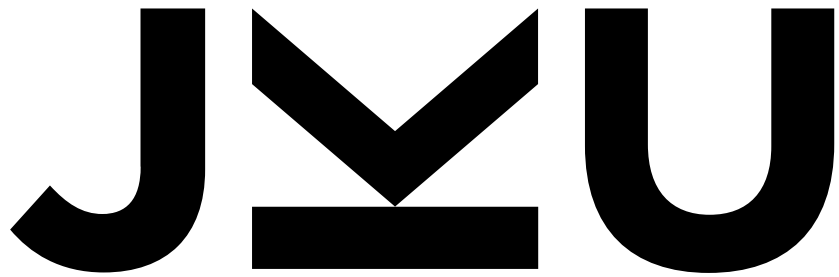


JOHANNES KEPLER
UNIVERSITY LINZ

UE MLPC 2025: INTRODUCTION AND DATA ANNOTATION TASK

Tara Jadidi, Paul Primus, Florian Schmid

2025-03-10

Institute of Computational Perception

JOHANNES KEPLER
UNIVERSITY LINZ

Agenda

- Who are we?
- Goals for this Class
- Project Vision
- Project Phases & Schedule
- Deliverables & Grading
- Task 1: Annotation Task
- Task 0: Form Teams
- Moodle

WHO ARE WE

Who are we?

- Tara Jadidi
- Paul Primus
- Florian Schmid

Who are we?

- Tara Jadidi
- Paul Primus
- Florian Schmid
- aaand . . . ?

Who are we?

Poll: What do you study?

- CS Bachelor
- CS Master
- AI Bachelor
- AI Master
- None of this
- All of this

Who are we?

Poll: What are you most comfortable with?

- Java
- Python
- Matlab
- None of this
- All of this

Who are we?

Poll: What do you know about machine learning?

- can explain overfitting
- can explain cross-validation
- have used SVMs
- have used random forests
- have used deep learning
- have heard of embedding models and representation learning
- not a lot yet

GOALS FOR THIS CLASS

Goals for this class

We would like you to:

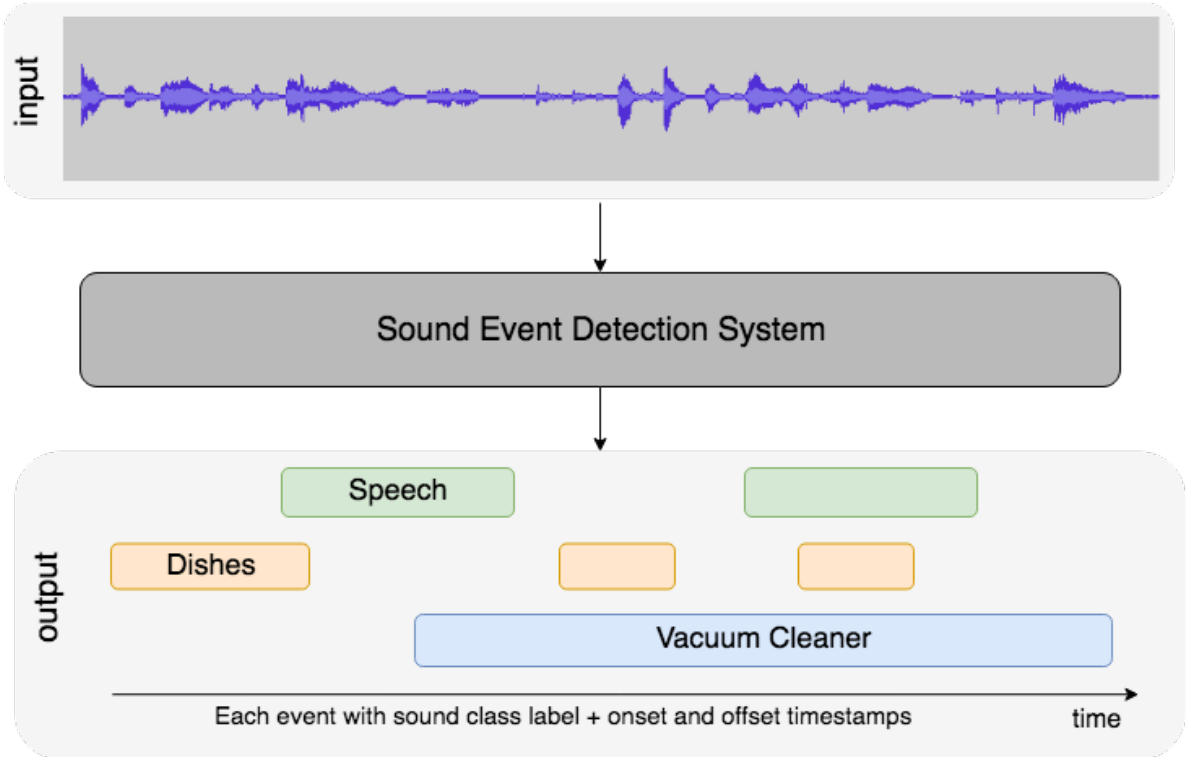
- become acquainted with several classification algorithms
- feel comfortable setting up a machine learning experiment from scratch
- experience the limits of machine learning in real-world tasks
- be able to communicate your results concisely

To achieve this, we will guide you through a machine learning project.

PROJECT VISION

The (fictitious) Story

- An innovative company called *Kepler Intelligent Audio Labs (KIAL)* is developing Sound Event Detection Systems



Customers

Hospital

Harbor Operator

Autonomous Vehicle

Terminology

- **Sound Event**: Any perceptible sound in an audio recording that lasts for a certain period, from a short moment to the entire recording.
- **Sound Event Detection System**: A piece of software that automatically identifies and localizes instances of a pre-defined set of sound event classes within an audio recording.
- **Sound Event Class** and **Sound Event Label** are used interchangeably in the following, referring to the types of **sound events** such as *dog barking*, *rainfall*, or *footsteps*.

On the next slides, we will ...

- ... explore the steps to build an ML pipeline for a *single* customer.
- ... discuss the high cost of collecting a dataset for each new customer.
- ... define what a shared, reusable dataset could look like.
- ... investigate how to leverage a general dataset for a specific customer.

A Single Customer: Events of Interest

The Harbor Operator ...

- ... wants to automate the monitoring of incoming and outgoing goods
- by tracking different types of vehicles.
- **Event Classes of Interest:** Ship, Train, Helicopter, Truck

A Single Customer: Requirements

The Harbor Operator ...

- ... approaches *KIAL* and hands over **requirements document:** *"A piece of software that detects all specified* ***Event Classes of Interest*** *in a continuous audio stream."*

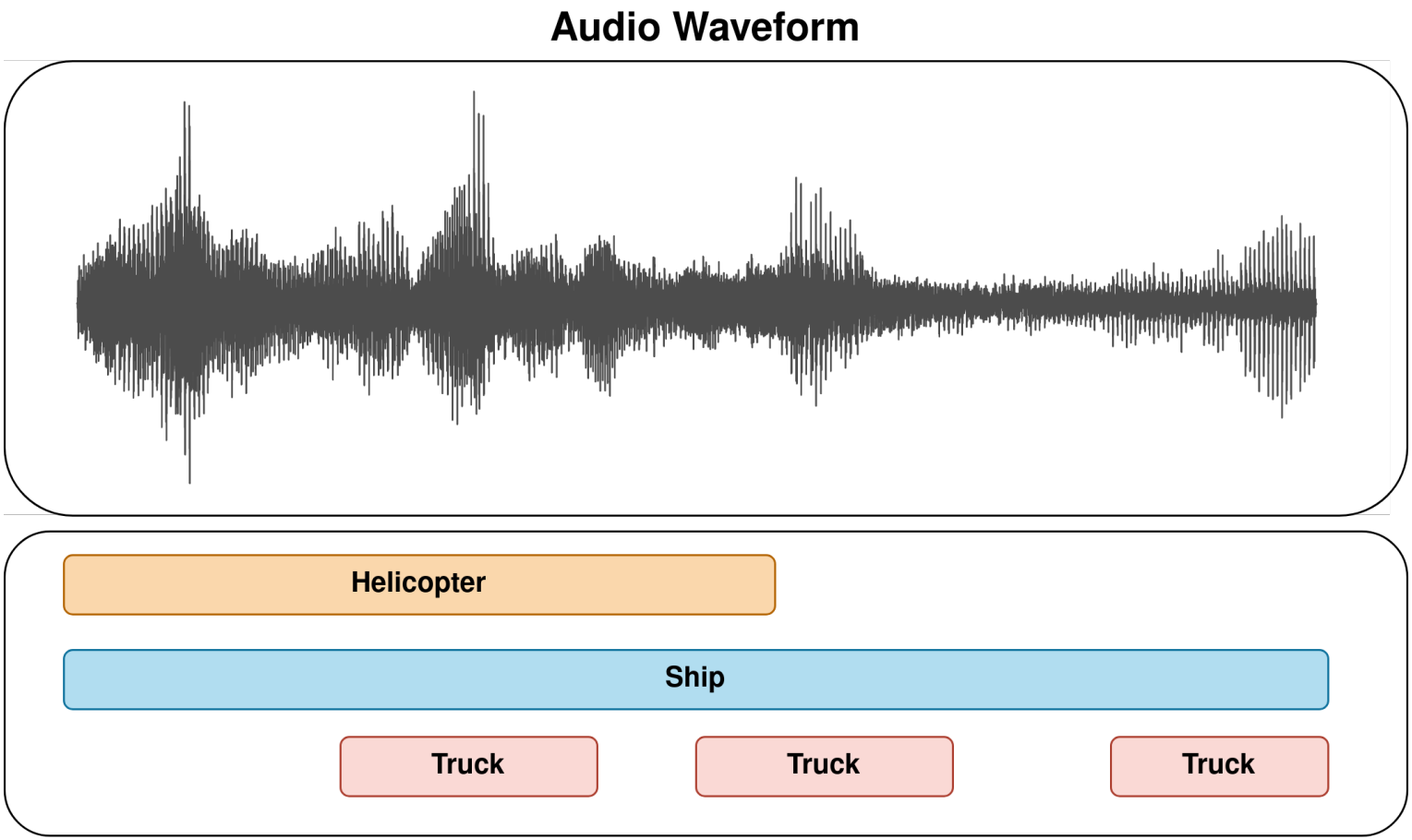
KIAL assigns *you* the project lead role for machine learning.
How do you plan to approach this task?

A Single Customer: ML Pipeline Steps

Events of Interest:
Ship, Train, Helicoptor, Truck

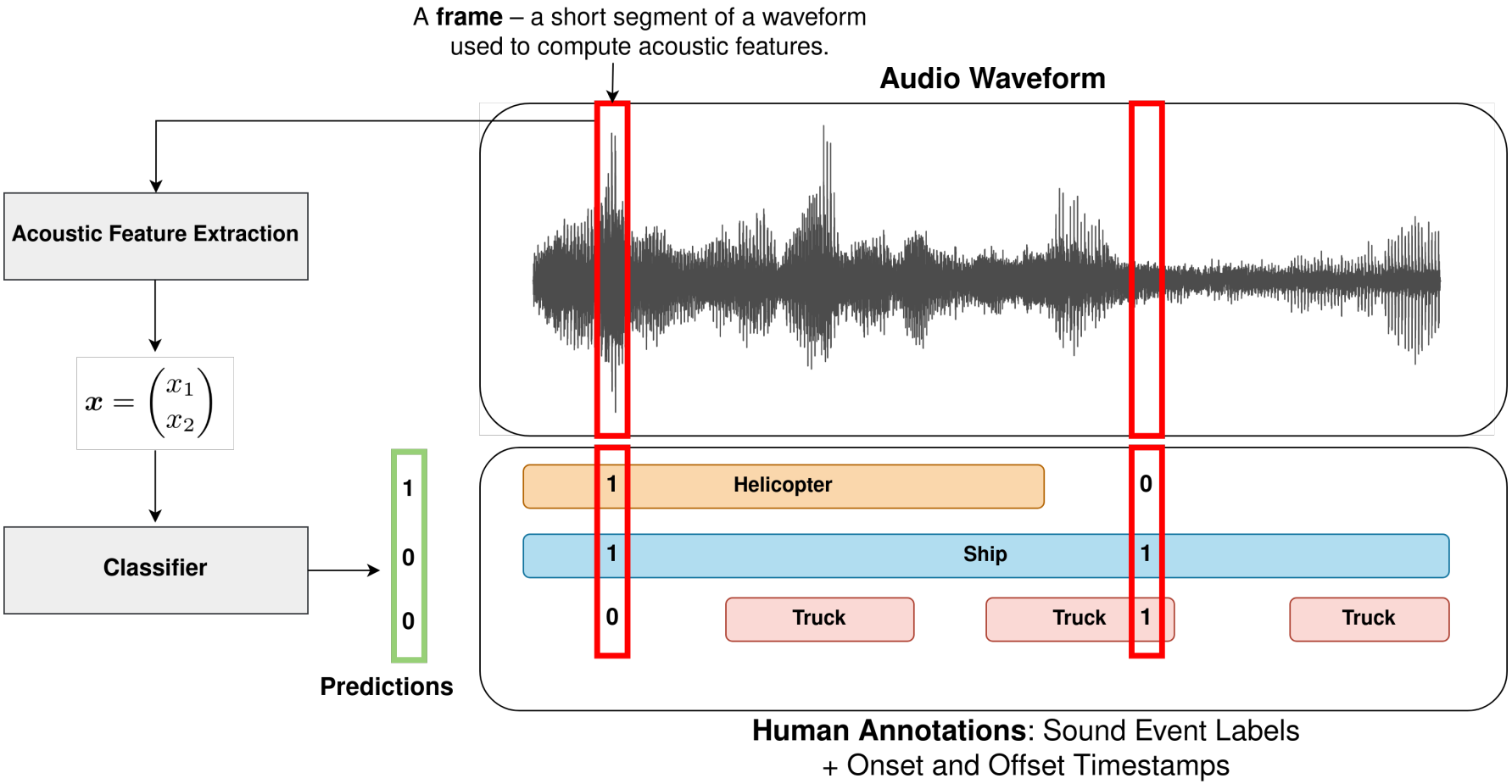
- Collect Dataset
- Preprocess Dataset
- Annotate Dataset
- Dataset Analyis
- Feature Engineering
- Model Development
- Model Deployment
- Monitoring

A Single Customer: Dataset



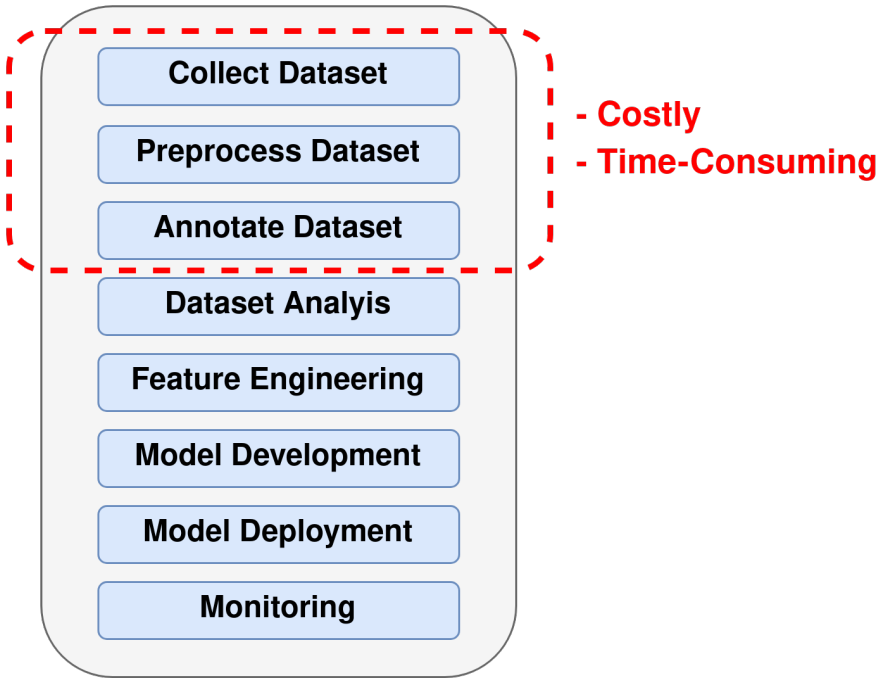
Human Annotations: Sound Event Labels
+ Onset and Offset Timestamps

A Single Customer: Classifier



A Single Customer: ML Pipeline Steps

Events of Interest:
Ship, Train, Helicoptor, Truck

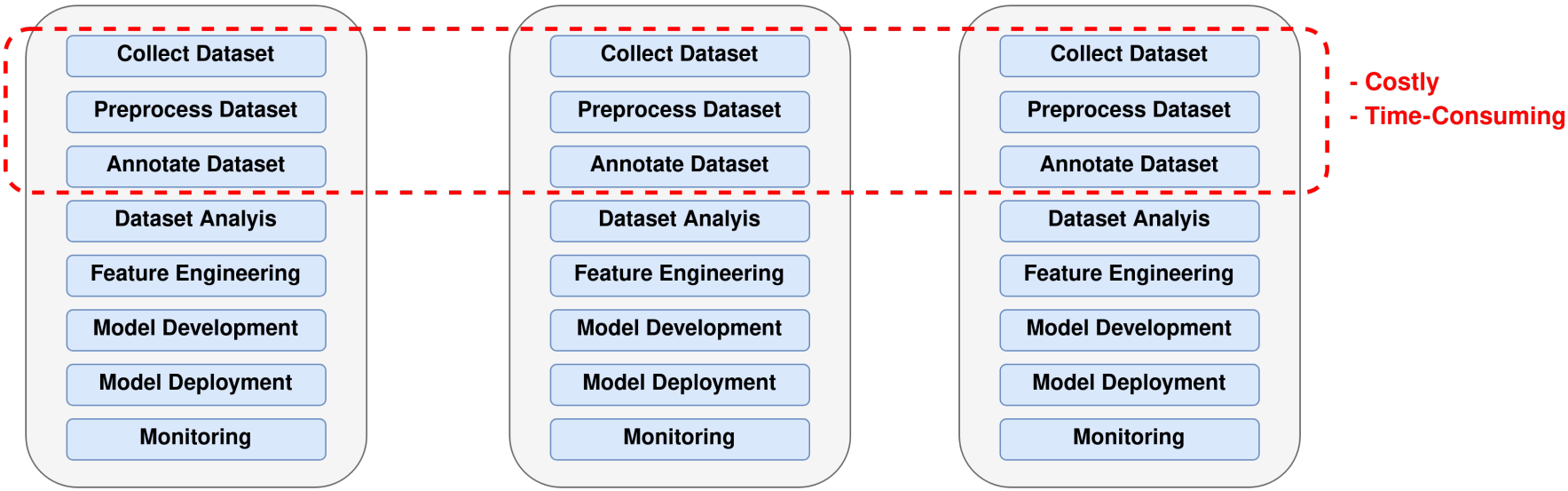


Multiple Customers: ML Pipeline Steps

Events of Interest:
Coughing, Sneezing, Footsteps, Snoring

Events of Interest:
Ship, Train, Helicoptor, Truck

Events of Interest:
Horn Honk, Siren, Shouting, Vehicle

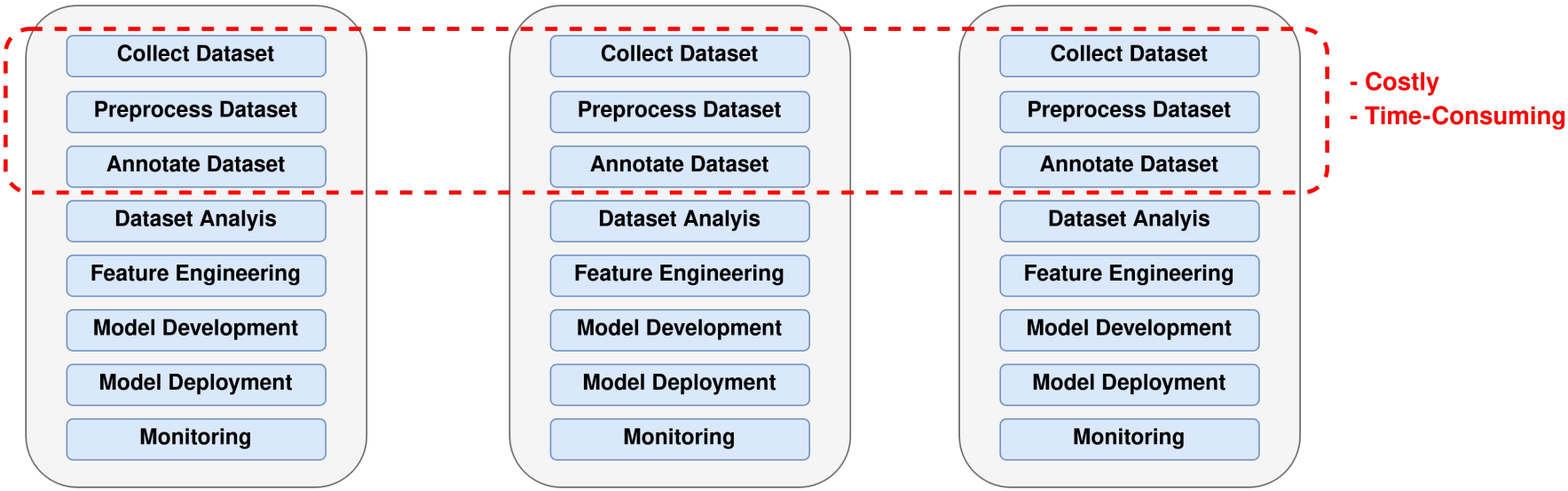


Multiple Customers: ML Pipeline Steps

Events of Interest:
Coughing, Sneezing, Footsteps, Snoring

Events of Interest:
Ship, Train, Helicoptor, Truck

Events of Interest:
Horn Honk, Siren, Shouting, Vehicle

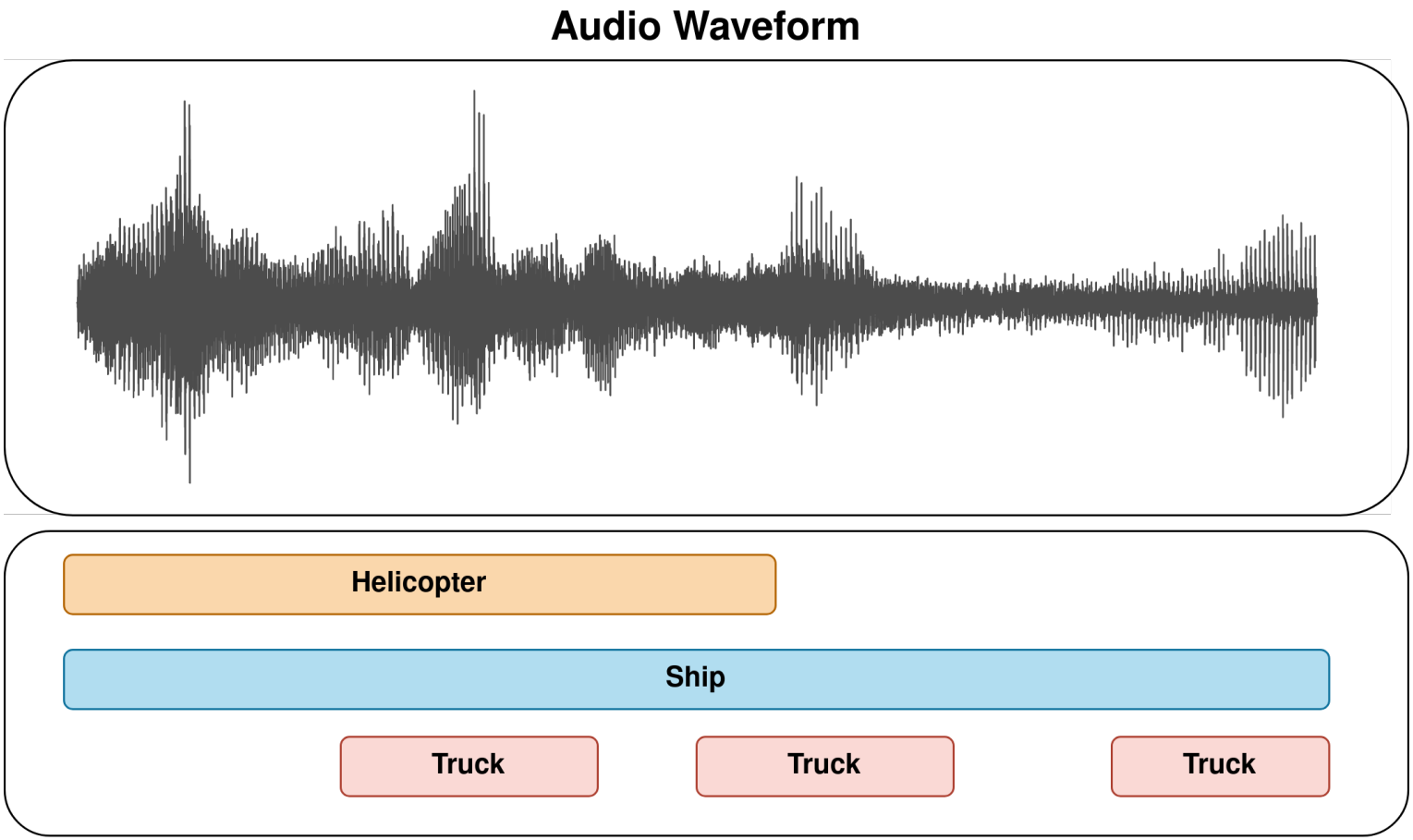


Can we unify datasets across customers?

Multiple Customers: A General Dataset

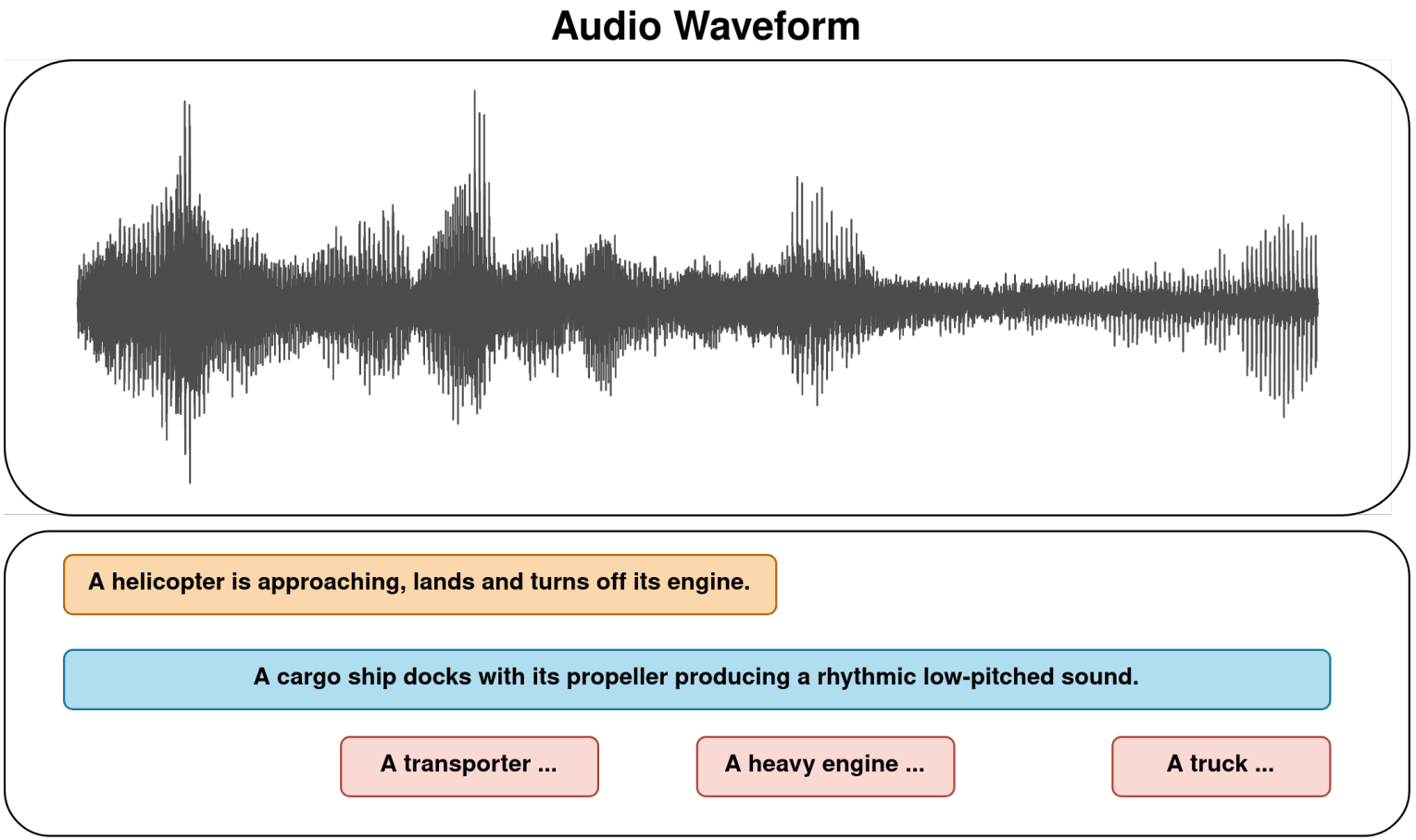
- Can we collect a general dataset that is costumer-independent?
- **Idea**:
  - Collect a diverse audio dataset (*not* specific to one customer)
  - Use free-text annotations instead of a fixed event classes
  - For a specific customer:
    - automatically extract related annotations and corresponding audios from general dataset
    - based on similarity of customer's event classes of interest and free-text annotations

Towards A General Dataset (1/4)



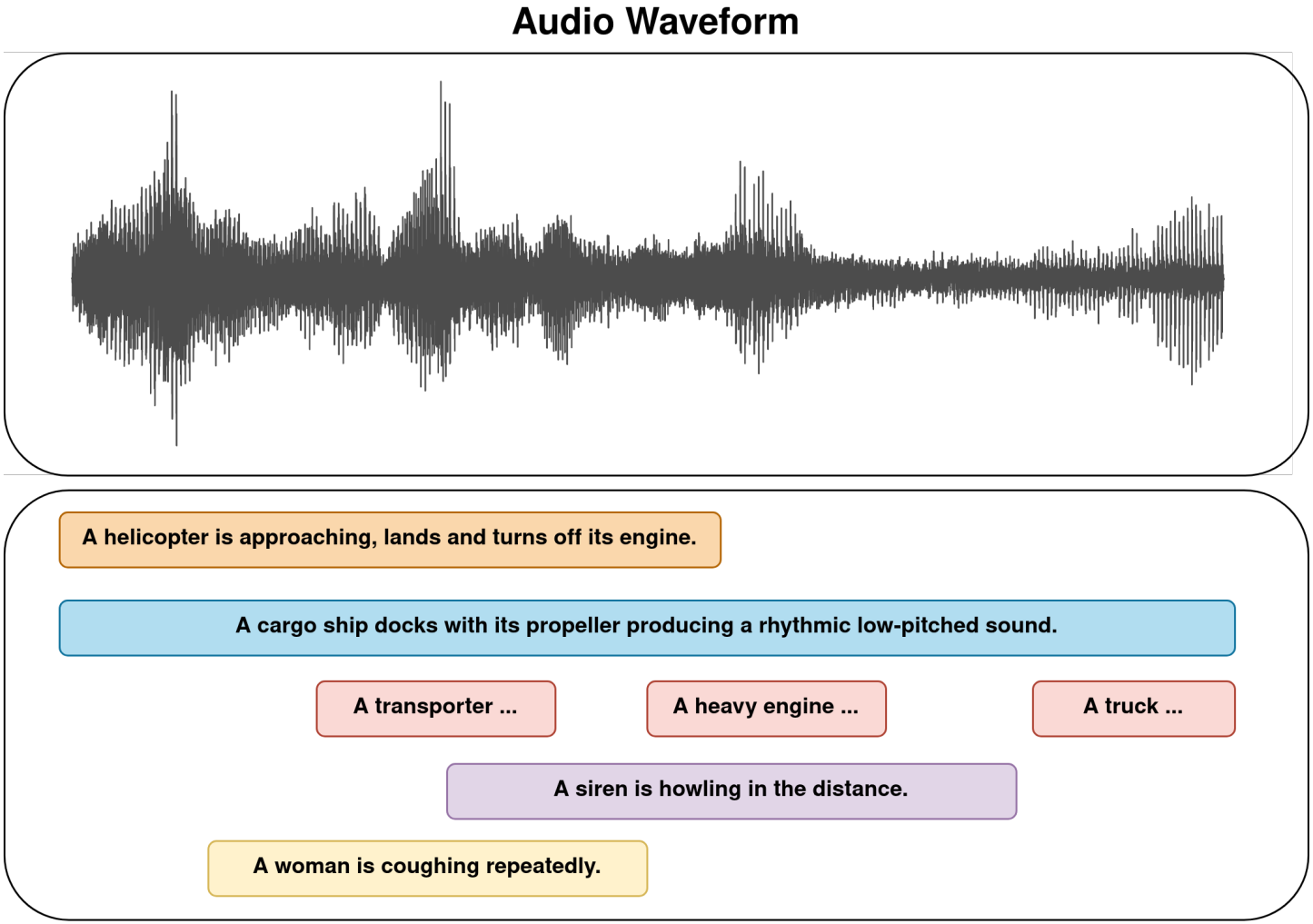
Human Annotations: Sound Event Labels
+ Onset and Offset Timestamps

Towards A General Dataset (2/4)



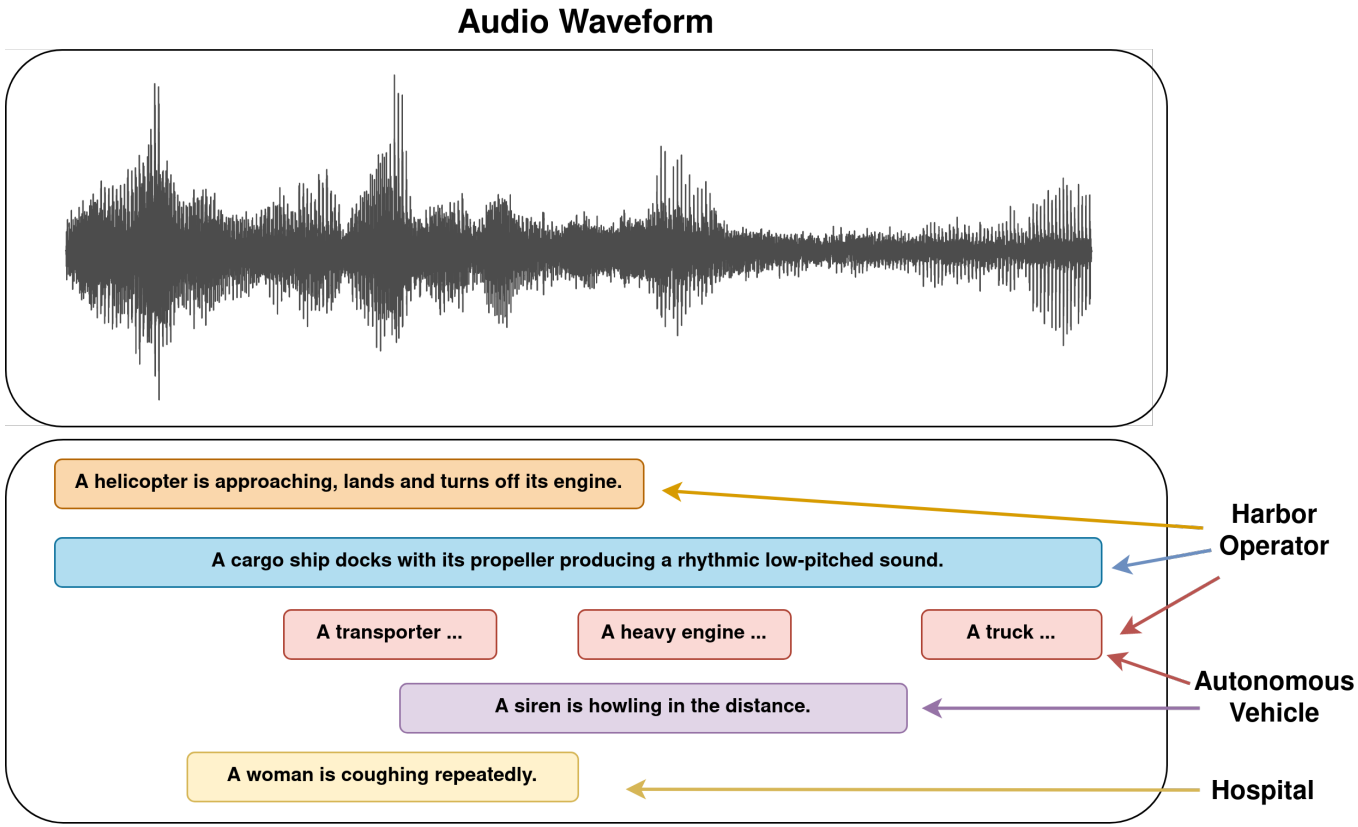
Human Annotations: Free-Text Annotations +
Onset and Offset Timestamps

Towards A General Dataset (3/4)



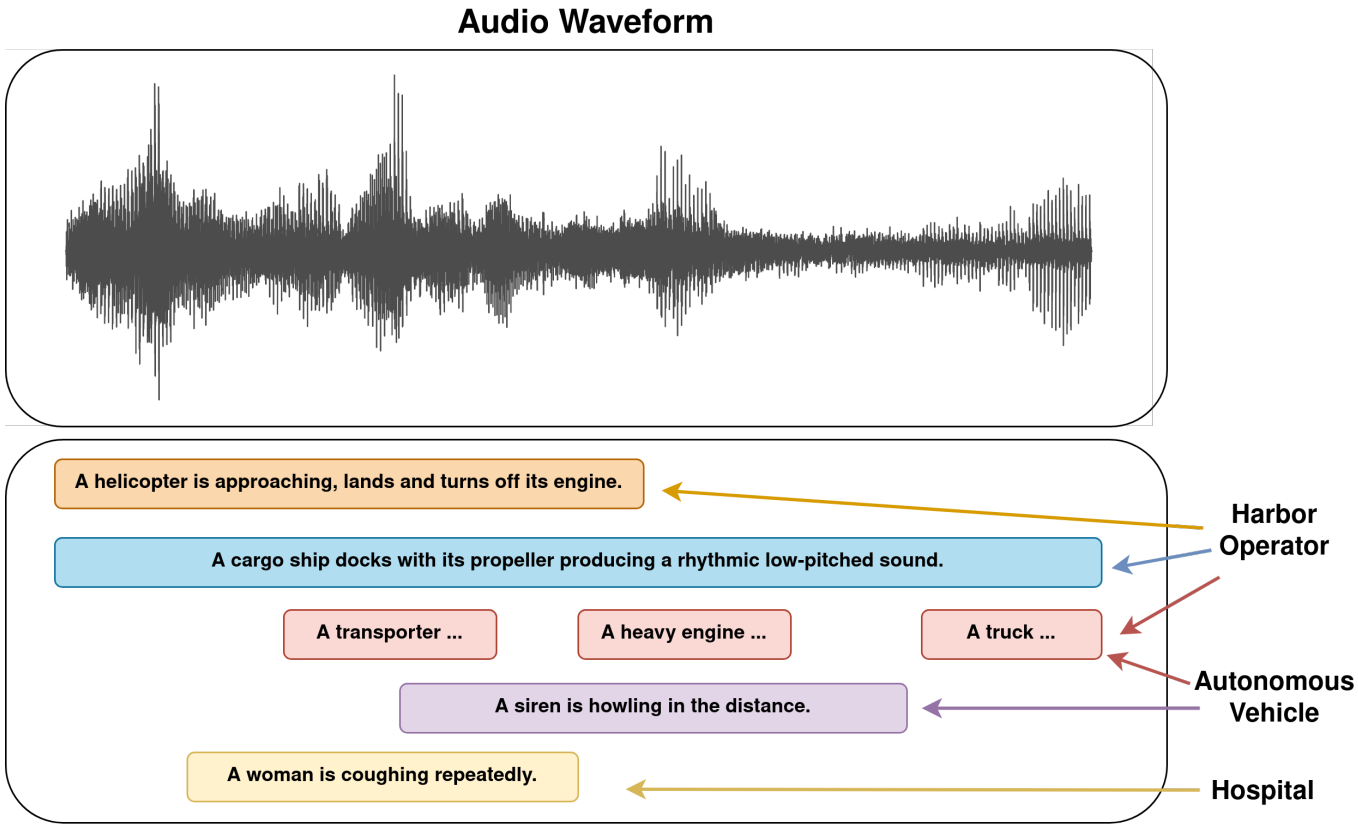
Human Annotations: Free-Text Annotations +
Onset and Offset Timestamps

Towards A General Dataset (4/4)



Human Annotations: Free-Text Annotations + Onset and Offset Timestamps

Towards A General Dataset (4/4)

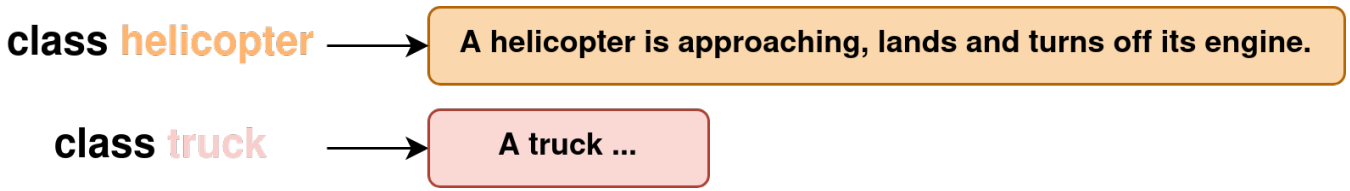


Human Annotations: Free-Text Annotations + Onset and Offset Timestamps

How can we automatically match a customer's event labels with free-text annotations?

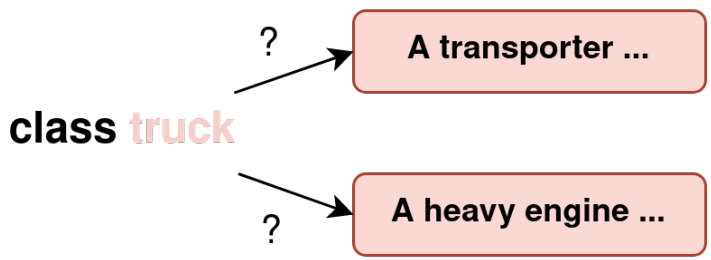
Matching Labels to Annotations, Approach 1

- Event label appears in related free-text annotation



→ exact match - easy!

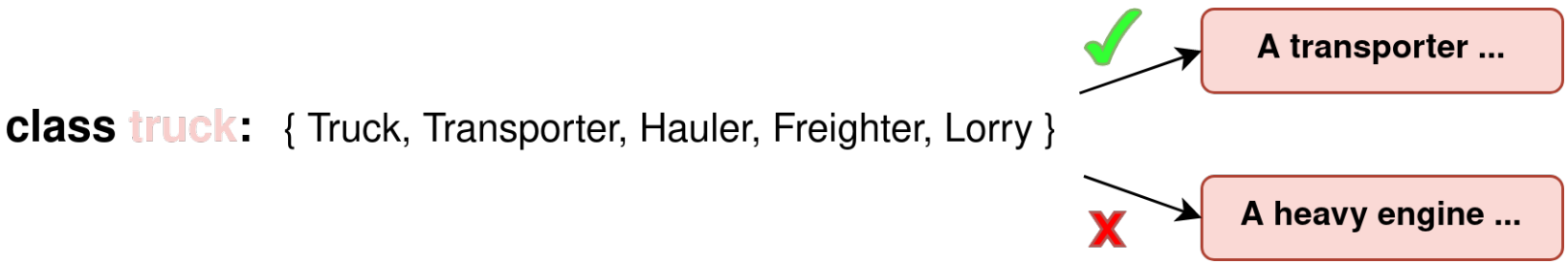
- Event label does not appear in related free-text annotation



→ no exact match, have to rely on *similarity*!

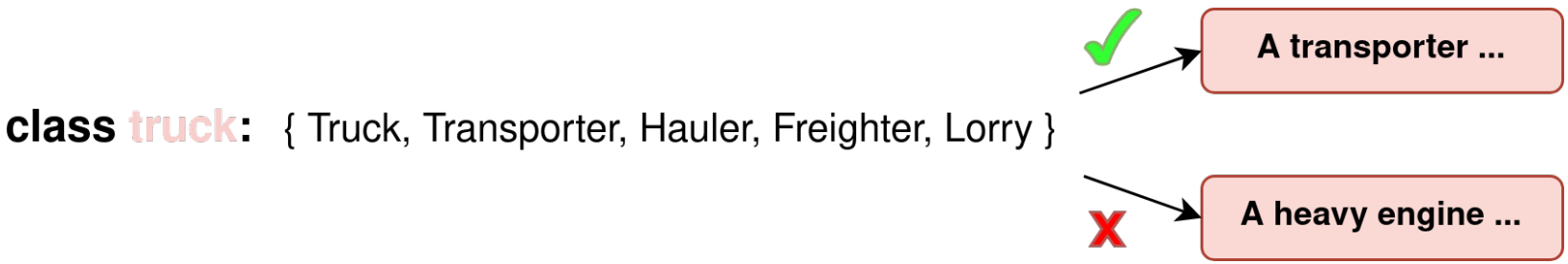
Matching Labels to Annotations, Approach 2

- Define a set of *synonyms* / *similar words* and rely on exact match



Matching Labels to Annotations, Approach 2

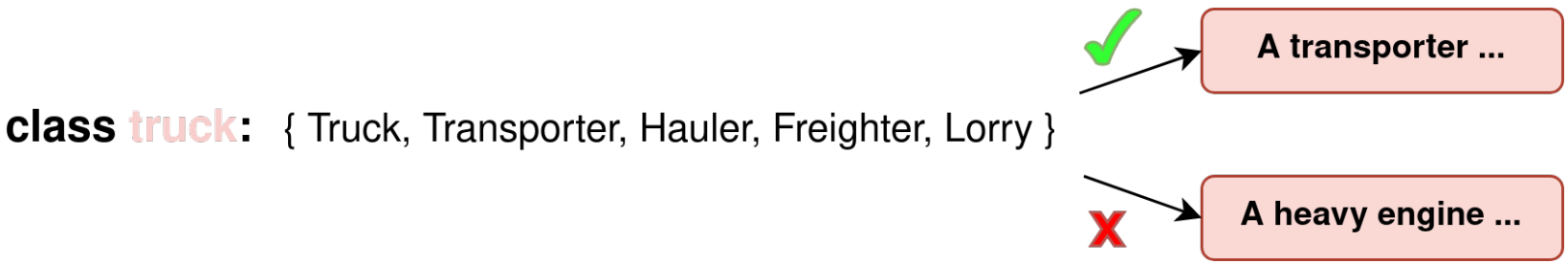
- Define a set of *synonyms* / *similar words* and rely on exact match



Can you think of an approach that does not rely not exact match?

Matching Labels to Annotations, Approach 2

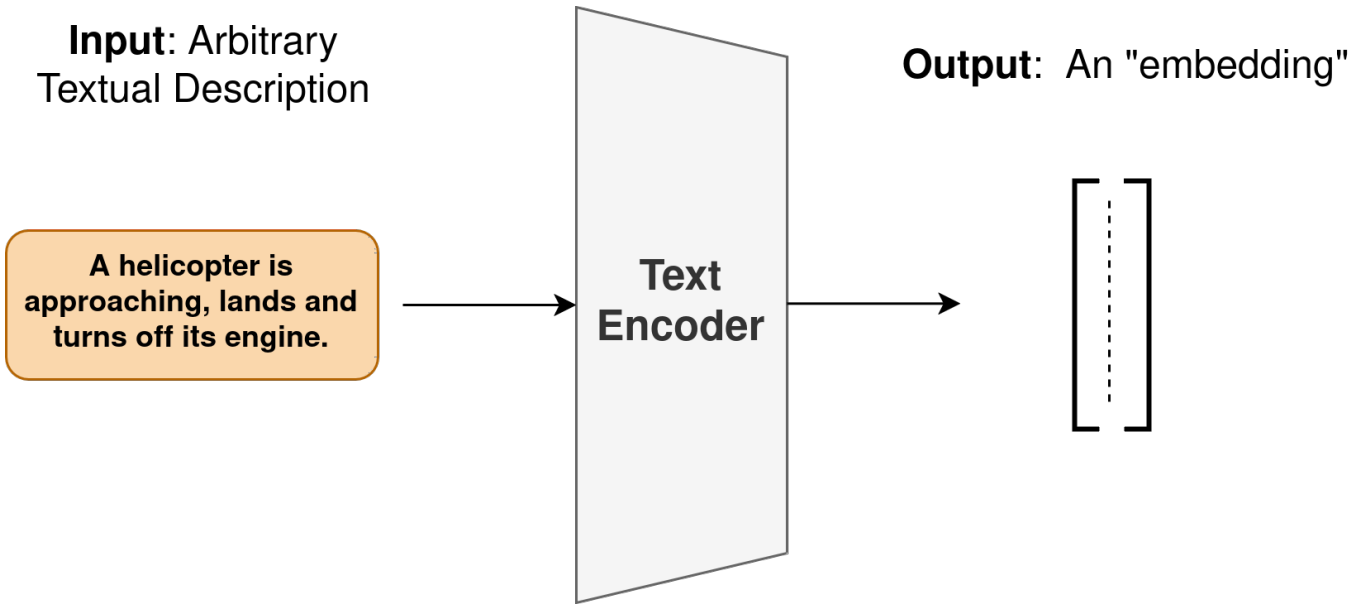
- Define a set of *synonyms* / *similar words* and rely on exact match



Can you think of an approach that does not rely not exact match?

→ Comparing Text Embeddings

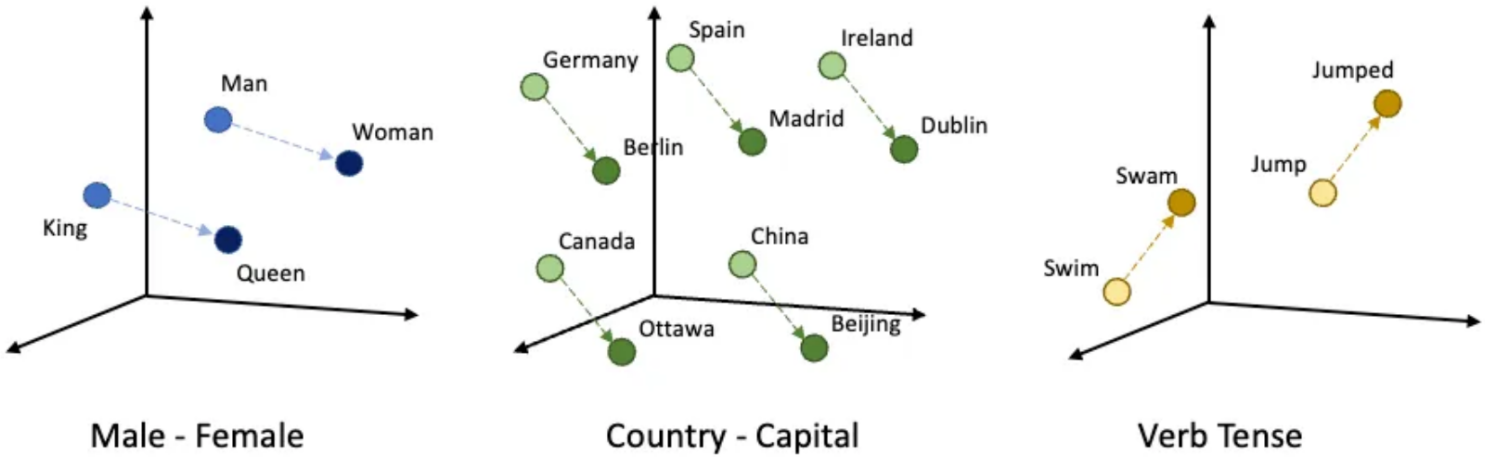
Interlude: Text Embeddings (1/2)



- A computer program that takes a piece of text and creates a compressed representation (**embedding**)
- Embedding: fixed-dimensional vector of real numbers
- Embeddings occupy the **embedding space**

Interlude: Text Embeddings (2/2)

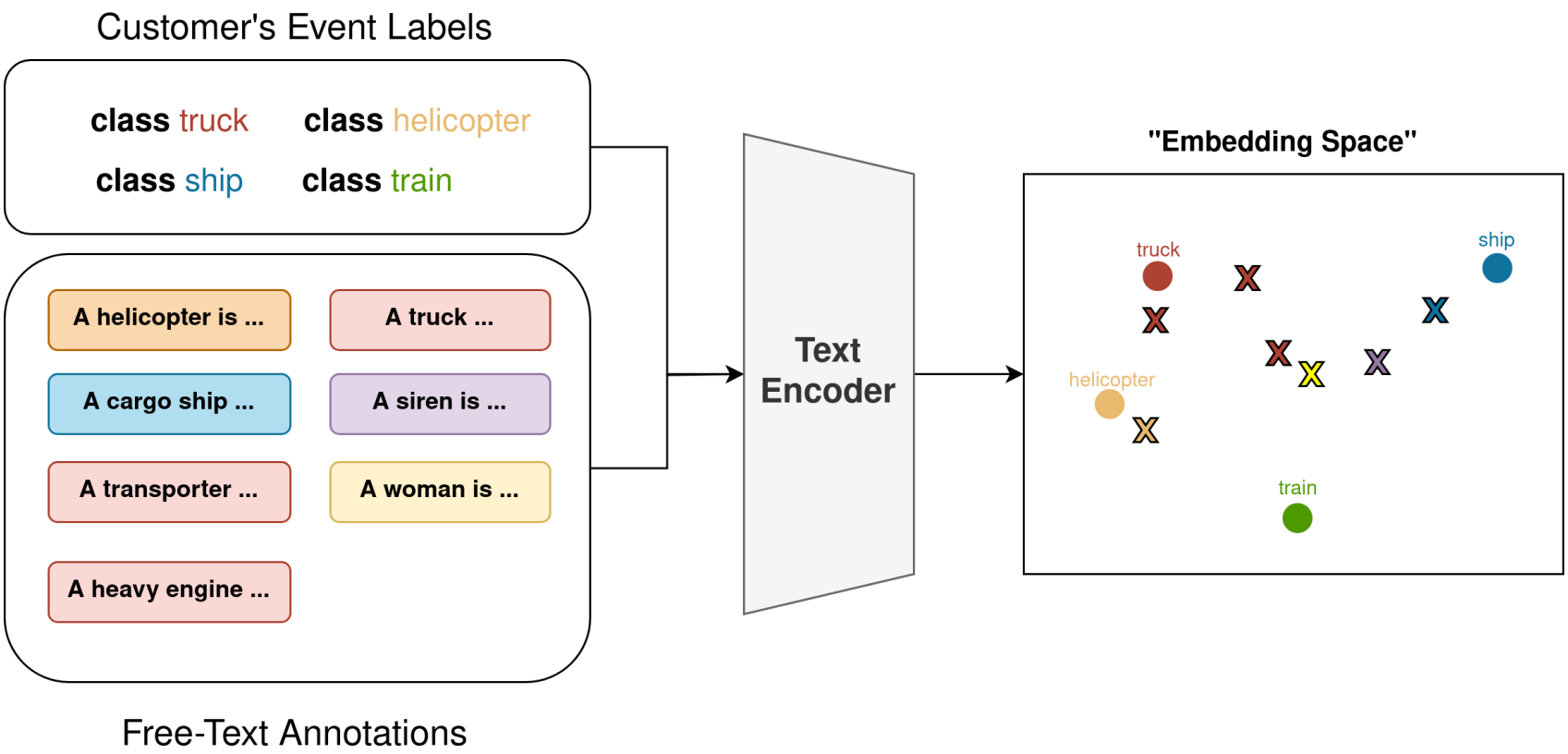
- **Embedding space**: A high-dimensional space where similar texts map to nearby points.



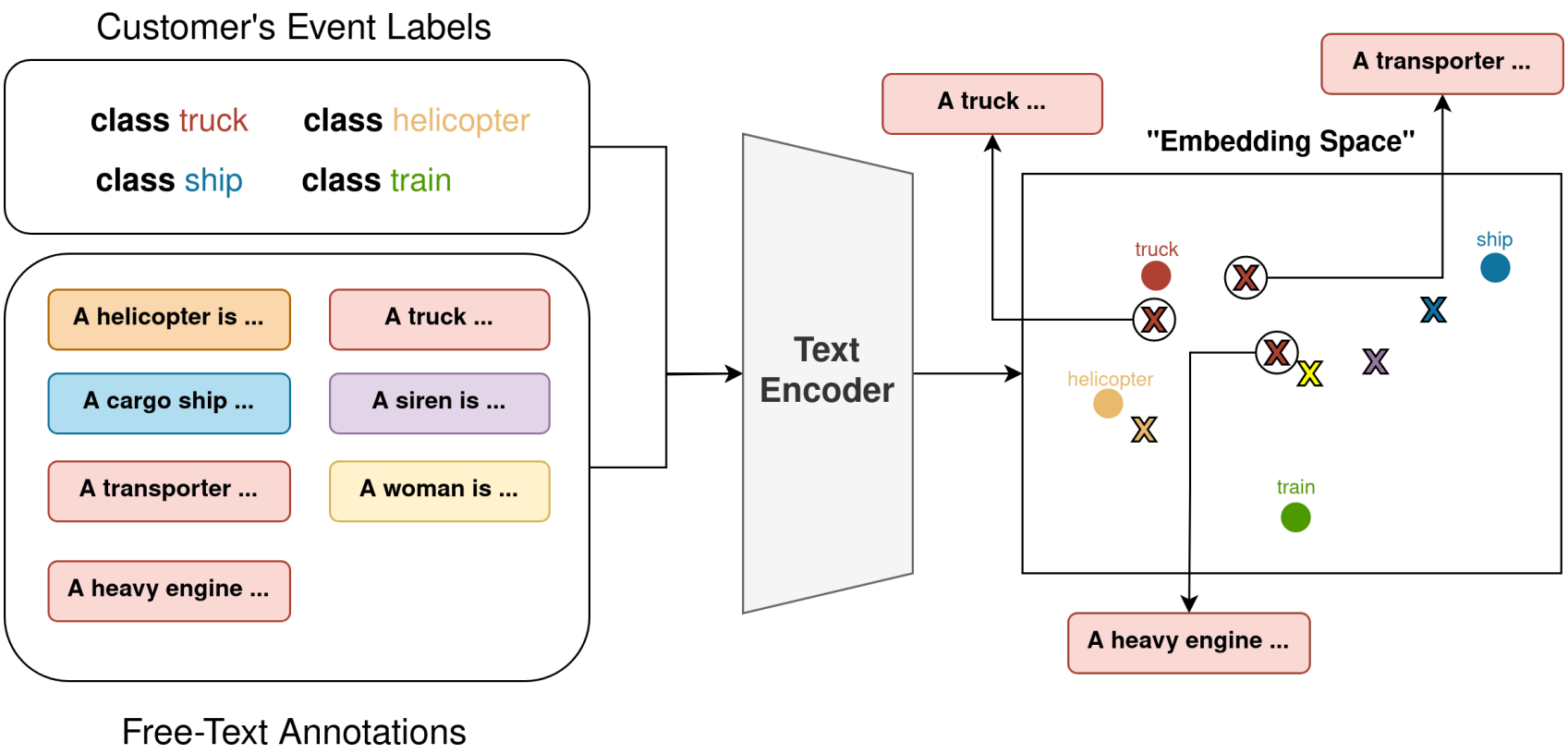
https://medium.com/towards-data-science/exploring-the-power-of-embeddings-in-machine-learning-18a601238d6b

Note: The embedding vectors here are 3-dimensional for visualization purposes only; in practice, they reside in much higher-dimensional spaces.

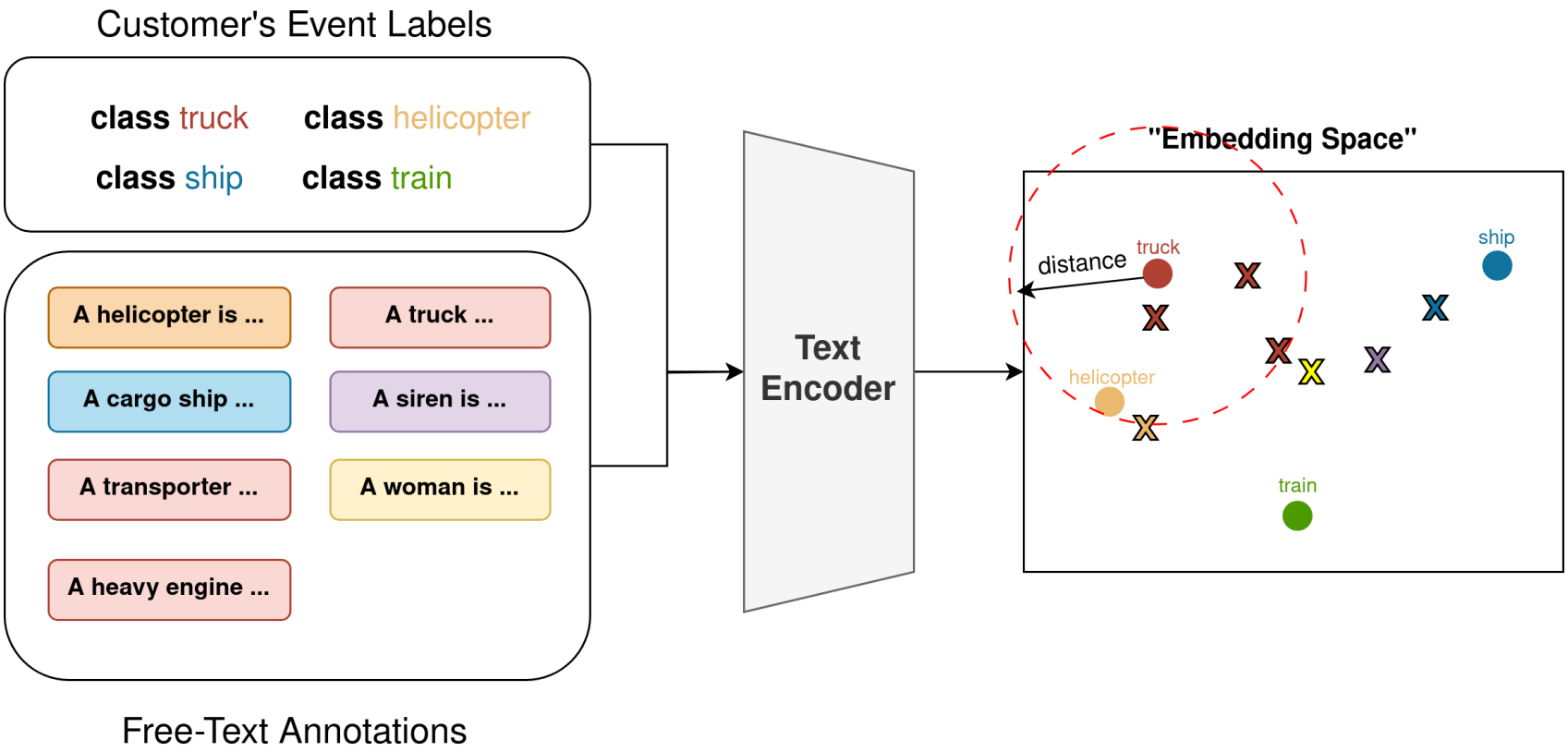
Matching Labels to Annotations, Approach 3



Matching Labels to Annotations, Approach 3

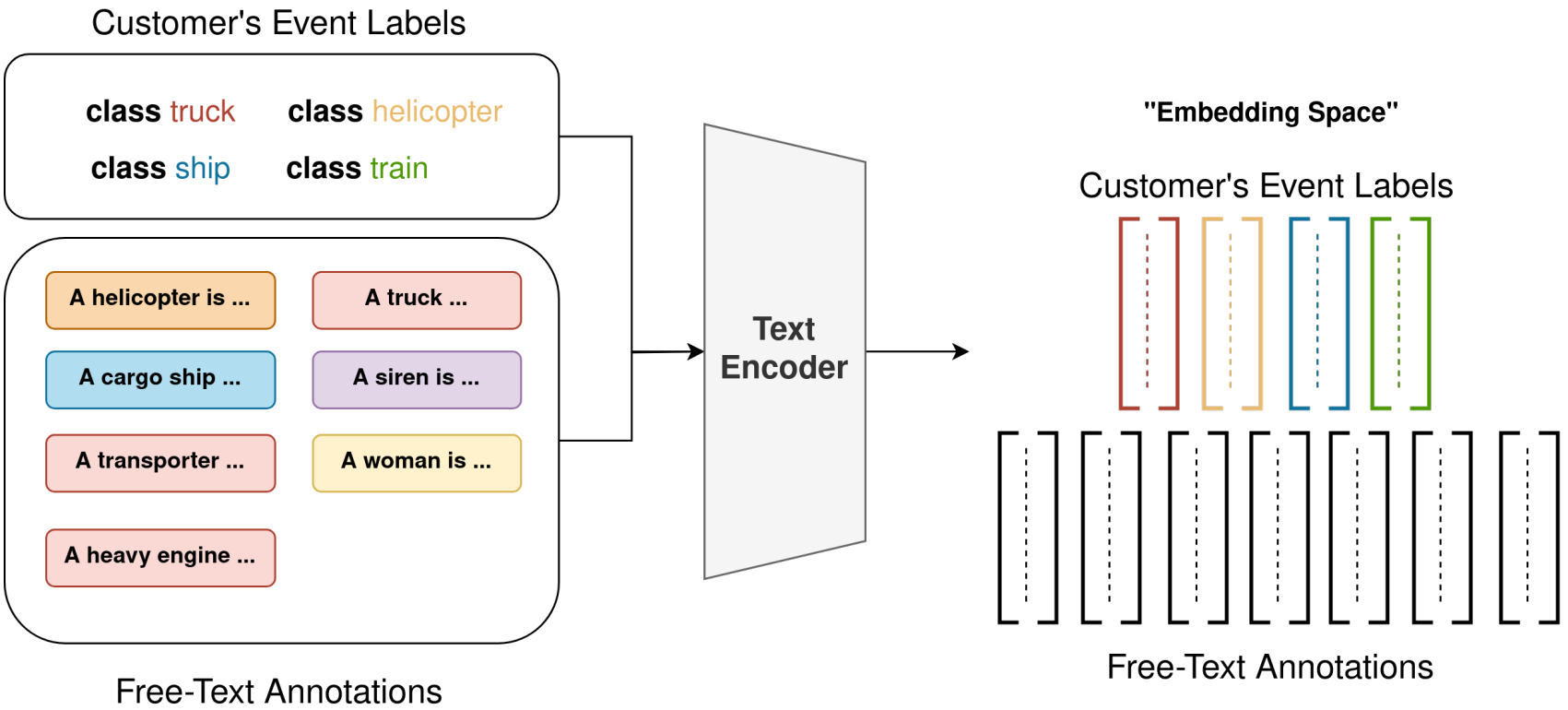


Matching Labels to Annotations, Approach 3



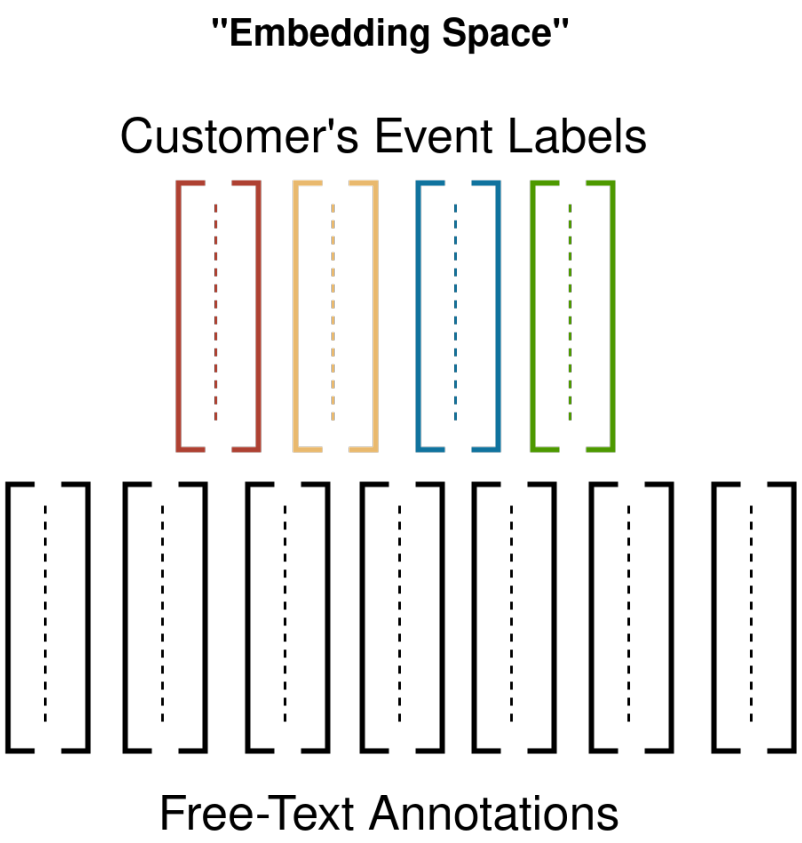
- Distance too small: end up with only a few annotations
- Distance too large: retrieve lots of unrelated annotations

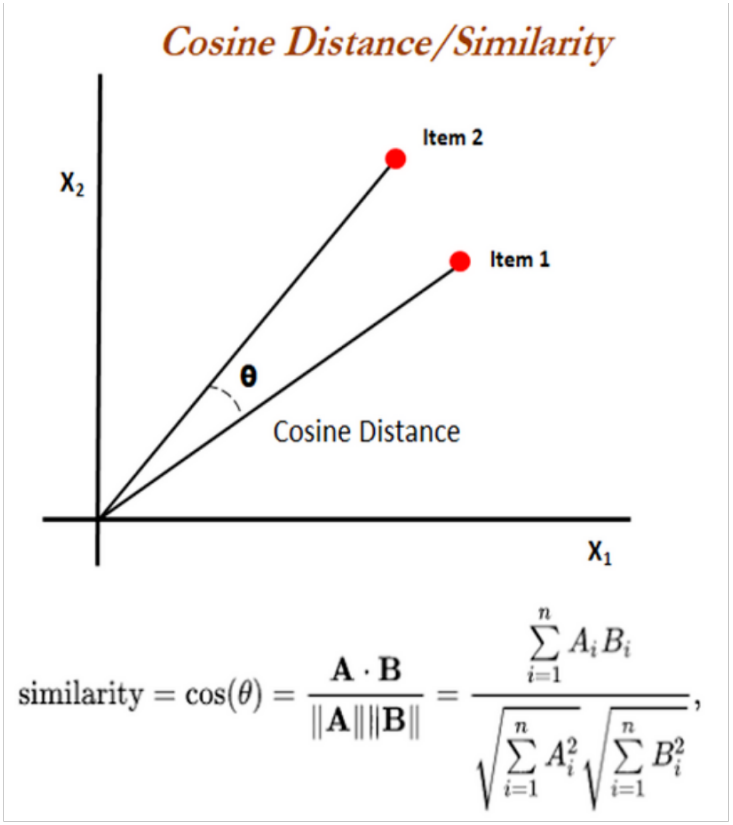
Matching Labels to Annotations, Approach 3



- Embedding space typically has many dimensions (not only 2 ...)

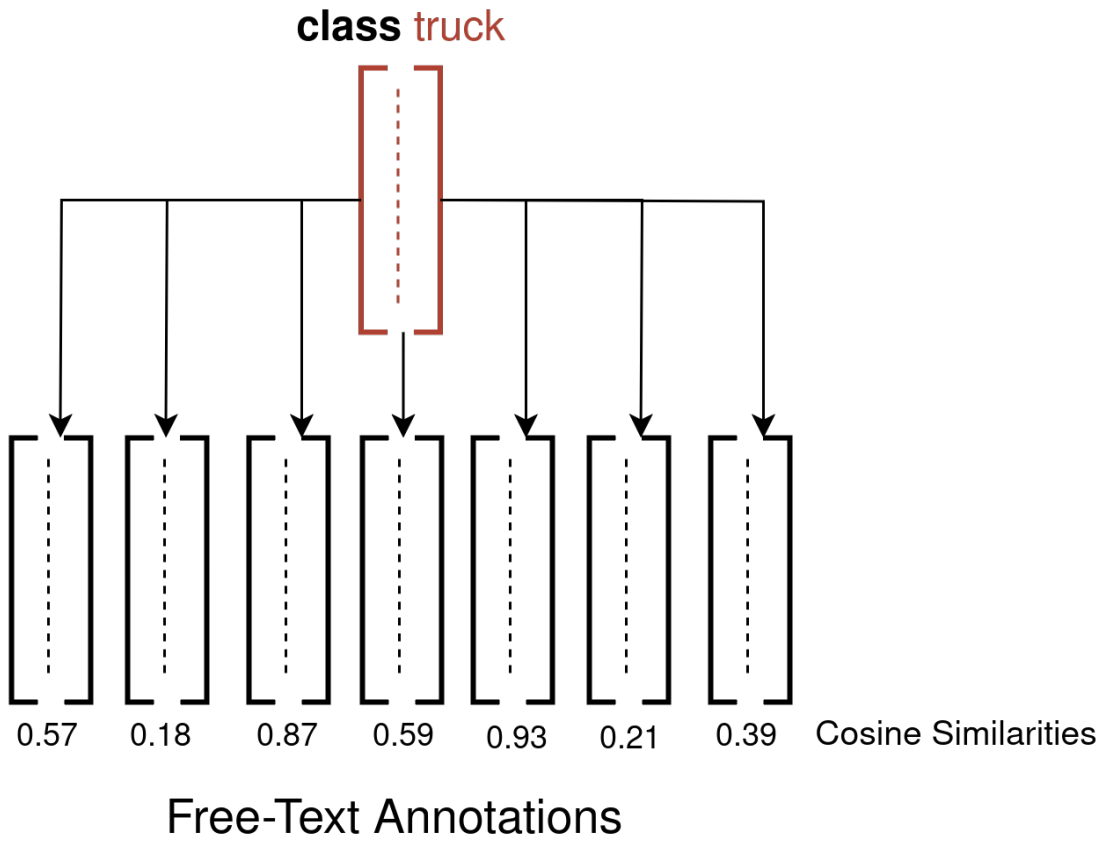
Matching Labels to Annotations, Approach 3





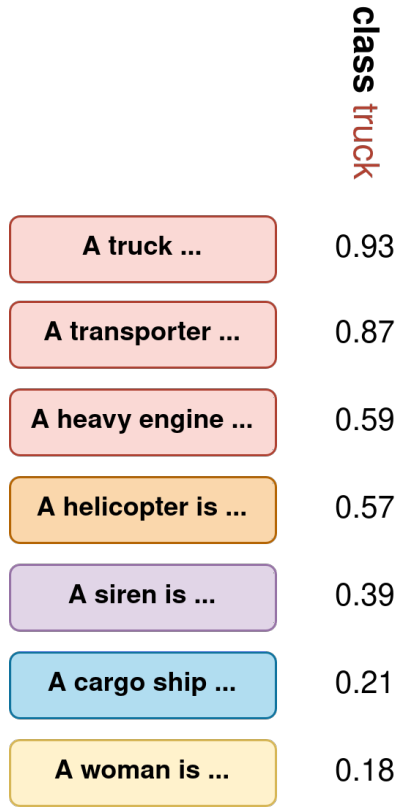
- Can’t visualize (directly), but can still calculate pairwise distances / similarities!

Matching Labels to Annotations, Approach 3



- Calculate Cosine Similarities between event label and free-text annotation embeddings

Matching Labels to Annotations, Approach 3



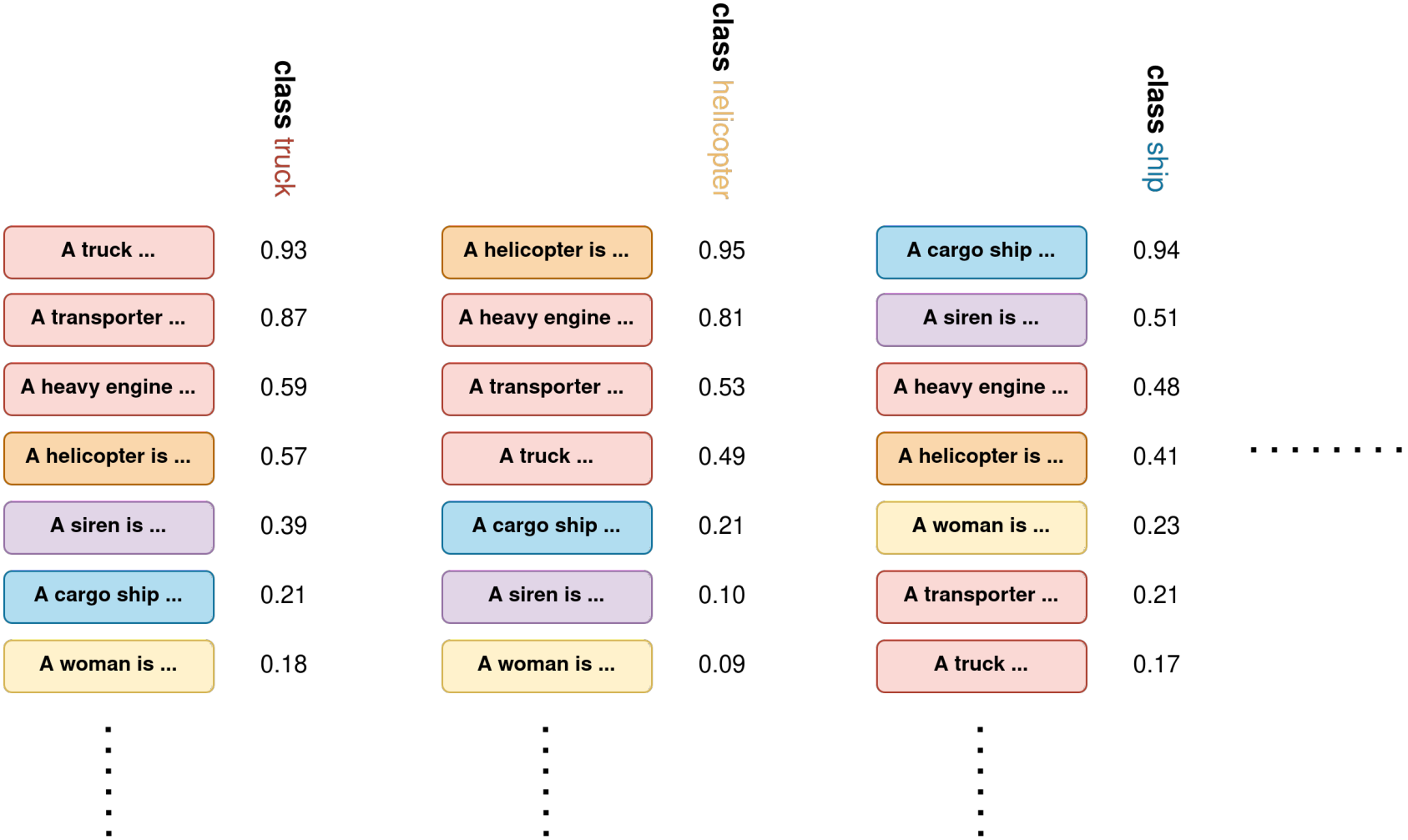
- Rank free-text annotations per class label according to similarities

Matching Labels to Annotations, Approach 3

class truck		class helicopter		class ship	
A truck ...	0.93	A helicopter is ...	0.95	A cargo ship ...	0.94
A transporter ...	0.87	A heavy engine ...	0.81	A siren is ...	0.51
A heavy engine ...	0.59	A transporter ...	0.53	A heavy engine ...	0.48
A helicopter is ...	0.57	A truck ...	0.49	A helicopter is ...	0.41
A siren is ...	0.39	A cargo ship ...	0.21	A woman is ...	0.23
A cargo ship ...	0.21	A siren is ...	0.10	A transporter ...	0.21
A woman is ...	0.18	A woman is ...	0.09	A truck ...	0.17

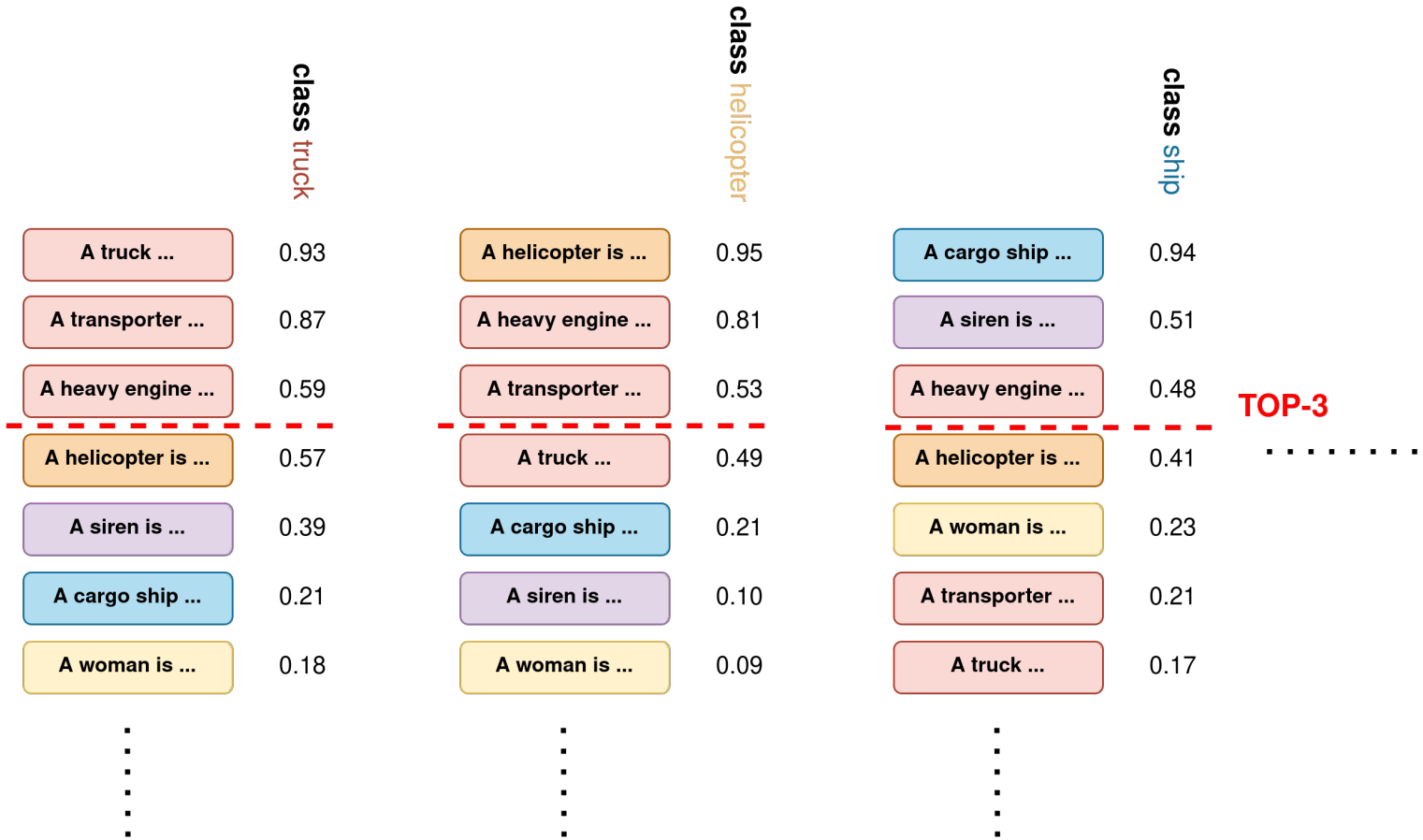
- Extends to arbitrary labels

Matching Labels to Annotations, Approach 3



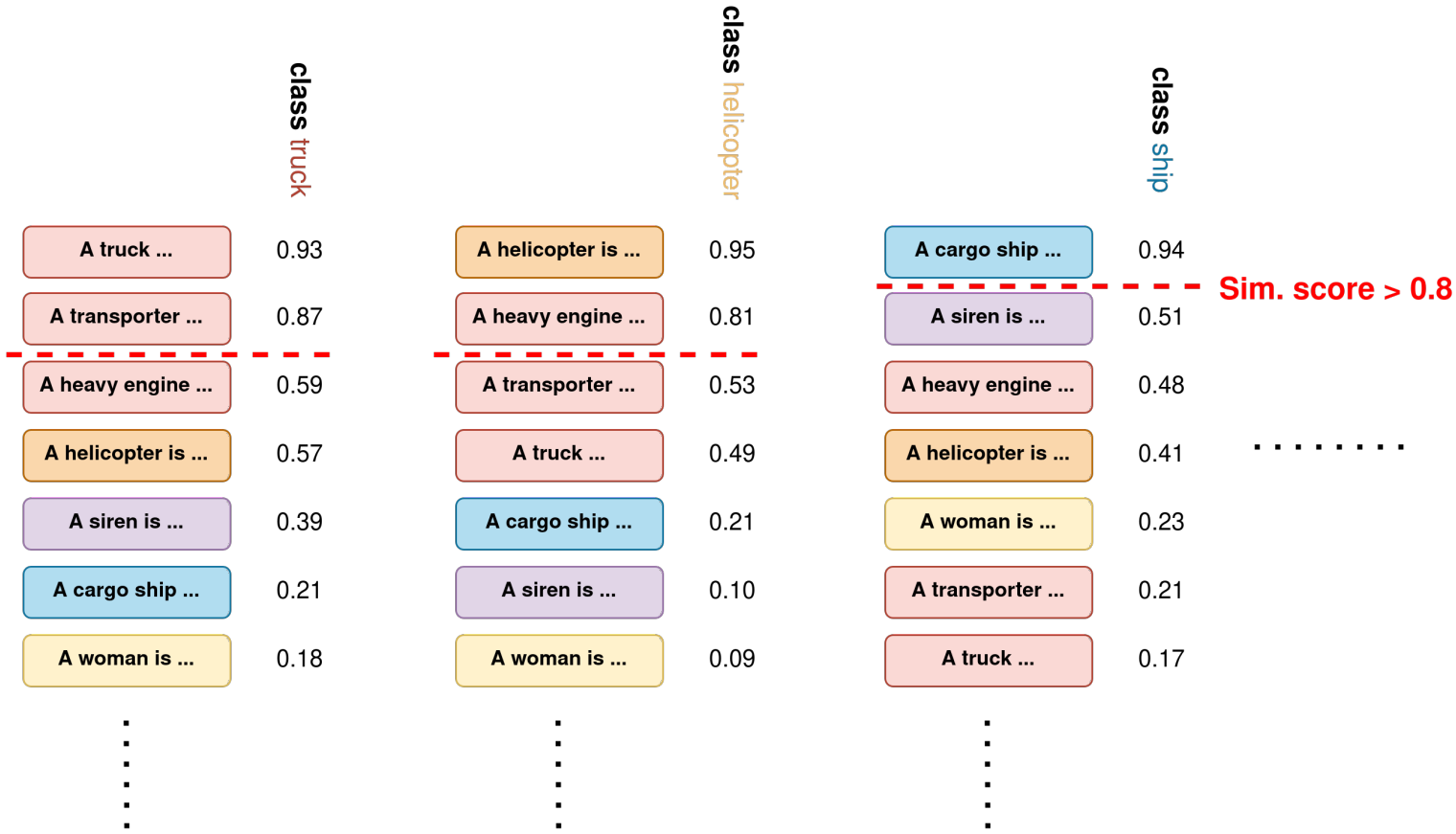
- Extends to an arbitrary number of free-text annotations

Matching Labels to Annotations, Approach 3



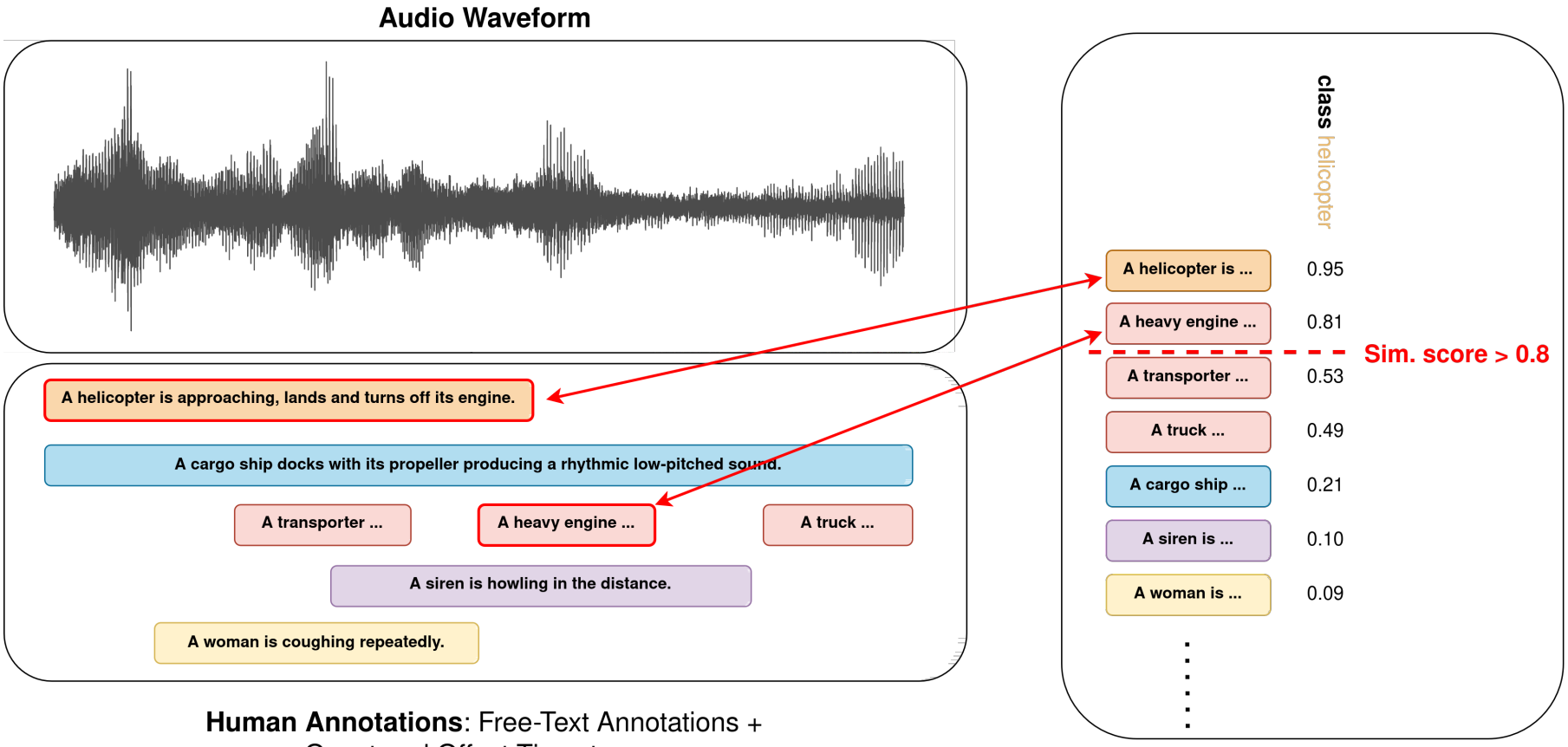
- How to select positive instances per event label?
- Top-K similar?

Matching Labels to Annotations, Approach 3



- How to select positive instances per event label?
- Similarity score $>$ threshold?

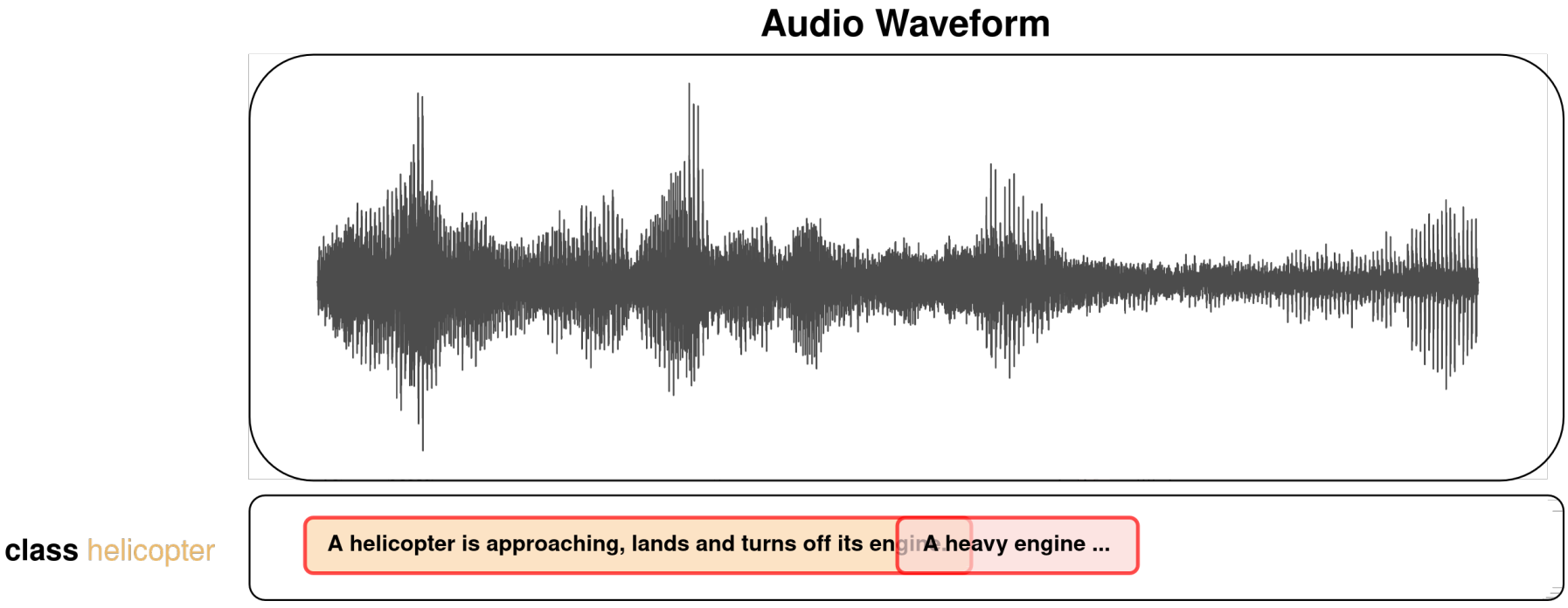
Matching Labels to Annotations, Approach 3



Human Annotations: Free-Text Annotations + Onset and Offset Timestamps

- We found two positive instances for class *helicopter* in this audio recording

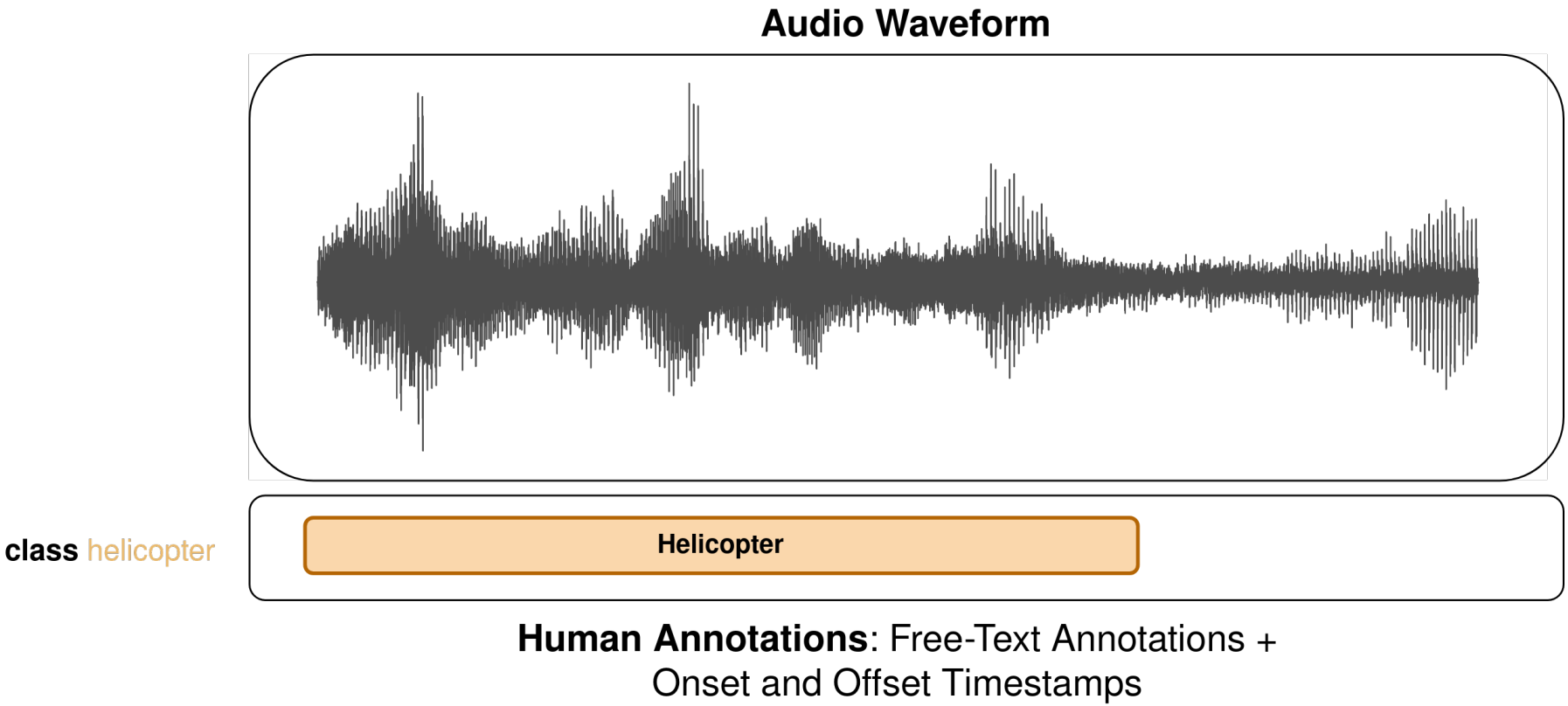
Matching Labels to Annotations, Approach 3



Human Annotations: Free-Text Annotations +
Onset and Offset Timestamps

- We found two related instances for class *helicopter* in this audio recording

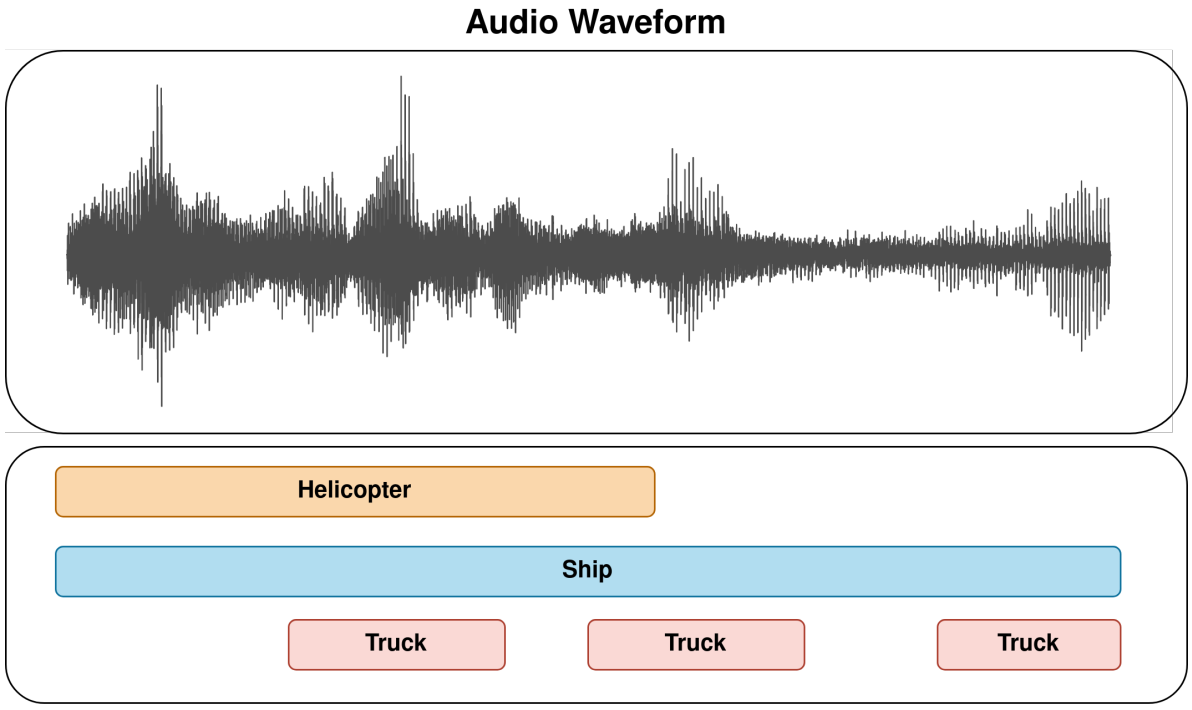
Matching Labels to Annotations, Approach 3



Human Annotations: Free-Text Annotations +
Onset and Offset Timestamps

- We can convert the two free-text annotations to positive instances for class *helicopter*

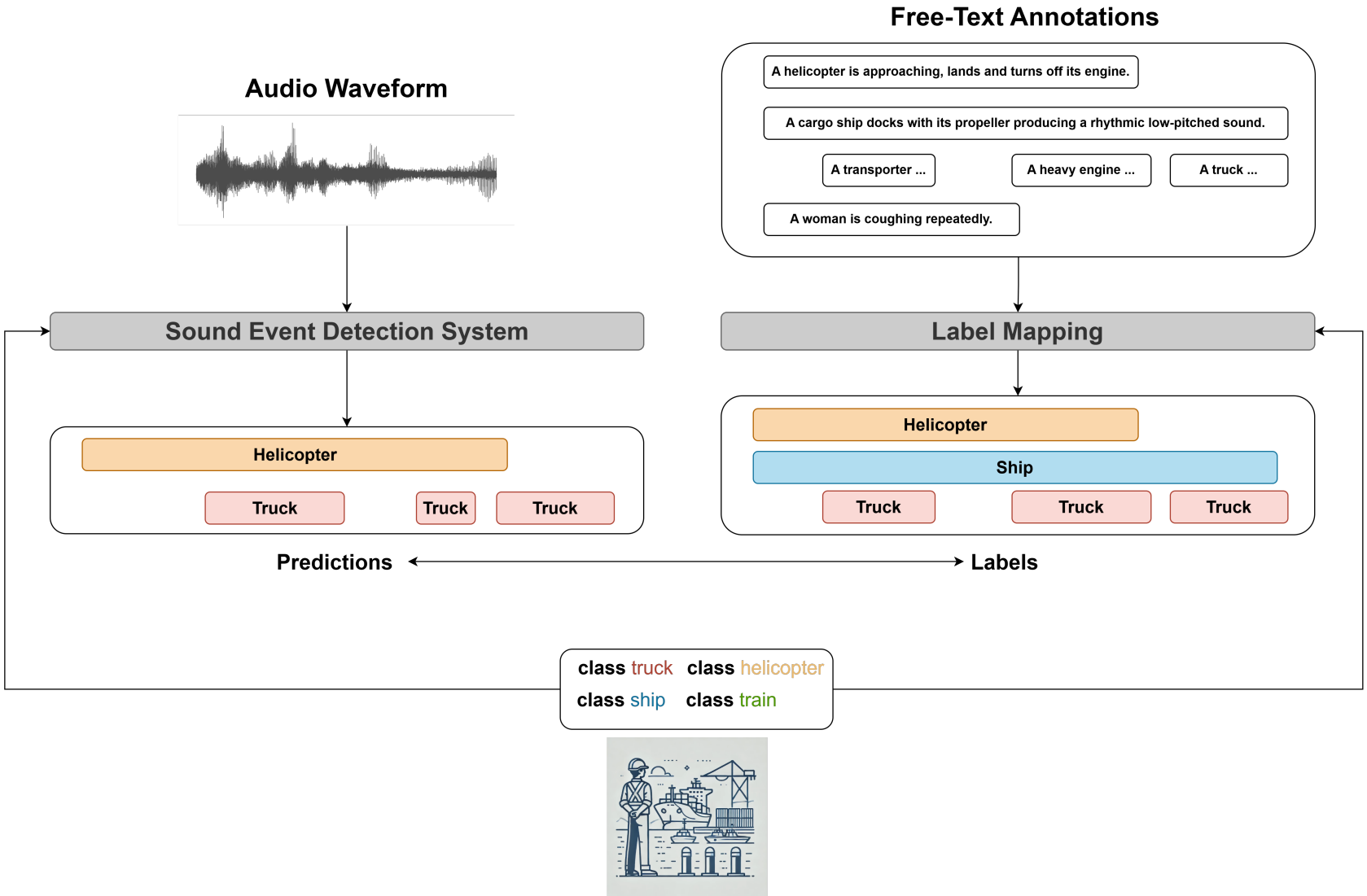
Matching Labels to Annotations, Approach 3



Human Annotations: Sound Event Labels
+ Onset and Offset Timestamps

- Back to where we started from!
- But: Extracted customer-specific from general dataset

The Full Picture



PROJECT PHASES & SCHEDULE

Project Phases

The project is structured in four phases:

1. Data Annotation: Each of you will annotate 40 audio recordings (between 15 and 30 seconds) with free text, including onset and offset timestamps.

Project Phases

The project is structured in four phases:

1. Data Annotation: Each of you will annotate 40 audio recordings (between 15 and 30 seconds) with free text, including onset and offset timestamps.
2. Data Exploration: We hand out precomputed audio features, annotations, and their corresponding embeddings. In teams of 4, you will perform an initial explorative analysis.

Project Phases

The project is structured in four phases:

1. Data Annotation: Each of you will annotate 40 audio recordings (between 15 and 30 seconds) with free text, including onset and offset timestamps.
2. Data Exploration: We hand out precomputed audio features, annotations, and their corresponding embeddings. In teams of 4, you will perform an initial explorative analysis.
3. Classification Phase: With your team, you will train, tune and compare different classifiers to detect events of interest for one particular customer.

Project Phases

The project is structured in four phases:

4. Challenge Phase: The costumer provides a cost matrix and a test set. You will try to recognize the events of interest as accurately as possible to minimize costs.

Project Schedule

		Date/Deadline
Meeting 1	Introduction, explain Tasks 0 and 1	March 10 ◀
Task 0	Form teams	March 24
Task 1	Data Annotation	March 24
Meeting 2	Release dataset, explain Task 2	April 7
Task 2	Data Exploration	April 24
Meeting 3	Discuss results, explain Task 3	April 28
Task 3	Classification Experiments	May 22
Meeting 4	Present results, release test data, explain Task 4	May 26
Task 4	The Challenge	June 18
Meeting 5	Final presentations	June 23

Tentative Schedule for Tutorial Sessions

		Date/Deadline
Tutorial 1	Python, NumPy, Matplotlib	March 31
Tutorial 2	data splits, Sklearn classifiers	May 05
Tutorial 3	PyTorch, advanced topics	June 02

DELIVERABLES & GRADING

Deliverables

- All tasks are **mandatory** to pass the course
- Each team submits a **report** on each of the team tasks: exploration, classification, challenge
- You will get a **set of questions** that you will address in your reports
- We assign you one of the questions to briefly summarize on **slides**. Your team may be chosen to present these slides in the next exercise session.

Grading

Grades will be based on the reports and slides:

- Annotation: max. 10 points
- Exploration: max. 25 points
- Classification: max. 40 points
- Challenge: max. 30 points

Grade boundaries:

- $\geq$87.5% is a 1
- $\geq$75% is a 2
- $\geq$62.5% is a 3
- $\geq$50% is a 4

Report & Slide Details

Report:

- Answer different questions for a number of topics
- Brief statement about contribution of team members
- Point deductions for:
  late submissions, modified template, exceeded page-limits

Slides:

- Create slides about 1 assigned question or topic
- Points awarded for submitted slides

Evaluation of Report

For every topic, we look at:

1. Thoroughness & Completeness Have you thought about the problem and answered every question?
2. Clarity Are the ideas, features, algorithms, and results described clearly? Based on your descriptions, could the reader reproduce your experiments?
3. Presentation Did you select an appropriate way of communicating your results, e.g., use meaningful and helpful plots?
4. Correctness Is the proposed procedure / experiment sound, correct?

TASK 1: DATA ANNOTATION

Prerequisite: The (General) Dataset

- Downloaded from the *freesound* platform
- Specified a broad range of search terms to collect *diverse* audios
- Filtered part of low-quality, synthetic, or offensive recordings
- Ended up with 15,642 audio recordings of length between 15 and 30 seconds (~100 hours of audio)
- Downloaded also metadata such as *tags* and *descriptions* that are attached to audios

Data Annotation Task: Summary

Deadline: Monday, March 24th, 23:59 (so we can prepare the dataset until Monday, April 7th)

- Step 1: Read the annotation guidelines document on Moodle
- Step 2: Create a LabelStudio account **after** receiving an invitation via your student email
- Step 3: Familiarize yourself with all elements of the annotation interface
- Step 4: Annotate **the first** 40 audio recordings in your queue

Detailed instructions on Moodle!

Step 1: Annotation Guidelines

Describe all audible sound events in an audio recording and mark onset and offset timestamps.

> Sounds simple?

Step 1: Annotation Guidelines

Describe all audible sound events in an audio recording and mark onset and offset timestamps.

> Sounds simple?

Q: What counts as a sound event?

A: Everything that you can hear. If you would subtract all your free-text annotations from the audio file it should be silent.

Step 1: Annotation Guidelines

Q: How should I **structure** my free-text annotations?

A: A **single** sentence containing:

- **Source** (*mandatory*): Airplane, Dog, Person, ...
- **Action** (*mandatory*): flying overhead, barking, walking on gravel, ...
- **Descriptor**: loud, metallic, rhythmic, ...
- **Temporal**: fading, repeating, gradually increasing, ...
- **Context**: indoors, echoing, distant, ...

Example: *"A telephone rings loudly in a steady pattern inside a room."*

Step 1: Annotation Guidelines

Q: Should I use **general or specific** terminology? (Animal $\rightarrow$ Bird $\rightarrow$ Woodpecker)

A: Use the most specific term possible. However, if you're unsure about the more specific term, it's better to choose a more general one.

Step 1: Annotation Guidelines

Q: What should I do if I **can't identify** the sound source?

A: Take a look at the **metadata** (*tags* and *descriptions*) provided via the annotation interface. However, be **cautious**: metadata may be incomplete or incorrect.

Step 1: Annotation Guidelines

Q: How should I handle **overlapping** sound events?

A: Provide a separate description for each sound source. **Do not** combine them into a single description using words like *while*, *after*, or *before*.

- **Incorrect:** *"A dog is barking and people are talking."*
- **Correct:** *"A dog barking loudly."* and *"Several people talking indoors, with overlapping voices."*

Step 1: Annotation Guidelines

Q: How should I annotate a sound event that lasts for the **entire** audio file?

A: Set the **onset** at the very **beginning** and the **offset** at the very **end**. This is a common occurrence, not an exception.

Step 1: Annotation Guidelines

Q: How should I annotate a sound source that occurs **repeatedly**?

A: If a sound occurs repeatedly with perceptible pauses in between, each instance should ideally be annotated separately. However, if the pauses are shorter than 1 second, you may merge them into a single annotation.

Exceptions:

- Do **not** separate individual words within a spoken sentence.
- Do **not** split naturally interrupted events. For example, an alarm may consist of repeated tones with short pauses in between—these should be treated as a single event.

Step 1: Annotation Guidelines

Q: Is the annotation task **mandatory?**

A: **Yes**, you have to complete the annotation task to receive a positive grade.

Step 1: Annotation Guidelines

Q: Can I get **point deductions** on the annotation task?

A: You will receive full points if you stick to the annotation guidelines on Moodle and annotate in a best-effort manner.

Step 1: Annotation Guidelines

Q: Can I **fail** the annotation task?

A: **No**. However, if your annotations are clearly not "best-effort" or deviate from the instructions, we will ask you to repeat the *full* annotation task.

Step 1: Annotation Guidelines

Q: Can I skip audio files in my queue?

A: No, you must annotate the **first 40 files** in your queue. However, if you find an audio file **disturbing** or **problematic**, you may leave a comment and skip it. If you skip a file, you must annotate the **next one** in line (e.g., if you skip one, you must annotate the 41st file in your queue).

Step 3: Annotation Interface

Live Demonstration

Step 4: Annotate Recordings

Live Demonstration

TASK 0: FORM TEAMS

Teaming Up

- Except for Task 1, all tasks are to be done in teams of 4 people
- To find teammates, we have set up a separate forum on Moodle
- Once you found teammates, sign up as a team on Moodle

Teaming Up

- Except for Task 1, all tasks are to be done in teams of 4 people
- To find teammates, we have set up a separate forum on Moodle
- Once you found teammates, sign up as a team on Moodle
- Teams are predefined, simply choose one that is empty
- Deadline is Monday, March 24th, 23:59

MOODLE

Moodle

- Zoom link, recordings, slides: Moodle
- Task 1: Moodle
- Team registration: Moodle
- Report submission: Moodle
- Questions: Forum on Moodle
- Python intro: Moodle
- Inquiries that need to stay secret or cannot be of interest to anybody else: Email

Moodle

- Zoom link, recordings, slides: Moodle
- Task 1: Moodle
- Team registration: Moodle
- Report submission: Moodle
- Questions: Forum on Moodle
- Python intro: Moodle
- Inquiries that need to stay secret or cannot be of interest to anybody else: Email (listed on Moodle)