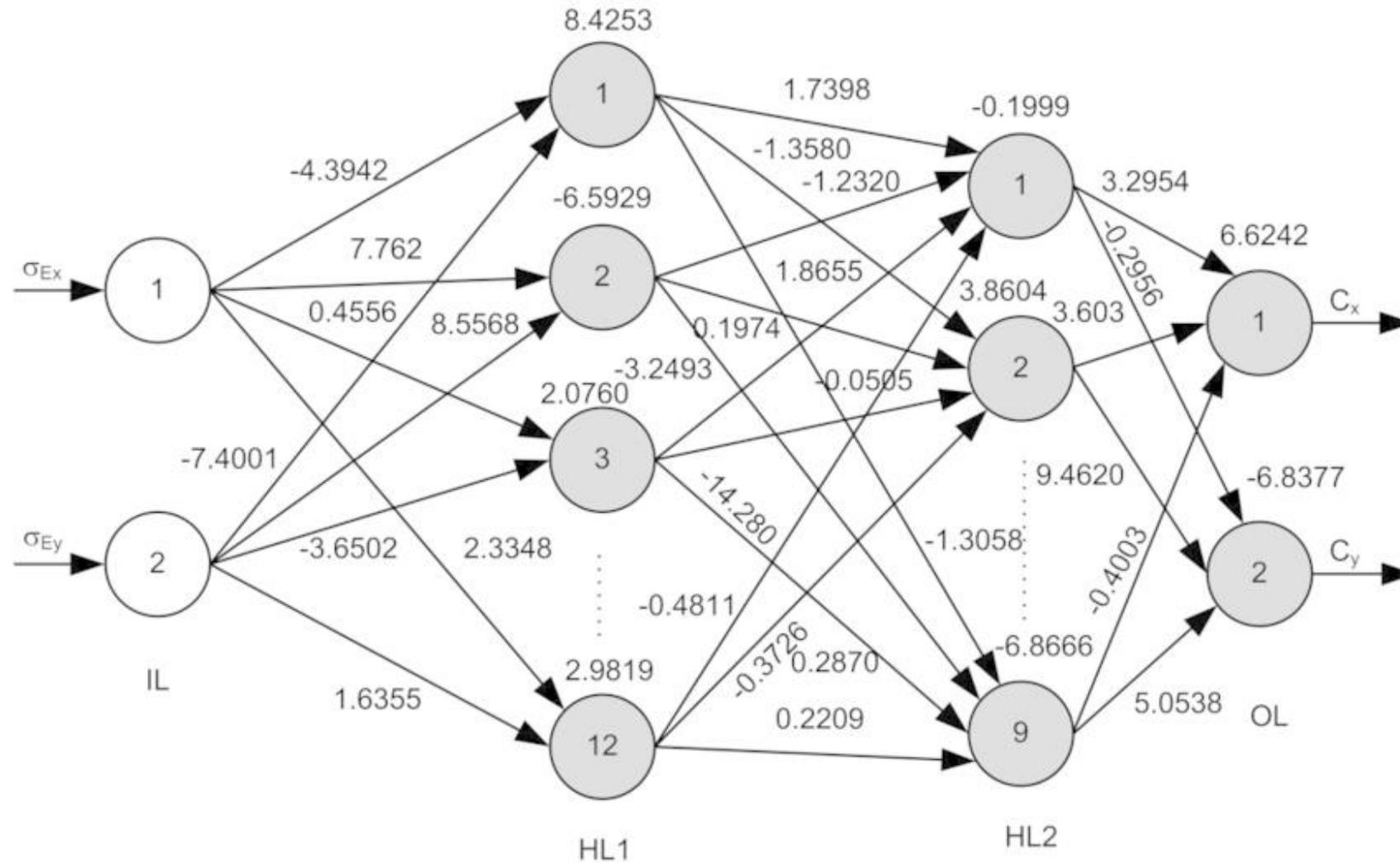
Traditional software development



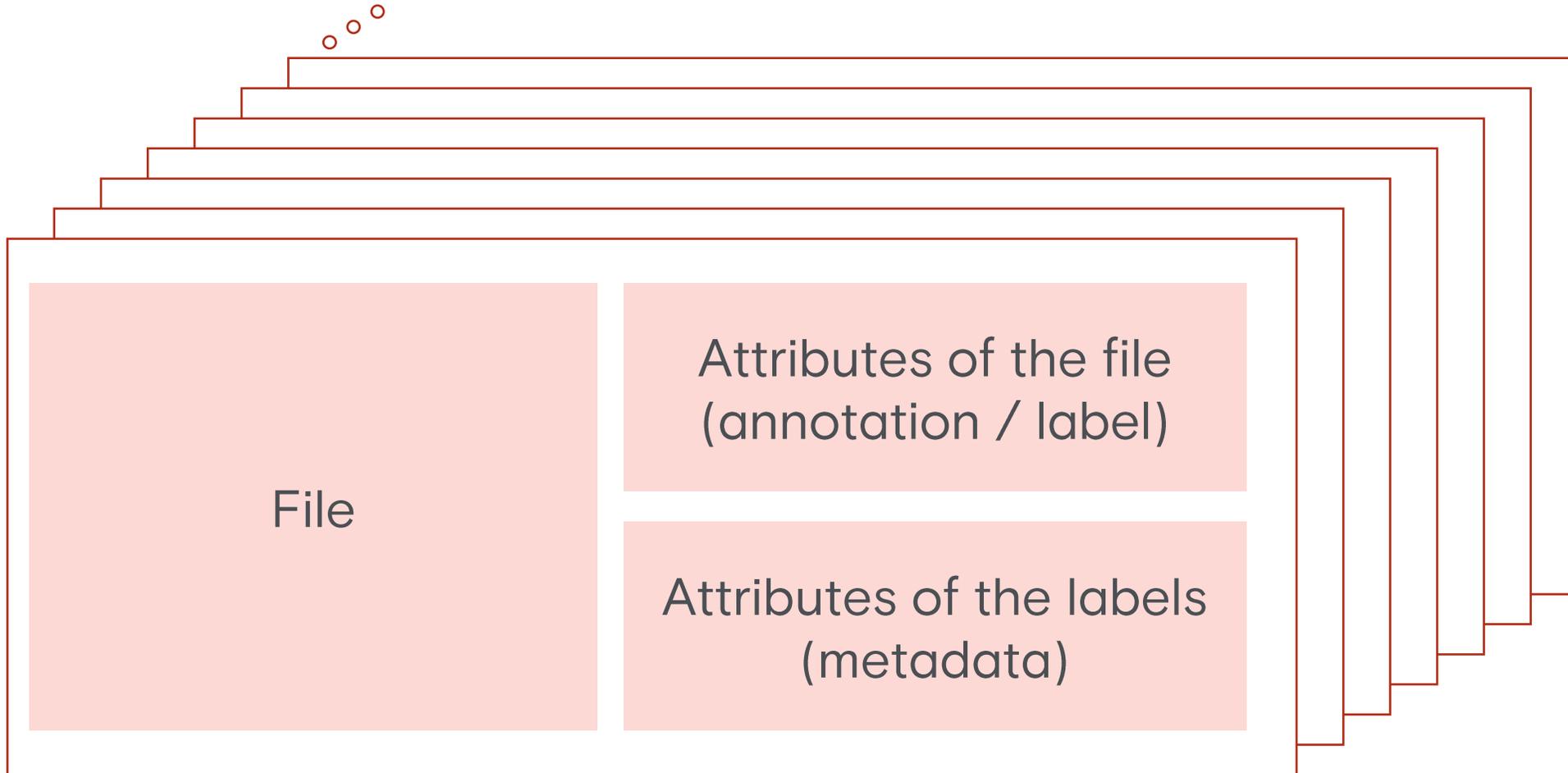
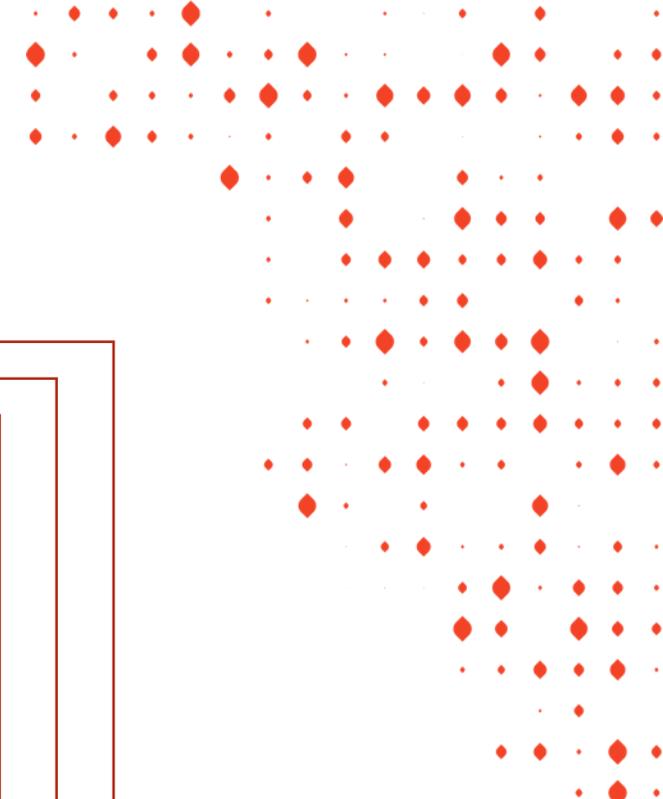
AI model development



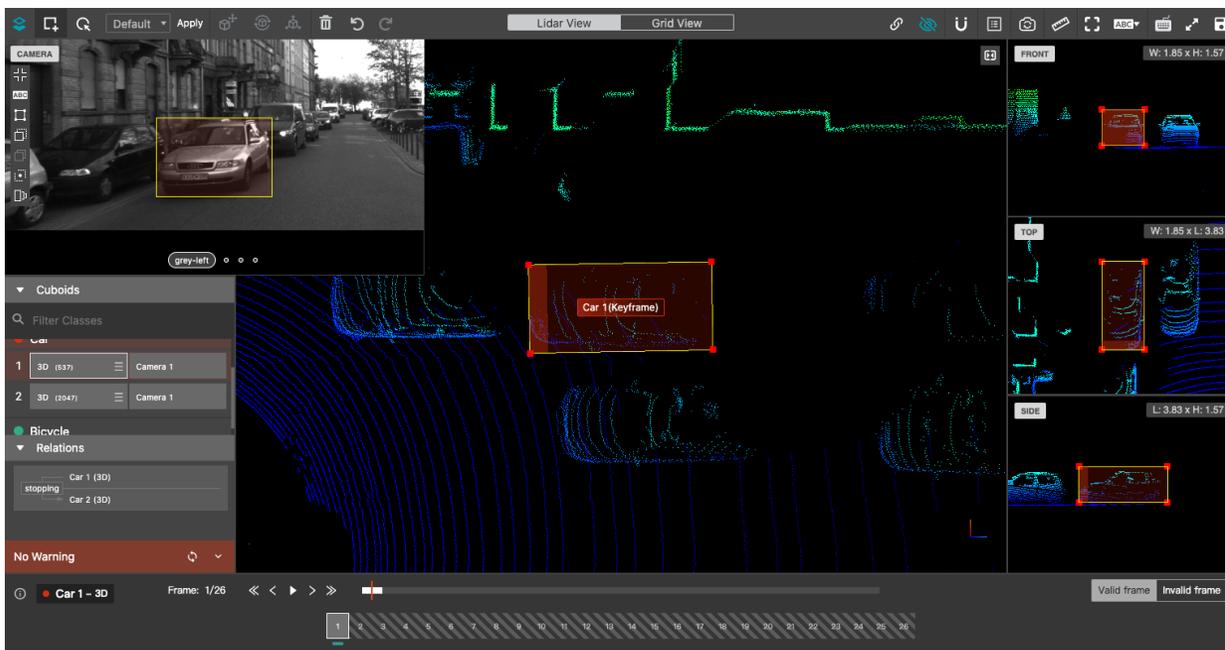
AI models are a series of nodes, weights and biases



Training data are the examples that “teach” the model



LIDAR training data example



"baseUrl": "https://sample-folder.com/",	folder containing point cloud source data
"frames": [list of frames in a scene
"items": [array of objects in the specified frame id
"id": "13f222fd-065c-...",	unique identifier (uuid) of object
"category": "Car",	object class
"number": 8,	object instance number
"position": {	x,y,z position of center of cuboid
"x": 66.49787120373375,	
"y": -37.28758690422451,	
"z": -4.426572264322248	
"rotation": {	rotation of cuboid in radians
"x": 0,	
"y": 0,	
"z": -1.5804235113355598	
"dimension": {	full height, width, depth of a cuboid in meters
"x": 1.86,	
"y": 4.43,	
"z": 1.86	
"locked": null,	not used, please ignore
"interpolated": true,	if the cuboid is a interpolated cuboid and never manually adjusted
"labels": {"size": "small", "occluded": "no"},	Cuboid attribute information
"isEmpty": false,	optional parameter to indicate if the cuboid is an empty cuboid
"pointCount": 120	count of points inside the cuboid
"isValid": true,	To mark a frame as valid or not for labelling
"image": "/image_00/image.png",	file path to image
"items": [array of 2D annotations
"id": "13f222fd-065c-...",	UUID of object, matches UUID of cuboid if it is the same object
"number": 1,	object instance
"category": "Car",	object class
"type": "RECT"	format of 2D annotation, a rectangle in this example
"position": {	top left corner of box
"x": 7.678934984761854,	
"y": 151.025760731091	
"dimension": {	full width and height of the rectangle in pixels
"x": 311.2317472514253,	
"y": 184.7346955567628	
"labels": {"size": "small", "occluded": "no"},	2D annotation attribute information
"isManual": true	the same meaning as the interpolated for 3D cuboid

Speech training data example

The screenshot shows a speech training interface with a vertical timeline on the left and a main content area on the right. The timeline has a play button, a refresh icon, and a progress indicator showing 00:00:15:80 / 00:04:20:84. The main content area contains four segments:

- Speaker B**: "Tell me how you're gonna break" with a noise marker at 1.720.
- Speaker B**: "How are you the schedule?" with a noise marker.
- Speaker B**: "On the actual the day before or the day of?" with a noise marker at 2.043. This segment is highlighted with a red border and includes an audio waveform with a play/pause button and a 00:00:02:90 timer.
- Speaker A**: "The day before, the day a"

```
{
  "id": "5",
  "layerId": "88620127-e109-4c6c-85a1-e92e39f0ff98",
  "startTime": 9.85503125,
  "endTime": 12.013343749999999,
  "ontologyName": "Speaker B",
  "metadata": {
    "transcription": {
      "text": "Tell me how you're gonna <11.575/> break <noise/> ",
      "annotatedBy": "human"
    },
    "original_text": "Tell me how you're going, they're great."
  },
  "nothingToTranscribe": false
},
{
  "id": "6",
  "layerId": "88620127-e109-4c6c-85a1-e92e39f0ff98",
  "startTime": 12.55671875,
  "endTime": 13.54221875,
  "ontologyName": "Speaker B",
  "metadata": {
    "transcription": {
      "text": "How are you <noise/> the schedule?",
      "annotatedBy": "human"
    },
    "original_text": "How are you the schedule?"
  },
  "nothingToTranscribe": false
},
{
  "id": "7",
  "layerId": "88620127-e109-4c6c-85a1-e92e39f0ff98",
  "startTime": 14.50578125,
  "endTime": 17.404906249999996,
  "ontologyName": "Speaker B",
  "metadata": {
    "transcription": {
      "text": "On the actual <noise/> the day before or the day of? <16.549/> ",
      "annotatedBy": "human"
    },
    "original_text": "And after the day before the day of dutiful with me."
  },
  "nothingToTranscribe": false
}
```

Training data quality is important

Low quality training data leads to poor performing models

Poor quality data is not always obvious



Three types of problems we see with training data

1.
Labelling
errors

2.
Unbalanced
training data

3.
Bias in
labelling
process

Task instructions: Draw a box around the cows



Task instructions: Draw a box around the cows



Intended output – a tight bounding box around each of the visible portions of the cows

1. Labelling errors – missing labels



Cow on the right was not labelled

1. Labelling errors – incorrect fit



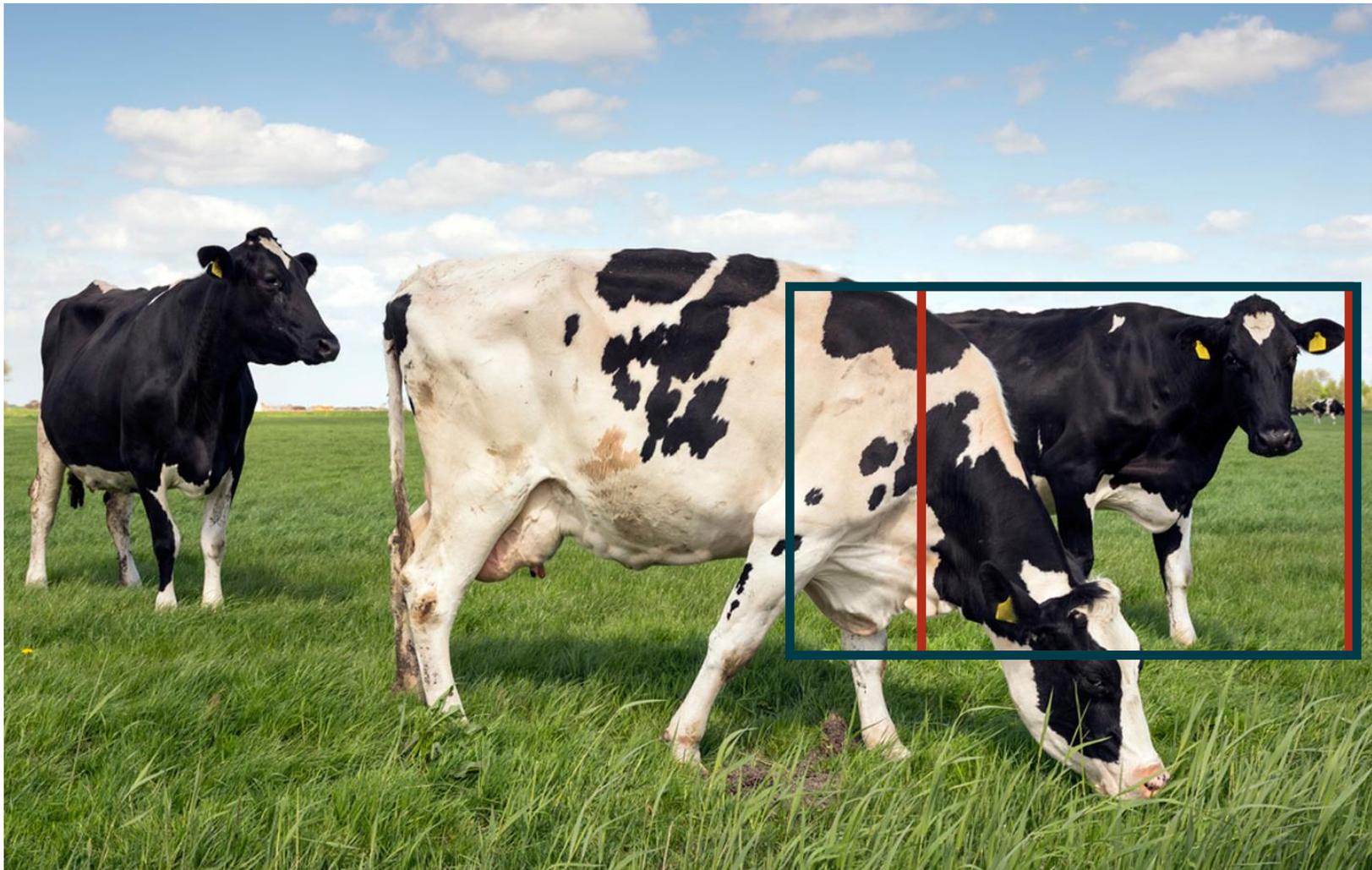
Bounding box is not “tight” enough around the cows

1. Labelling errors – misinterpreting instructions



One large box instead of three

1. Labelling errors – handling occlusion



Box is placed around the expected size of the occluded cow

2. Unbalanced training data – class imbalance



3. Bias in labelling process



Black Pudding



Hagelslag
(Sprinkles on toast)

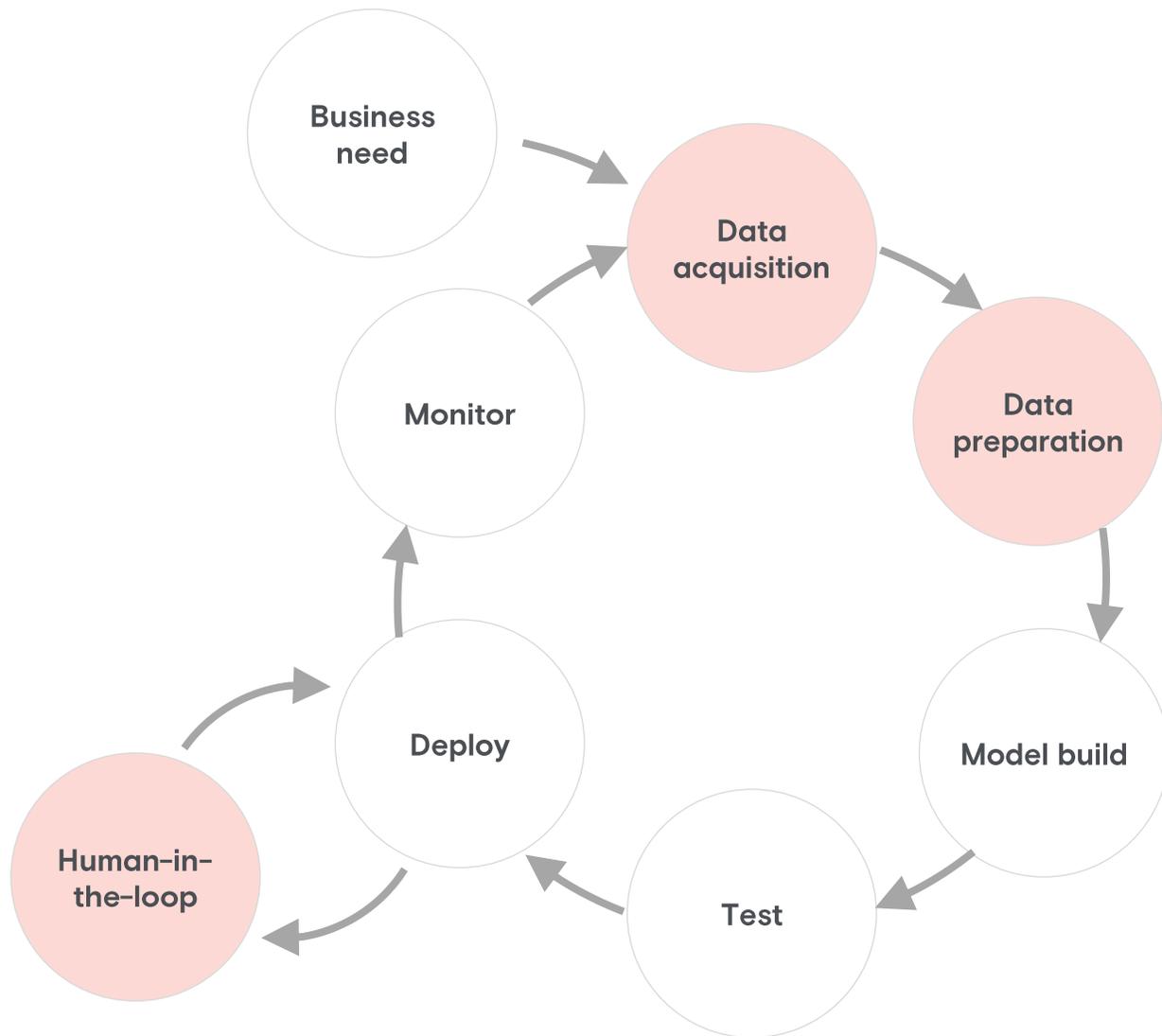


Vegemite

Some datasets require specific knowledge / context for accurate labelling

AI
model = Model
instructions + Training
Data

Good AI model = Model instructions + High Quality Training Data



 Appen focus areas

Our role in the AI lifecycle is to deliver high-quality training data

The AI training data market is rapidly evolving

Five market trends

1 Data remains a major roadblock for AI

2 AI use cases becoming narrower

3 Shift from model-centric to data-centric AI

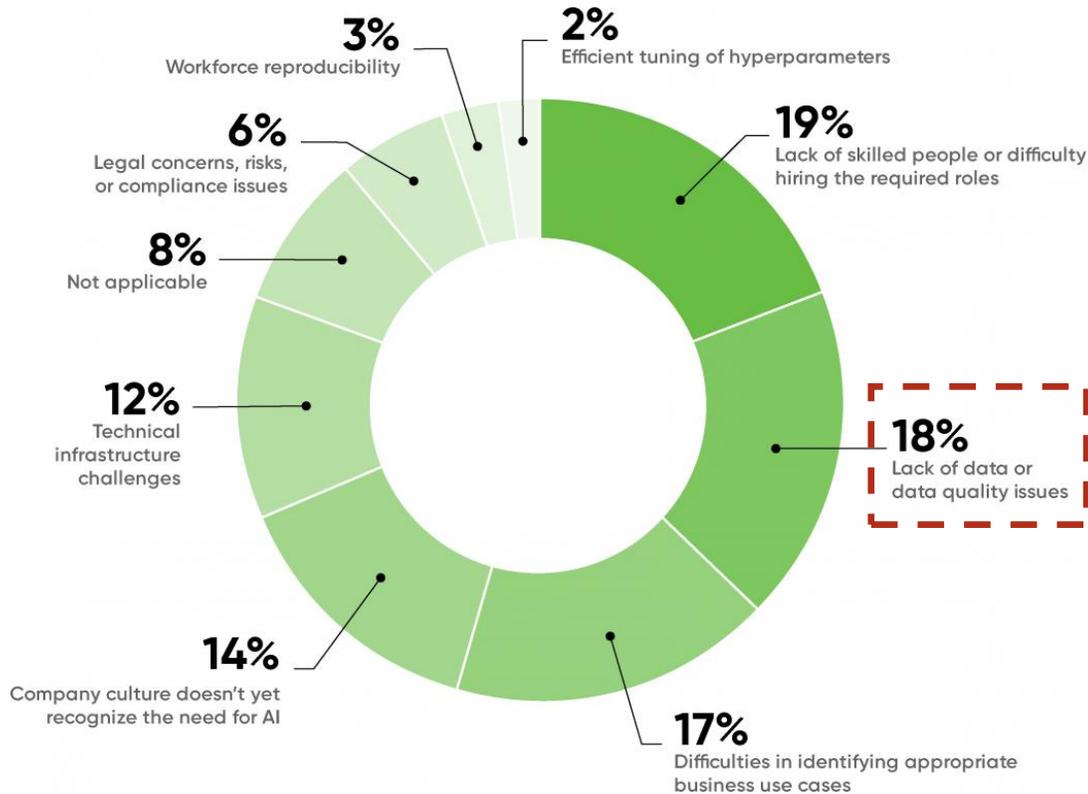
4 Emerging need for training data operations

5 AI assisted annotations increasing

1 Data remains a major obstacle for AI

In a recent survey, 18% of respondents see lack of data or data quality as the major bottleneck to AI

Many AI practitioners spend most of their time collecting and preparing data



*“... Airbnb have discovered that nearly **70% of the time a data scientist spends developing machine learning models – is not doing the actual modelling, but collecting data and feature engineering.**”*

② AI is becoming narrower

Real-world examples we have supported

Biz-speak

- AI model built to suggest improvements to common “biz-speak”
- Challenge: Biz-speak is highly nuanced
- Our task: Suggest alternative phrases for common biz-speak terms

Body movements

- AI models built to automate personal training
- Challenge: Movement profile changes with age
- Our task: capture and annotate video of seniors doing somersaults

Long-tail languages

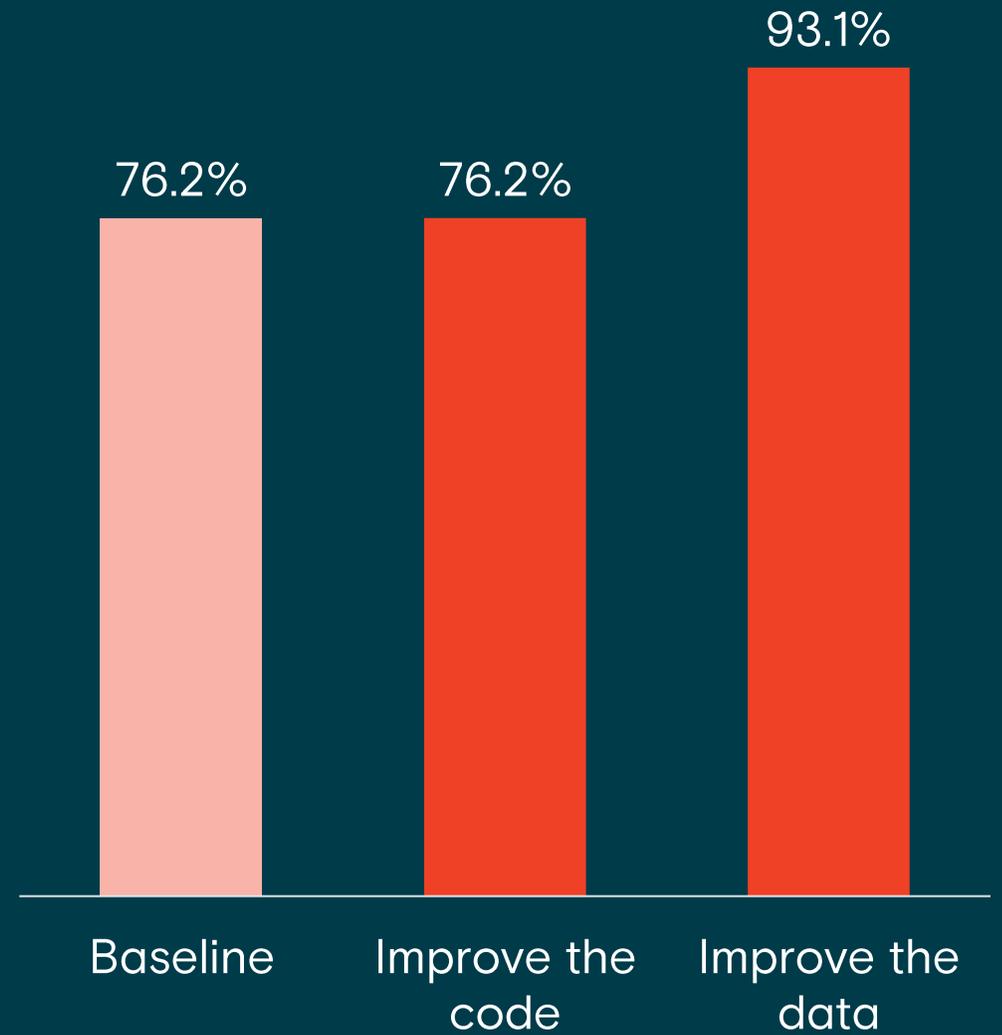
- Covid information needs to be shared globally
- Challenge: Translation tech doesn't support all languages
- Our task: Data collection/annotation for languages including Dari, Dinka, Hausa, Luganda, Pashto, and Zulu

Good AI model = Model instructions + High Quality Training Data

Is it better to improve code or training data?



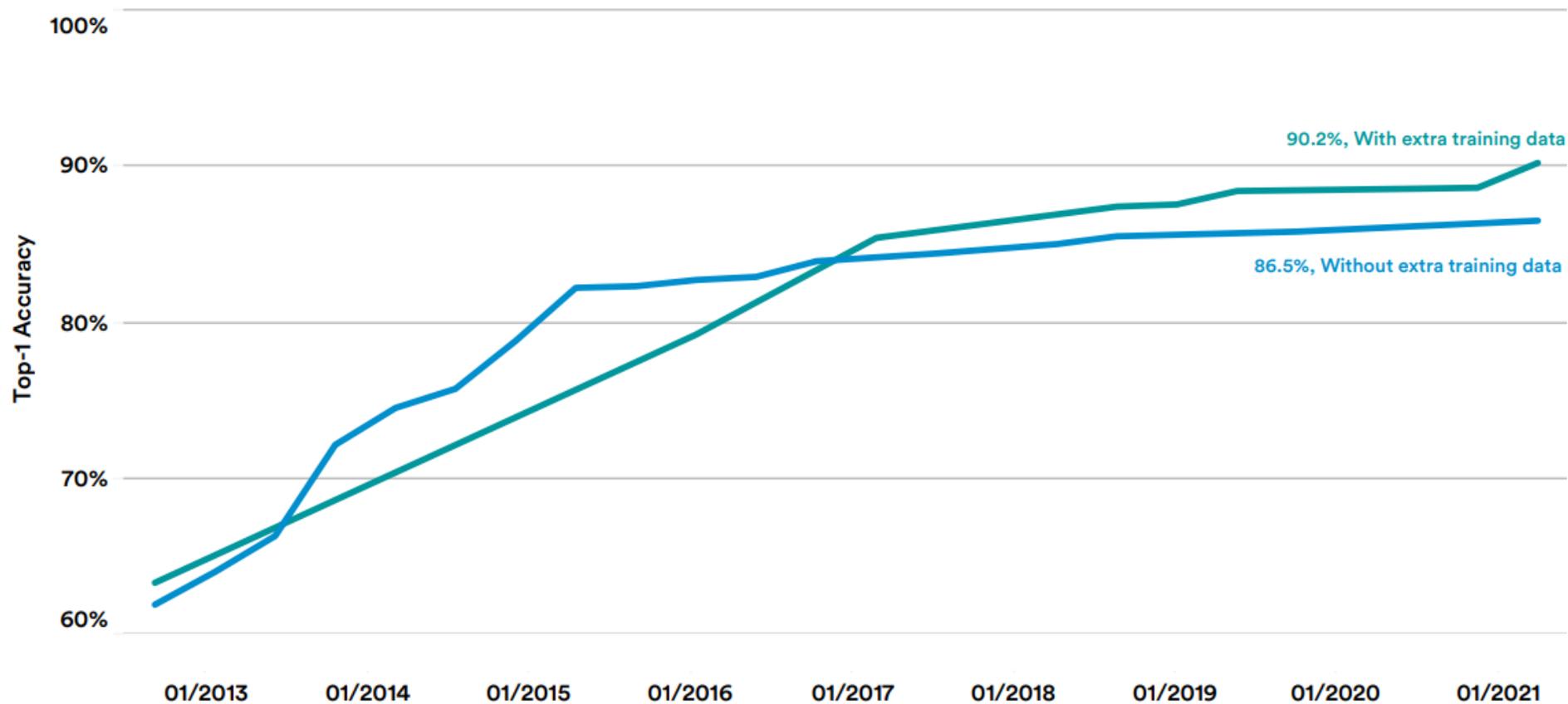
Steel defect detection model



3 ImageNet performance has capped without extra training data

IMAGENET CHALLENGE: TOP-1 ACCURACY

Source: Papers with Code, 2020; AI Index, 2021 | Chart: 2021 AI Index Report



Source: Artificial Intelligence Index Report 2021



③ There is a shift from “model-centric” to “data-centric” AI

Model-centric

Use available data

Develop a model that compensates for noise in the data

Hold the data fixed and interactively improve the model

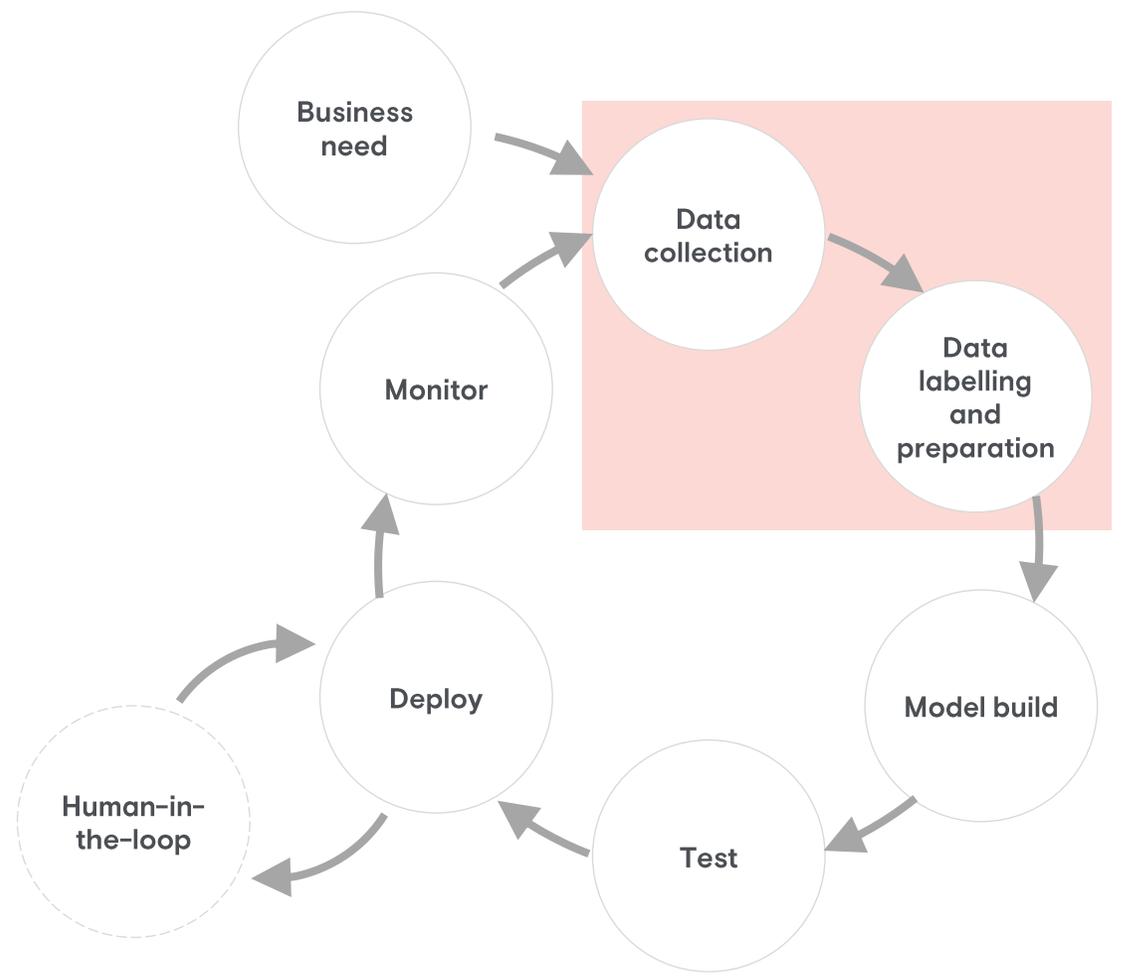
Data-centric

Improve volume and/or quality of the data

Try different models to achieve the best performance

Hold the model fixed and interactively improve the data

4 Need for training data management emerging



- Data collection
- Data ingestion
- Data exploration
- Data labelling
- Data validation
- Data preparation

- Version control
- Training data security
- Access controls
- Data pipeline monitoring
- Collaboration
- Meta-data insights

5 Automation critical to the labelling process



Pre-labelling

- AI performs an initial 'best guess' of the annotation
- Crowd workers check and correct the pre-label (if required)
- Significantly reduces annotation time



Speed labelling

- AI models that assist crowd workers by automating slow and/or manual tasks
- Works similar to an auto-complete function
- Significantly reduces annotation time



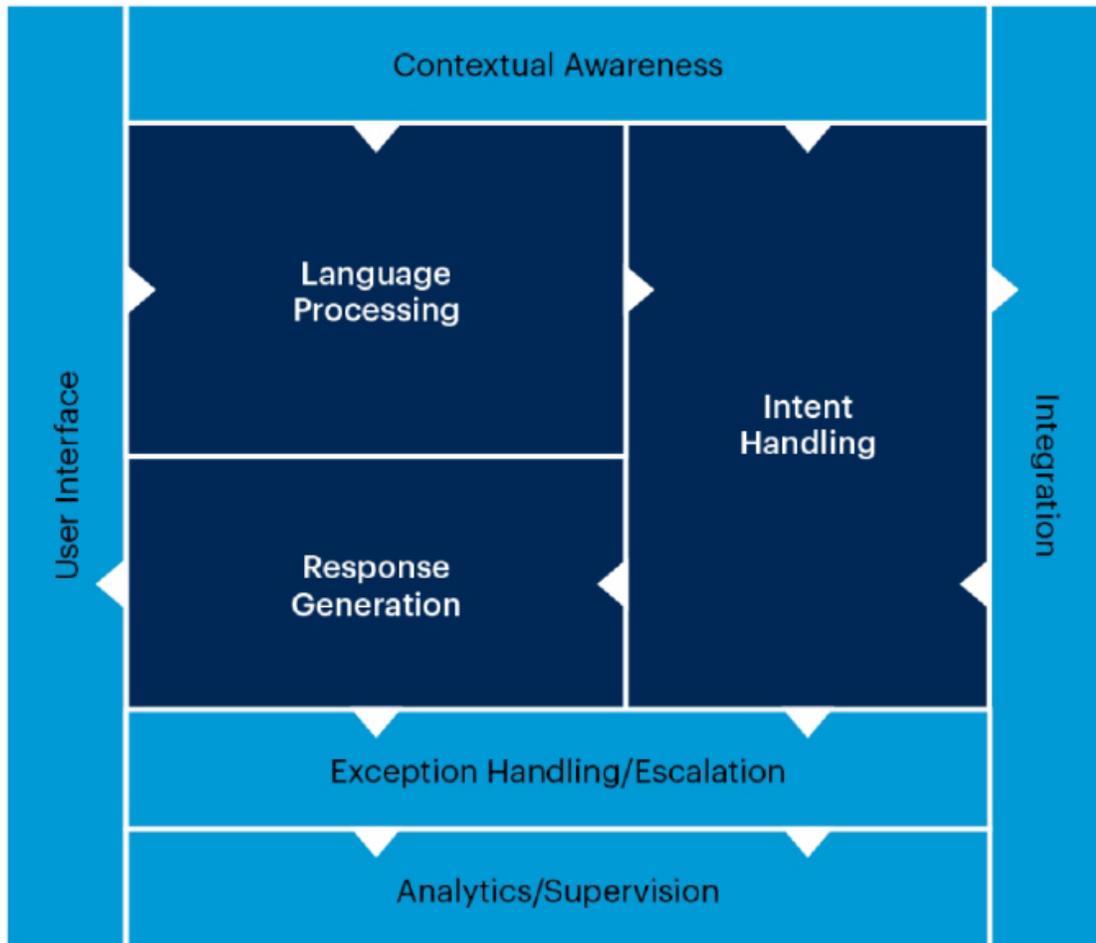
Smart Validators

- AI models that verify crowd output before they are submitted
- Crowd get notified if input is not within expected thresholds
- Improves data quality and overall worker performance

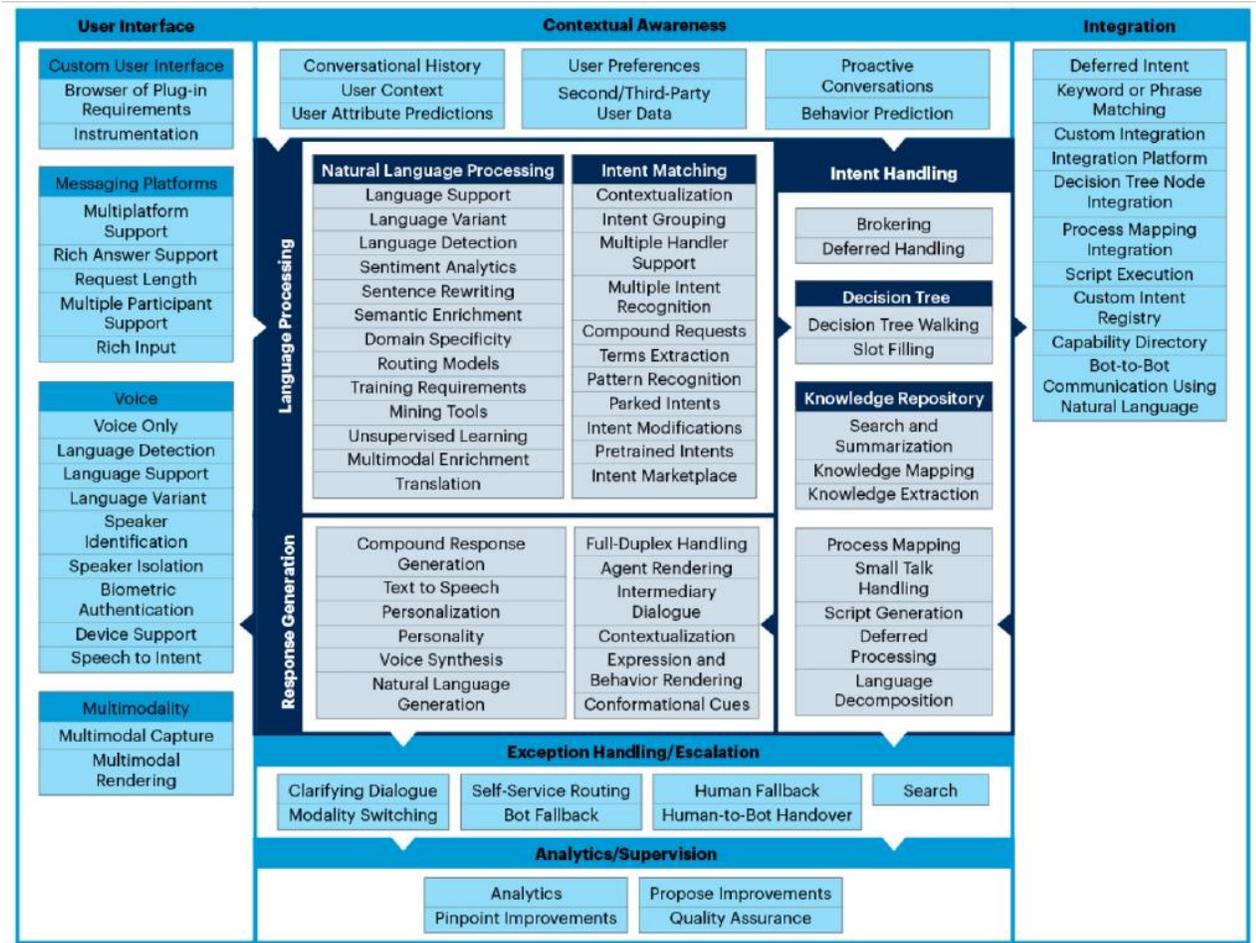
AI models rely on a
variety of training data
techniques

AI enabled applications typically involve many different models

Voice interface architecture

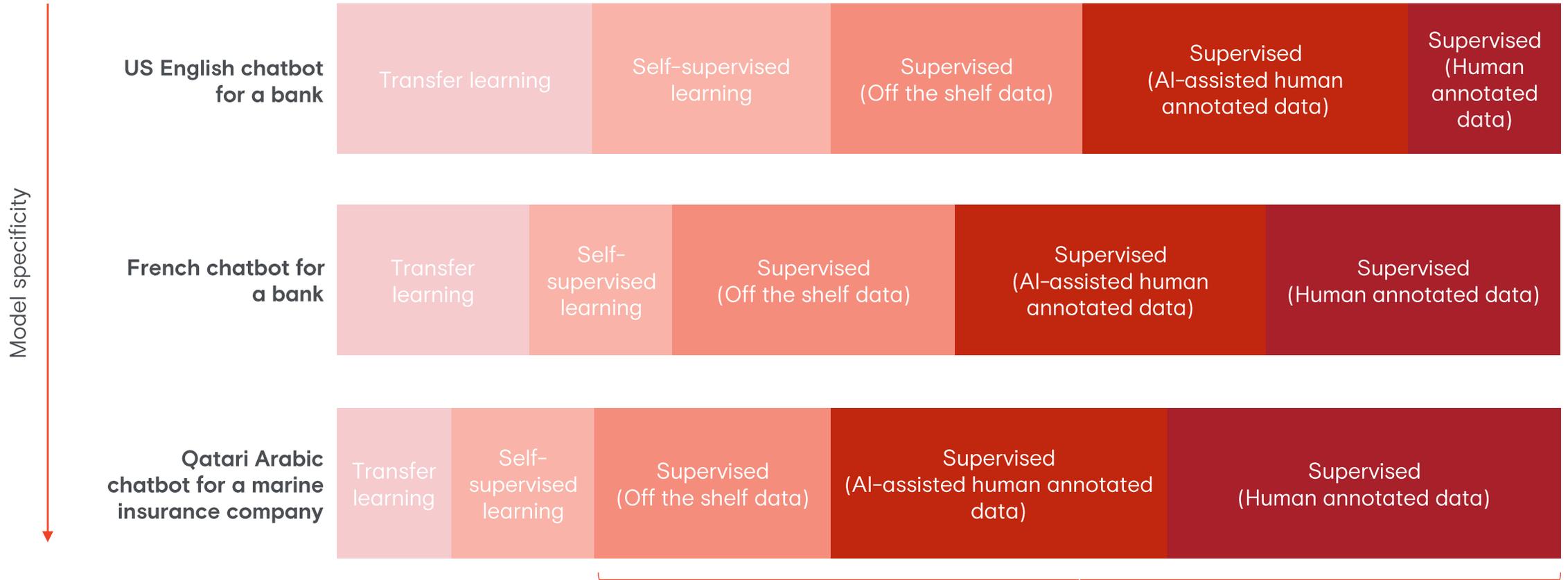


Detailed steps – many models and interfaces



Real-world models typically use a combination of data-sources and techniques

Contribution from different model/data approaches (Illustrative)



Appen AI training data solutions

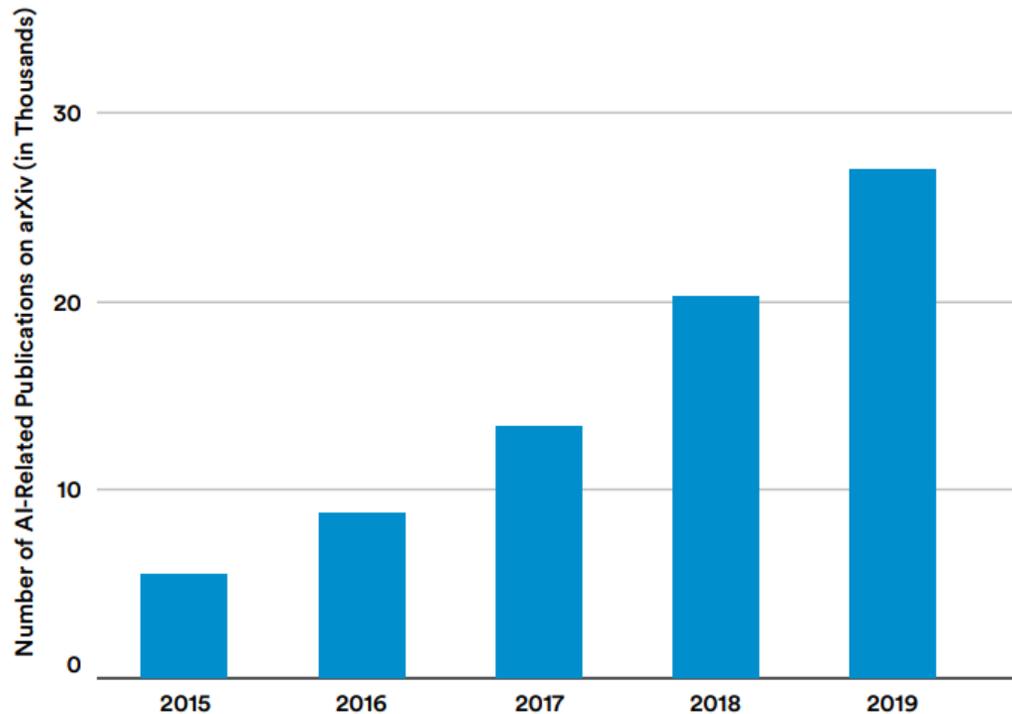
New AI approaches continue to emerge

~30k AI related publications on ARXIV in 2019

Many research teams exploring new areas

NUMBER of AI-RELATED PUBLICATIONS on ARXIV, 2015-20

Source: arXiv, 2020 | Chart: 2021 AI Index Report



Source: Artificial Intelligence Index Report 2021



ALIGN: Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision

Tuesday, May 11, 2021

Posted by Chao Jia and Yinfe Yang, Software Engineers, Google Research

Learning good visual and vision-language representations is critical to solving computer vision problems – image retrieval, image classification, video understanding – and can enable the development of tools and products that change people's daily lives. For example, a good vision-language matching model can help users find the most relevant images given a text description or an image input and help tools such as Google Lens find more fine-grained information about an image.

To learn such representations, current state-of-the-art (SotA) visual and vision-language models rely heavily on curated training datasets that require expert knowledge and extensive labels. For vision applications, representations are mostly learned on large-scale datasets with explicit class labels, such as ImageNet, OpenImages, and JFT-300M. For vision-language applications, popular pre-training datasets, such as Conceptual Captions and Visual Genome Dense Captions, all require non-trivial data collection and cleaning steps, limiting the size of datasets and thus hindering the scale of the trained models. In contrast, natural language processing (NLP) models have achieved SotA performance on GLUE and SuperGLUE benchmarks by utilizing large-scale pre-training on raw text without human labels.

RESEARCH | ML APPLICATIONS

Learning from videos to understand the world

Our project "Learning from Videos" is designed to build AI that automatically learns audio, textual and visual representations from publicly available videos on Facebook. Learn how this will improve AI-powered products – starting with Reels' recommendations:

March 12, 2021

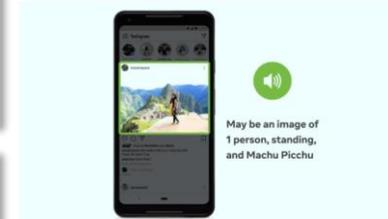
Research in Brief

RESEARCH | ML APPLICATIONS

Large-scale forecasting: Self-supervised learning framework for hyper-parameter tuning

Unlike existing hyperparameter tuning methods, our new self-supervised learning framework for hyperparameter tuning is not search-based. It uses time series features as inputs and produces optimal hyperparameters in 6-20x less time – without sacrificing accuracy.

April 05, 2021



ML APPLICATIONS

How Facebook is using AI to improve photo descriptions for people who are blind or visually impaired

In 2016 we debuted Automatic Alt Text, making Facebook more accessible by using computer vision tech to describe photos for people who are blind or visually impaired.

January 19, 2021

Google Research

PUBLICATIONS

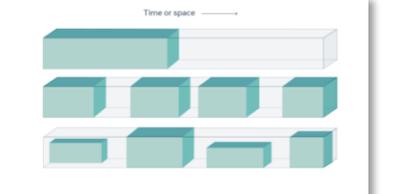
"Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI

Nithya Sambasivan, Shreeya Kasraji, Hannah Hyndriks, Dana Abney, Praveen Kumar Parthasarathy, Lara Moss Allouf, MSRCH, ACM (2021)

Download Google Scholar Copy BibTeX

Abstract

AI models are increasingly applied in high-stakes domains like health and conservation. Data quality carries an elevated significance in high-stakes AI due to its heightened downstream impact. Missing positions like cancer detection, wildlife poaching, and loan allocations. Paradoxically, data is the most under-valued and de-glamorized aspect of AI. In this paper, we report on data practices in high-stakes AI, from interviews with 53 AI practitioners in India, East and West African countries, and USA. We define, identify, and present empirical evidence on Data Cascades – compounding errors during regime, downstream effects from data issues – figured by conventional AI/ML practices that undervalue data quality. Data cascades are pervasive (92% prevalence), invisible, delayed, but often avoidable. We discuss HCI opportunities in designing and incentivizing data excellence as a first-class citizen of AI, resulting in safer and more robust systems for all.



RESEARCH

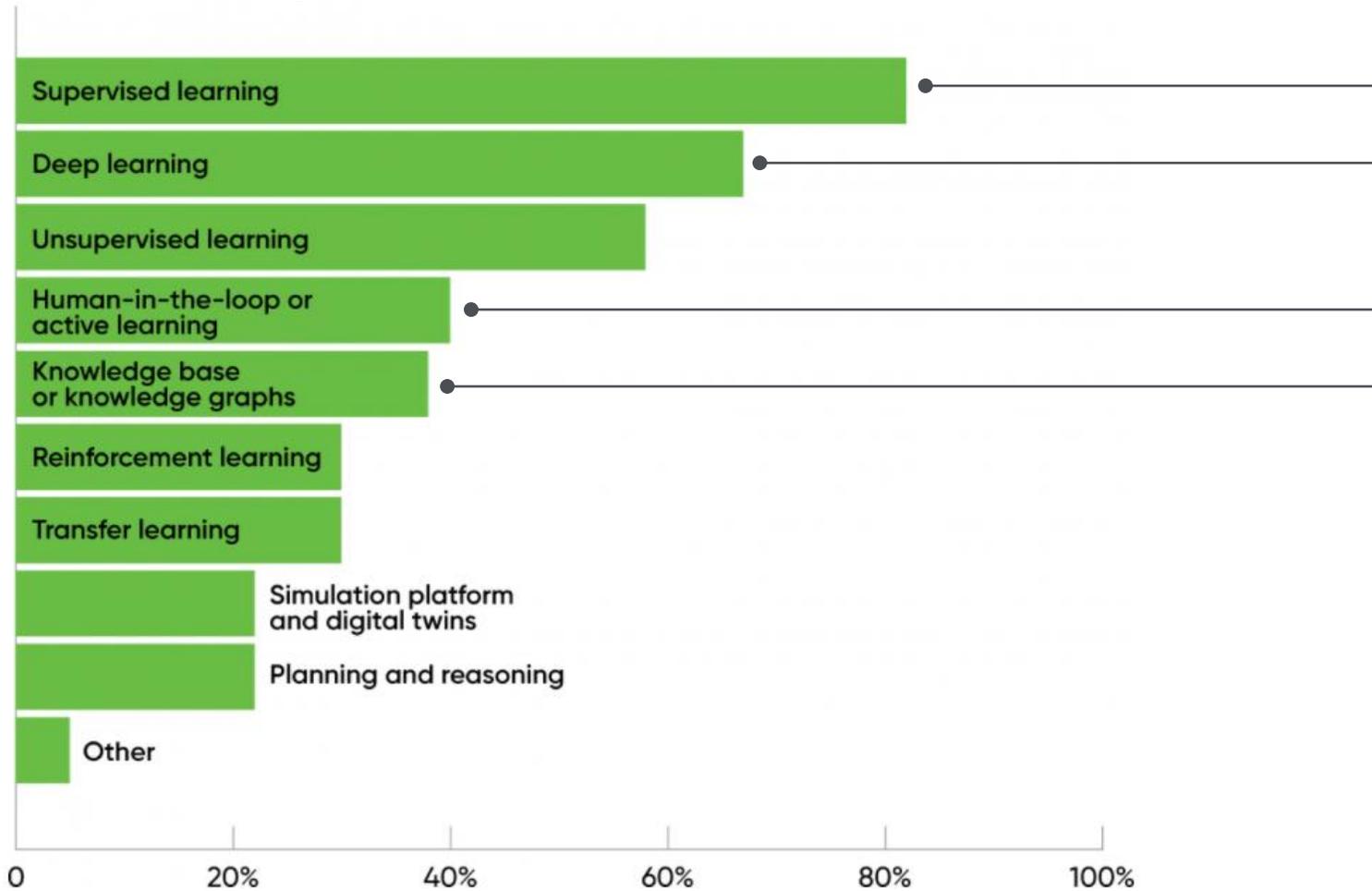
Self-supervised learning: The dark matter of intelligence

How can we build machines with human-level intelligence? There's a limit to how far the field of AI can go with supervised learning alone. Here's why self-supervised learning is one of the most promising ways to make significant progress in AI.

March 04, 2021

Popular AI techniques rely on human involvement

AI technologies used in mature practices



Techniques that typically require some level of human annotation and/or preparation

Source: O'Reilly: AI Adoption in the Enterprise 2021. Survey n = 3,574

Appen is accelerating
its transformation into
an AI powered provider
of AI data and solutions

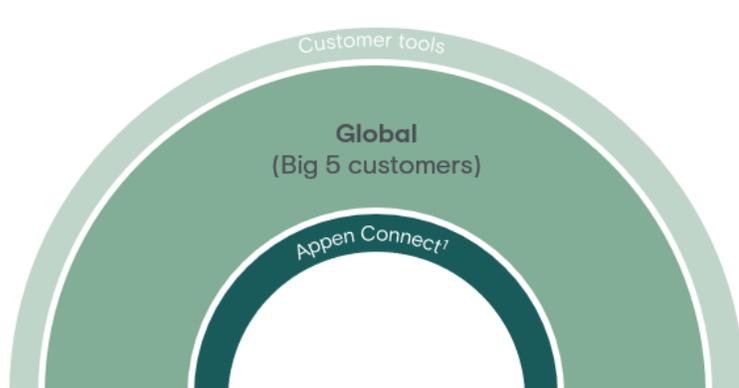
Appen continues to evolve

	From	To
Data type	Language data	AI data
Delivery model	Service led	Product led
Revenue	Project based	Committed
Customer	Concentration	Diversification
Org structure	Functional alignment	Customer alignment
Reporting	Data modality, AUD	Strategy led, USD

Appen's evolution

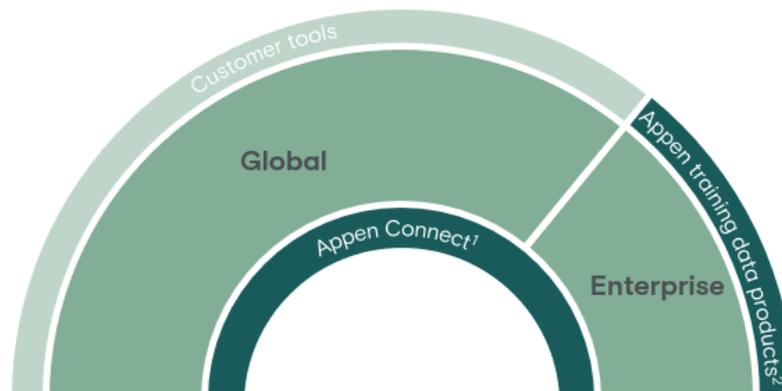
Phase 1: Dominate the AI data annotation services market

Acquired Butler Hill 2010
Acquired Leapforce 2017



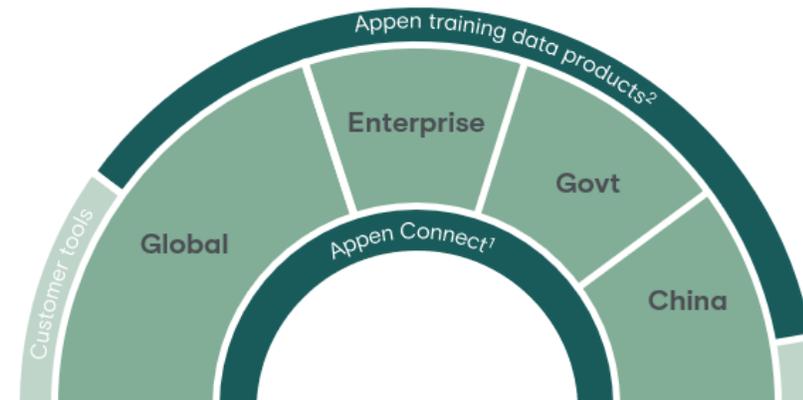
Phase 2: Build leading data annotation products

Acquired Figure Eight 2019
Ongoing product investment
New Enterprise customer base



Phase 3: Product-led expansion

Increased sales and marketing
China and Govt BUs established
Global customers on Appen products



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



The Appen Product Suite



Appen Connect

Match our global crowd workforce to annotation tasks



Appen Data Annotation Platform

Collect and annotate training data



Appen Intelligence

Powers Appen products with proprietary ML models



Appen In-Platform Audit

Organize and analyze training data to identify quality, distribution & bias

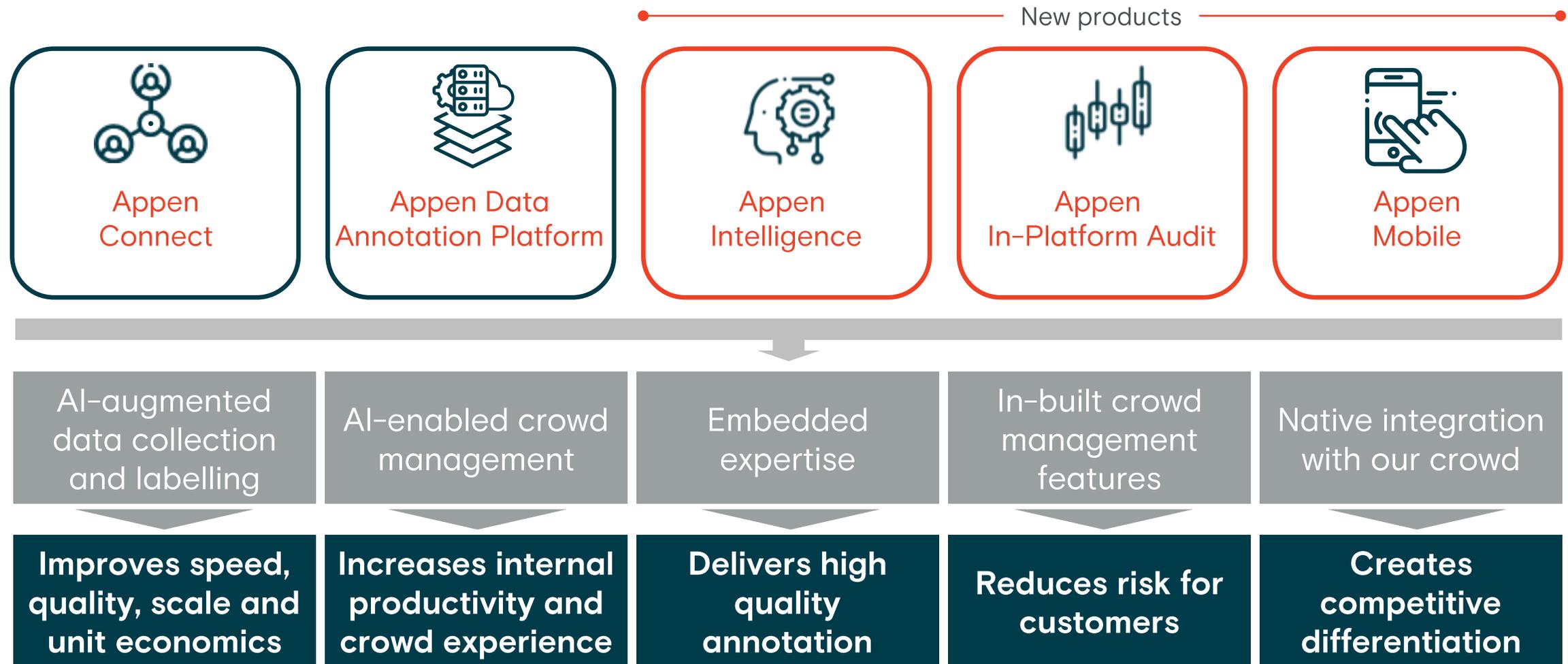


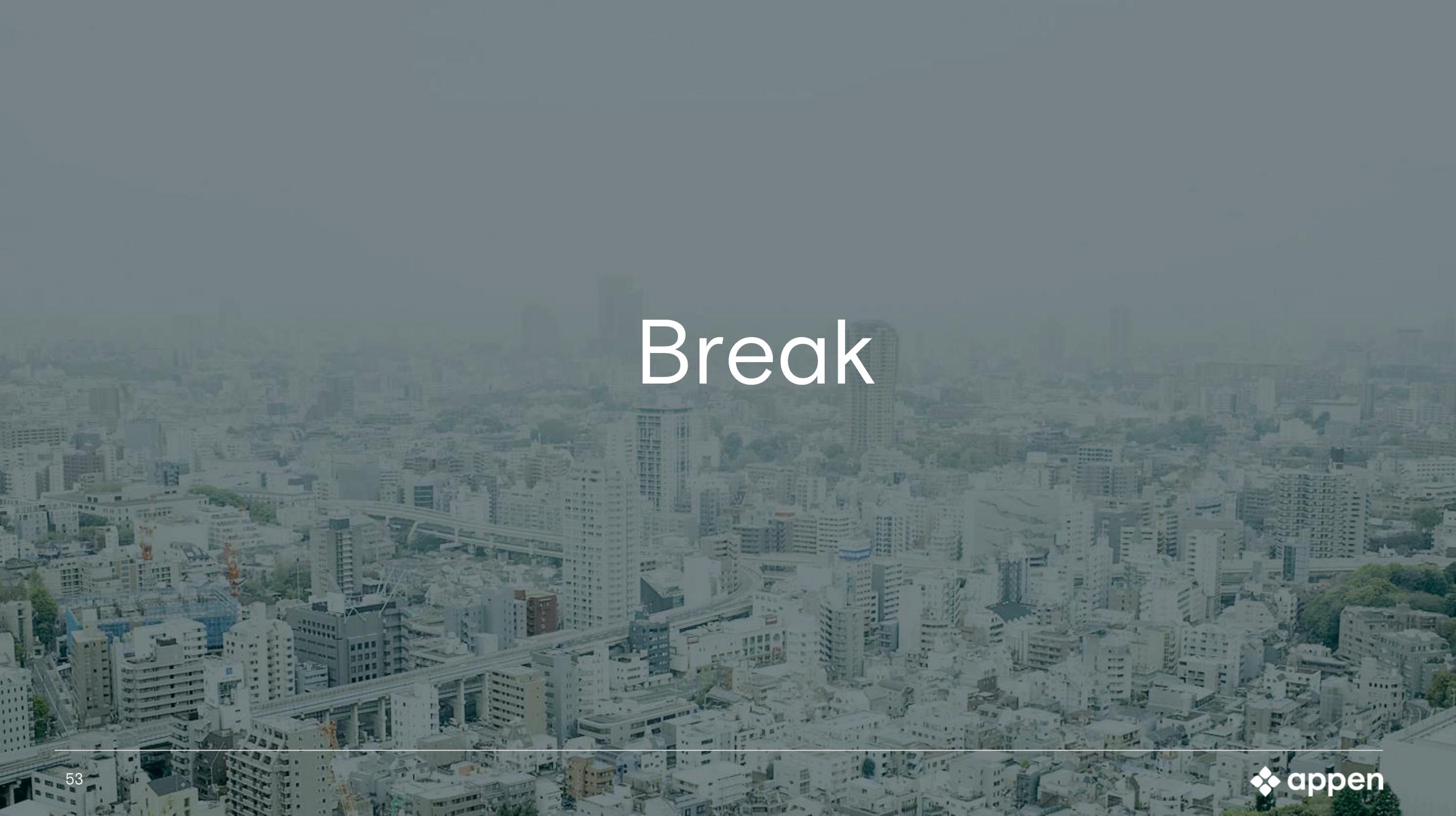
Appen Mobile

Engages, enables and expands crowd

New products

Value created from our products



An aerial photograph of a dense urban area, likely Tokyo, showing a mix of high-rise buildings and lower residential structures. The image is semi-transparent, serving as a background for the text.

Break



Appen Product Suite

Enables the creation of high-quality training data, faster and at scale, using innovative technology

The Appen Product Suite



Appen Connect

Match our global crowd workforce to annotation tasks

- Assigns tasks to our global workforce of 1m+ working in 170+ countries and 235 languages
- Pays 70k+ crowd workers per month
- Quality management and fraud detection



Appen Data Annotation Platform

Collect and annotate training data

- Full suite of data collection and annotation tools
- AI-enabled/automation capabilities
- Ability to support complex, multi-step annotations
- Wide array of data security options



Appen Intelligence

Empower Appen products with proprietary ML models

- Over 20 proprietary models to improve crowd productivity and data labelling quality
- Used extensively to automate internal crowd management tasks



Appen In-Platform Audit

Organize and analyze training data to identify quality, distribution & bias

- Enables our customers to navigate and troubleshoot training data



Appen Mobile

Engages, enables and expands crowd

- Simple and intuitive user experience
- Crowd can interact with Appen anywhere and anytime
- Reaches people who don't have access to a desktop or prefer mobile
- Enables location-based data collection

New products

Appen Connect



Appen Connect

Match our global crowd workforce to annotation tasks



Appen Data Annotation Platform

Collect and annotate training data



Appen Intelligence

Powers Appen products with proprietary ML models



Appen In-Platform Audit

Organize and analyze training data to identify quality, distribution & bias



Appen Mobile

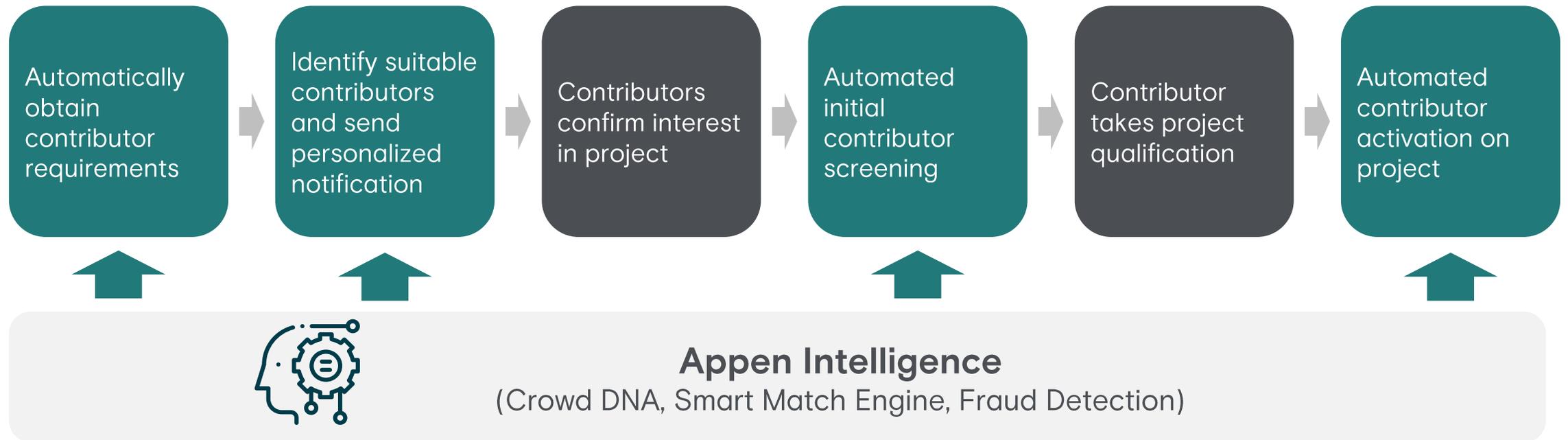
Engages, enables and expands crowd

Appen Connect – an intelligent marketplace to match our crowd with customer tasks



Highly automated process to match contributors to projects

Process to match crowd contributors to projects

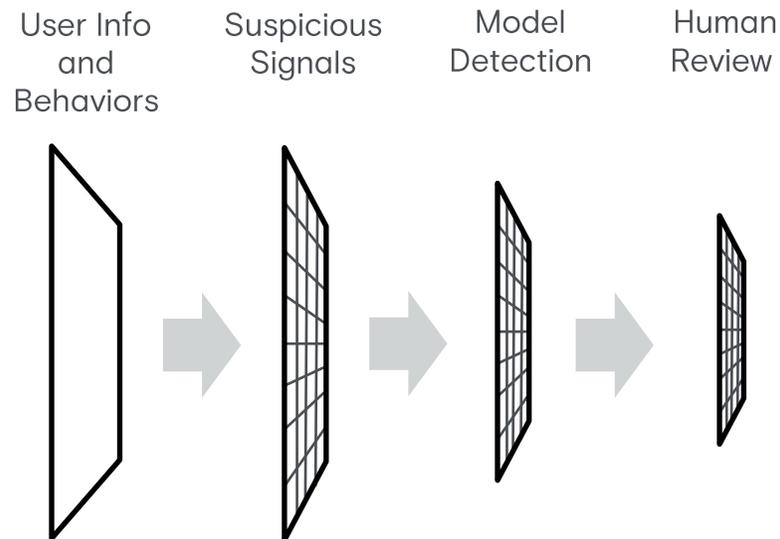


Automatic identification of fraudulent contributors

Example

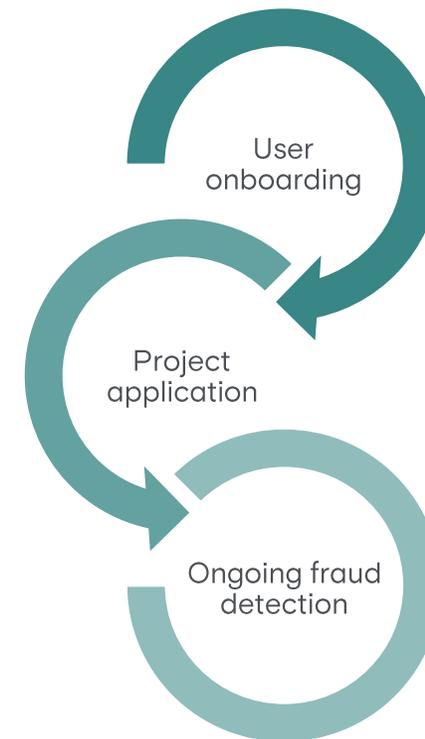
Account	Country	IP	Active Time
10029xx	Brazil	10.21.368.xx	8am-6pm
40483xx	US	10.21.368.xx	3am-5am

Methodology



- Processes 1 million+ profiles every day
- Leverages 200+ signals
- Results in ~95% precision

Scenarios





Benefits for our customers

- Connects customers to our global crowd of contributors
- Reduces project overhead cost by automating internal effort
- Supports future growth with scalability enabled by AI

Future direction

- Continued optimisation of user experience
- Further automation of project management effort

Appen Data Annotation Platform



Appen Connect

Match our global crowd workforce to annotation tasks



Appen Data Annotation Platform

Collect and annotate training data



Appen Intelligence

Powers Appen products with proprietary ML models



Appen In-Platform Audit

Organize and analyze training data to identify quality, distribution & bias



Appen Mobile

Engages, enables and expands crowd

Our annotation tools support a wide variety of AI use cases



Data Collection

Telephony

Studio recording

In-car recording



Content Relevance

Search relevance

Sentiment analysis

Data categorization

Data validation



Audio & Speech

Voice transcription

Translation



Text & Language

Name entity extraction

Text relationship



Image & Video

Image transcription

Image annotation

Pixel level semantic segmentation

Video annotation



3D Point Cloud

LiDAR annotation

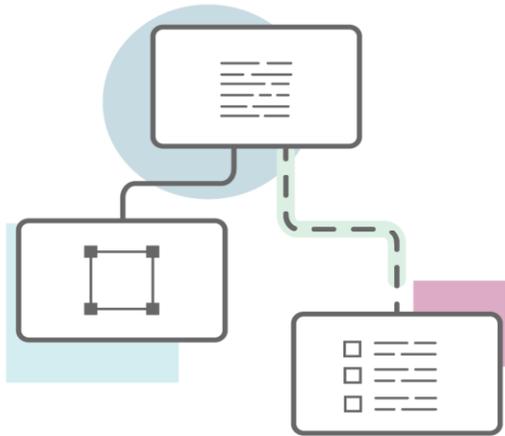
LiDAR semantic annotation



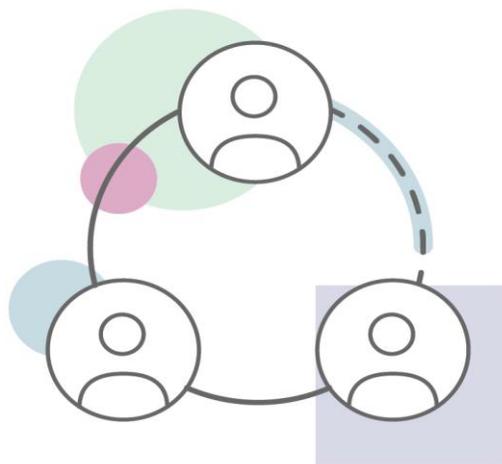
Job Designer / CML (Custom Markup Language)

Appen Workflows enables complicated AI data use cases

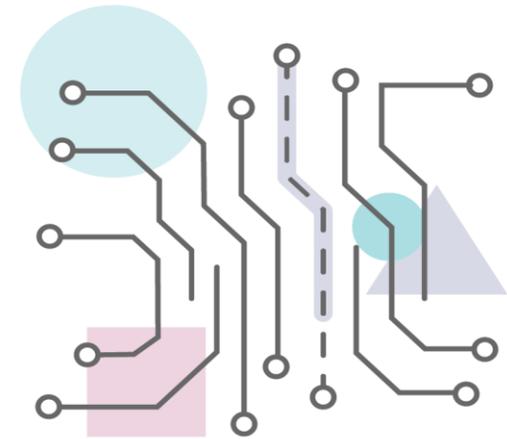
Simple user interface to build and automate multi-step data annotation projects



Simple user interface



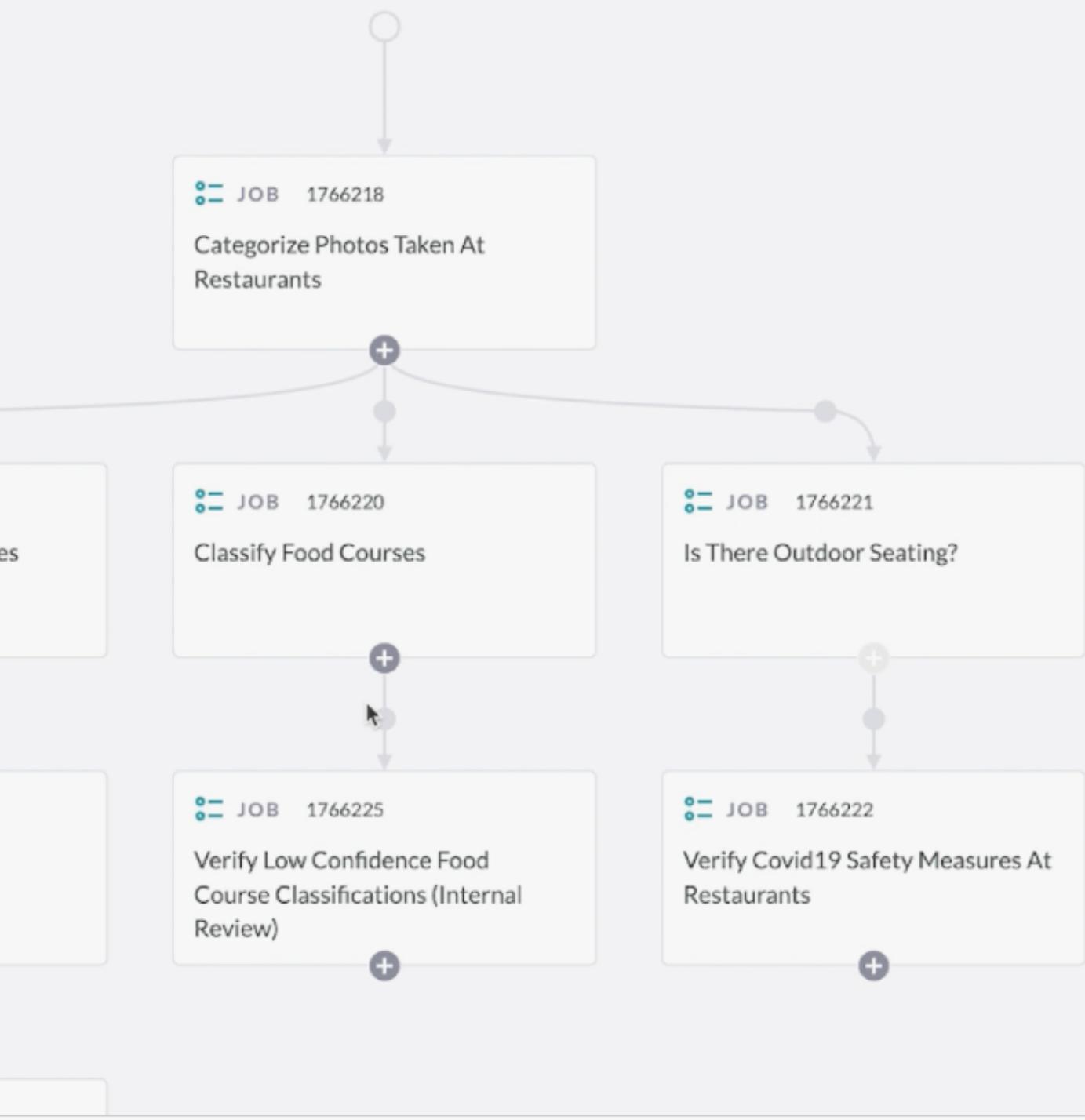
Human & Machine
Intelligence



Flexible settings

Demo 1

Appen Workflows



Built-in quality controls

Test Questions

Ground truth that tests contributors' understanding of a job both before and while they work

Configurable Test Questions

Enable Quiz Mode

Keep total questions same as Rows per Page in Job

TOTAL TEST QUESTIONS IN QUIZ
5 Test Questions

Enable Work Mode

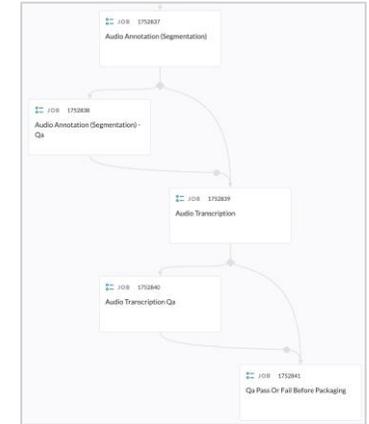
Enable Test Questions in Work Mode

FREQUENCY OF TEST QUESTIONS
1 Test Questions every 10 Rows

Save & Close Cancel

Quality Assurance

Highly qualified contributors review and correct annotations within the workflow



Dynamic Judgments

Leverage majority vote, achieve high confidence annotations while minimizing the number of required judgments per unit

Minimum Confidence

We'll stop collecting judgments on a row when the row's confidence score is above this value or when max judgments is reached

0.8

Recommended: at least 0.7

ML Validation

Leverage machine learning model predictions to validate human results

Appen ASR - (en-uk) Speech Rec **PRETRAINED**

Appen ASR - (en-us) Speech Rec **PRETRAINED**

Blur Faces in Images
Given a source image, return the image with pictured faces blurred. **PRETRAINED**

Blur Faces in Videos
Given a source video, return the video with pictured faces blurred. **PRETRAINED**

Box and Transcribe Words
Given a box around a line of text in an image, box and transcribe the words within the line of text. **PRETRAINED**

Exif Rotation
Rotate image based on Exif tag. **PRETRAINED**

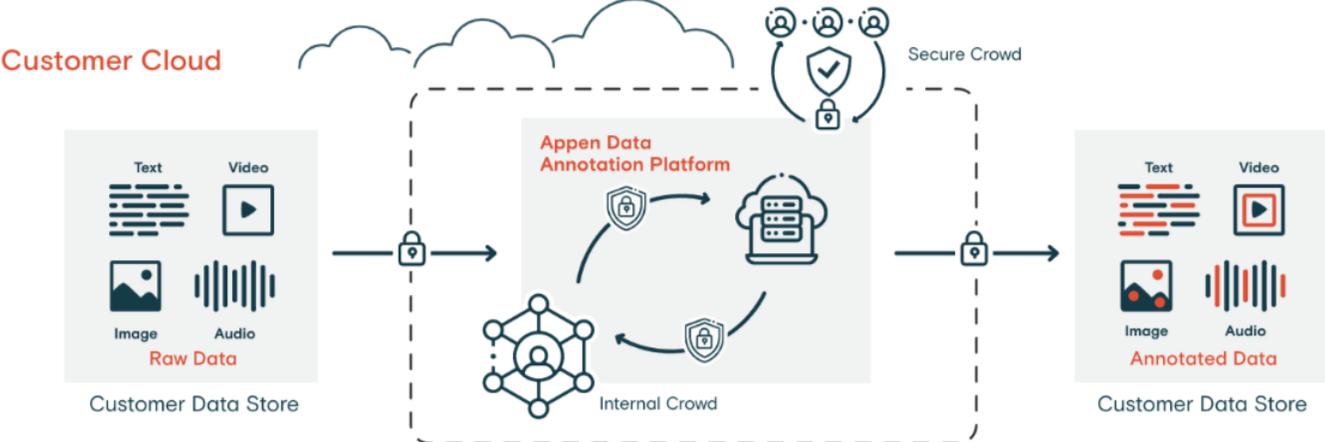
Four Point Transform
Transform image based on four point method. **PRETRAINED**

Identify Face Landmarks
Predict the locations of key points within faces pictured in the image. **PRETRAINED**

Label Bounding Boxes in Street Scene Images
Generate instance segmentation bounding boxes with labels for street objects. **PRETRAINED**

Industry leading security solutions

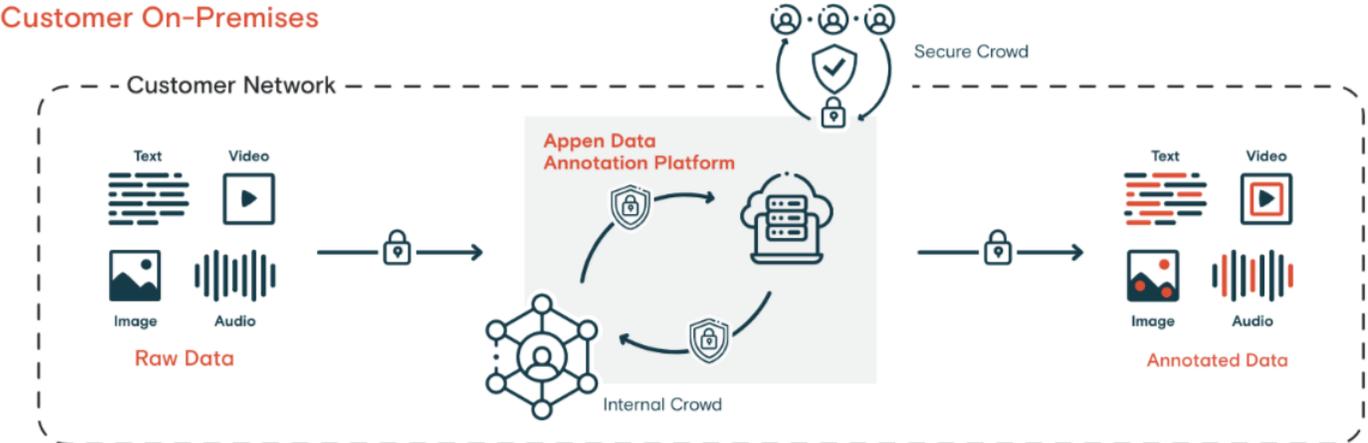
Flexible Deployment Options



Compliance & Certification

Comply with legal obligations related to data security and privacy to protect the health, safety and welfare of customers.

Standards: SOC 2, GDPR, HIPAA



Secure Data Access

Customers can use secure data buckets (s3 and azure) to completely own and govern their own data. Appen Data Annotation Platform only renders the data when contributor works on the annotation.

Benefits for our customers

- Full suite of data annotation tools and flexibility to customise tasks
- Appen Workflows supports complex, multi-step annotation
- Built-in quality control to guarantee high quality
- Security options to protect customer data

Future direction

- Ongoing evolution of annotation tools
- Tighter integration with customers' internal systems
- Improved Workflows features to support more complex annotation tasks
- Ongoing evolution of quality and security features

Appen Intelligence



Appen Connect

Match our global crowd workforce to annotation tasks



Appen Data Annotation Platform

Collect and annotate training data



Appen Intelligence

Powers Appen products with proprietary ML models



Appen In-Platform Audit

Organize and analyze training data to identify quality, distribution & bias



Appen Mobile

Engages, enables and expands crowd

Appen Intelligence – automated pre-labelling and quality control

Audio & Speech



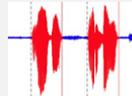
Speaker diarization



Language Detection



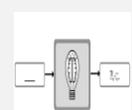
Speech Recognition



Audio Segmentation



Gender Detection



Punctuation Restoration

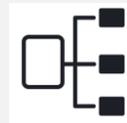
Text & Language



Text Tokenization



Gibberish Detection



Text Classification



Name Entity Extraction



Grammar & Spell Checking



Computer-Aided Translation

Image & Video



OCR Transcription



Object Detection



Object Tracking



Semantic Segmentation



Blurring



Face Landmark

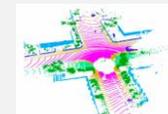
3D Point Cloud



3d Object Detection



3d Object Tracking



3d Semantic Segmentation



3d Lane Detection

Demo 2

Optical Character Recognition (OCR) Transcription

wordbox 1
CONTRACT FOR THE SALE OF REAL ESTATE.

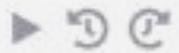
THIS MEMORANDUM OF AGREEMENT made and entered into this 1st day of May, 1947, by and between A. C. Dunning, party of the first part, and George Thelen, party of the second part, both of the County of Douglas, Nebraska, WITNESSETH:

That in consideration of the sum of EIGHT THOUSAND DOLLARS (\$8,000.00) - - - - - the first part hereby agrees to sell and convey to the party of the second part, the real estate located in Shelby, Nebraska, known as the Hotel Building and Barber Shop, located on a part of Lot Eight (8), and probably Lot Nine (9), in Block (8), Original Town of Shelby, Nebraska, and the party of the second part hereby agrees to purchase said above described real estate and pay the sum of \$8,000.00 for the same as follows:

\$1,000.00 at and before the ensealing and delivery of these presents, the receipt whereof is acknowledged by first party and a balance of \$7,000.00 in cash upon the approval of the abstract of title.

It is assumed that the abstract can be submitted and examined on or before September 1st, 1947. If the title is approved prior to that date, the balance of the consideration shall be payable upon the approval of the title.

May 20, 2021



00:00:02:83 / 00:04:20:84

● Speaker B



Play/Pause



00:00:03:42

So what would you like to talk about today?

● Speaker A

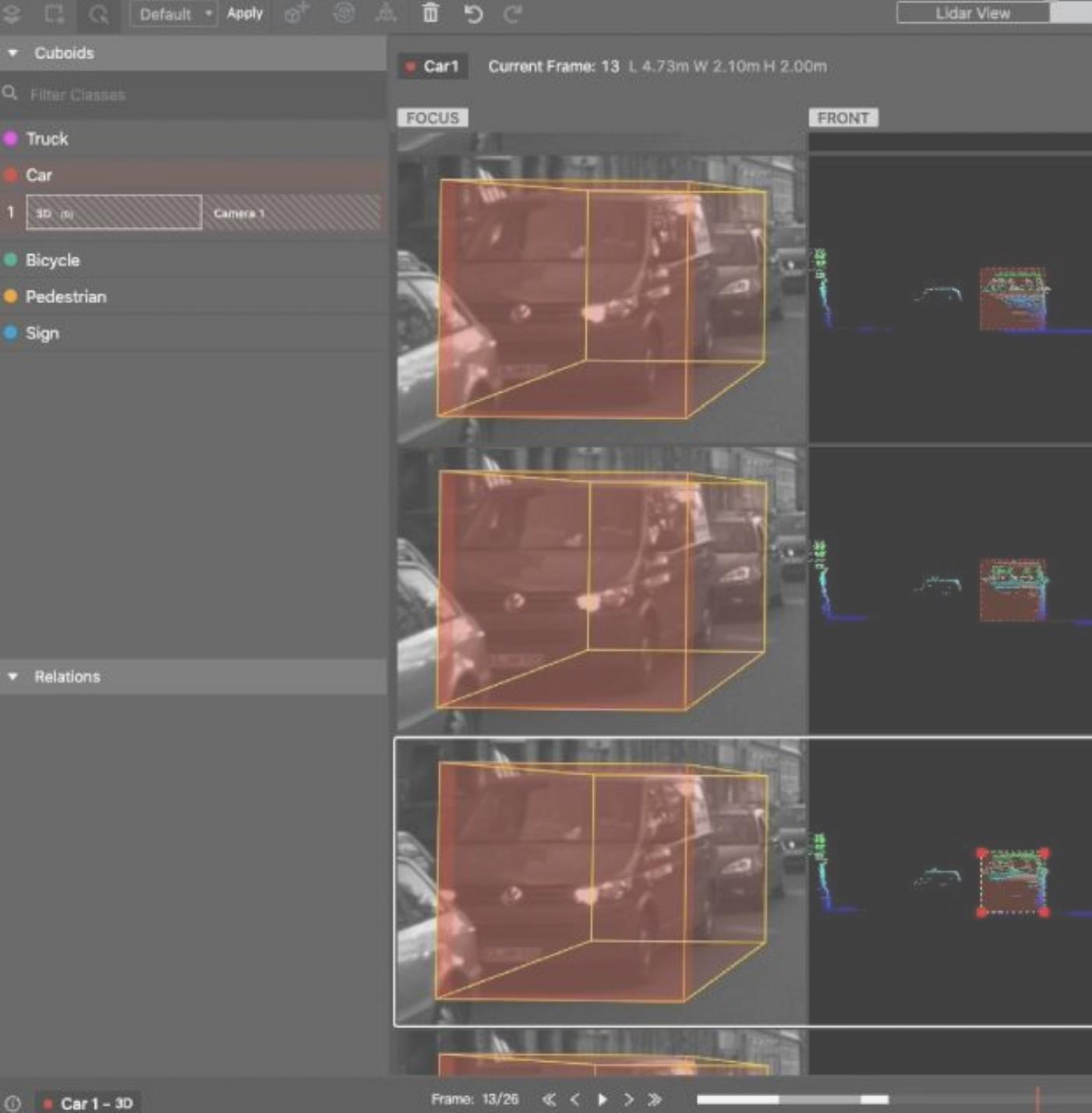
I think we're gonna talk about reading.

● Speaker B

Okay.

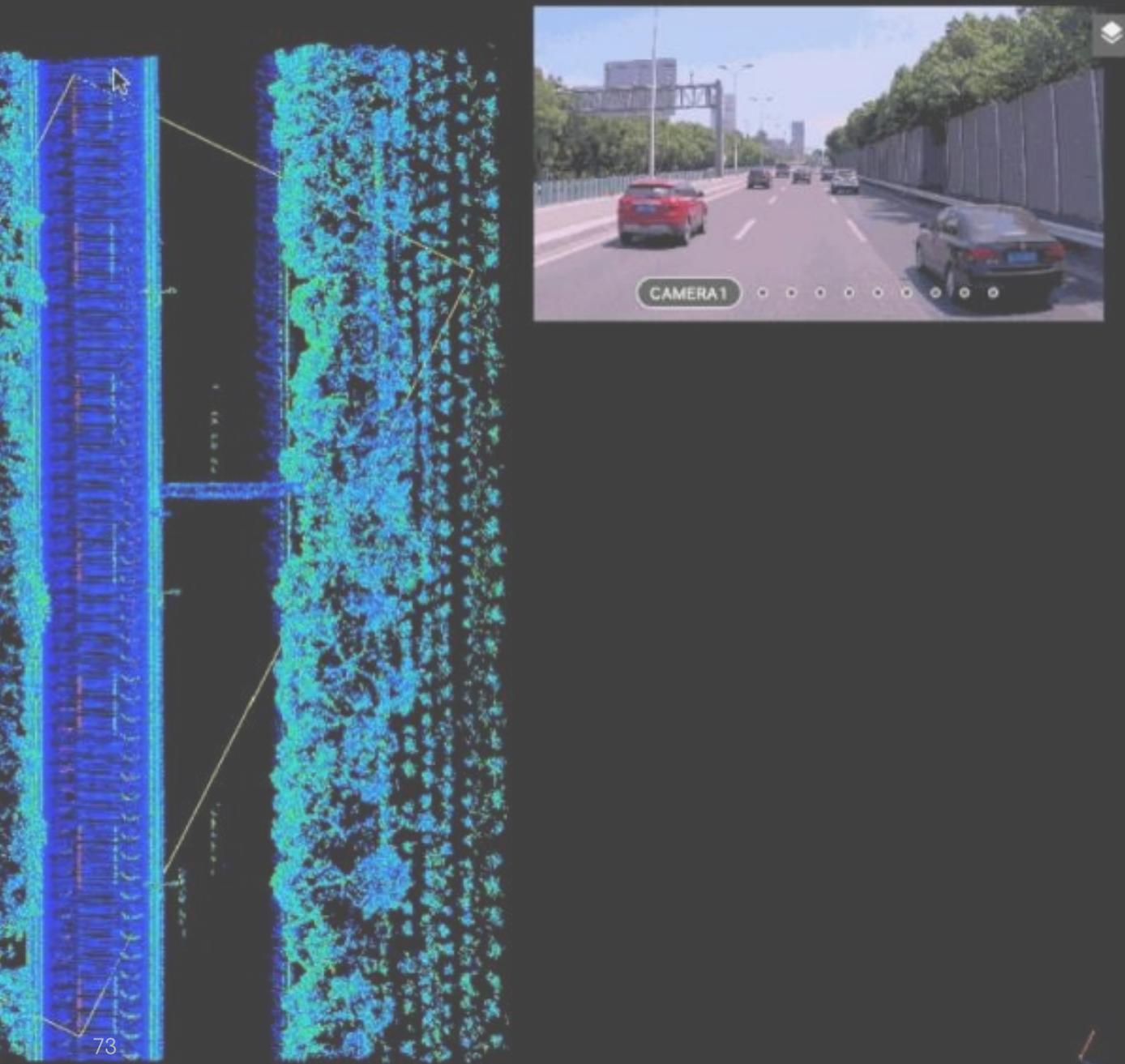
Demo 3

Audio Transcription



Demo 4

LiDAR Annotation



Demo 5

LiDAR Lane Line Segmentation

ML assisted annotation improves annotation time by up to 600%

		ML assisted annotation time improvement vs human annotation
Audio & Speech	Audio Transcription	Up to 90%
	Translation	Up to 160%
Image & Video	2D Image Bounding Box	Up to 30%
	2D Image Semantic Segmentation	Up to 90%
	Video Annotation	Up to 500%
	OCR Transcription	Up to 600%
3D Annotation	3D Point Cloud Bounding Box	Up to 400%
	3D Semantic Segmentation	Up to 600%
Content Relevance	Content Relevance	Tasks are subjective and difficult to automate

Note: These are results observed from A/B testing and certain production projects. The results can vary based on the actual project data.

Benefits for our customers

- Improved speed of annotation enables customers to obtain more data at lower unit cost
- Quality is typically improved compared to standalone human annotation
- Faster time to obtain training data shortens model iteration cycles
- Project management overhead is saved and crowd is managed by platform

Future direction

- Expansion of models to support a wider variety of automation use cases
- Tighter integration with other products

Appen In-Platform Audit



Appen Connect

Match our global crowd workforce to annotation tasks



Appen Data Annotation Platform

Collect and annotate training data



Appen Intelligence

Powers Appen products with proprietary ML models



Appen In-Platform Audit

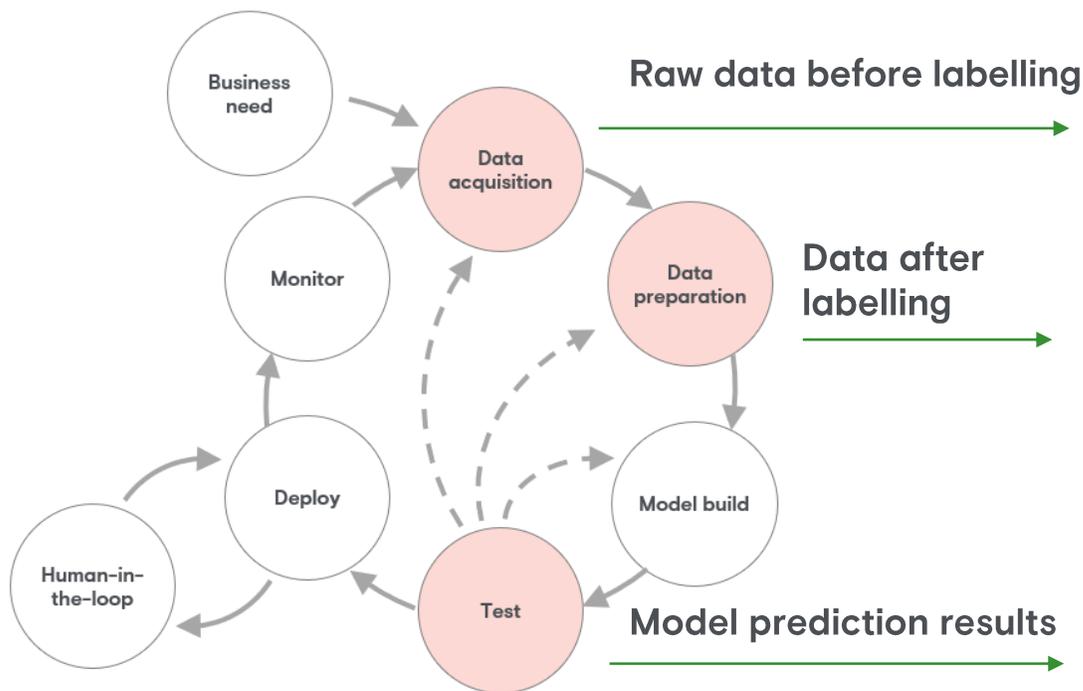
Organize and analyze training data to identify quality, distribution & bias



Appen Mobile

Engages, enables and expands crowd

In-Platform Audit helps to analyze and debug data



**Appen
In-Platform Audit**

Analyse training data or
ML model through human
or machine learning

Class imbalance

- Number of examples in the raw dataset for each class is not balanced

Accuracy

- Training dataset label accuracy

Label imbalance

- Number of examples in the training dataset for each class label is not balanced

Example benefit of training data insights

Attempt 1

9 million Tweets from males

1 million Tweets from females

This diagram illustrates a class imbalance problem. It consists of two rounded rectangular boxes. The top box is dark grey and contains the text '9 million Tweets from males'. The bottom box is light red and contains the text '1 million Tweets from females'.

Class imbalance problem

Attempt 2

9m Tweets labelled correctly

1m positive Tweets labelled as negative

This diagram illustrates an accuracy problem. It consists of two rounded rectangular boxes. The top box is dark grey and contains the text '9m Tweets labelled correctly'. The bottom box is light red and contains the text '1m positive Tweets labelled as negative'.

Accuracy problem

Attempt 3

5 million Tweets from males

4 million positive 1 million negative

5 million Tweets from females

This diagram illustrates a label imbalance problem. It consists of three rounded rectangular boxes. The top box is dark teal and contains the text '5 million Tweets from males'. Below it are two smaller boxes: a grey one on the left containing '4 million positive' and a light red one on the right containing '1 million negative'. The bottom box is dark teal and contains the text '5 million Tweets from females'.

Label imbalance problem

Benefits for our customers

- Ability to visually explore data to identify labelling errors
- Enables identification of training data composition and imbalances
- Support to debug and improve training data to improve AI performance

Future direction

- Evaluate model performance
- Support more training data management needs

Appen Mobile



Appen Connect

Match our global crowd workforce to annotation tasks



Appen Data Annotation Platform

Collect and annotate training data



Appen Intelligence

Empower Appen products with proprietary ML models



Appen In-Platform Audit

Organize and analyze training data to identify quality, distribution & bias

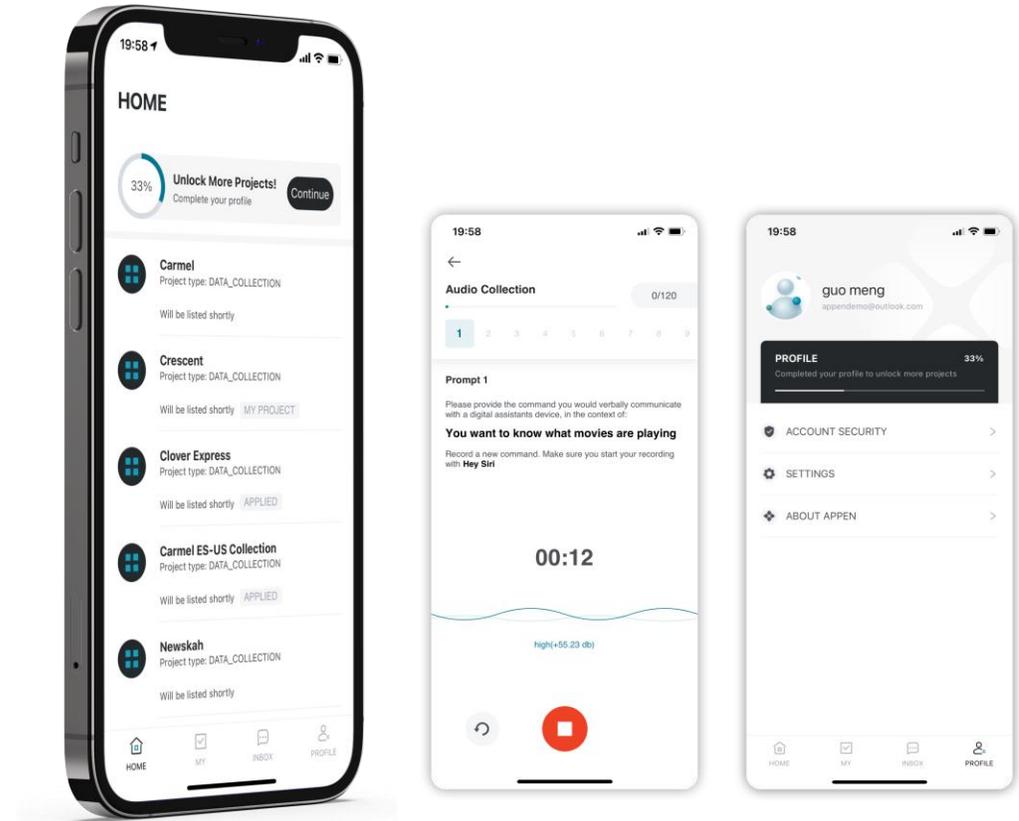


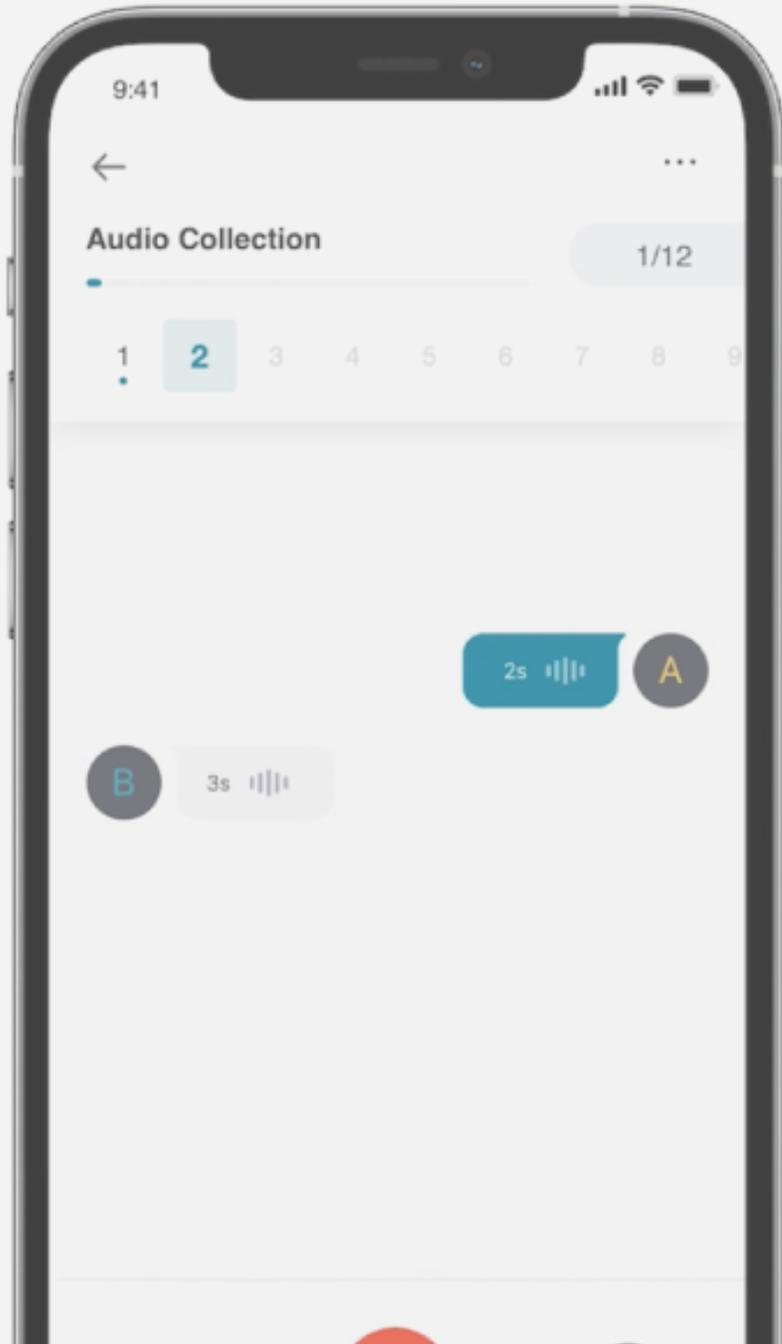
Appen Mobile

Engages, enables and expands crowd

Appen Mobile provides an enhanced crowd experience and enables new use cases

- Simple and intuitive user experience
- Supports collecting many data modalities including audio, video, text, conversation, eye-tracking, and hand-written text
- Enables location-based data collection, heightened during pandemic
- Allows access to a broader set of mobile-only crowd workers – 65% of Global 2020 Internet traffic





Demo 6

Appen Mobile

Benefits for our customers

- Improved sign-on and task experience for contributors leads to a broader crowd base
- Access to mobile-only workforce
- Mobile form factor provides enhanced data collection capabilities

Future direction

- Collaborate with customers to identify new data collection techniques
- Support annotation tasks beyond data collection
- Actively expand crowd to mobile-only areas, supporting diversity and impact sourcing

Recap and closing

The Appen Product Suite



Appen Connect

Match our global crowd workforce to annotation tasks



Appen Data Annotation Platform

Collect and annotate training data



Appen Intelligence

Empower Appen products with proprietary ML models



Appen In-Platform Audit

Organize and analyze training data to identify quality, distribution & bias

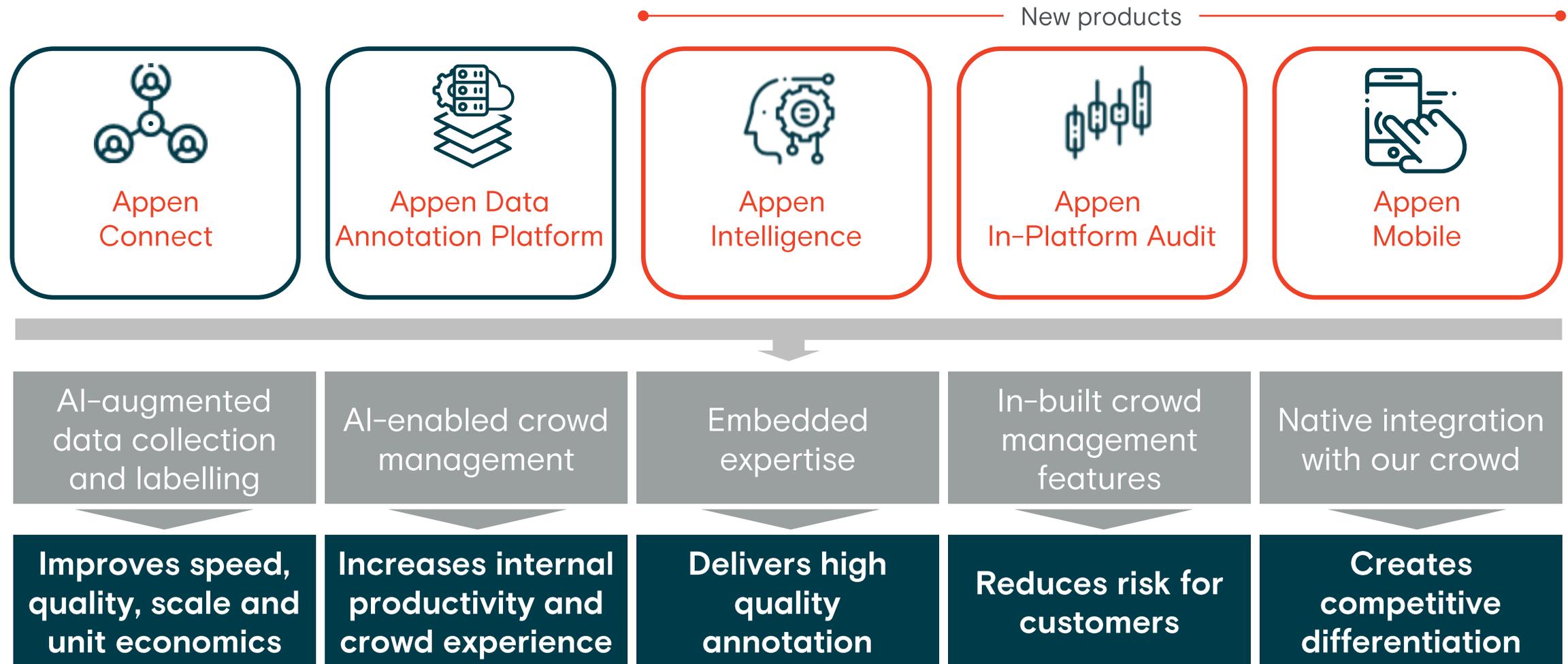


Appen Mobile

Engages, enables and expands crowd

New products

Value created from our products



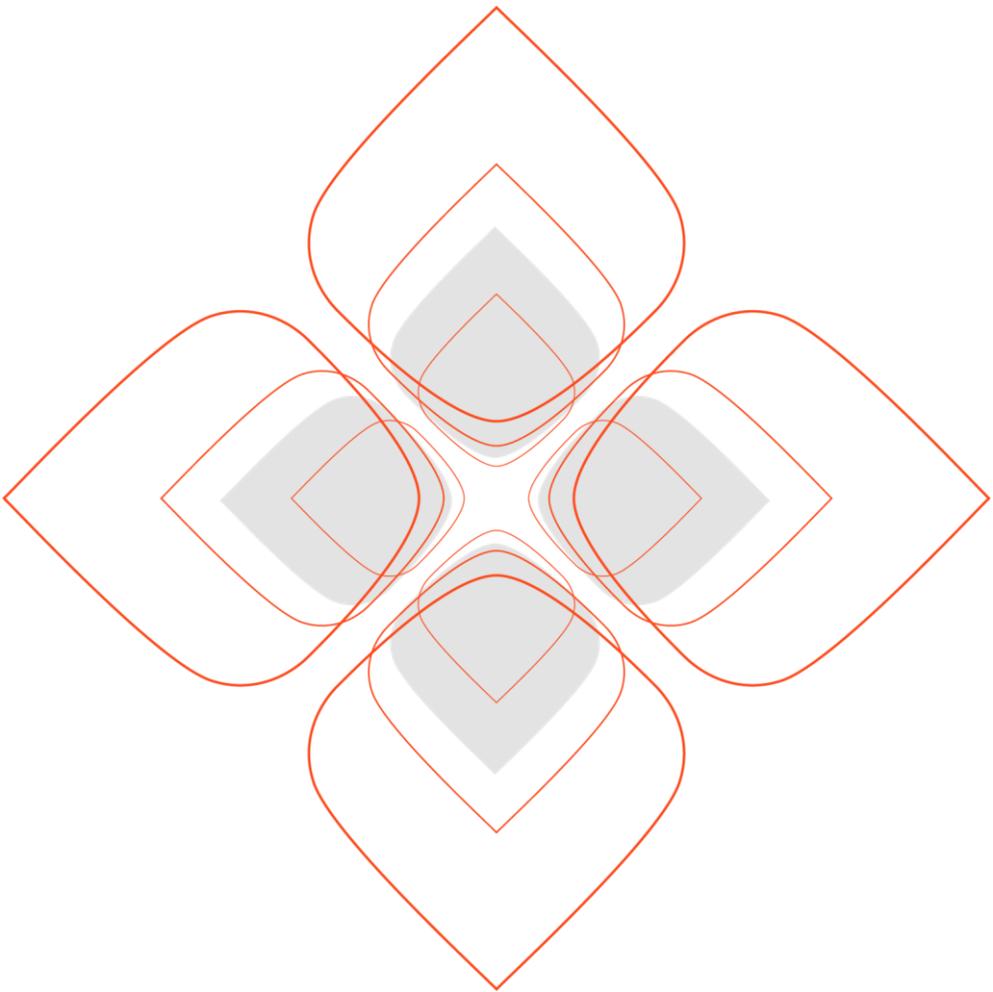
Appen has the unique combination of industry leading technology, crowd of over 1m and deep internal expertise



Q&A



Appen is accelerating
its transformation into
an AI powered provider
of AI data and solutions



End

