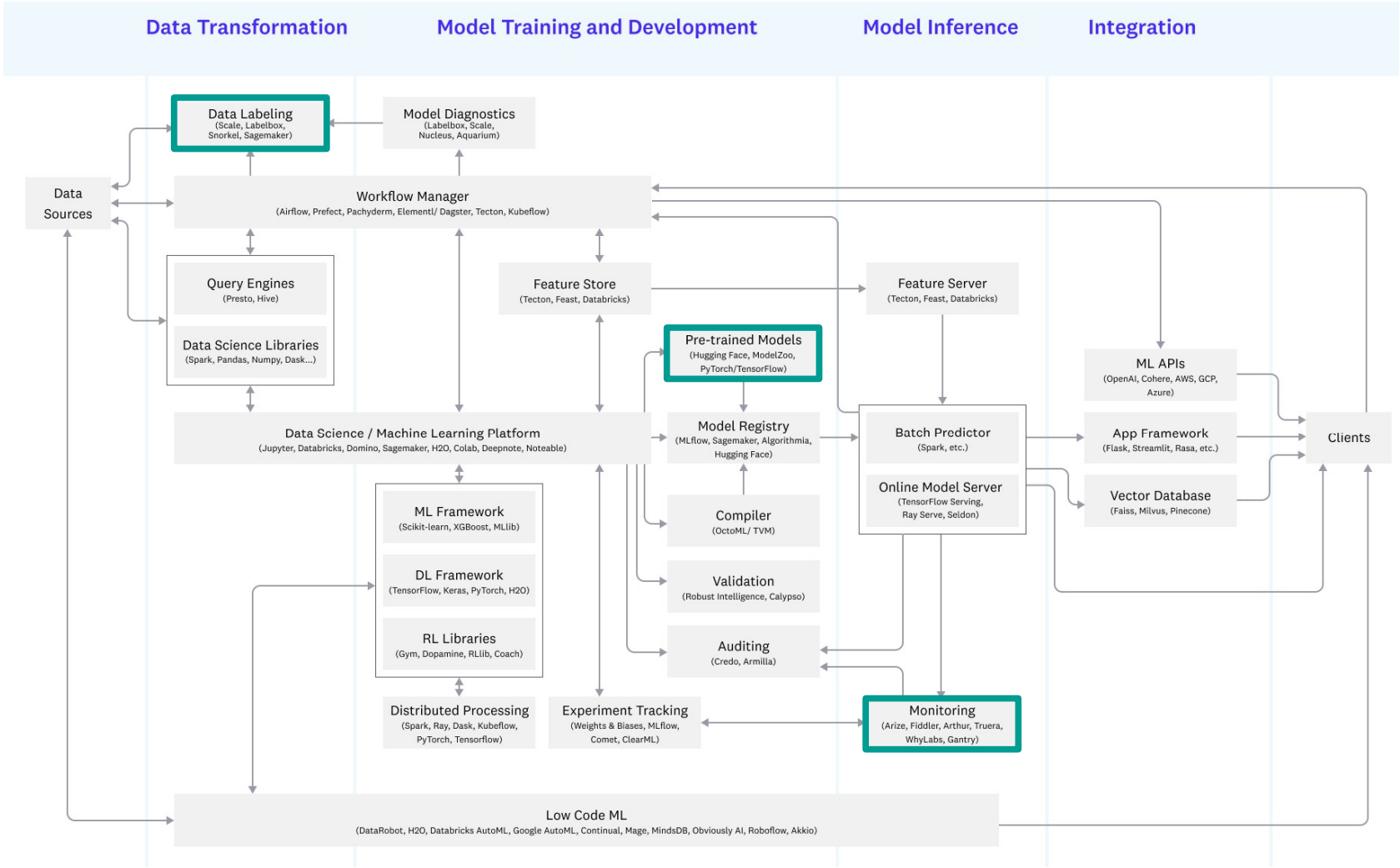
Hochschule Konstanz
Fakultät Informatik

MSI-Seminar MLOps

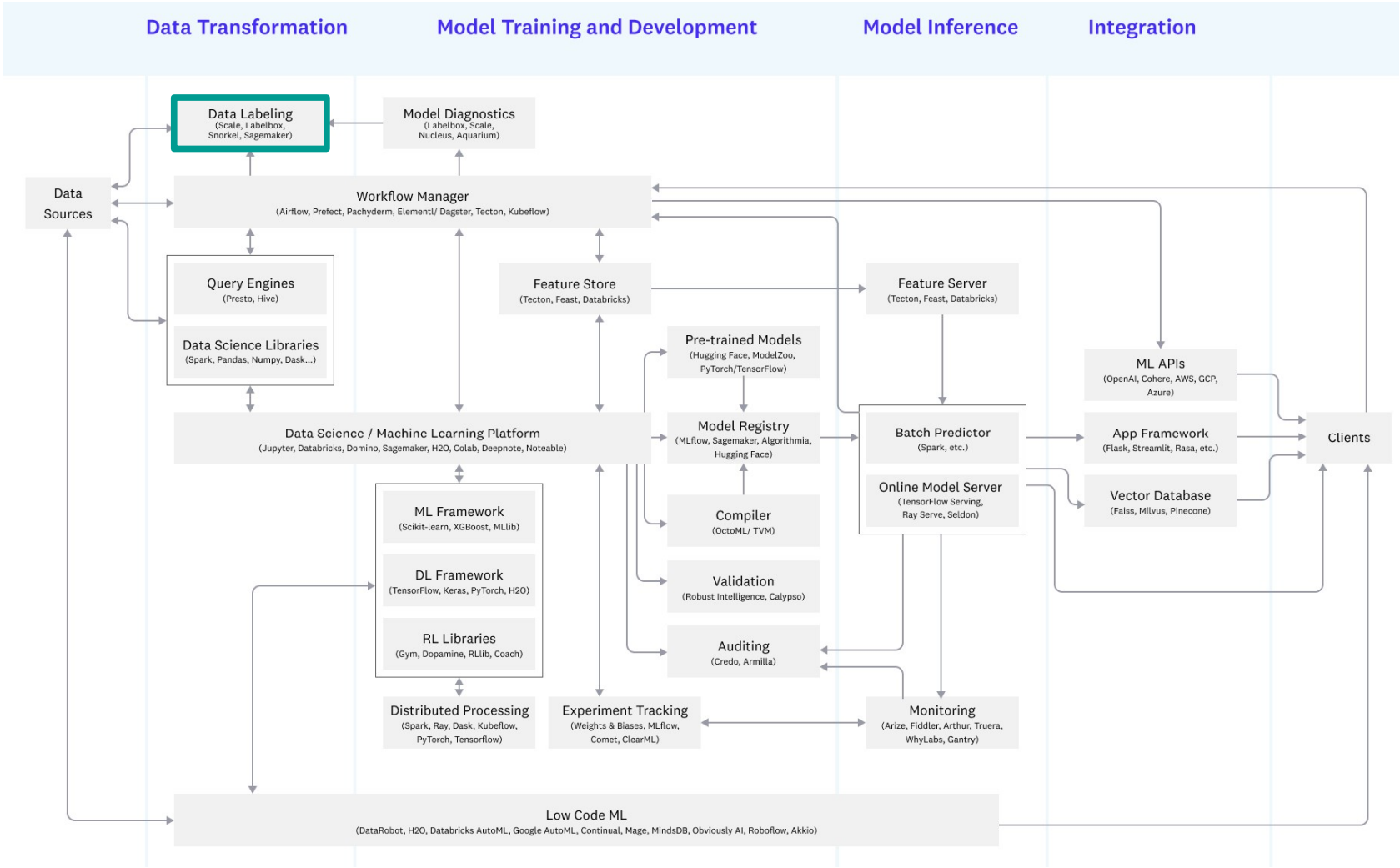
Data Labeling

Leon Auer

Machine Learning Infrastructure (2.0)



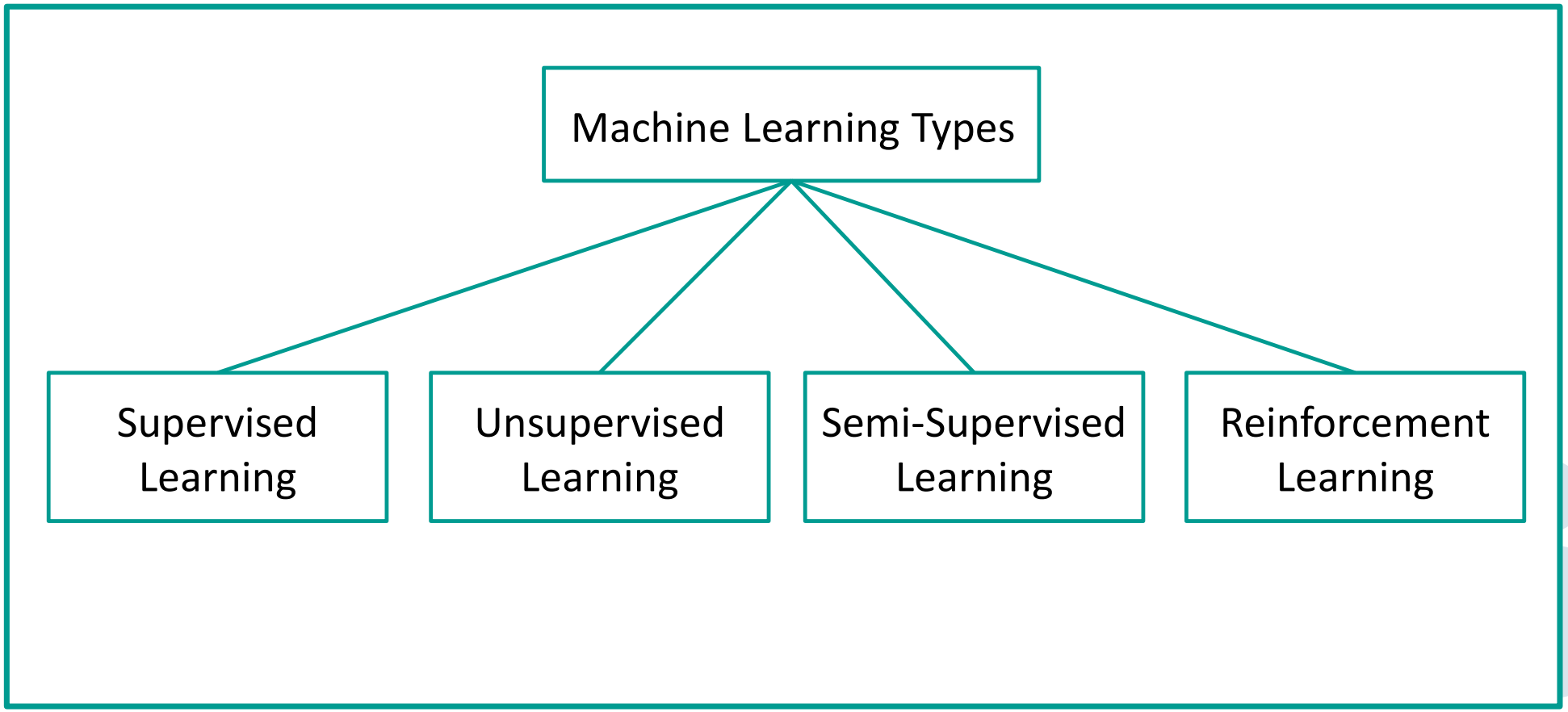
Machine Learning Infrastructure (2.0)



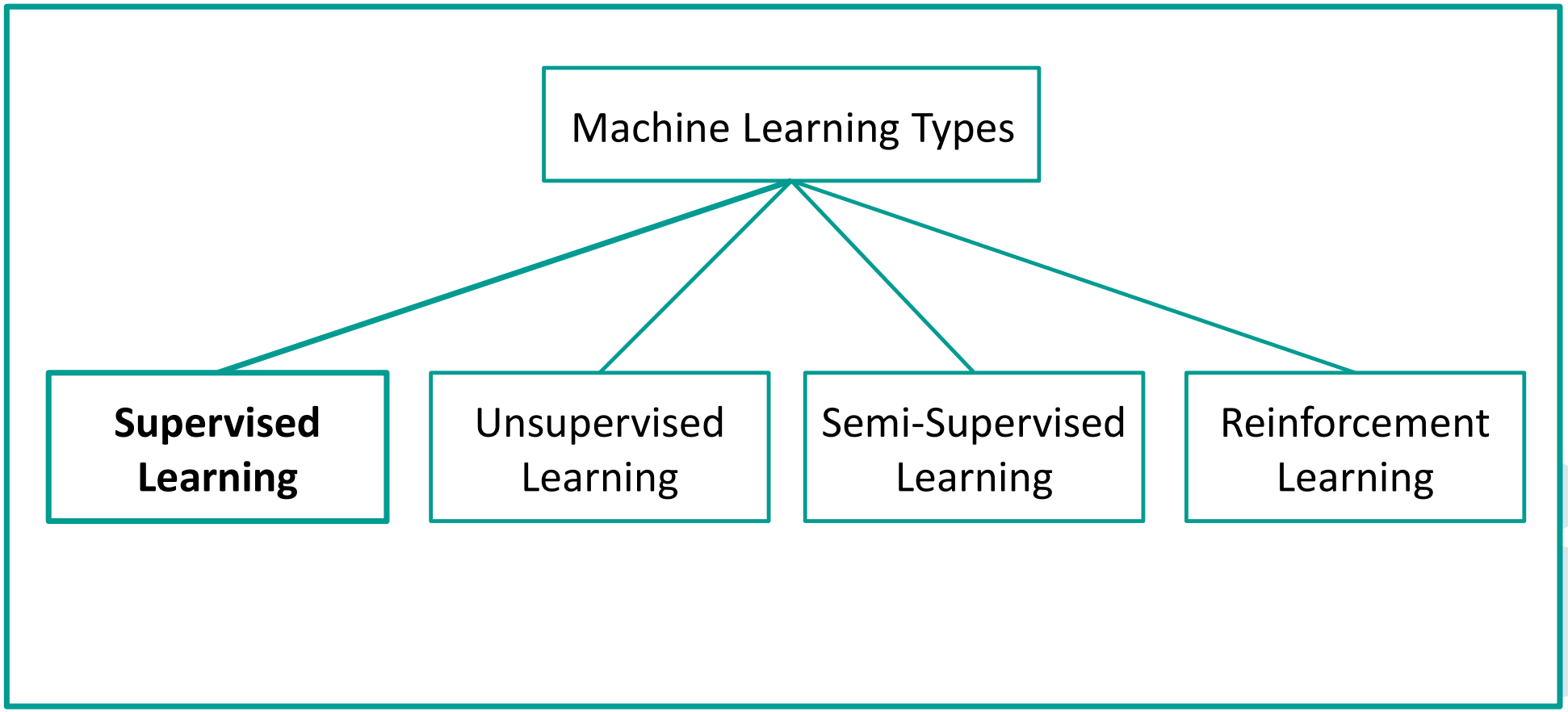
Agenda


Basics &
Key steps
Types &
Approaches
Challenges &
Best practices
Tools &
Demo

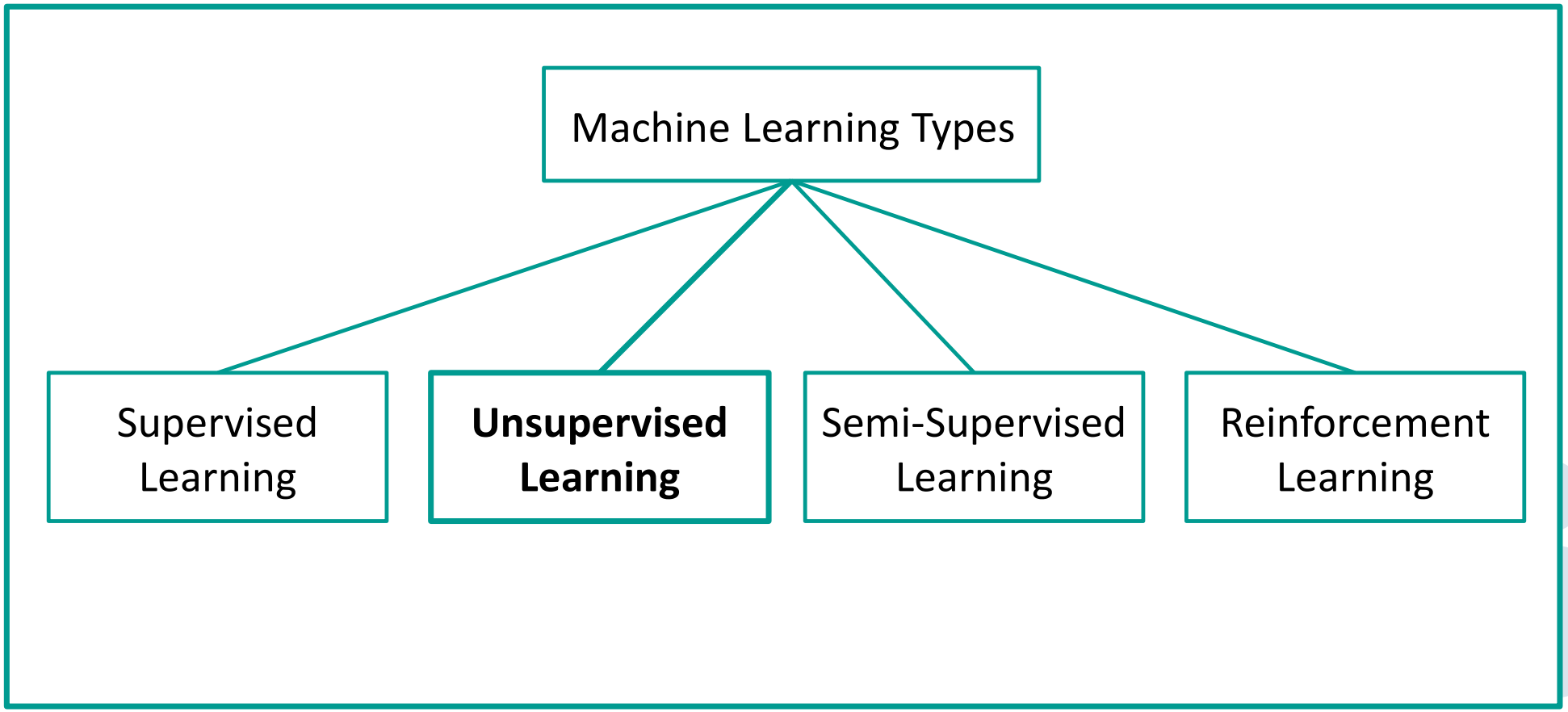
Types of Machine Learning



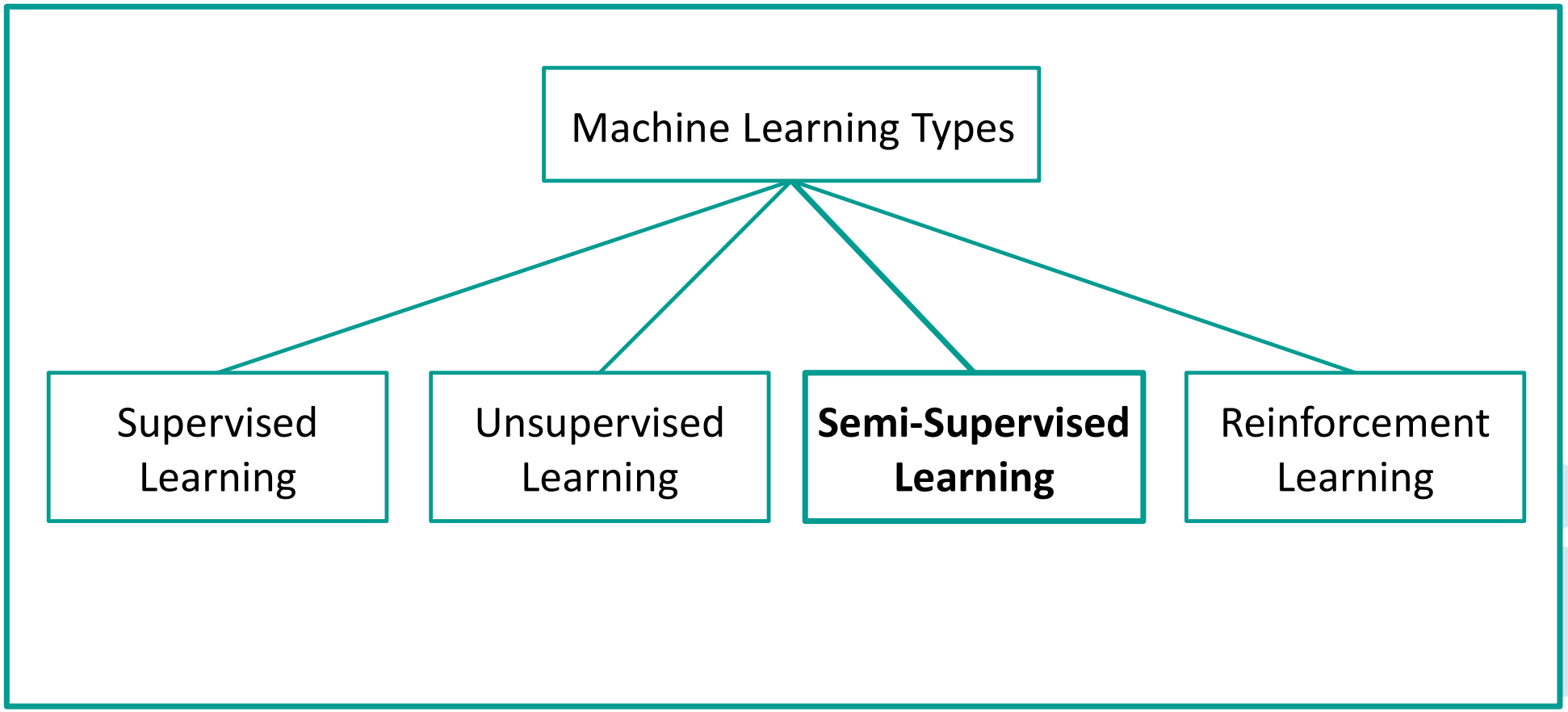
Types of Machine Learning



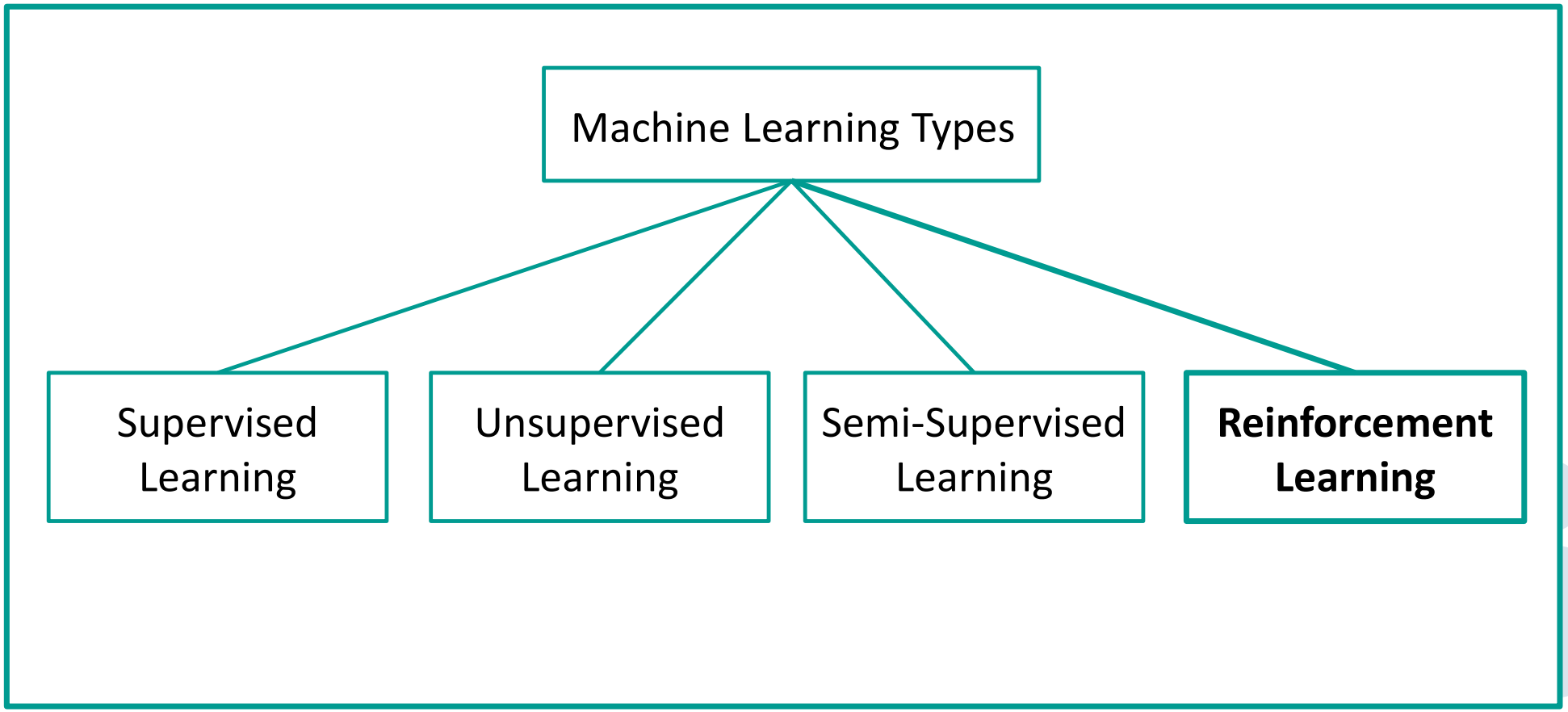
Types of Machine Learning



Types of Machine Learning

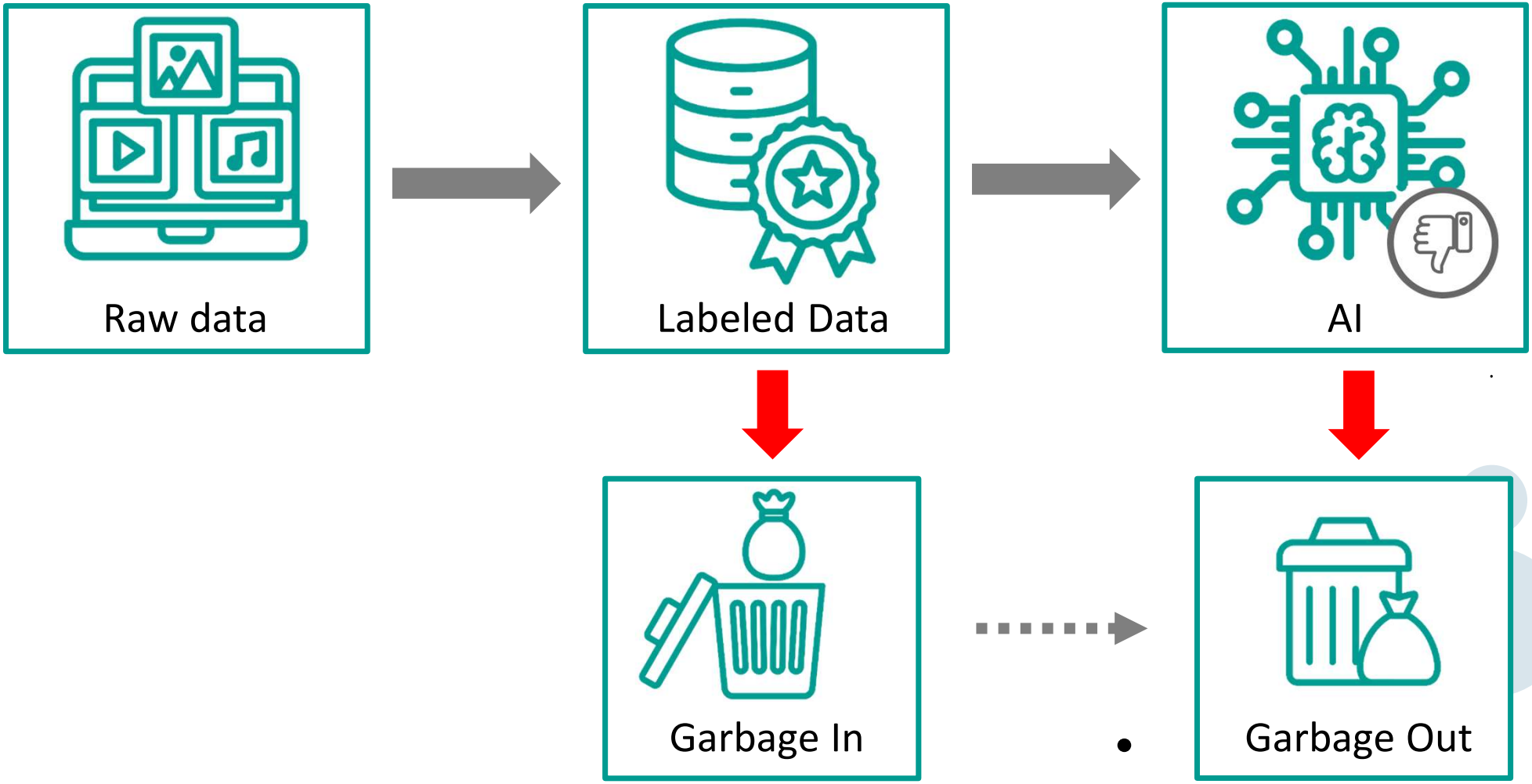


Types of Machine Learning



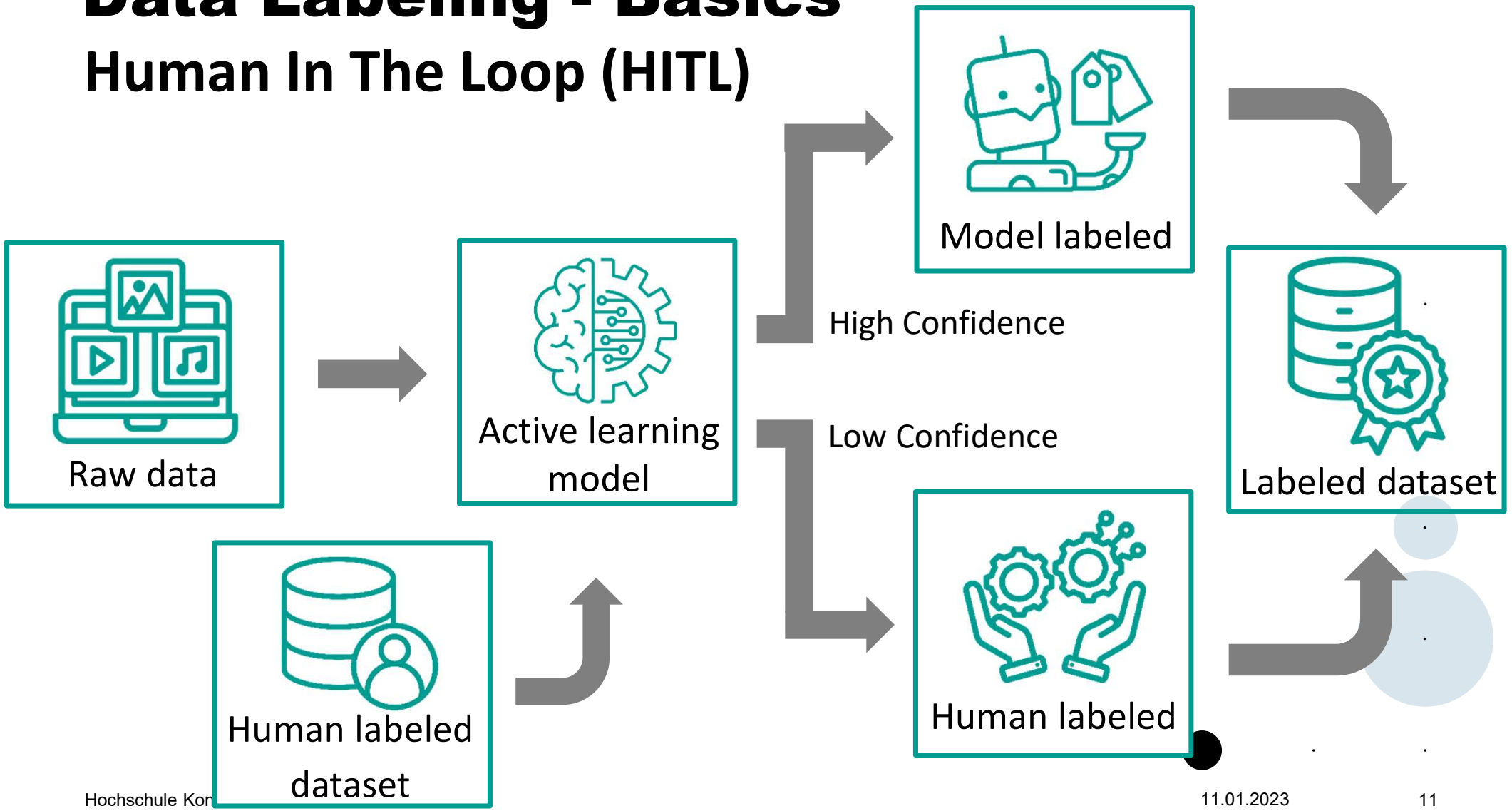
Data Labeling - Basics

Data Labeling - Basics

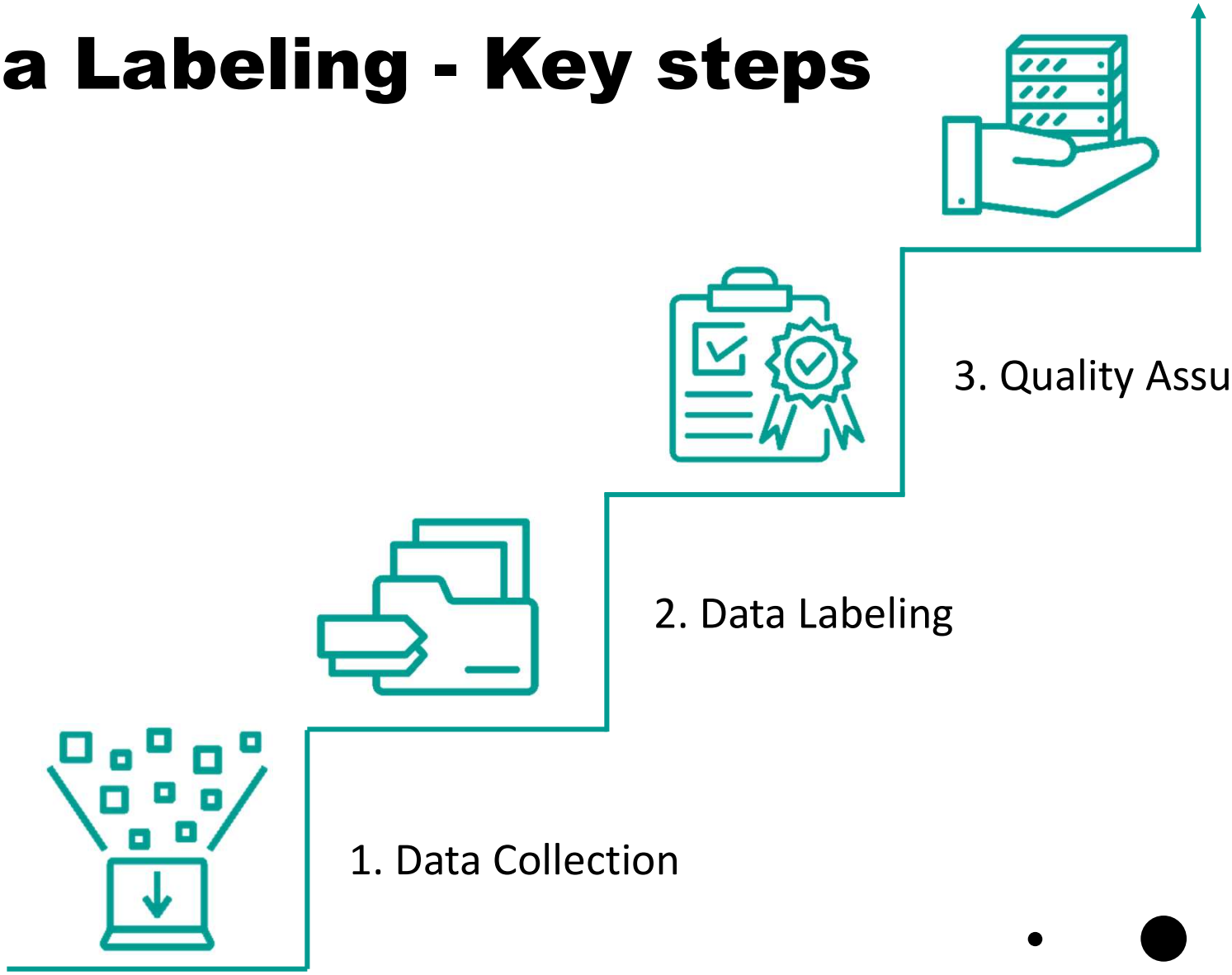


Data Labeling - Basics

Human In The Loop (HITL)

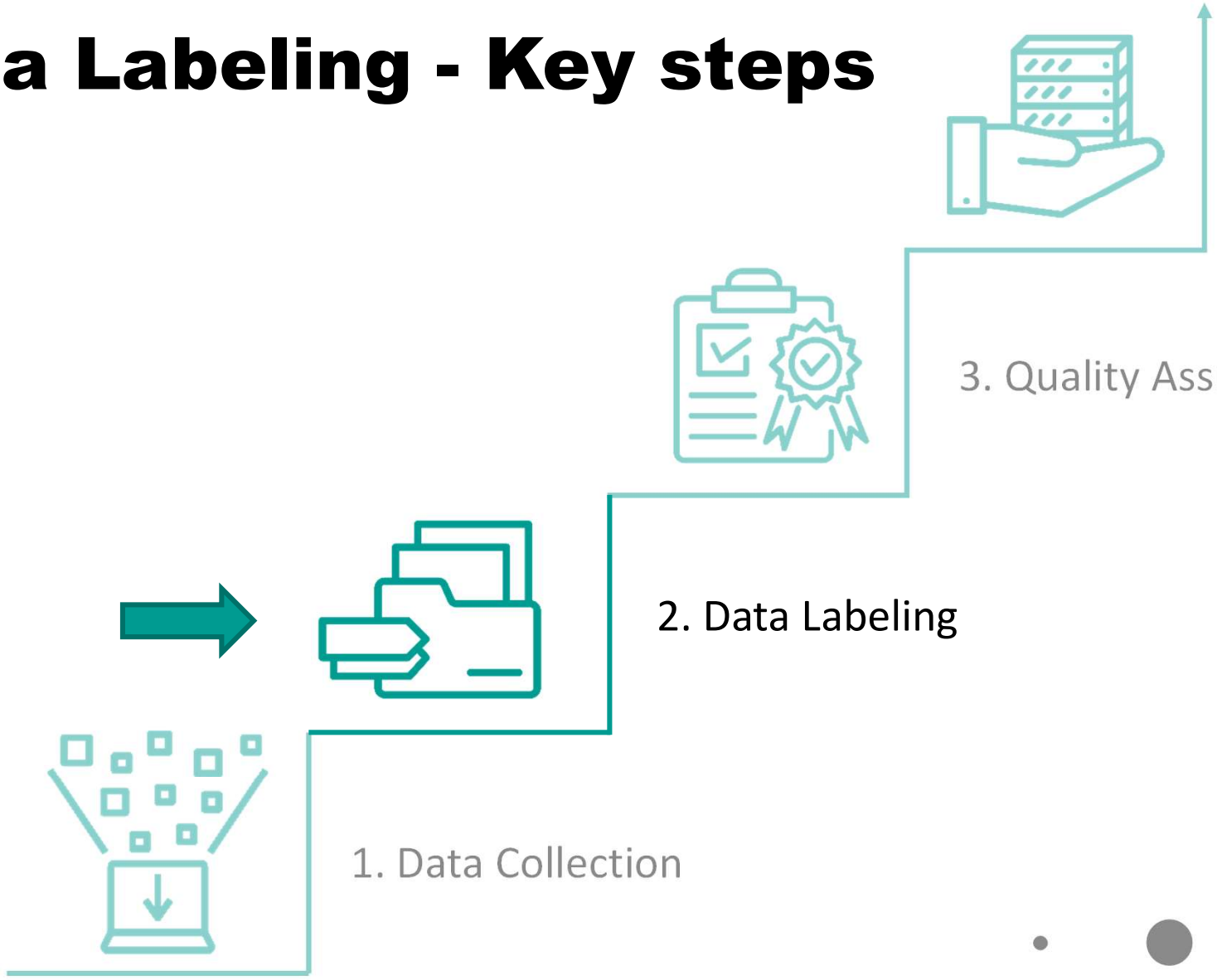


Data Labeling - Key steps

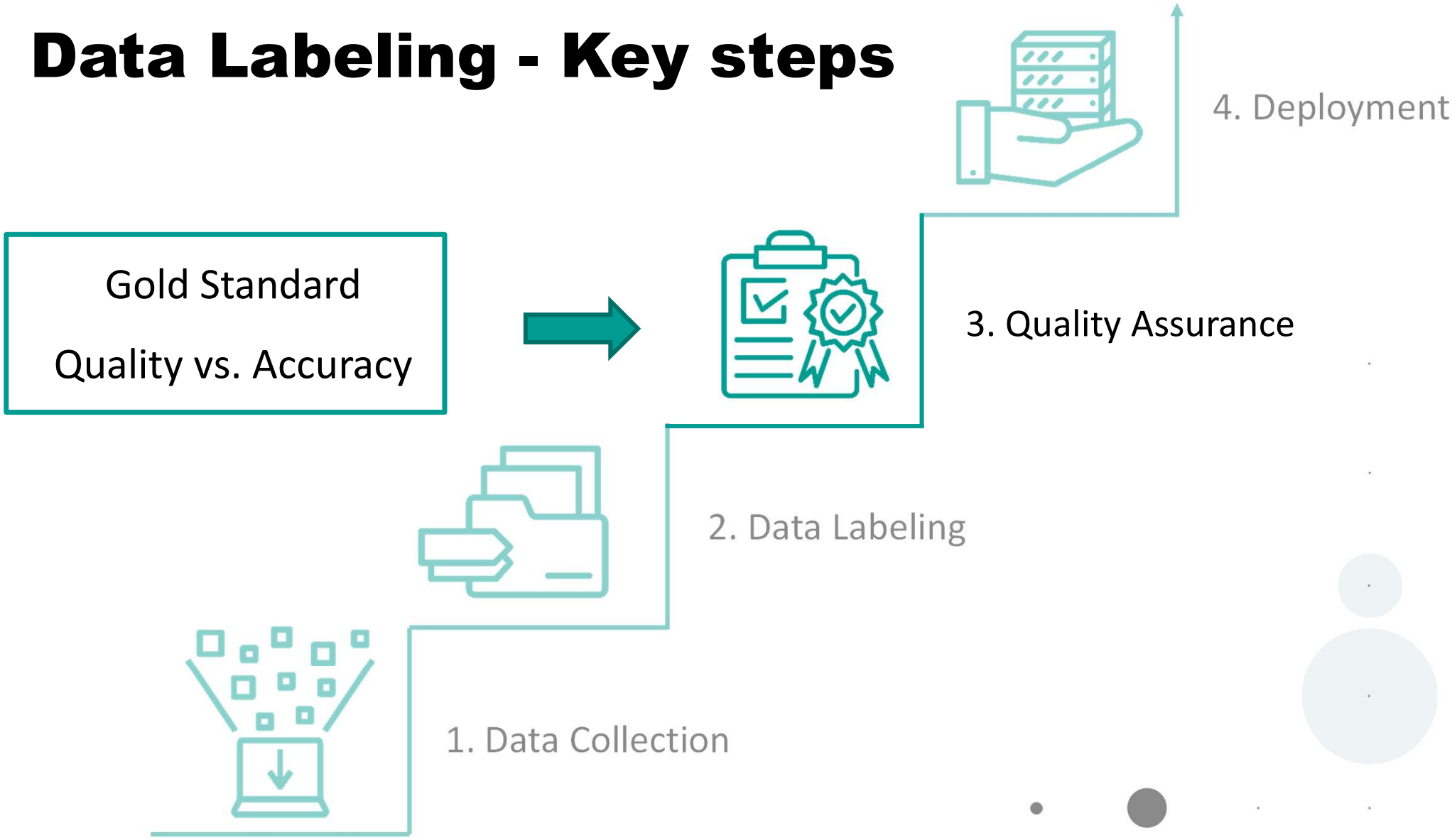


Data Labeling - Key steps

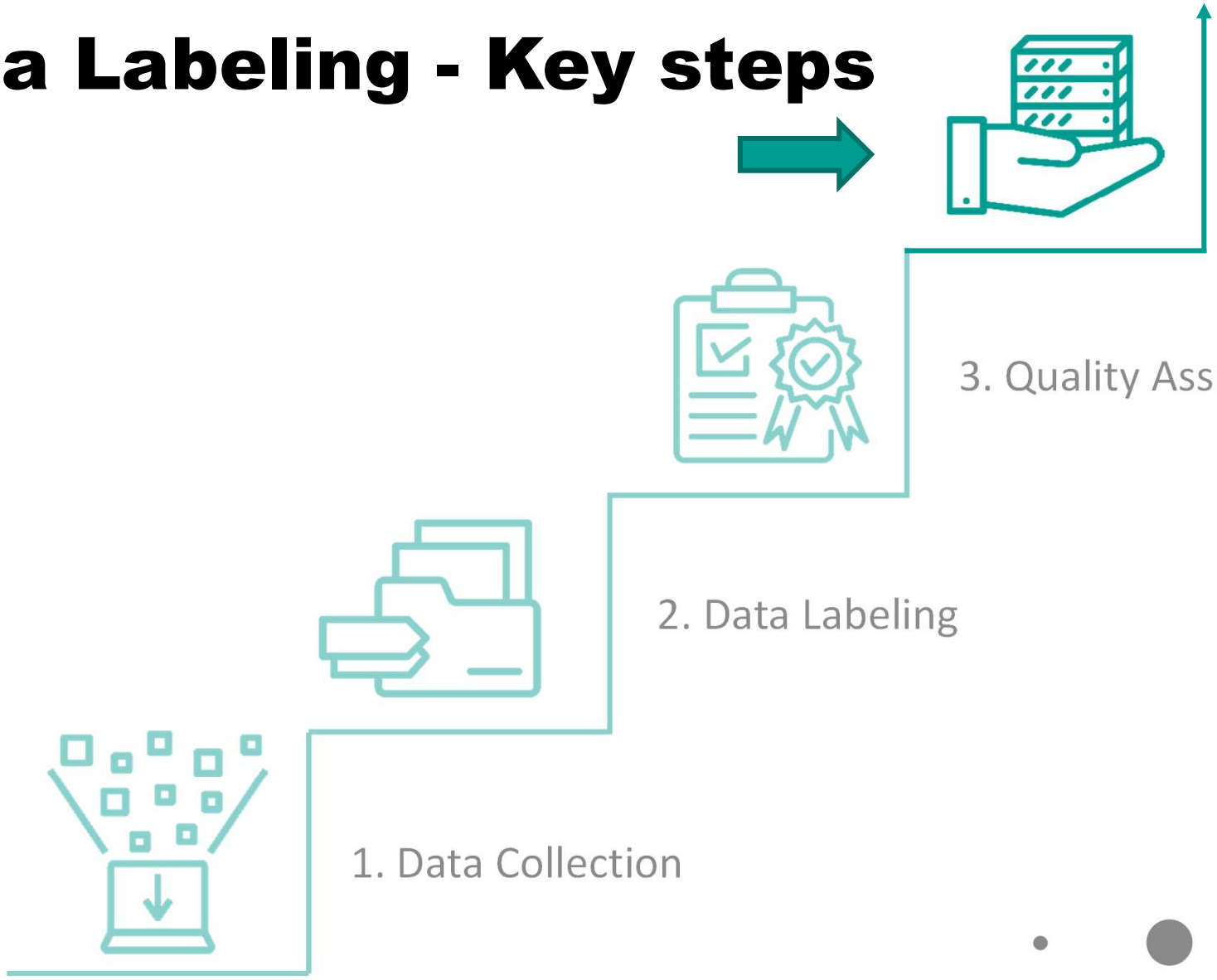
Data Labeling - Key steps



Data Labeling - Key steps

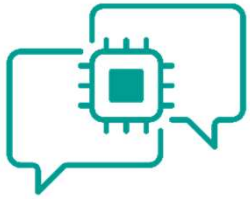


Data Labeling - Key steps



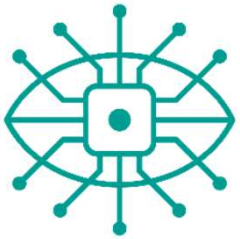
Data Labeling - Types

Natural Language Processing (NLP)



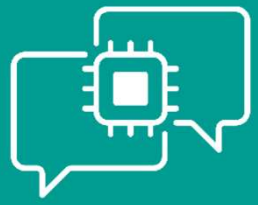
Audio Processing

Computer Vision (CV)



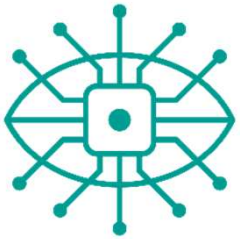
Data Labeling - Types

Natural Language Processing (NLP)



Audio Processing

Computer Vision (CV)

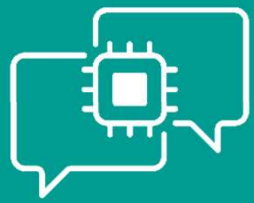


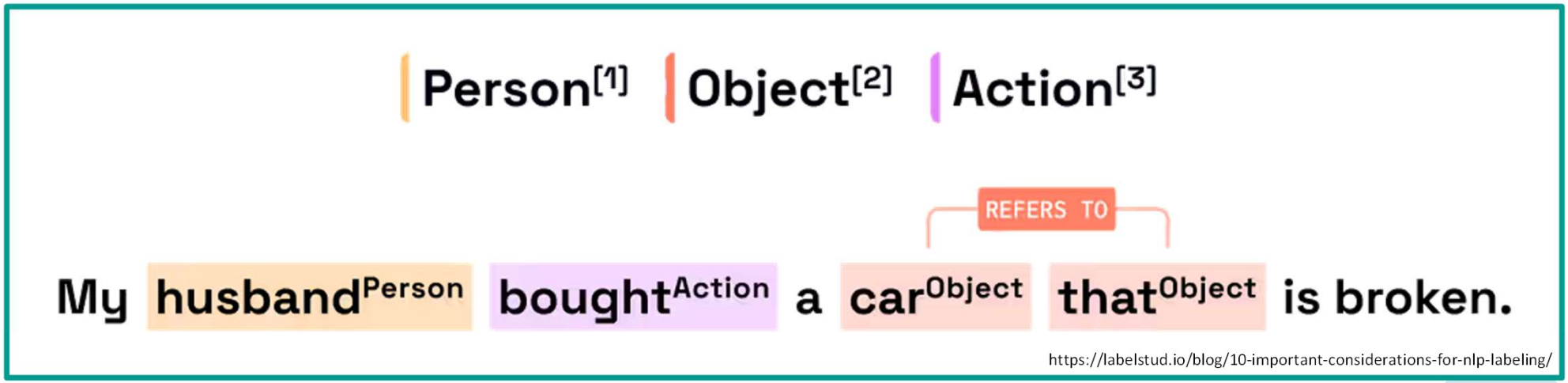
Text Labeling

- Text Classification
- Phonetic Labeling
- Entity Labeling & Linking

Data Labeling - Types

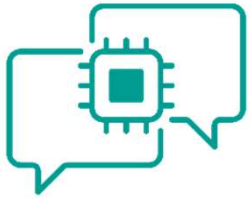
Natural Language Processing (NLP)





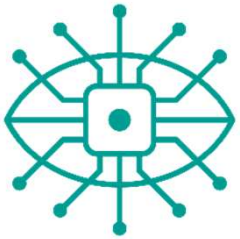
Data Labeling - Types

Natural Language Processing (NLP)



Audio Processing

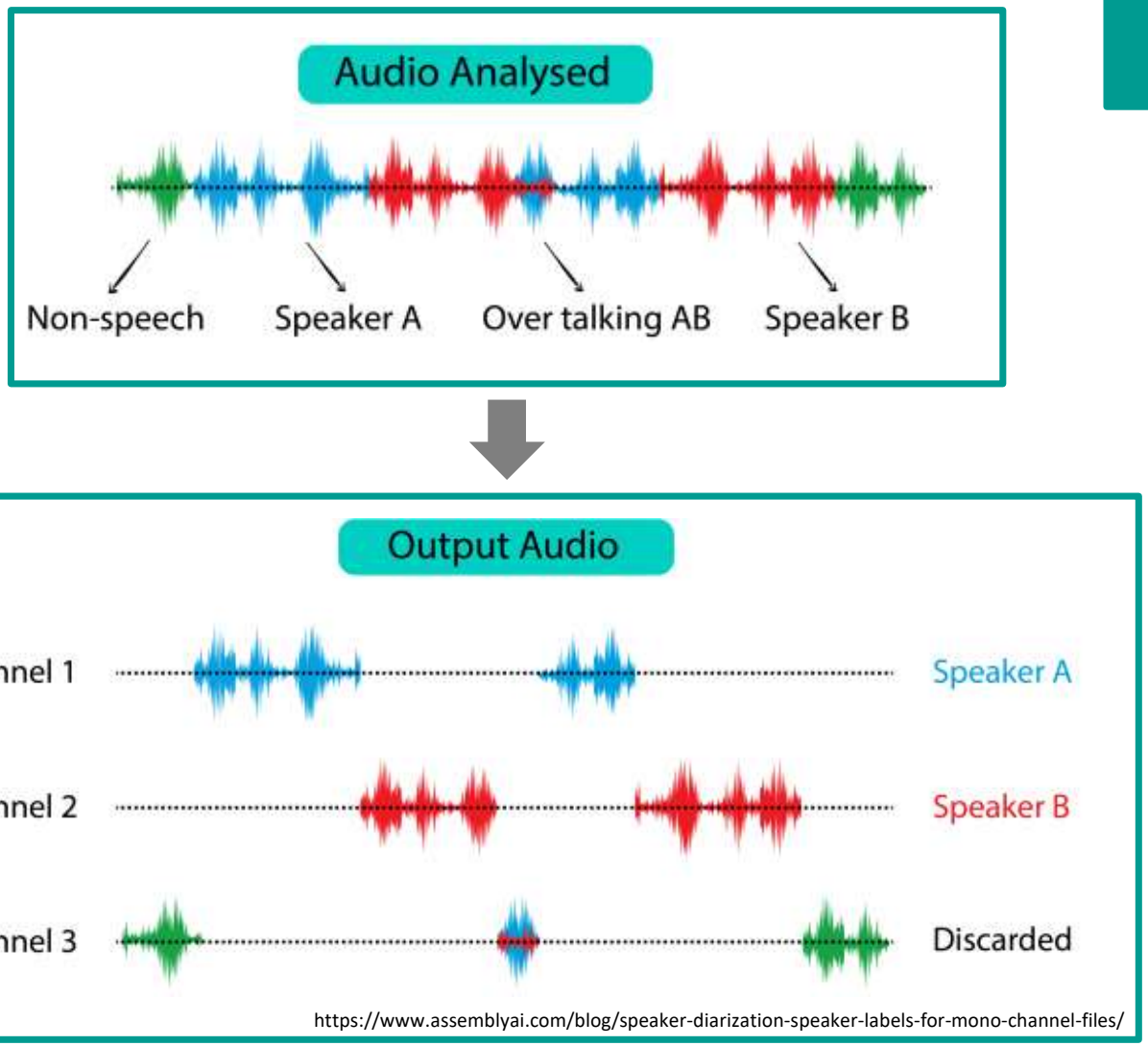
Computer Vision (CV)



Audio Labeling

- Audio Classification
- Audio Segmentation
- Audio Transcription

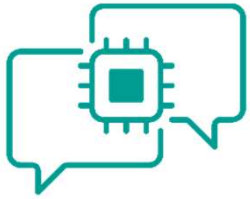
Data Labeling - Types



https://www.assemblyai.com/blog/speaker-diarization-speaker-labels-for-mono-channel-files/

Data Labeling - Types

Natural Language Processing (NLP)



Audio Processing

Computer Vision (CV)

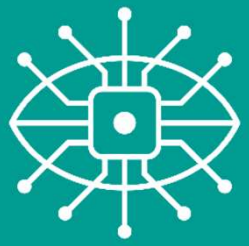


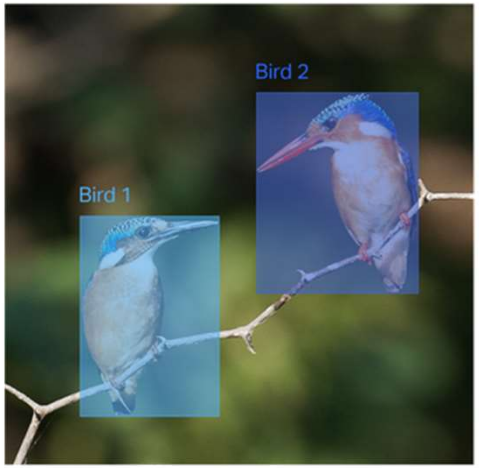
Image Labeling

Video Labeling

3D Labeling

Data Labeling - Types

Classification



Object detection

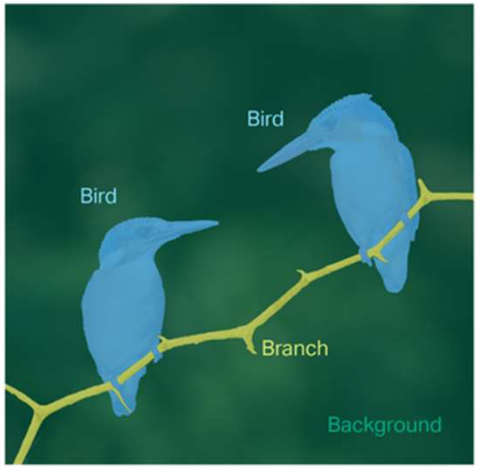


Image segmentation

https://labelbox.com/guides/image-segmentation/

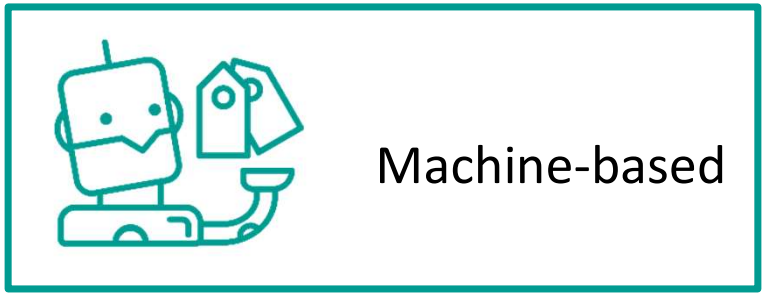
Data Labeling - Types

Computer Vision (CV)

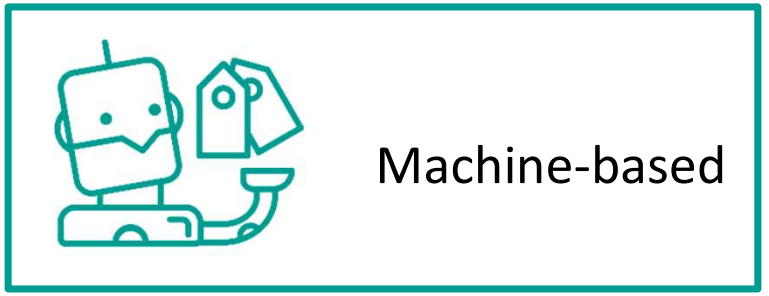
https://www.anolytics.ai/3d-cuboid-annotation/

https://developers.google.com/ml-kit/vision/pose-detection

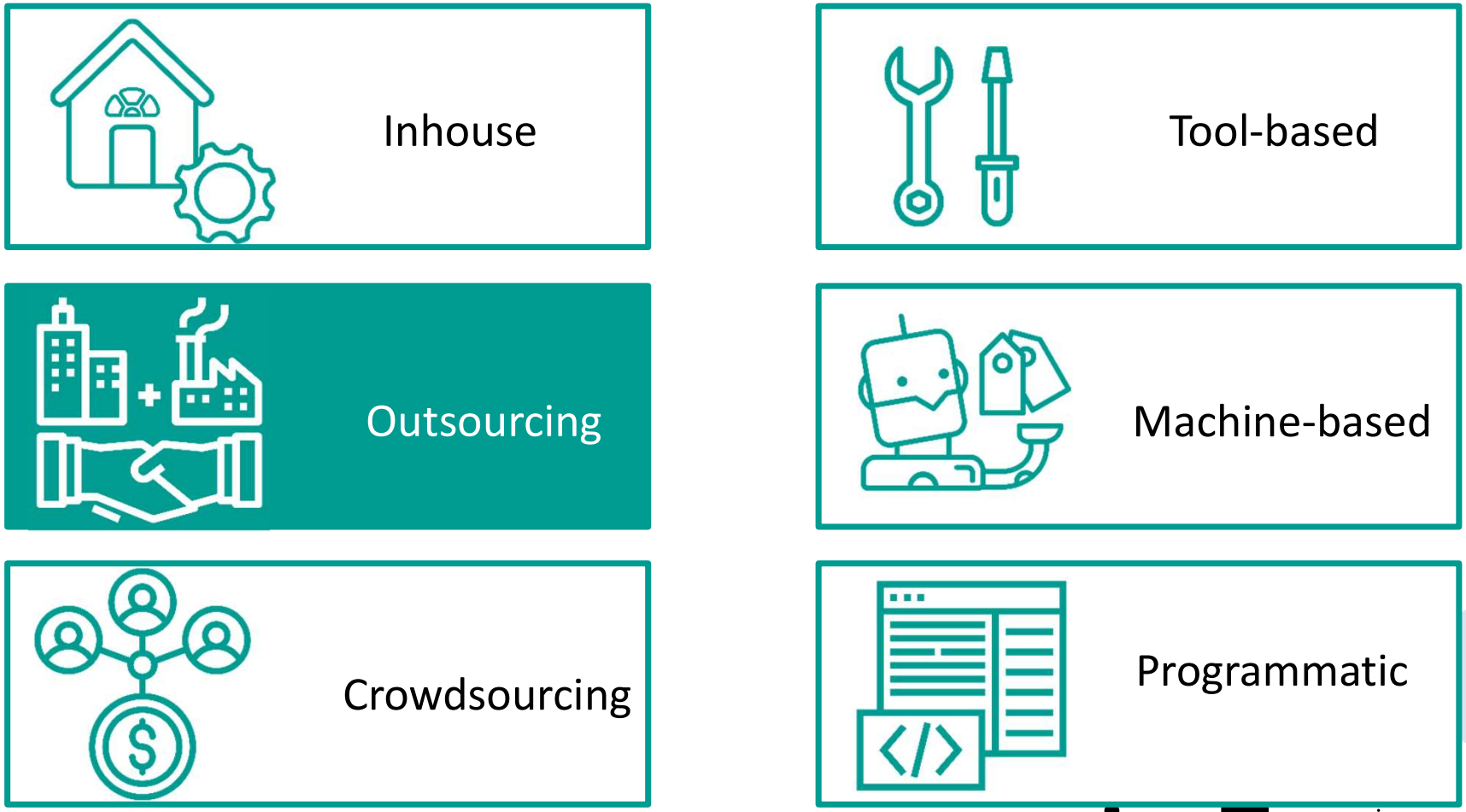
Data Labeling - Approaches



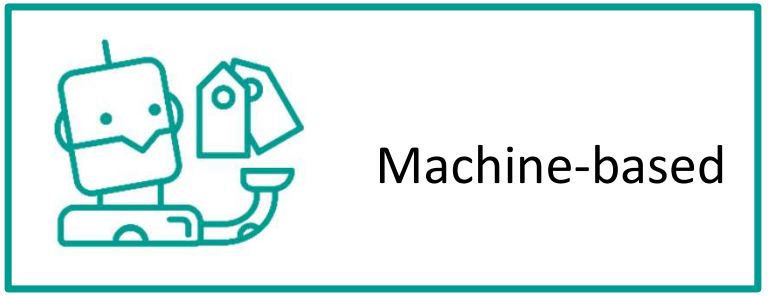
Data Labeling - Approaches



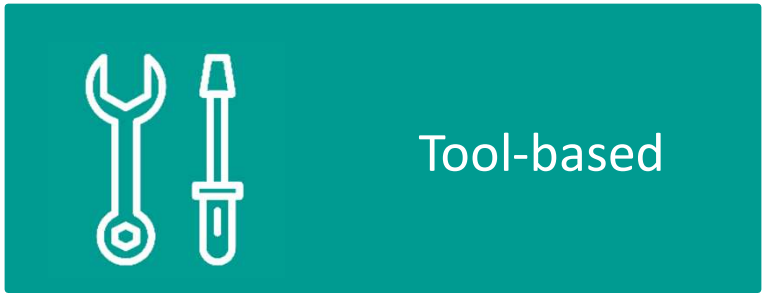
Data Labeling - Approaches

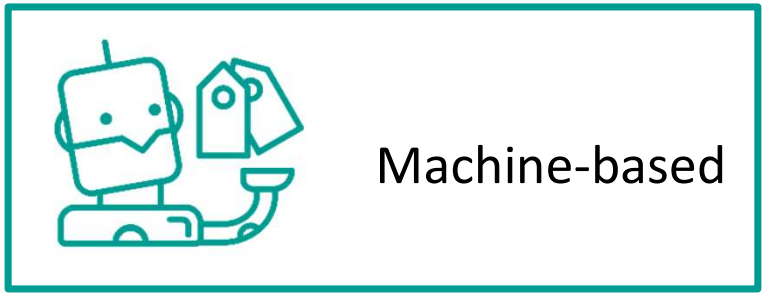


Data Labeling - Approaches

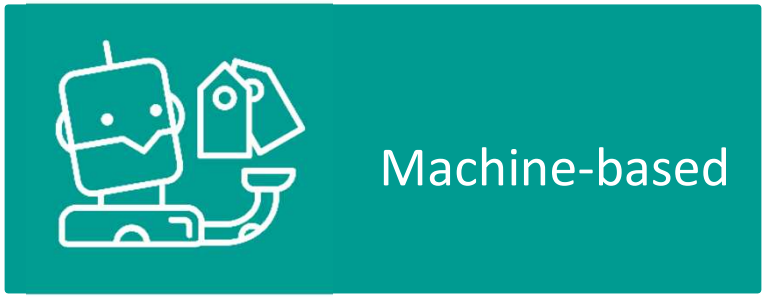


Data Labeling - Approaches

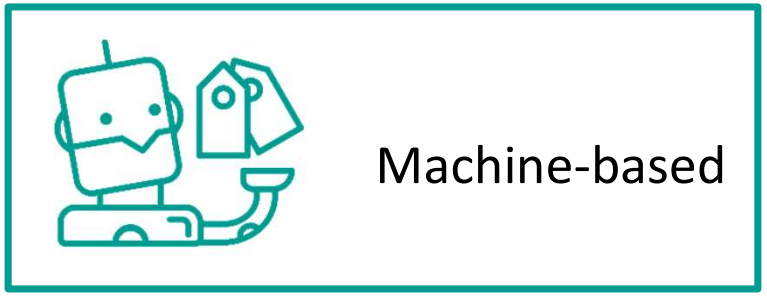




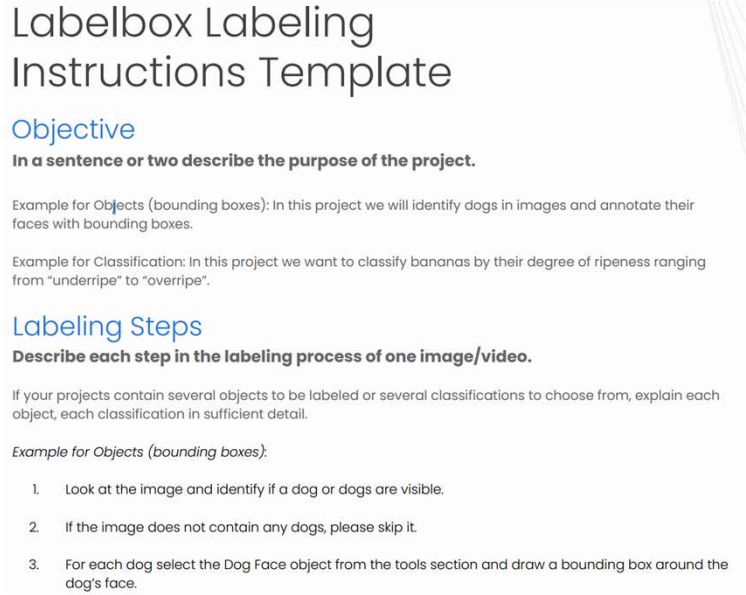
Data Labeling - Approaches



Data Labeling - Approaches



Data Labeling - Best practices



Labelbox Labeling Instructions Template

Objective

In a sentence or two describe the purpose of the project.

Example for Objects (bounding boxes): In this project we will identify dogs in images and annotate their faces with bounding boxes.

Example for Classification: In this project we want to classify bananas by their degree of ripeness ranging from "underripe" to "overripe".

Labeling Steps

Describe each step in the labeling process of one image/video.

If your projects contain several objects to be labeled or several classifications to choose from, explain each object, each classification in sufficient detail.

Example for Objects (bounding boxes):

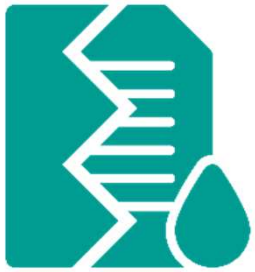
1. Look at the image and identify if a dog or dogs are visible.
2. If the image does not contain any dogs, please skip it.
3. For each dog select the Dog Face object from the tools section and draw a bounding box around the dog's face.

Labeling instructions

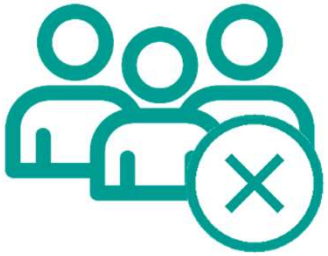
Managment of data labelers

& QA checks

Data Labeling - Challenges



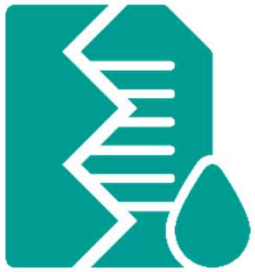
Label leakages



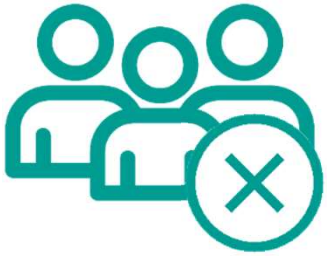
Human error &
Inconsistency

Data Labeling - Challenges

Data Labeling - Challenges



Label leakages



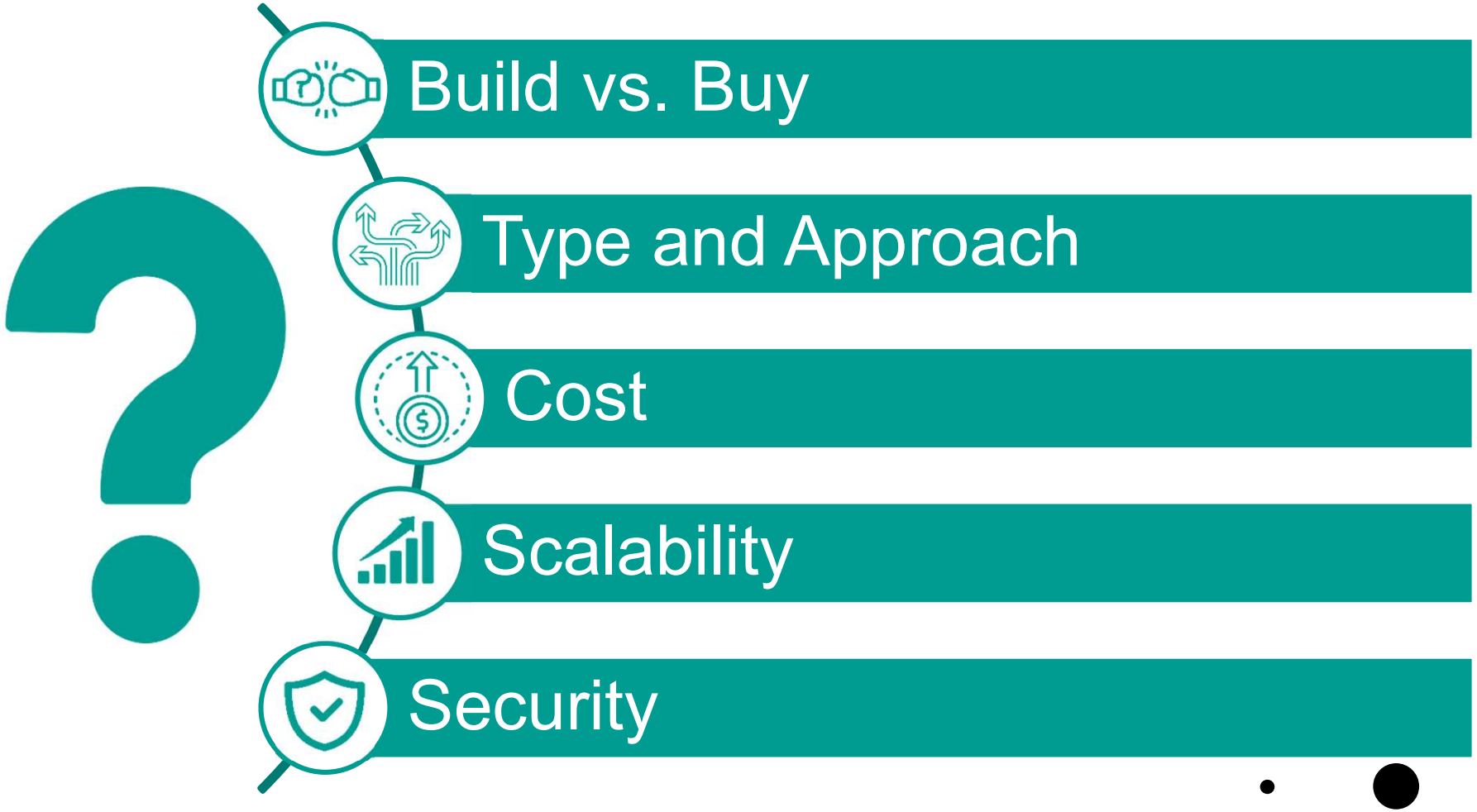
Human error & Inconsistency

Domain expertise

Expensive & time-consuming

Data Labeling - Security / DSGVO

Data Labeling - Tool decision



Data Labeling - Tools

Company	Hosted	Programmatic	Out-/Crowdsourcing
scale	X	X	
snorkel		X	
cloudfactory			X
Labelbox / appen	X	X	X
Amazon SageMaker / amazon mechanical turk	X	X	X

Data Labeling - Tools

Company	Hosted	Programmatic	Out-/Crowdsourcing
scale	X	X	
snorkel		X	
cloudfactory			X
Labelbox / appen	X	X	X
Amazon SageMaker / amazon mechanical turk	X	X	X

Data Labeling - Tools

Company	Hosted	Programmatic	Out-/Crowdsourcing
scale	X	X	
snorkel		X	
cloudfactory			X
Labelbox / appen	X	X	X
Amazon SageMaker / amazon mechanical turk	X	X	X

Data Labeling - Tools

Company	Hosted	Programmatic	Out-/Crowdsourcing
scale	X	X	
snorkel		X	
cloudfactory			X
Labelbox / appen	X	X	X
Amazon SageMaker / amazon mechanical turk	X	X	X

Data Labeling - Tools

Company	Hosted	Programmatic	Out-/Crowdsourcing
scale	X	X	
snorkel		X	
cloudfactory			X
Labelbox / appen	X	X	X
Amazon SageMaker / amazon mechanical turk	X	X	X

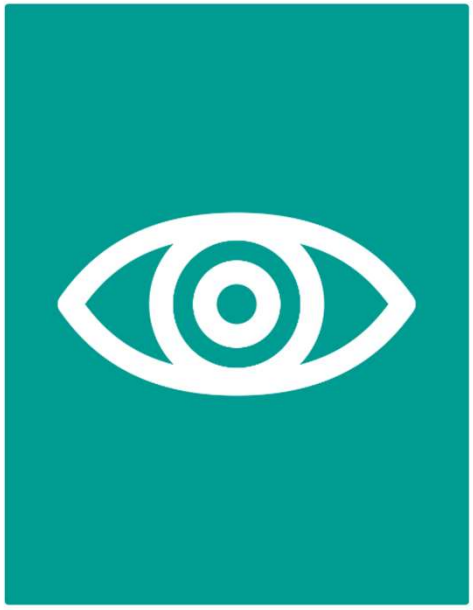
Data Labeling - Tools

Company	Hosted	Programmatic	Out-/Crowdsourcing
scale	X	X	
snorkel		X	
cloudfactory			X
Labelbox / appen	X	X	X
Amazon SageMaker / amazon mechanical turk	X	X	X

Data Labeling - Pricing

Hosted	Inhouse	Programmatic	Out-/ Crowdsourcing
			

Data Labeling - Pricing

Hosted	Inhouse	Programmatic	Out-/ Crowdsourcing
			

Data Labeling - Pricing

Hosted	Inhouse	Programmatic	Out-/ Crowdsourcing
			

Data Labeling - Demo

Computer Vision	NLP (Programmatic)
Labelbox	snorkel

Data Labeling - Demo

https://app.labelbox.com/signin

Benutzer: labelbox1@proton.me

Passwort: XXXXXXX

Benutzer: labelbox2@proton.me

Passwort: XXXXXXX

Benutzer: labelbox3@proton.me

Passwort: XXXXXXX

Vielen Dank

für Ihre Aufmerksamkeit