Vandex



Crowdsourcing Practice for Efficient Data Labeling: Aggregation, Incremental Relabeling, and Pricing

Alexey Drutsa, Valentina Fedorova, Dmitry Ustalov, Olga Megorskaya, Evfrosiniya Zerminova, Daria Baidakova

Part IV

Setting up and running label collection projects

Daria Baidakova, Project Manager

Yandex. Toloka is a service of Swiss company Yandex Services AG

Tutorial schedule

Introduction: 15 min

Part I: 20 min
Main Components

Part II: 10 min Introduction to Crowd Platform

Part III: 15 min
Brainstorming
pipeline

Part IV: 60 min Set & Run Projects

Part V: 25 min
Theory on
Aggregation

Break: 30 min

Part VI: 20 min Set & Run Projects cont.

Part VII: 10 min Results & Conclusions

What you need for the practice session

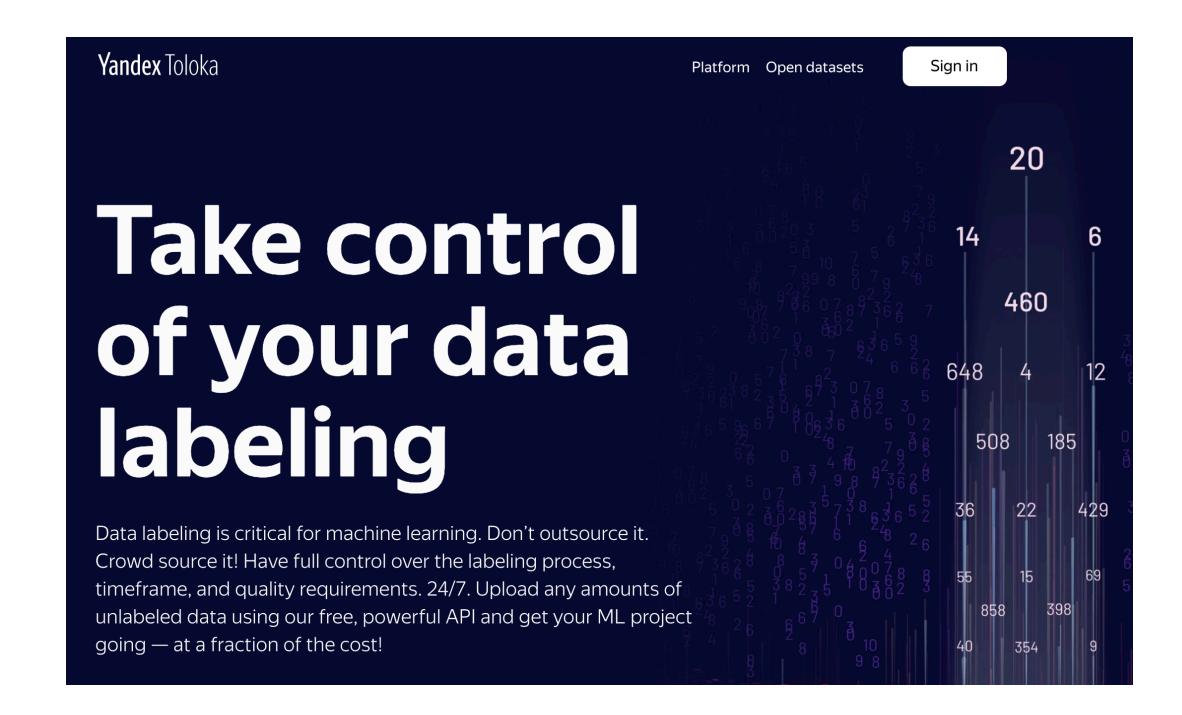
We are starting the practice session

We give you links to:

- > A step-by-step instruction to configure and run our crowd projects
- > A dataset with photos that should be labeled
- > Login+Password to sign in Yandex. Toloka as a requester

Sign in Yandex. Toloka as a requester

- 1. Go to https://toloka.ai
- 2. Click on "Sign in" in the top-right corner
- 3. Use received Login+Password to sign in



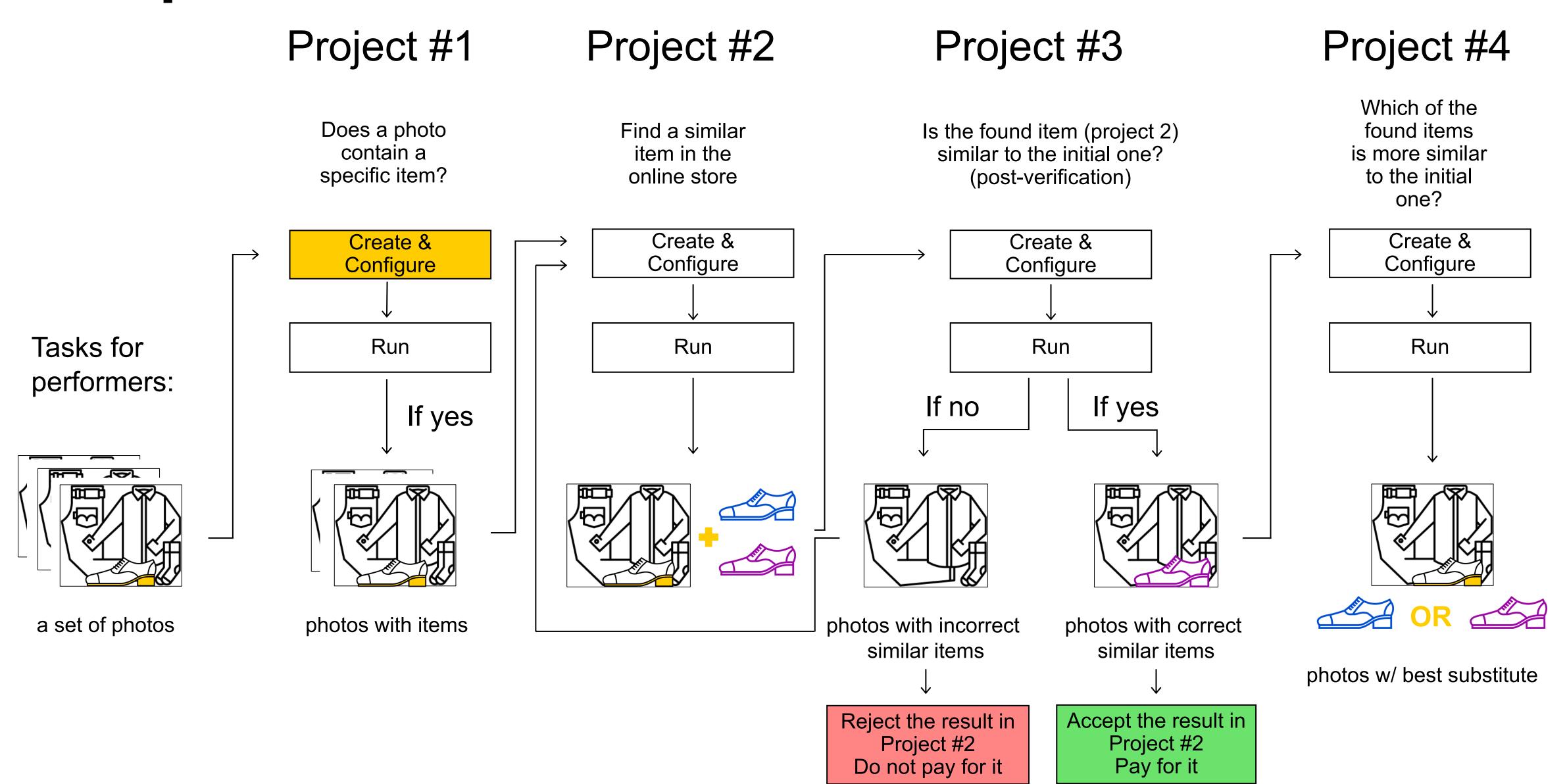
Practice: creating a real crowdsourcing pipeline

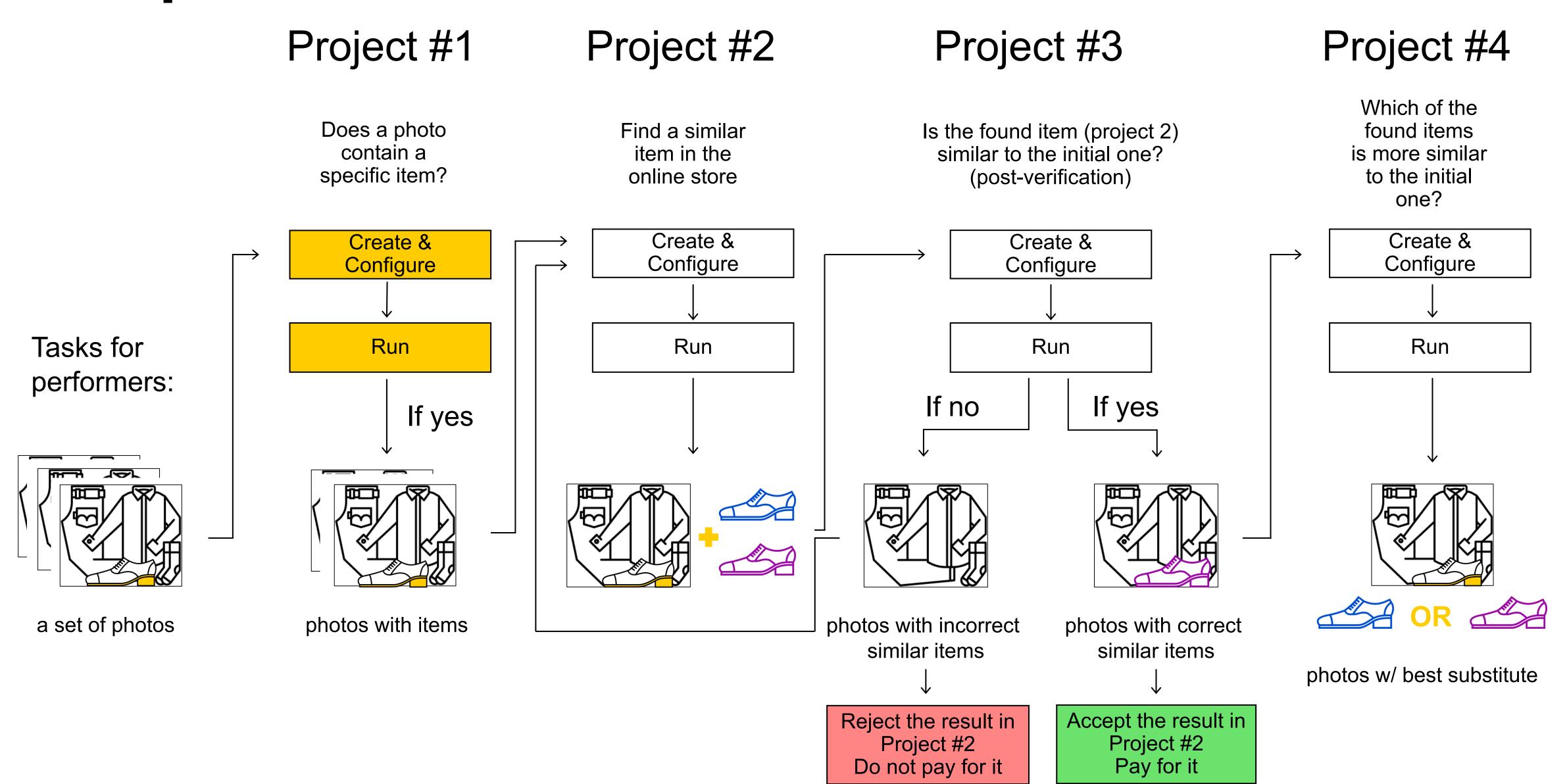
Now we will create a real simplified crowdsourcing pipeline

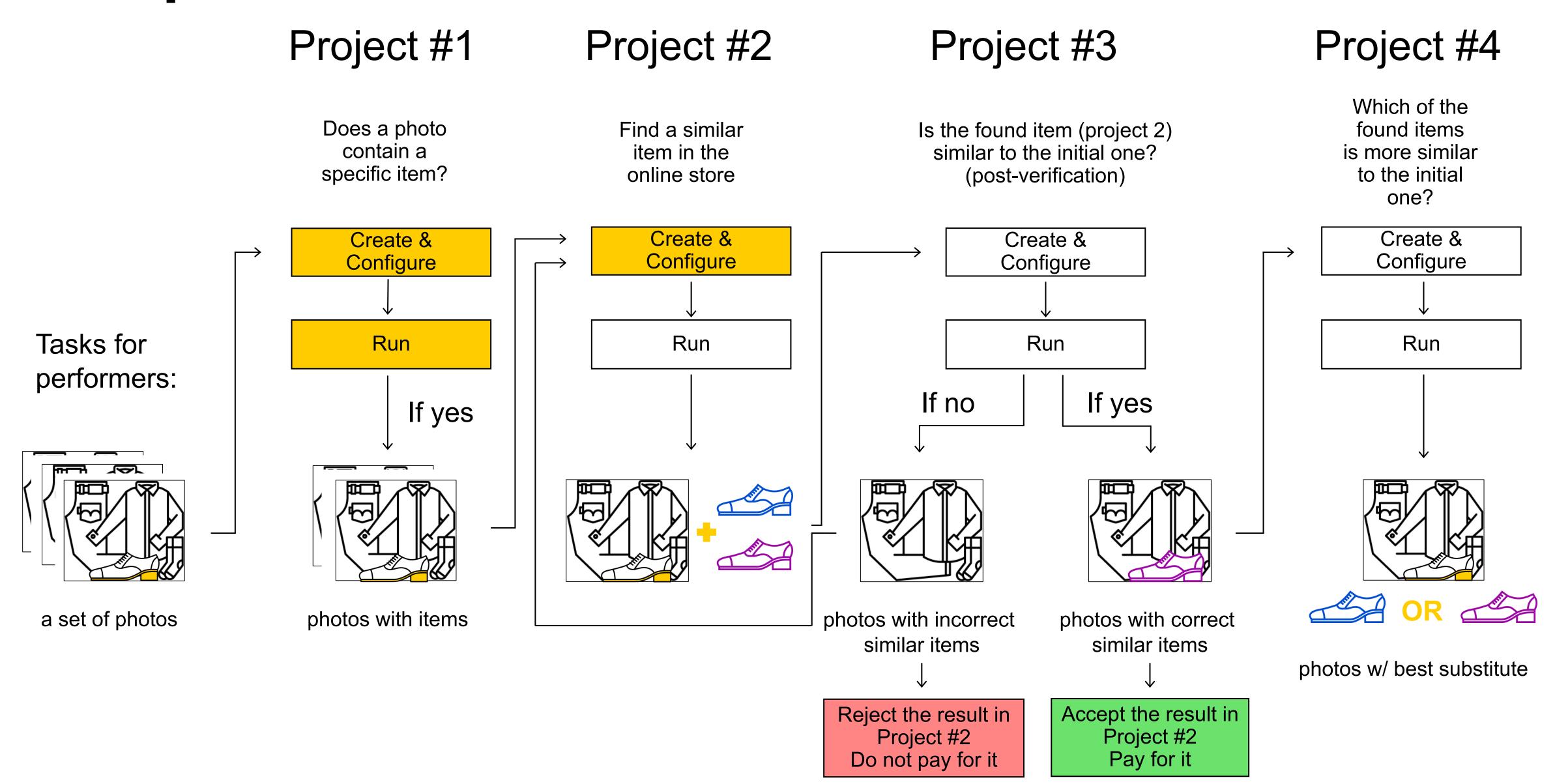
- To simplify the task, we ask you to:
- > Finding a substitute for one type of item
- Choose any item you want to find the best substitute for. For example, Shoes

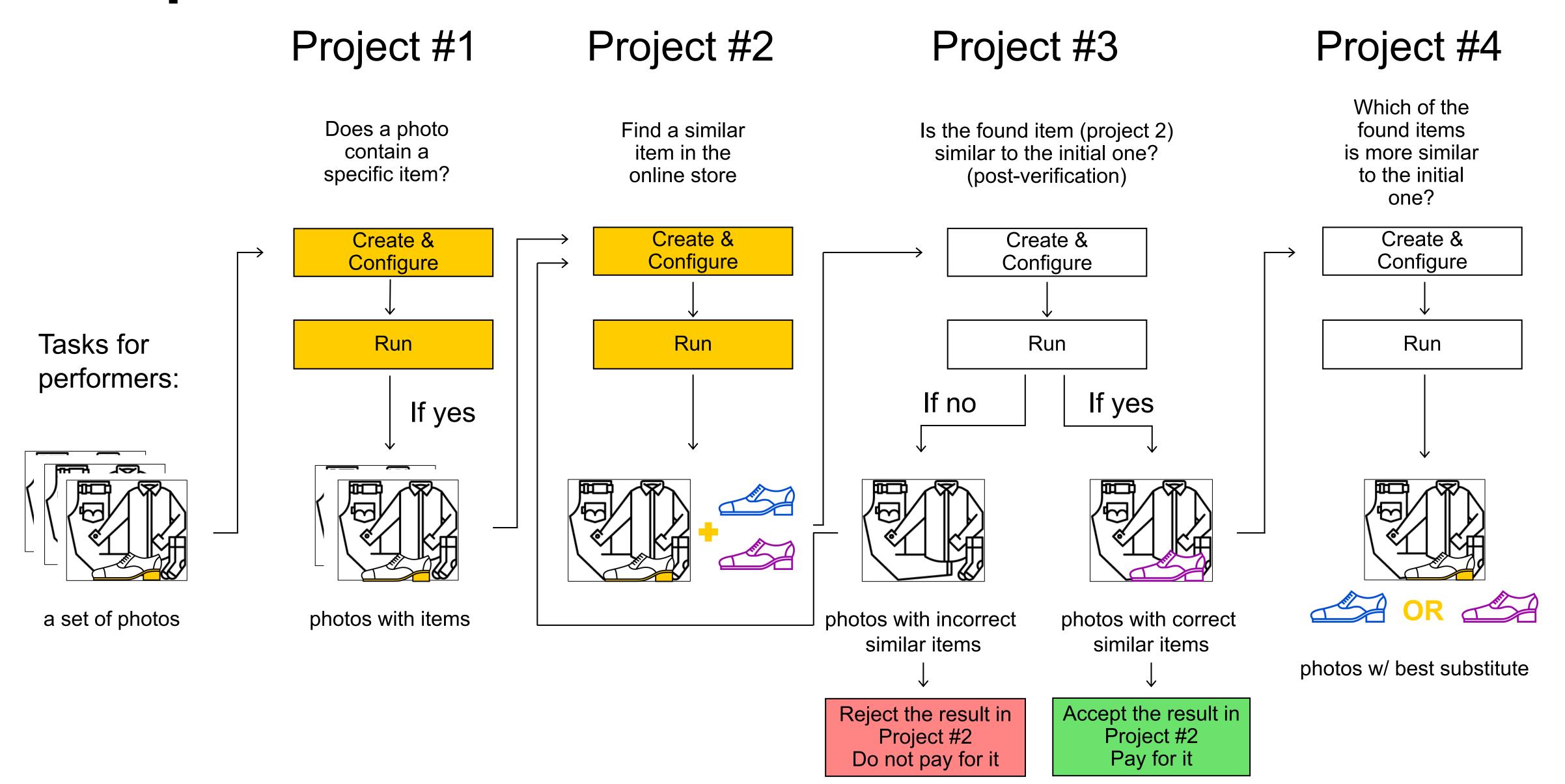
Reminder: we implement and run our pipeline

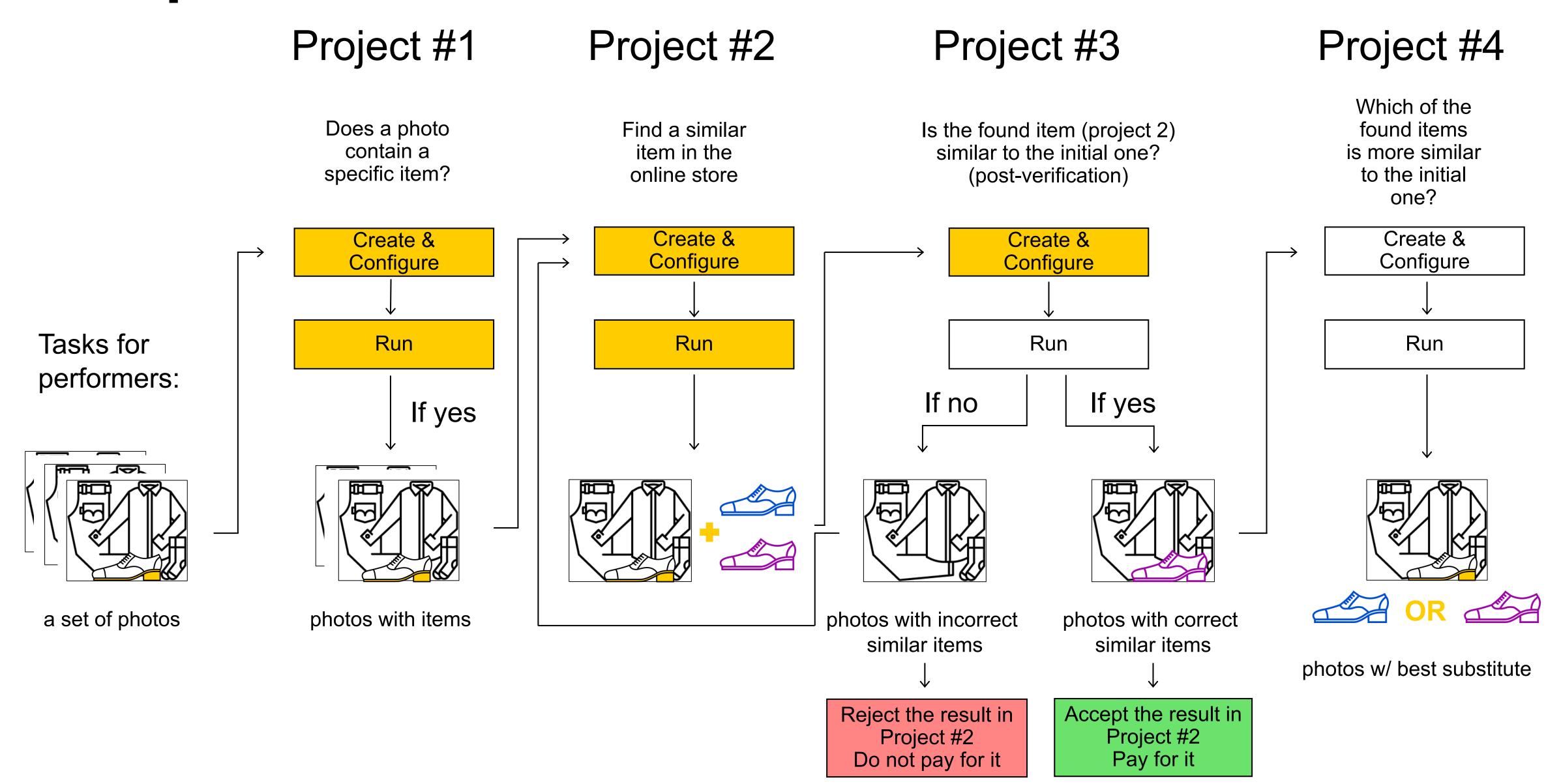
Project #2 Project #1 Project #3 Project #4 Which of the Does a photo Is the found item (project 2) found items Find a similar contain a similar to the initial one? is more similar item in the specific item? to the initial online store (post-verification) one? Create & Create & Create & Create & Configure Configure Configure Configure Tasks for Run Run Run Run performers: If no If yes If yes ₹ P photos with items photos with incorrect photos with correct a set of photos similar items similar items photos w/ best substitute Accept the result in Reject the result in Project #2 Project #2 Pay for it Do not pay for it

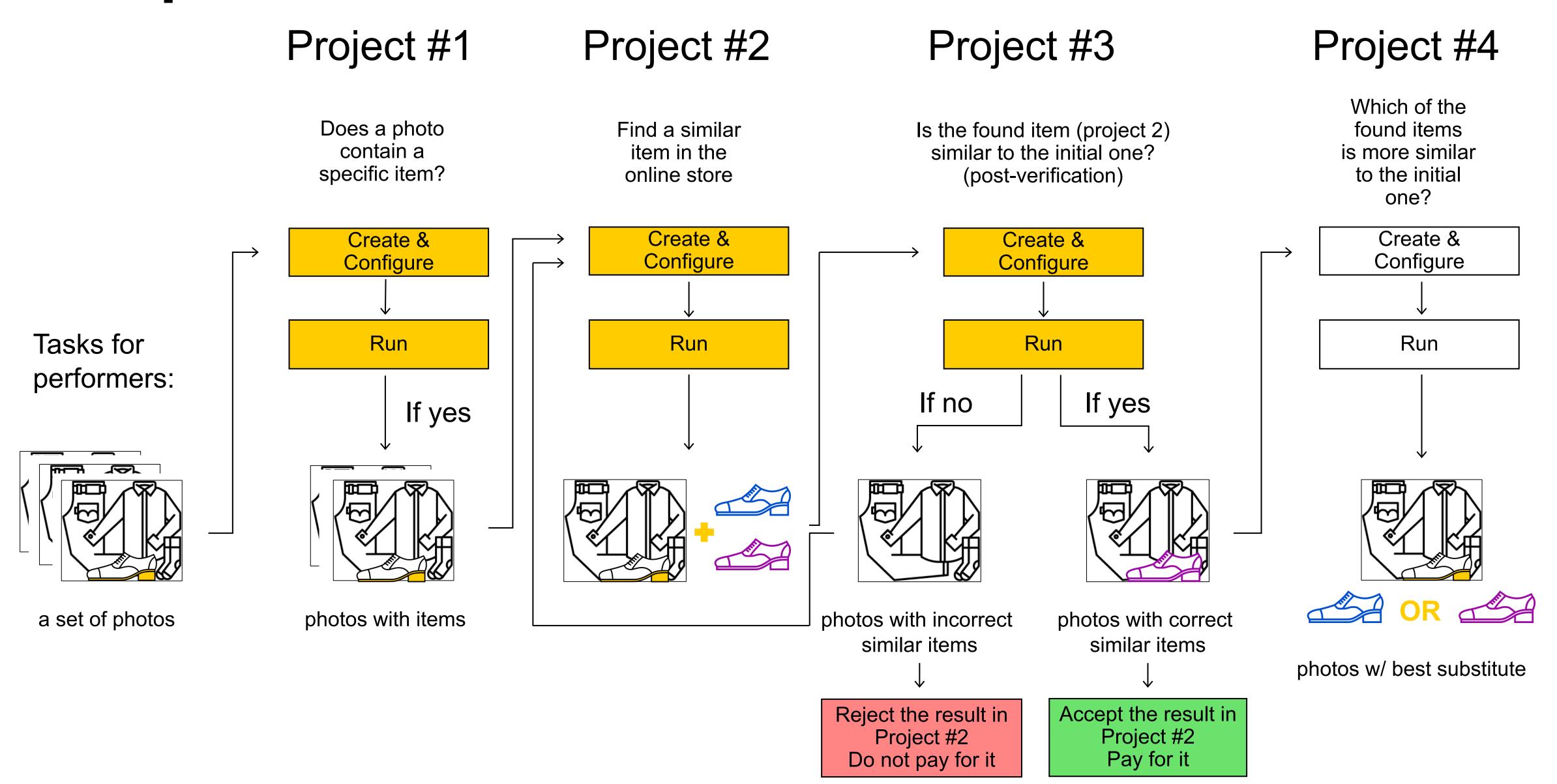


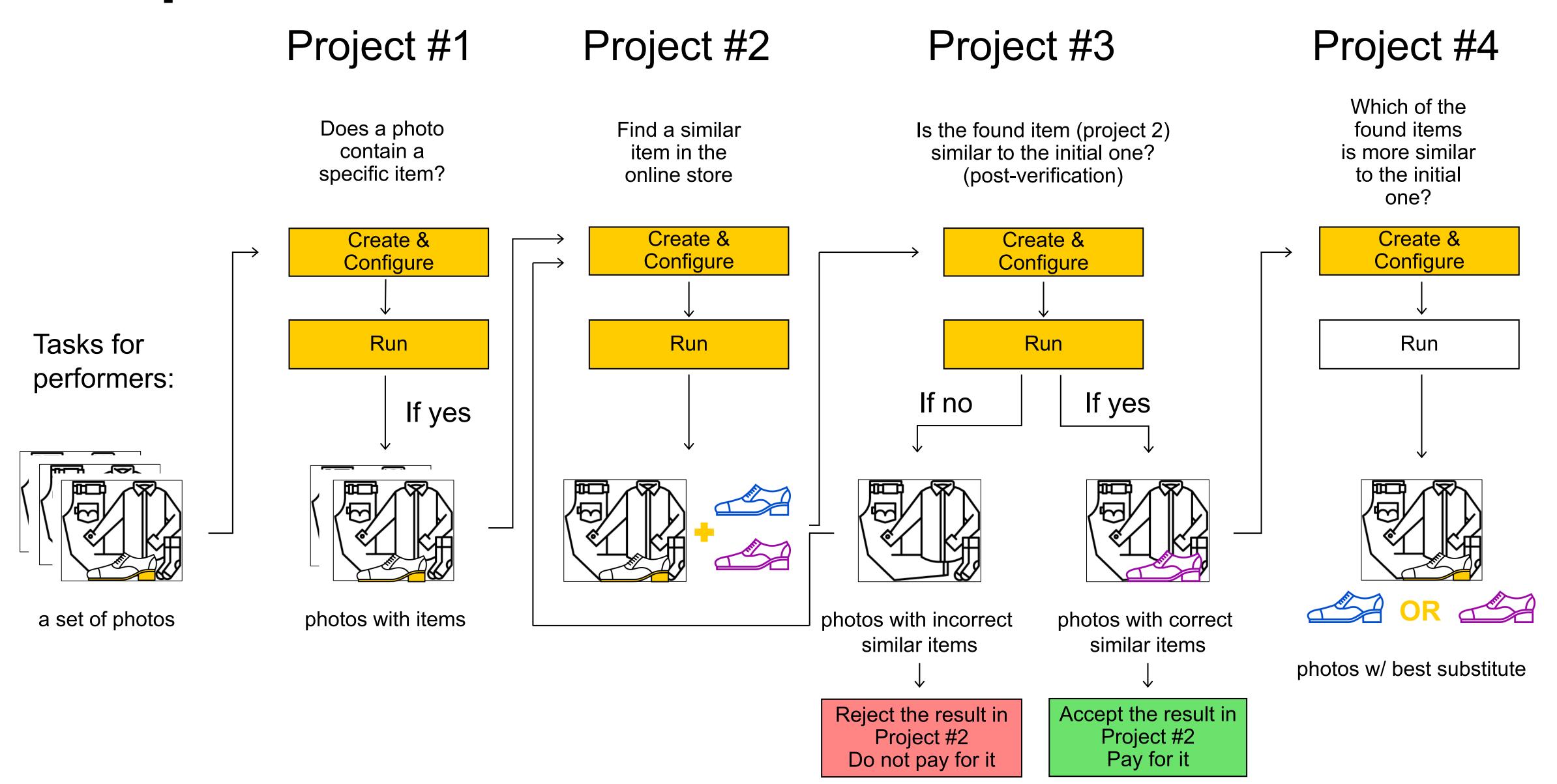


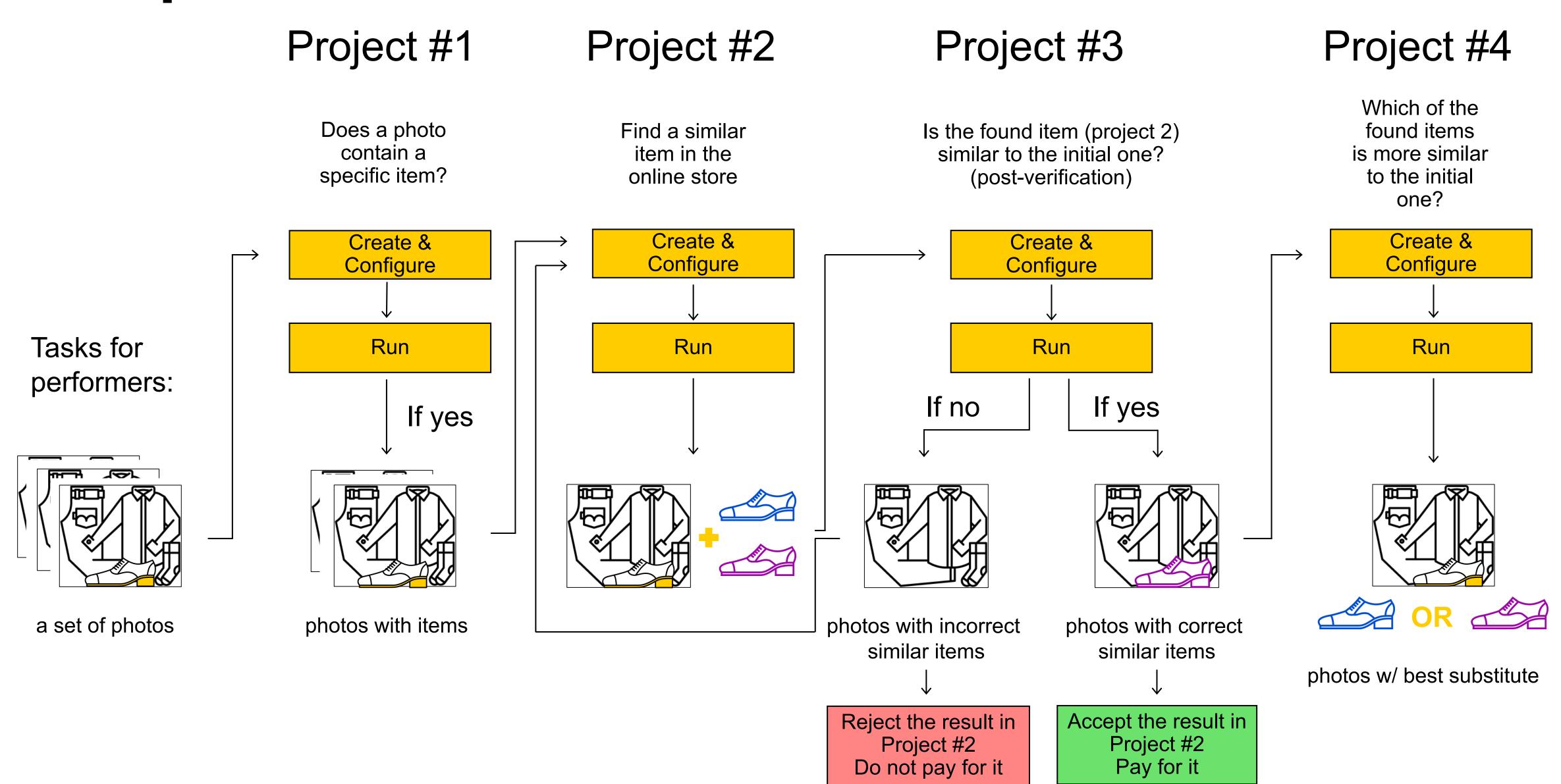












In this part we create 2 projects:

You configure crowd projects

You connect them in a pipeline

You run them on real performers

You have:

- The instruction with all details https://tlk.s3.yandex.net/relsubstitutes/SIGMOD_instruction.pdf
- > The dataset with photos
 http://tlk.s3.yandex.net/relsubstitutes/dataset_y.tsv
 where Y is any number from 1 to 10



Project #1 Project #2 Project #3 Project #4 Which of the Does a photo Find a similar found items Is the found item (project 2) similar to the initial one? contain a item in the is more similar specific item? to the initial online store (post-verification) one? Create & Create & Create & Create & Configure Configure Configure Configure Tasks for Run Run Run Run performers: If no If yes If yes **₩** OR S a set of photos photos with items photos with incorrect photos with correct similar items similar items photos w/ best substitute Accept the result in Reject the result in Project #2 Project #2

Do not pay for it

Pay for it

Project #1 Project #2 Project #3 Project #4 Which of the Does a photo Find a similar found items Is the found item (project 2) similar to the initial one? contain a item in the is more similar specific item? to the initial online store (post-verification) one? Create & Create & Create & Create & Configure Configure Configure Configure Tasks for Run Run Run Run performers: If no If yes If yes **₩** OR S a set of photos photos with items photos with incorrect photos with correct similar items similar items photos w/ best substitute Accept the result in Reject the result in Project #2 Project #2 Do not pay for it Pay for it

Project #1 Project #2 Project #3 Project #4 Which of the Does a photo Find a similar found items Is the found item (project 2) similar to the initial one? contain a item in the is more similar specific item? to the initial online store (post-verification) one? Create & Create & Create & Create & Configure Configure Configure Configure Tasks for Run Run Run Run performers: If no If yes If yes ₹ T OR S a set of photos photos with items photos with incorrect photos with correct similar items similar items photos w/ best substitute Accept the result in Reject the result in Project #2 Project #2 Pay for it Do not pay for it

Project #1 Project #2 Project #3 Project #4 Which of the Does a photo Find a similar found items Is the found item (project 2) similar to the initial one? contain a item in the is more similar specific item? to the initial online store (post-verification) one? Create & Create & Create & Create & Configure Configure Configure Configure Tasks for Run Run Run Run performers: If no If yes If yes ₹ T OR S a set of photos photos with items photos with incorrect photos with correct similar items similar items photos w/ best substitute Accept the result in Reject the result in Project #2 Project #2 Pay for it Do not pay for it

Project #1 Project #2 Project #3 Project #4 Which of the Does a photo Find a similar found items Is the found item (project 2) similar to the initial one? contain a item in the is more similar specific item? to the initial online store (post-verification) one? Create & Create & Create & Create & Configure Configure Configure Configure Tasks for Run Run Run Run performers: If no If yes If yes ₹ T OR S a set of photos photos with items photos with incorrect photos with correct similar items similar items photos w/ best substitute Accept the result in Reject the result in Project #2 Project #2 Pay for it Do not pay for it

Yandex

Thank you! Questions?

Daria Baidakova

Project Manager





dbaidakova@yandex-team.ru



https://research.yandex.com/tutorials/crowd/sigmod-2020