

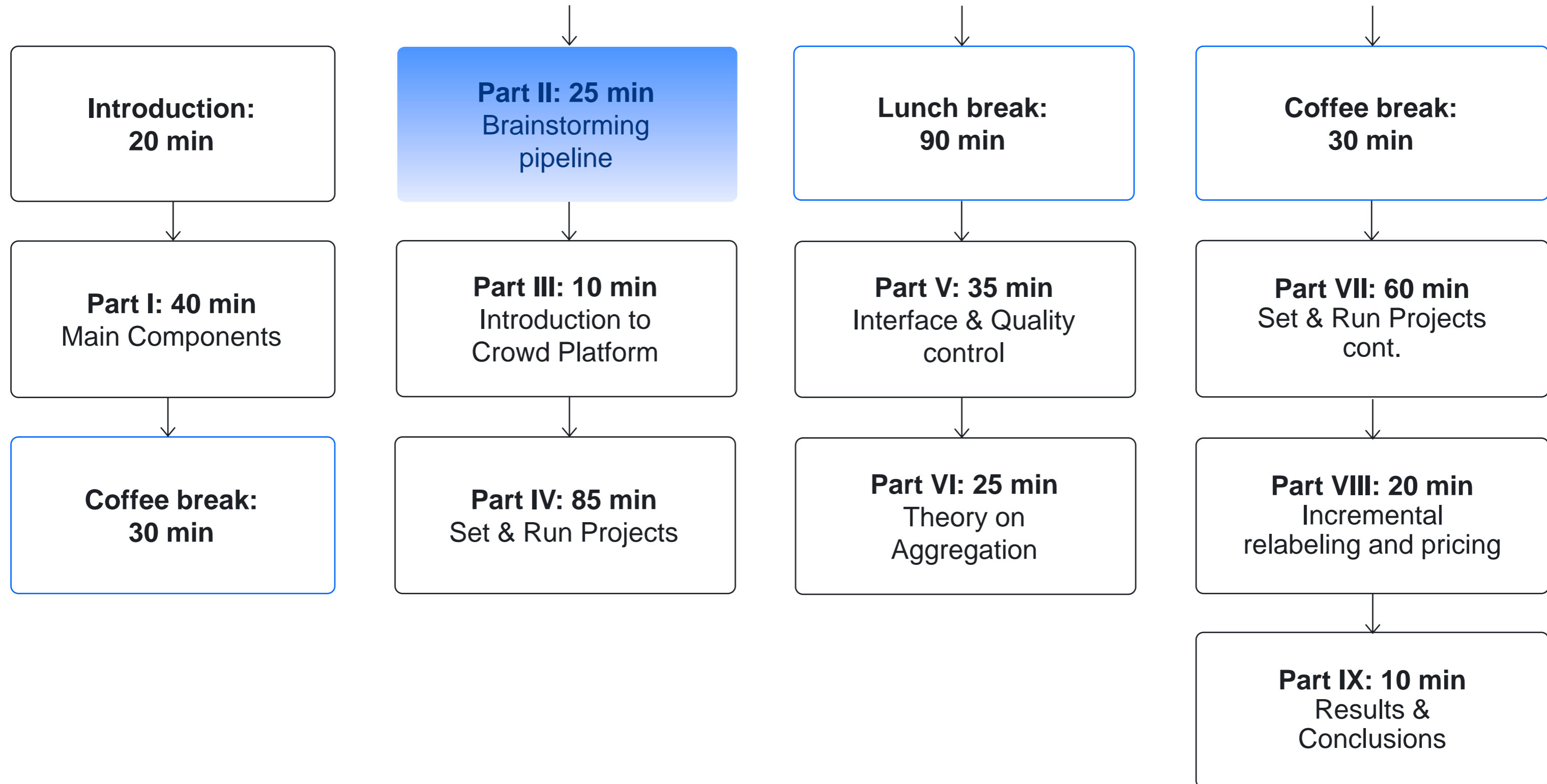
Part II

# Label collection projects to be done

Daria Baidakova,  
Project Manager

Toloka

# Tutorial schedule



# Practice session

Our practice session will consist of three parts:

## Part I (now)

Think and discuss in groups how you would design a crowdsourcing pipeline

## Part II (in 35 min)

Run the best-practice pipeline on a real crowd on Toloka

## Part III (in 145 min)

Complete the pipeline on Toloka

# Practice session: scope

Imagine that you develop a machine learning pipeline to help improve search quality at an online store to find substitutes

The screenshot shows the ASOS website interface. At the top, there is a navigation bar with 'ASOS' logo, 'WOMEN' and 'MEN' tabs, a search bar, and icons for user profile, heart, and shopping bag. Below the navigation bar, there are category links: 'New in', 'Clothing', 'Shoes', 'Accessories', 'Activewear', 'Face + Body', 'Living + Gifts', 'Brands', 'Sale' (highlighted in red), 'Marketplace', and 'Inspiration'. A 'UTLET' button is also visible. The breadcrumb trail reads: 'Home > Women > Dresses > Wednesday's Girl midi dress in smudge spot print'.

The main product image shows a woman wearing a blue midi-length dress with a white polka-dot pattern. To the left of the main image are four smaller thumbnail images of the same dress from different angles. Below the thumbnails is a 'VIDEO' icon. At the bottom left of the main image is a 'SHARE' icon, and at the bottom right is a 'SELLING FAST' badge. Navigation arrows are present on both sides of the main image.

The 'YOU MIGHT ALSO LIKE' section features four product cards, each with a heart icon and the word 'EXCLUSIVE' in a grey box:

- Wednesday's Girl midi dress in smudge spot print**  
£22.00
- Wednesday's Girl mini smock dress in daisy print**  
£20.00
- ASOS DESIGN tiered long sleeve smock maxi dress ...**  
£42.00
- New Look tiered smock midi dress in multi colour...**  
£20.50 ~~£25.99~~


At the bottom of the 'YOU MIGHT ALSO LIKE' section, there are four dots indicating the current position in the carousel.

# Practice session: scope

Imagine that you develop a machine learning pipeline to help improve search quality at an online store to find substitutes

- ▶ You have a dataset of pictures with people wearing different clothes
- ▶ You need to find a better substitute for the initial item in an image
- ▶ These collected data will further be used to train a search algorithm

This is your goal  
for the practice  
session of our tutorial



# Dataset under study: pictures of people wearing different clothing items



# Items to be matched in photos

Each photo may contain clothing items of different types, for example:

- ▶ Hats
- ▶ Shirts
- ▶ Jackets
- ▶ Coats
- ▶ Jeans
- ▶ Pants (trousers)
- ▶ Bags
- ▶ Sunglasses
- ▶ Other items

## **During your practice:**

Choose one type of items you want to find substitutes for in the photos

For example: Shoes

# Formal setup: find the best substitute item

- ▶ Each clothing item of a selected type in each photo from the dataset needs to be matched by a substitute item
- ▶ **Let us do it via crowdsourcing**

Example: we decide to find the best substitute for the shoes, so our pipeline would be like..

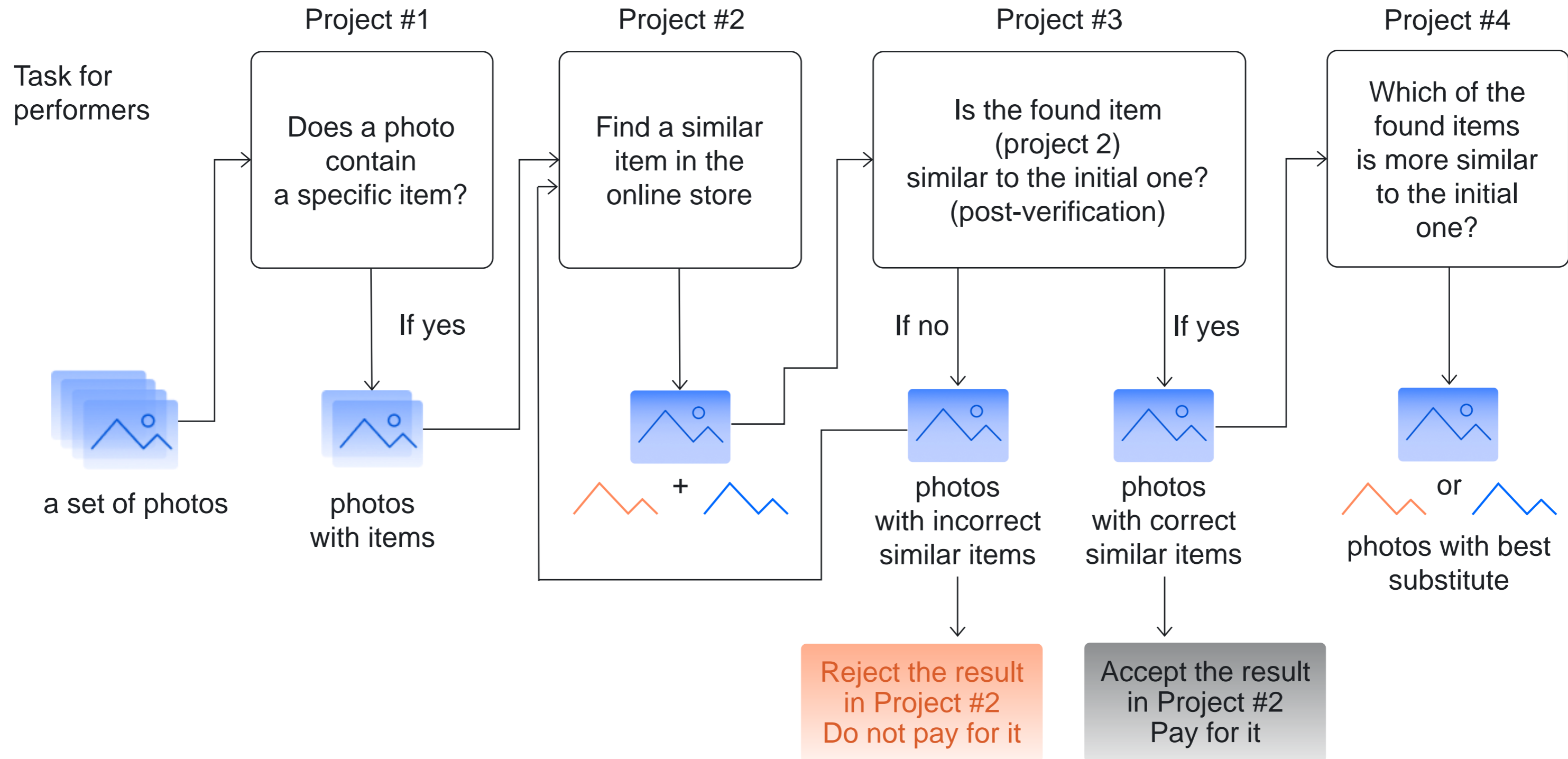
## During your practice:

Discuss in groups how you would design a crowd pipeline to find the best substitute!

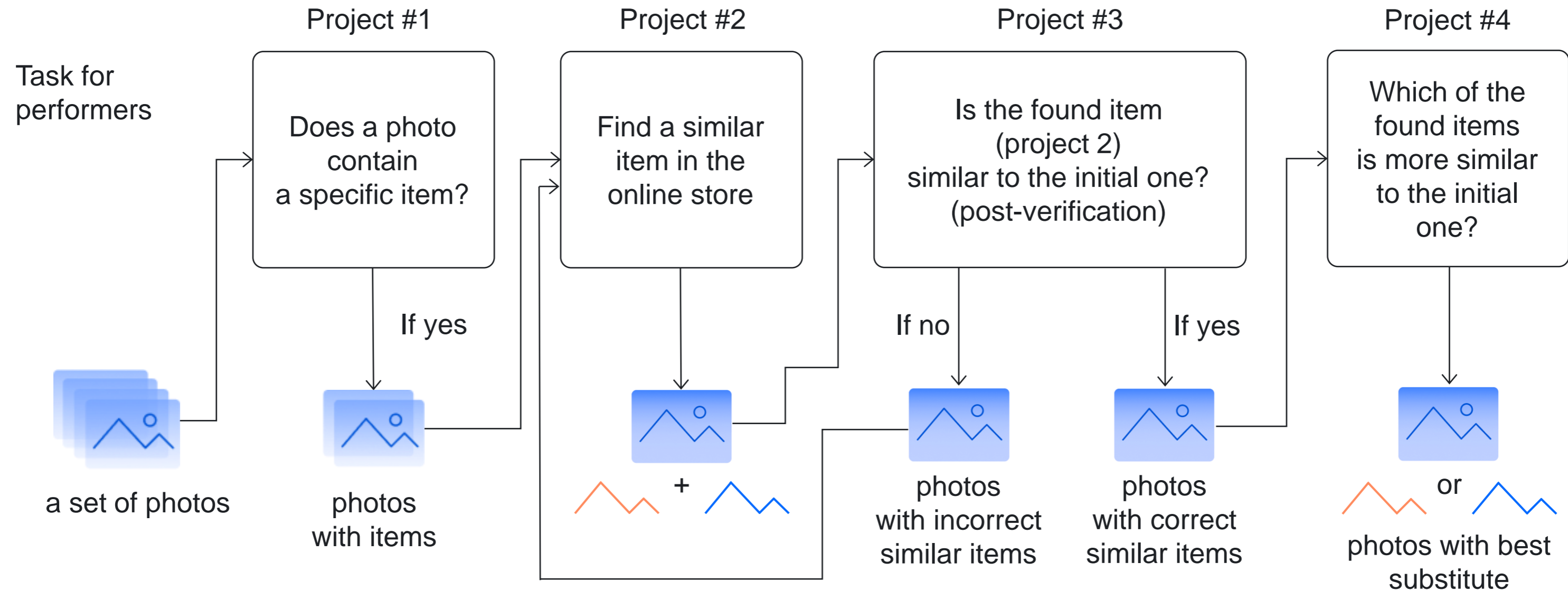


# Suggested pipeline

# We suggest the following pipeline



# We suggest the following pipeline



During the practical session we will help you implement and run this pipeline

# Project #1: Filter out photos without objects

## Task

- ▶ Does a photo contain an item of desired type?

## Key setting

- ▶ Type: classification
- ▶ Quality control: golden set
- ▶ Overlap: 3 answers per photo
- ▶ Pay: \$0.01 per a suite of 10 photo

## Why?

- ▶ Save money: no need to process further photos without desired objects



Are there **shoes** in the picture?

Yes  No  Picture not found

# Project #2: Searching for similar items on the online store

## Task

- ▶ Find a similar item on the internet

## Key setting

- ▶ Type: product photo search
- ▶ Quality control: post verification
- ▶ Overlap: 3 answers per photo
- ▶ Pay: \$0.02 per 1 photo

## Peculiar properties

- ▶ Hard to use golden set and consensus
- ▶ Results will be verified in Project #3



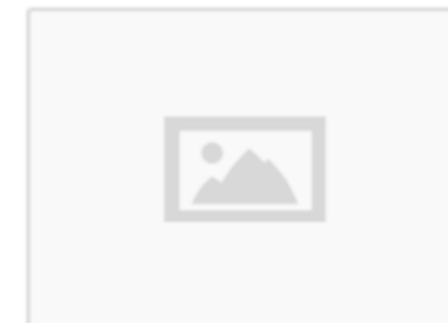
Find the same **shoes** on ASOS

ASOS

Shoes must be the same color and the same style.

Paste the link here

Upload the image here. The image should show the shoes you found.



# Project #3: Accept correctness of items found

## Task

- ▶ Does an image contain a requested item?

## Key setting

- ▶ Type: classification
- ▶ Quality control: consensus
- ▶ Overlap: 3 answers per photo
- ▶ Pay: \$0.01 per a suite of 10 photo

## Why?

- ▶ Need to verify the results obtained from Project #2



Check that the uploaded image matches the product in the store.

[Check the item](#)

Are these **shoes** similar to each other?

Shoes must be the same color and the same style.

Yes No

# Project #4: Decide which substitute works best

## Task

- ▶ Which of the items is similar to the initial one?

## Key setting

- ▶ Type: side-by-side image comparison
- ▶ Quality control: consensus
- ▶ Overlap: 3 answers per photo
- ▶ Pay: \$0.01 per a task suite of 10 photo

## Why?

- ▶ Need to understand which substitute fits best

