

**Y**andex

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

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

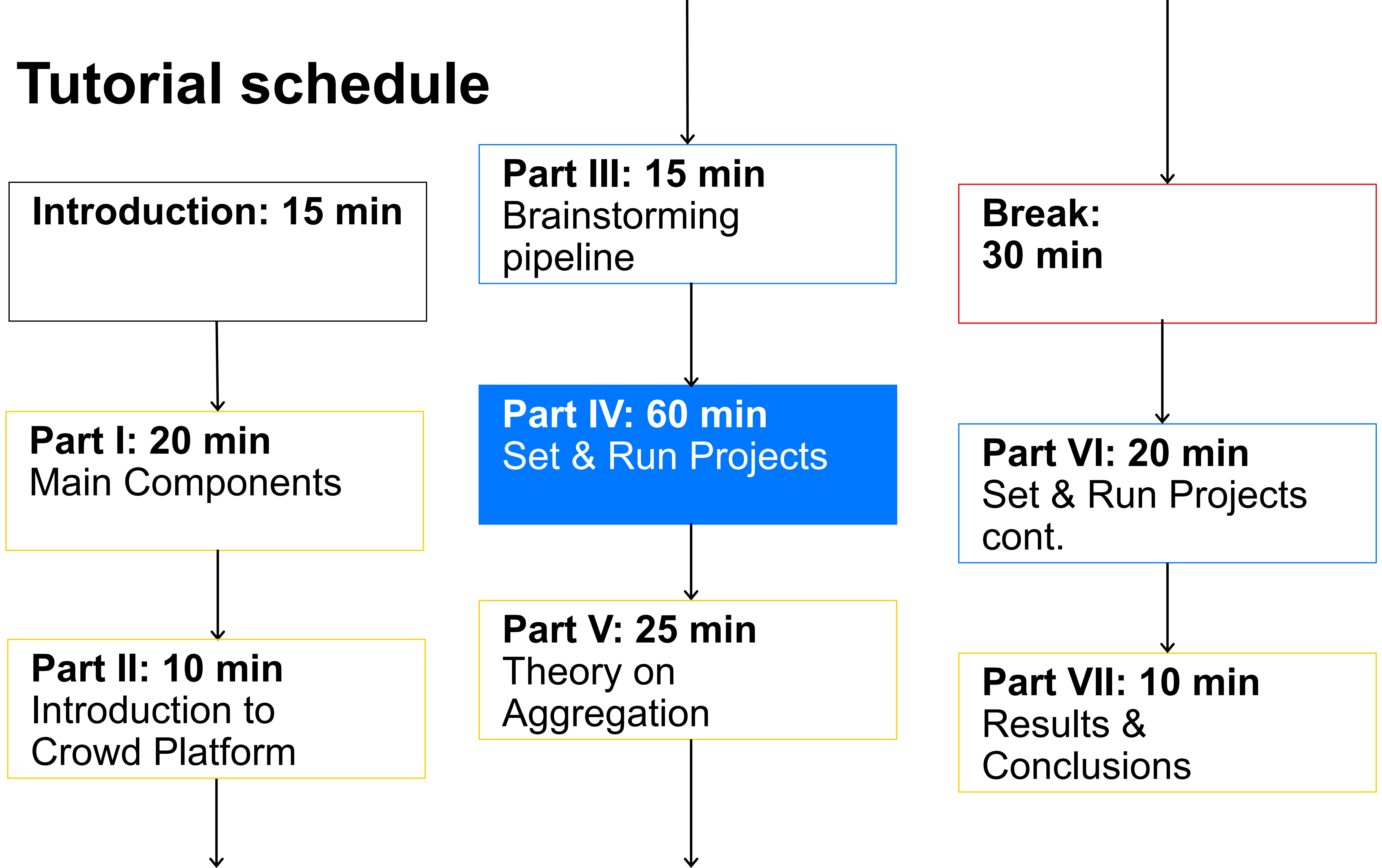
# Part IV

# **Setting up and running label collection projects**

Daria Baidakova,  
Project Manager

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

# Tutorial schedule



# What you need for the practice session

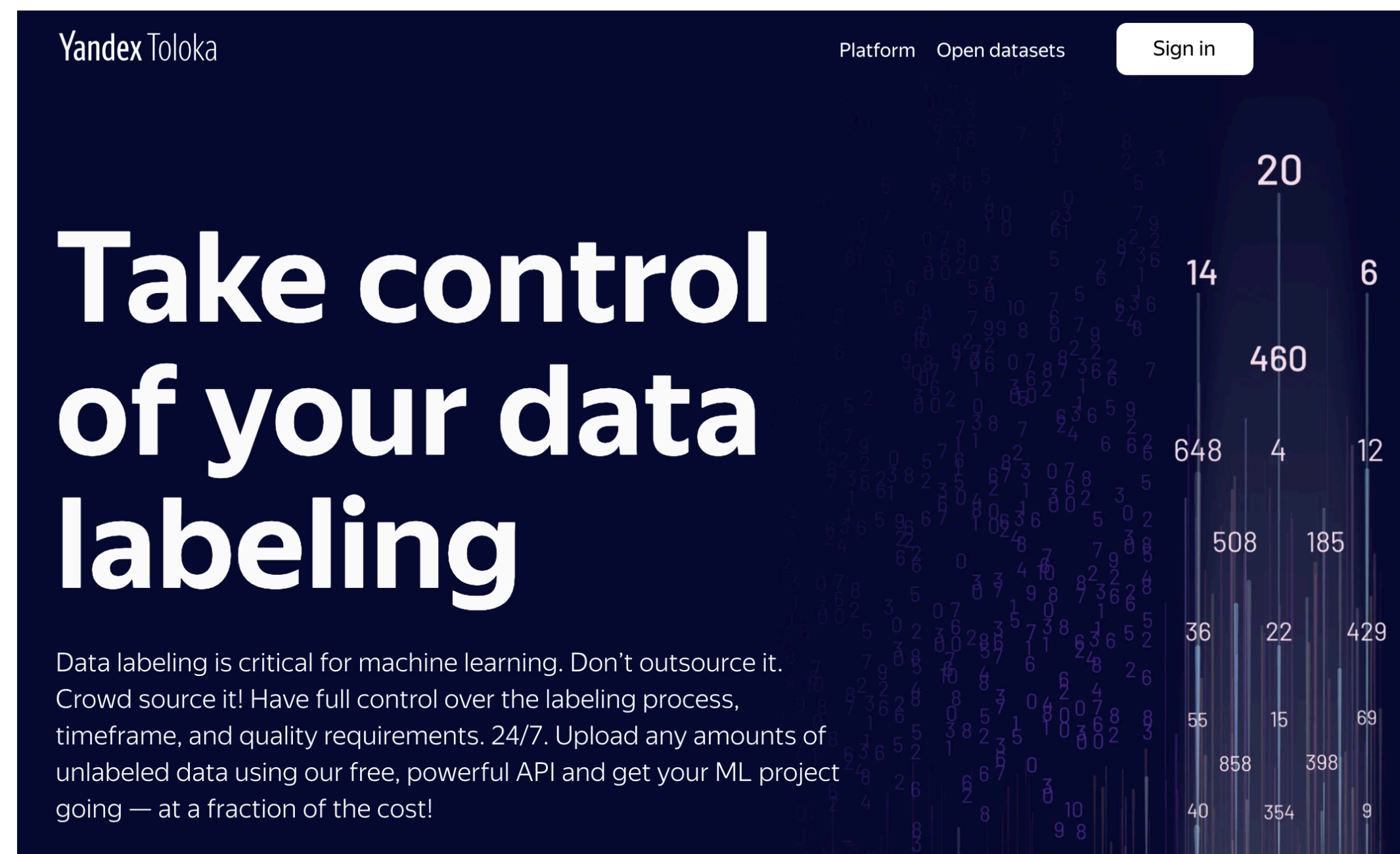
**We are starting the practice session**

**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

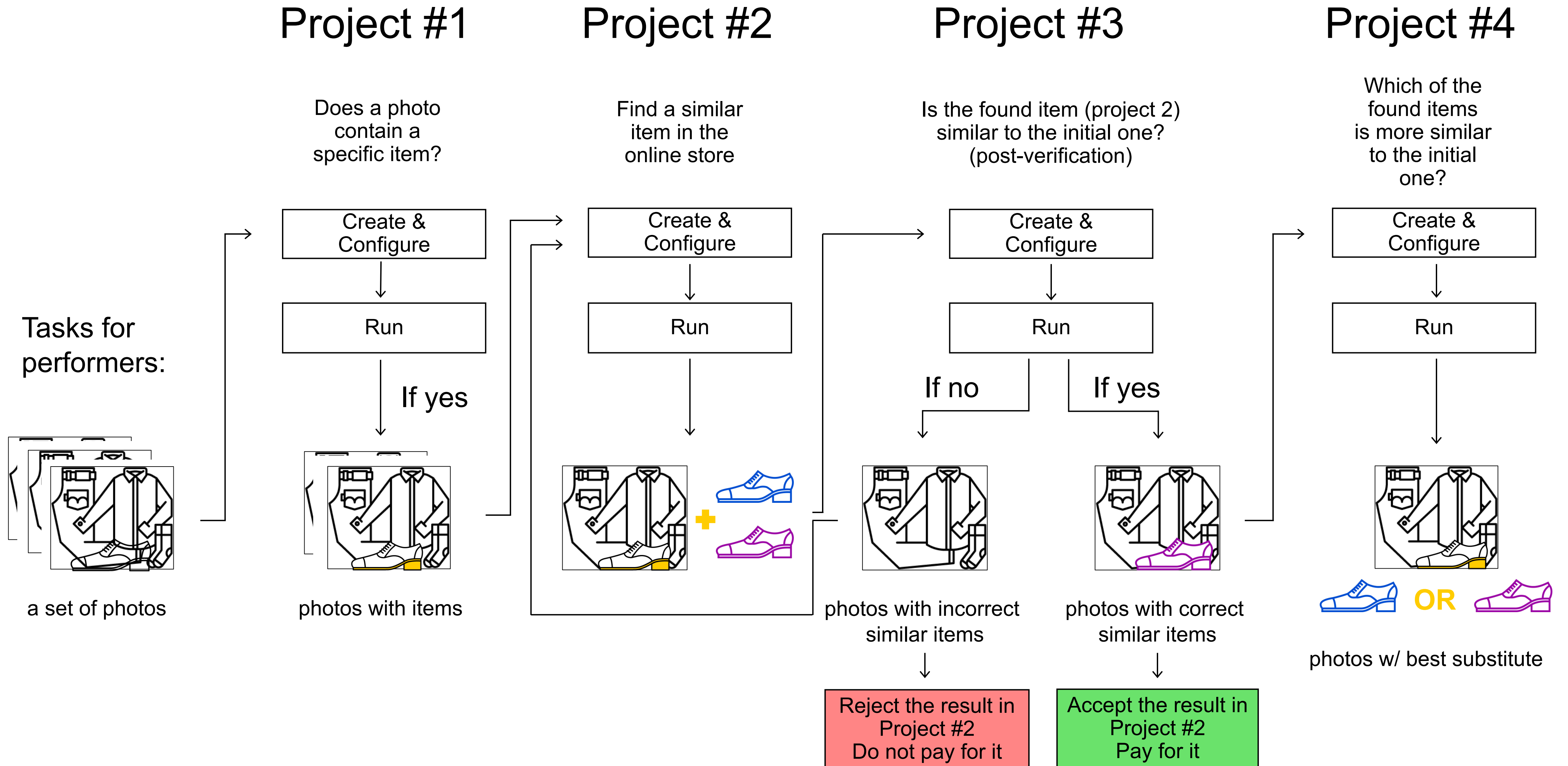


# Practice: creating a real crowdsourcing pipeline

Now we will create a real simplified crowdsourcing pipeline

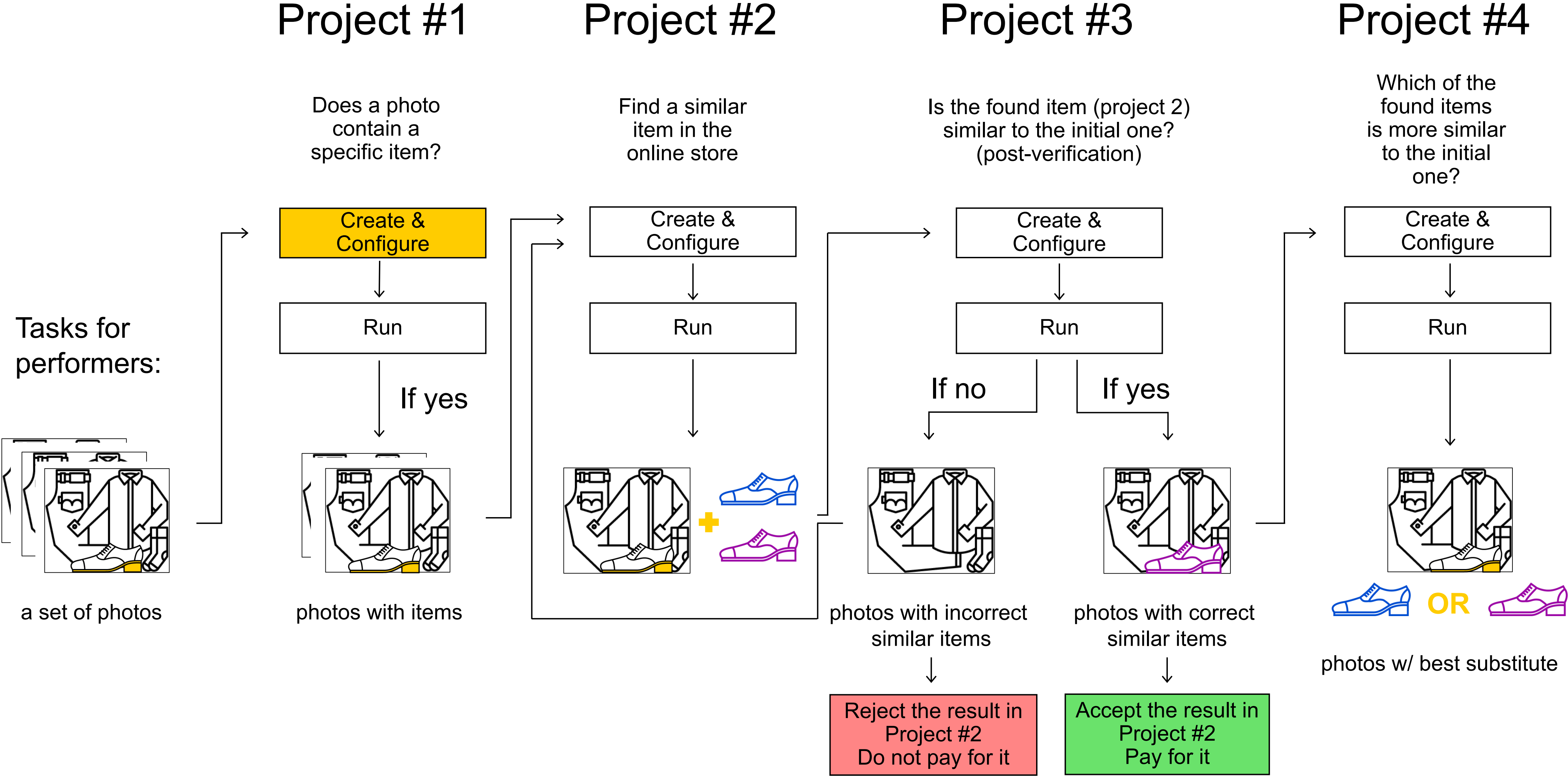
- | To simplify the task, we ask you to:**
  - › Finding a substitute for **one type** of item
  - › Choose any item you want to find the best substitute for. For example, **Shoes**

# Reminder: we implement and run our pipeline

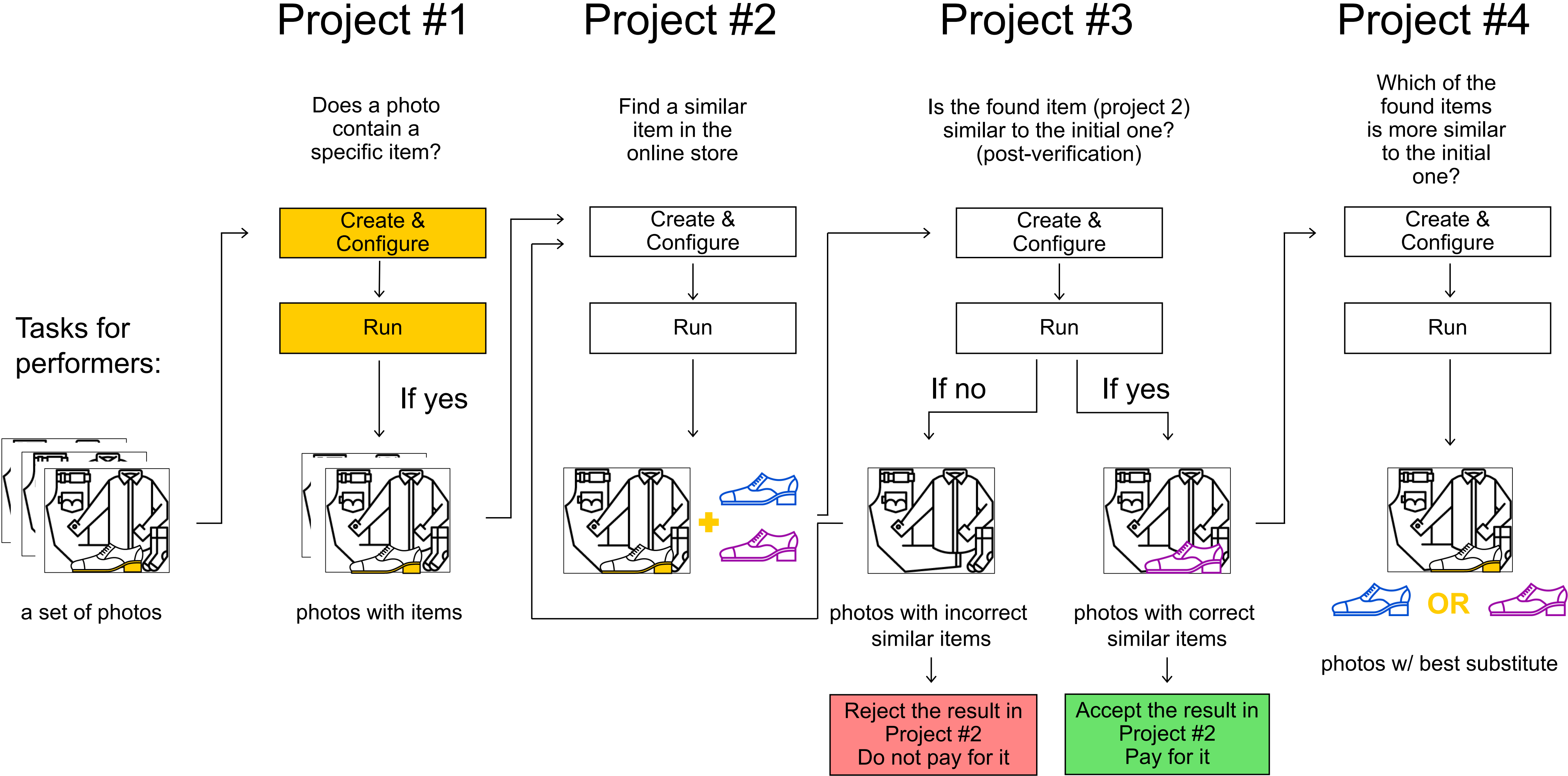




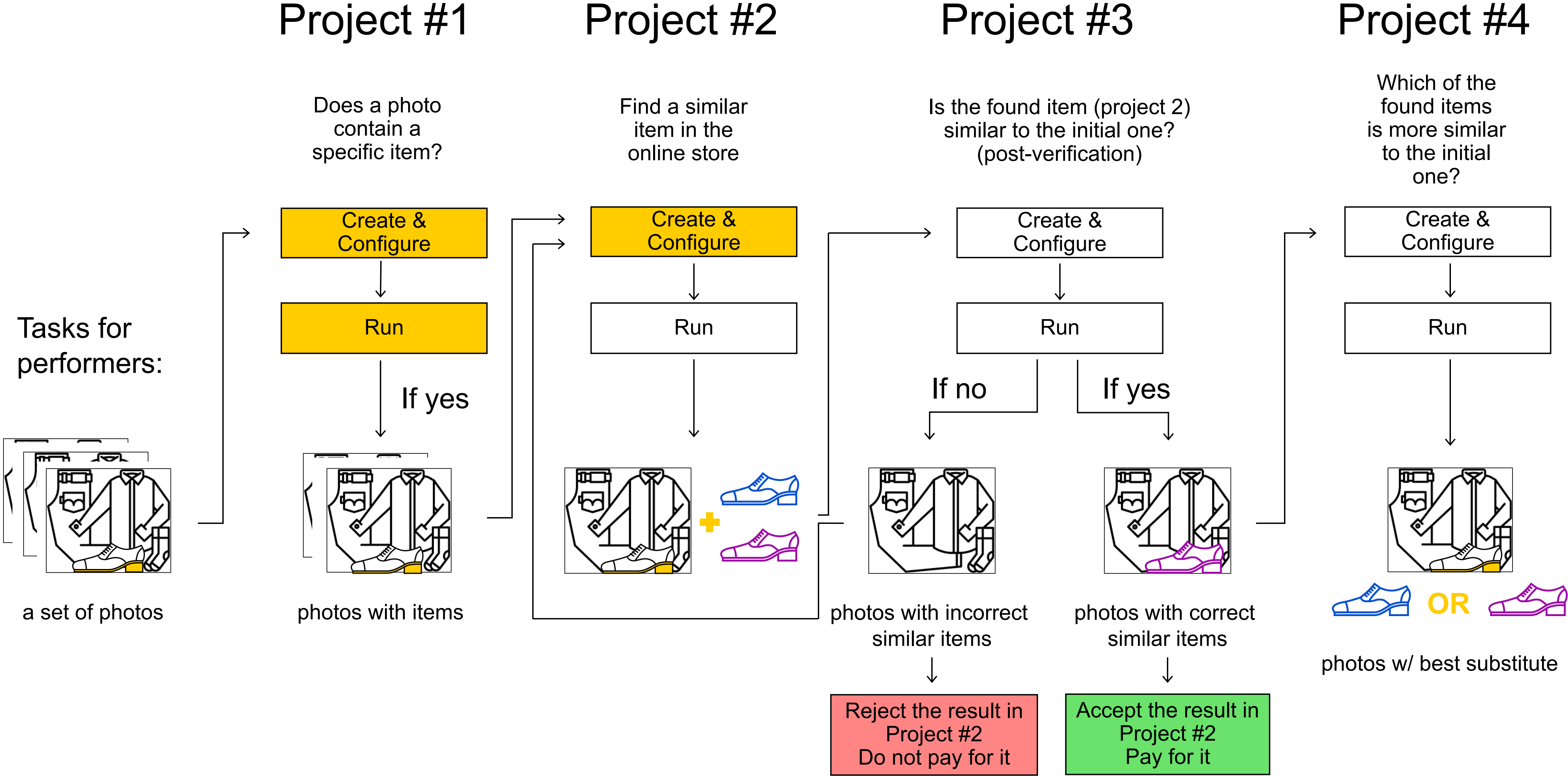
# Step #1



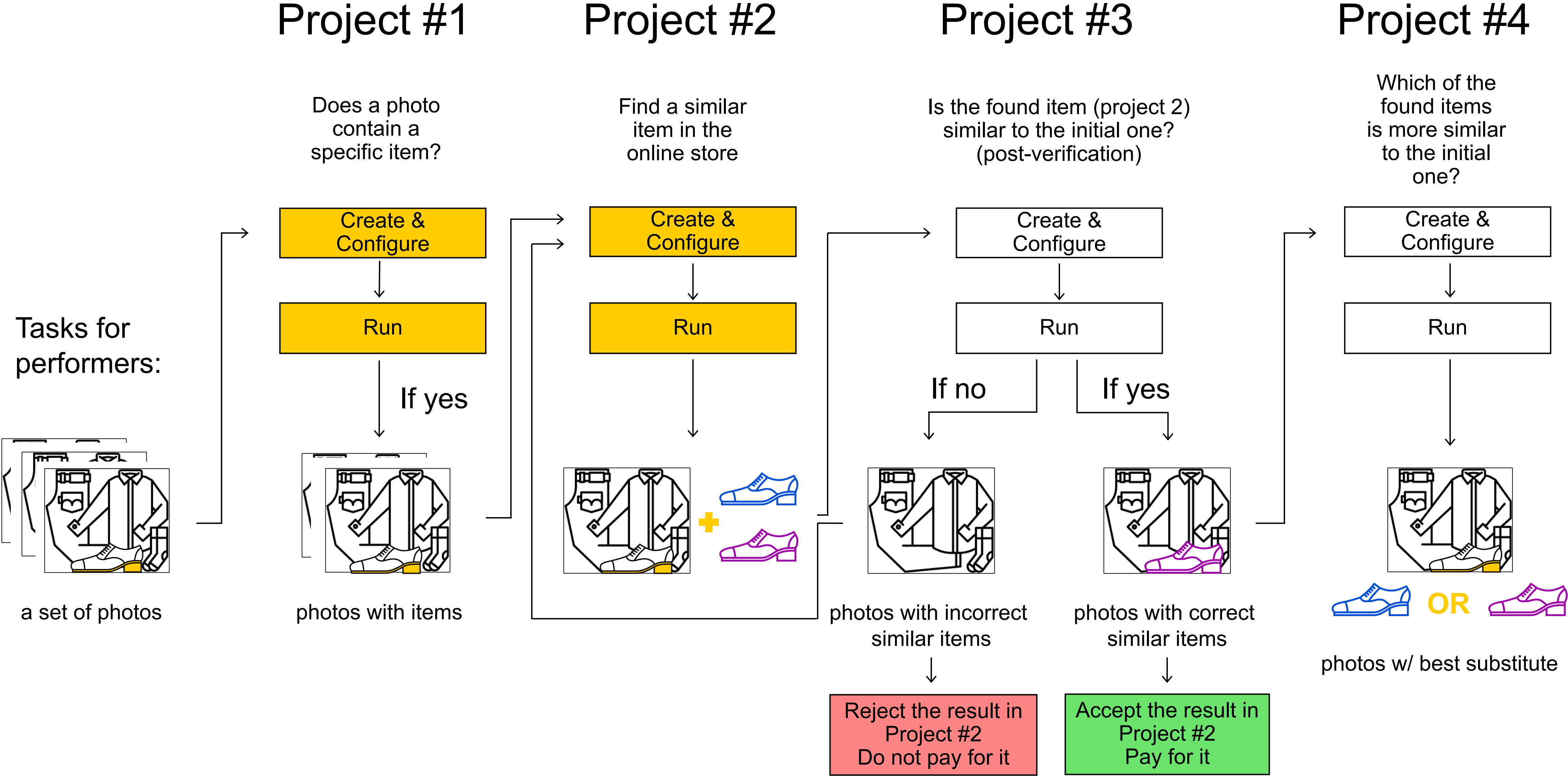
# Step #2



# Step #3

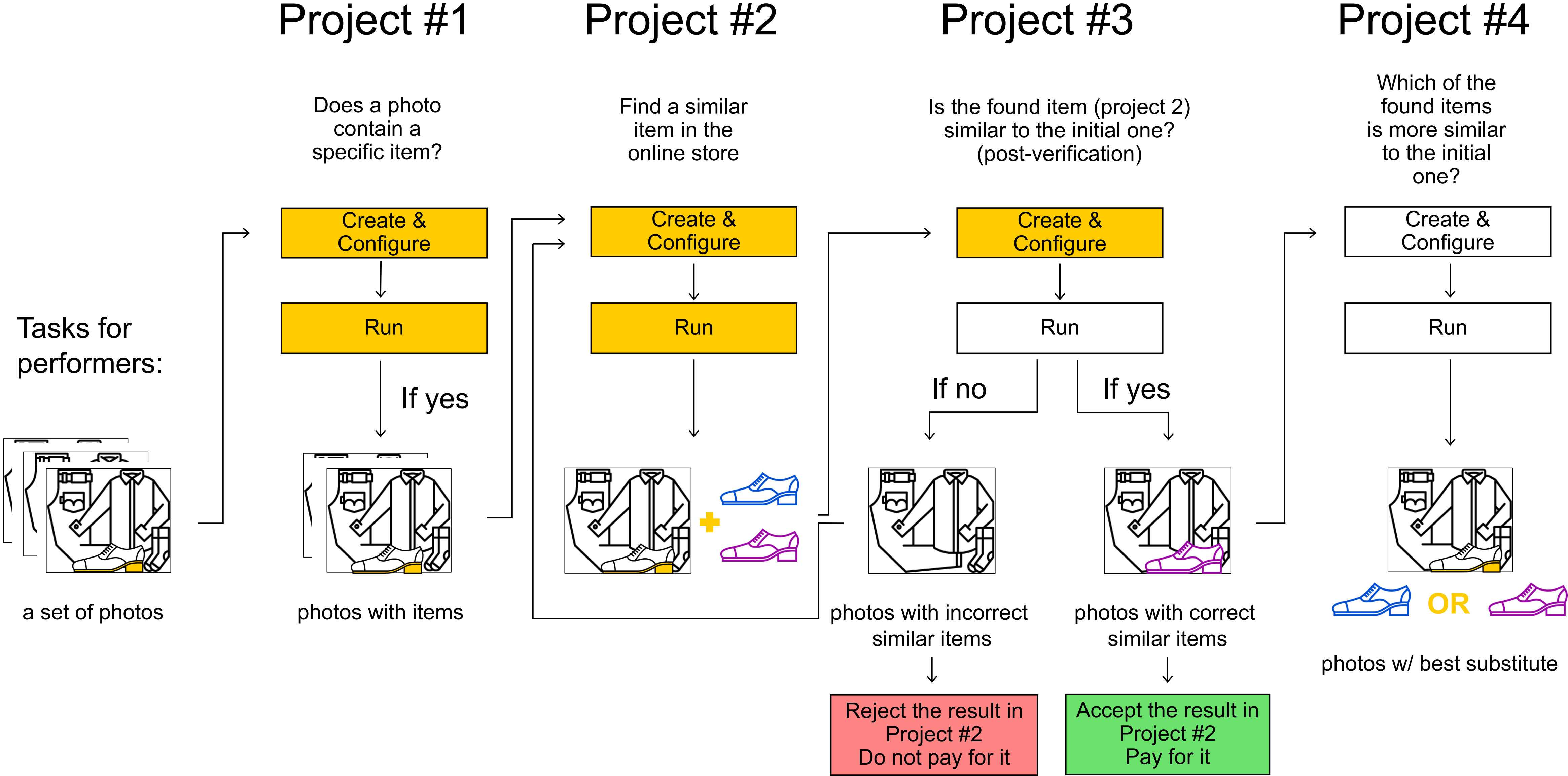


# Step #4

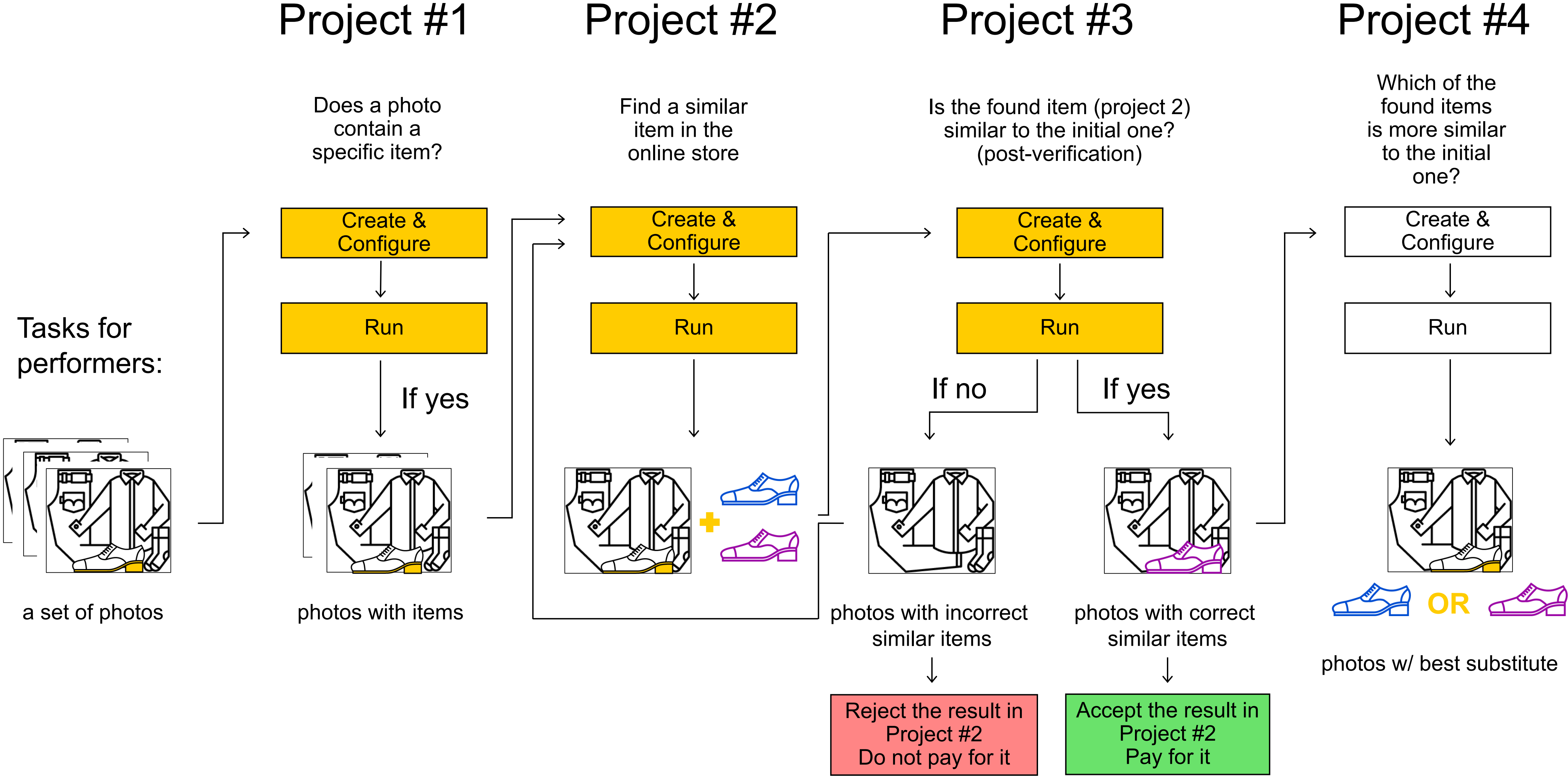




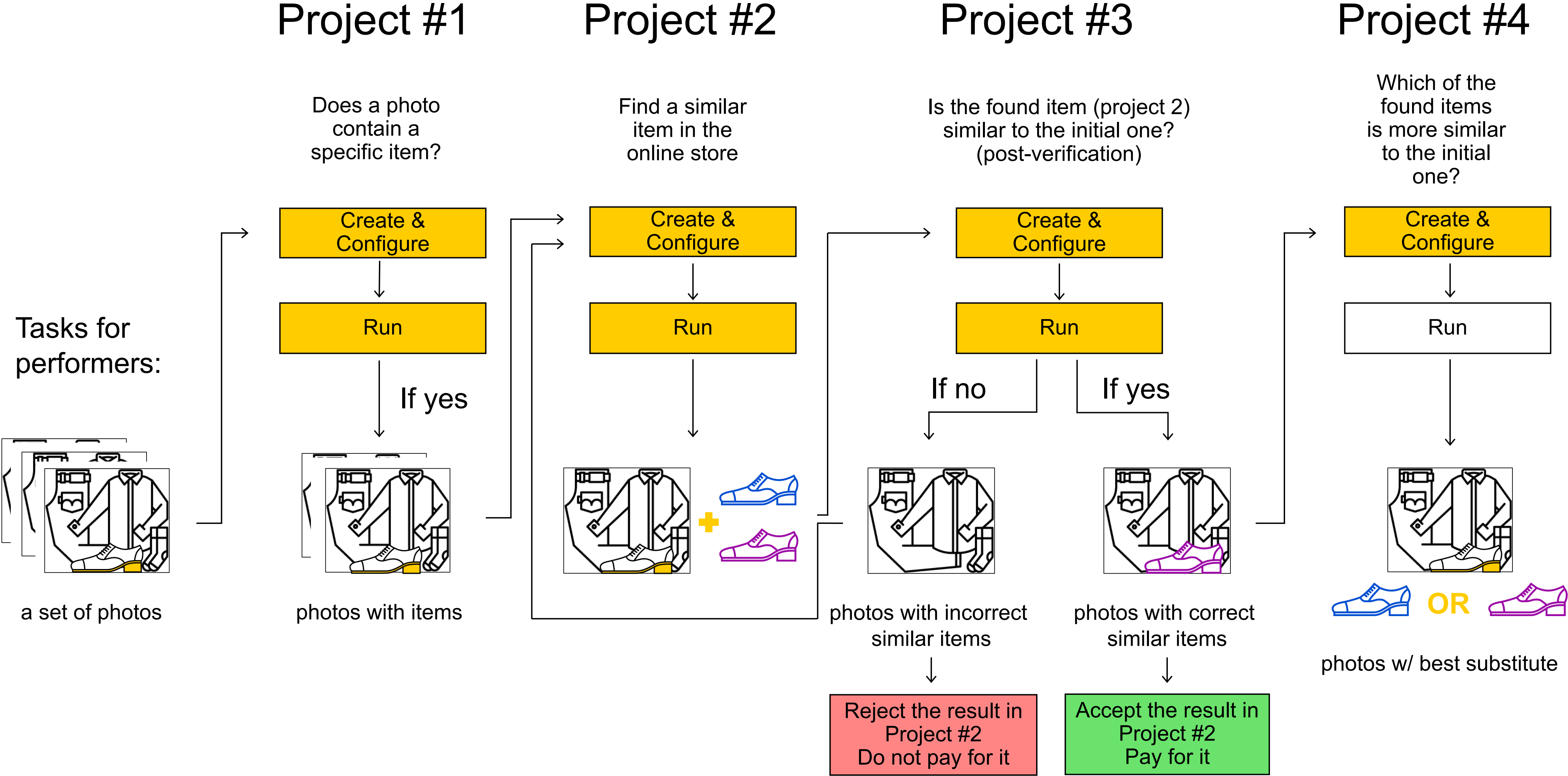
# Step #5



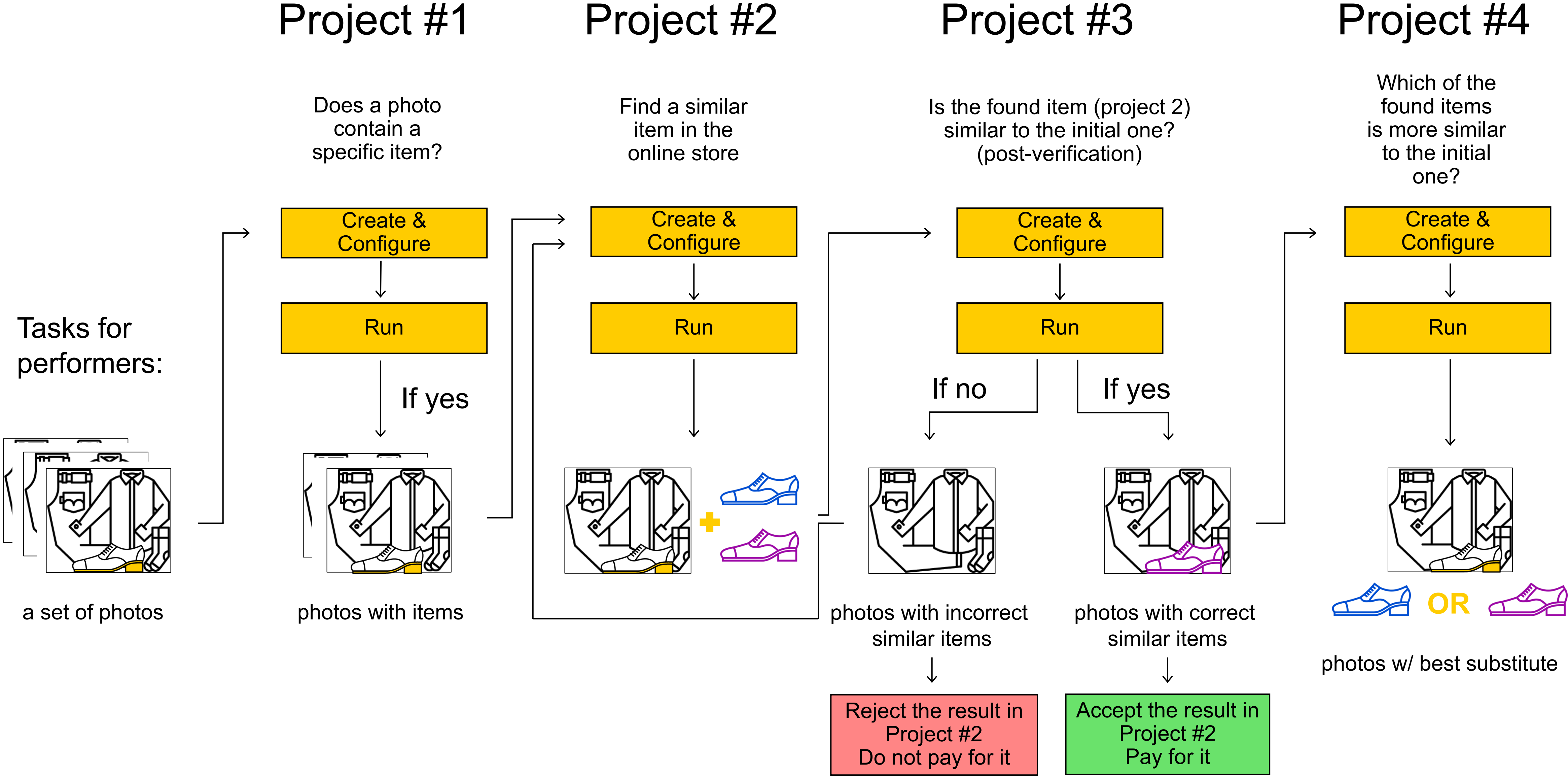
# Step #6



# Step #7



# Step #8





# 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

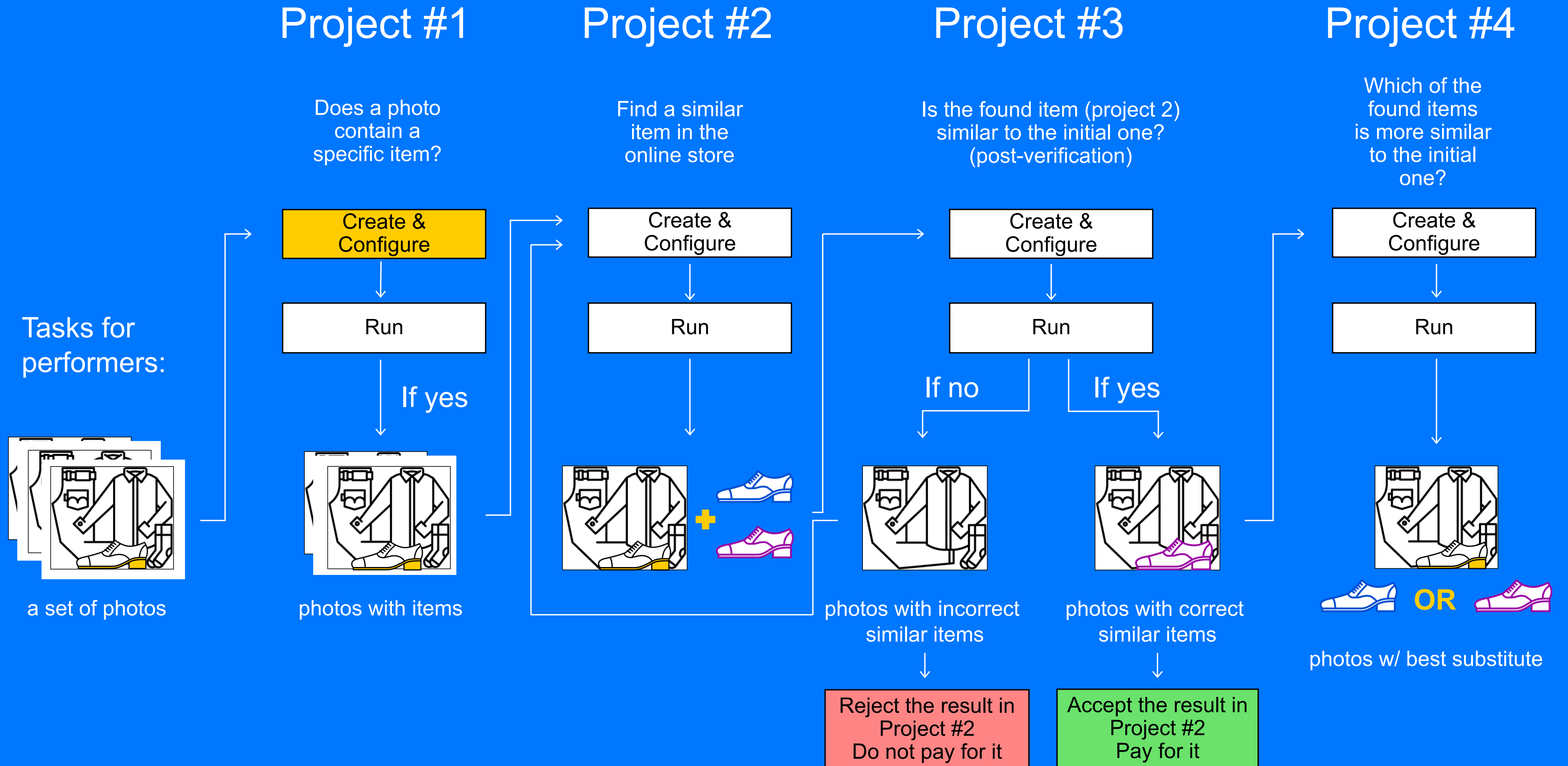
[https://tlk.s3.yandex.net/re substitutes/SIGMOD\\_instruction.pdf](https://tlk.s3.yandex.net/re substitutes/SIGMOD_instruction.pdf)

- The dataset with photos

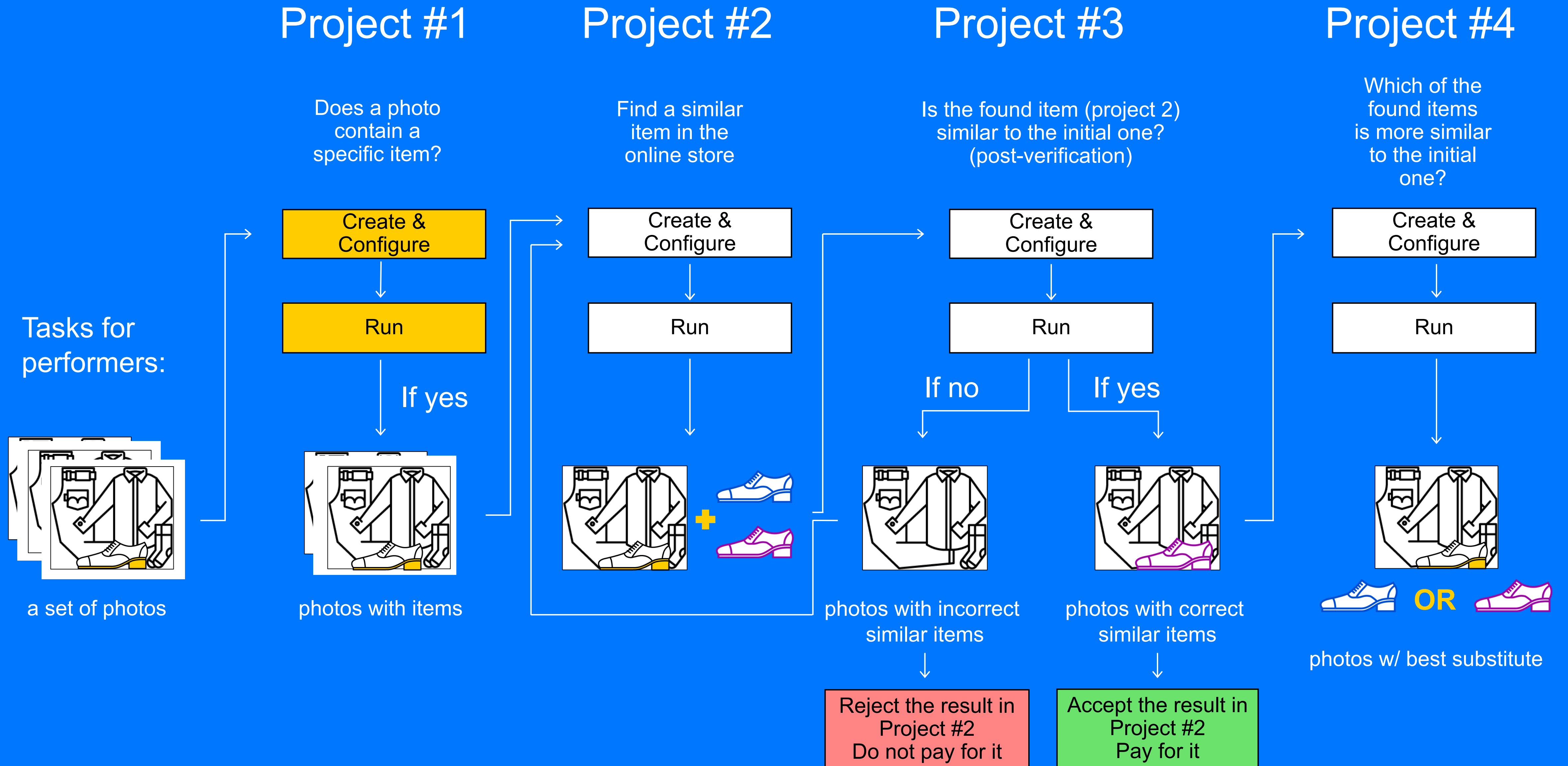
[http://tlk.s3.yandex.net/re substitutes/dataset\\_Y.tsv](http://tlk.s3.yandex.net/re substitutes/dataset_Y.tsv)

where **Y** is any number from 1 to 10

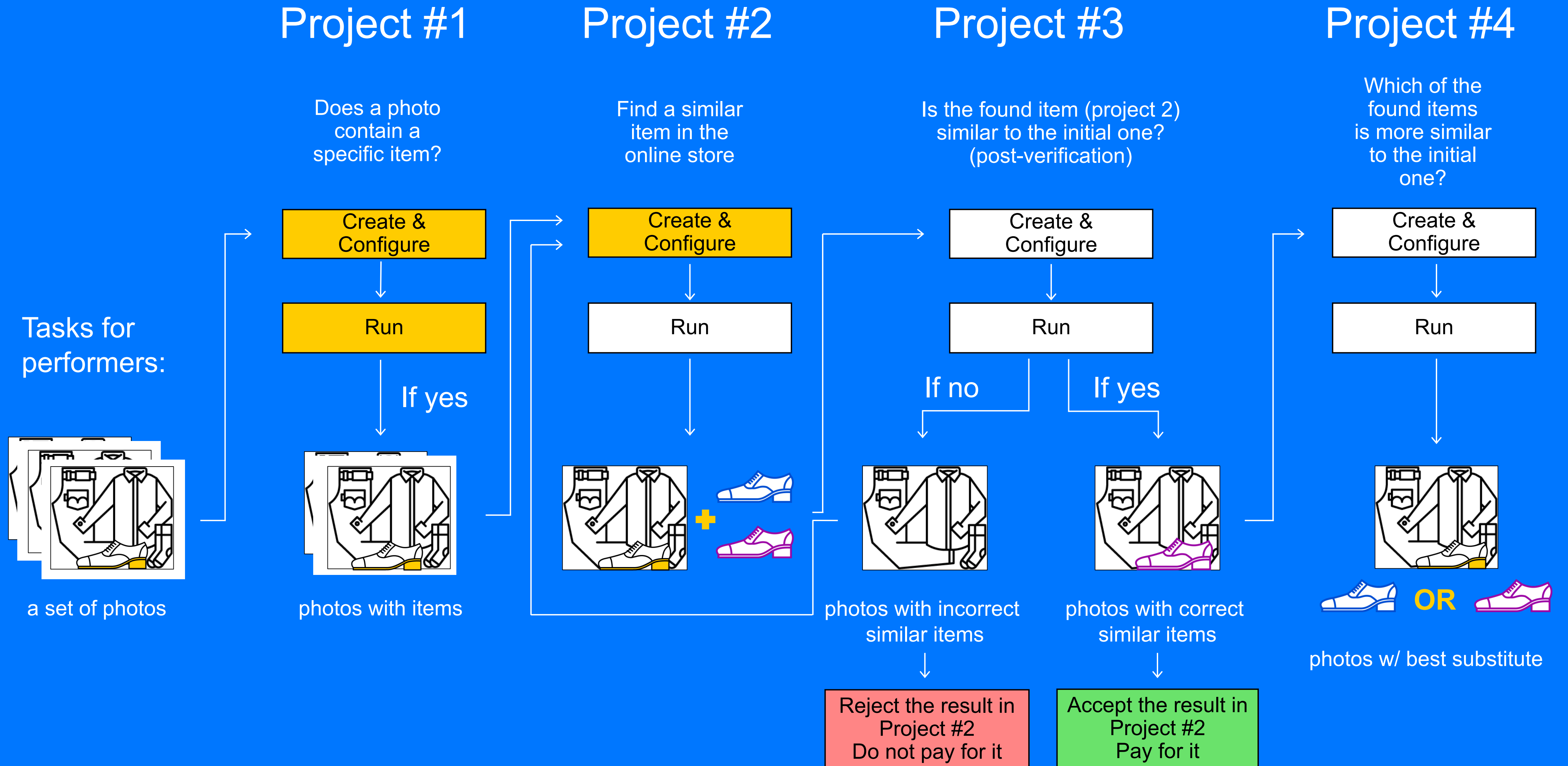
# Most of us are at this step



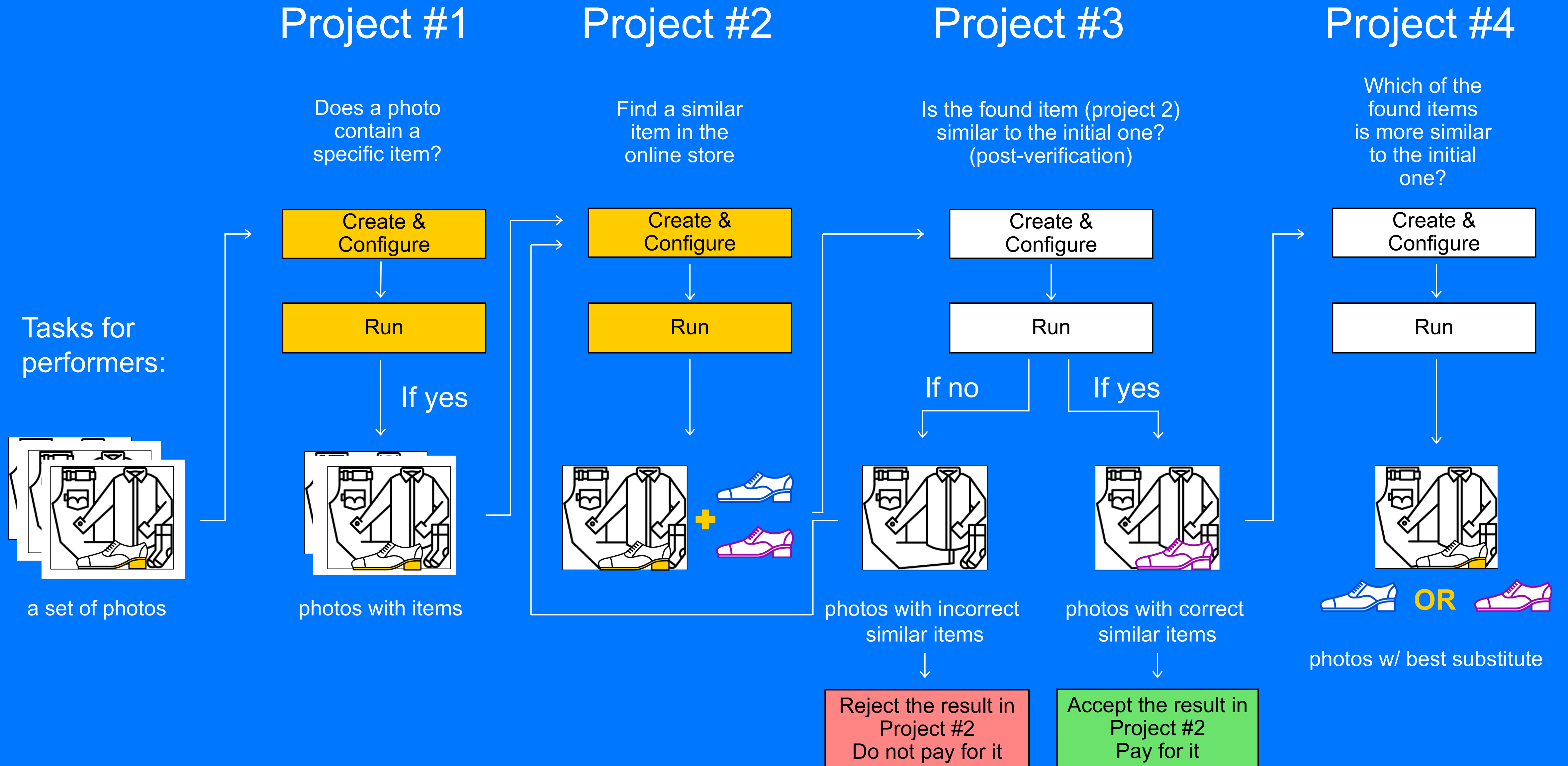
# Most of us are at this step



# Most of us are at this step

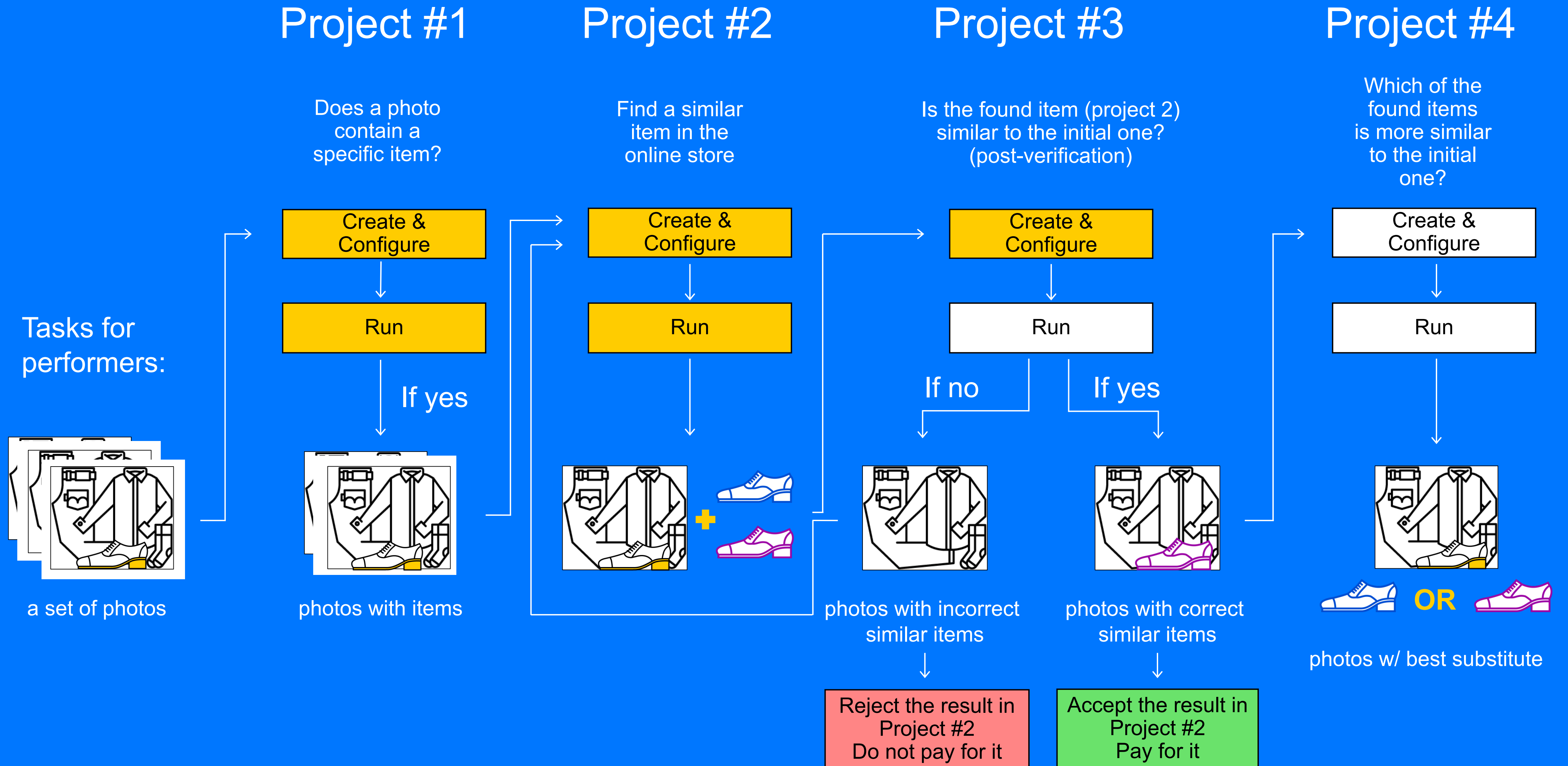


# Most of us are at this step





# Most of us are at this step



**Y**andex

**Thank you!**  
**Questions?**

**Daria Baidakova**  
Project Manager



[dbaidakova@yandex-team.ru](mailto:dbaidakova@yandex-team.ru)



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