

Yandex

Practice of Efficient Data Collection via Crowdsourcing: Aggregation, Incremental Relabelling, 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 outline

Introduction: 20 min

Part I: 40 min
Main Components

Coffee break: 30 min

Part II: 25 min
Brainstorming pipeline

Part III: 10 min
Introduction to Crowd Platform

Part IV: 85 min
Set & Run Projects

Lunch break: 90 min

Part V: 35 min
Interface & Quality control

Part VI: 25 min
Theory on Aggregation

Coffee break: 30 min

Part VI: 60 min
Set & Run Projects cont.

Part VII: 20 min
Incremental relabeling and pricing

Part VIII: 10 min
Results & Conclusions

What you need for the practical session

We are starting the practical session

We give you a card with information and 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

We also provide several copies of a printed version of the instruction

Did everybody receive this card?

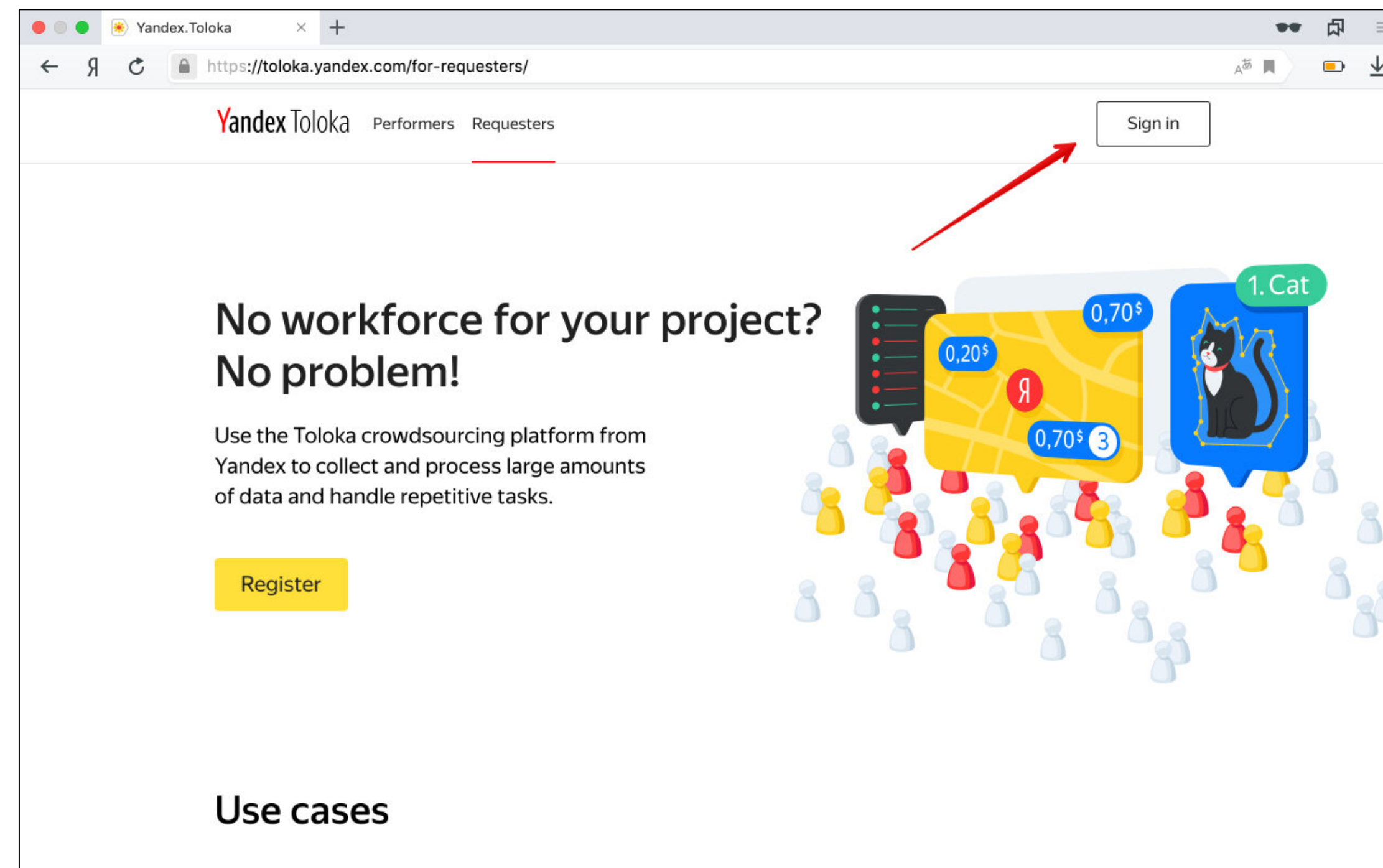
Requester account that you received

You have Login+Password to sign in Yandex.Toloka as a requester

- | The same account is given for several participants (a group)**
 - › So, you can divide work on the project configuration within this group
 - › Or, each member of a group may work individually and create the whole pipeline by her/himself

Sign in Yandex.Toloka as a requester

1. Go to <https://ya.cc/t1k>
2. Click on “Sign in” in the top-right corner
3. Use received Login+Password to sign in



Requester account that you received

You have Login+Password to sign in Yandex.Toloka as a requester

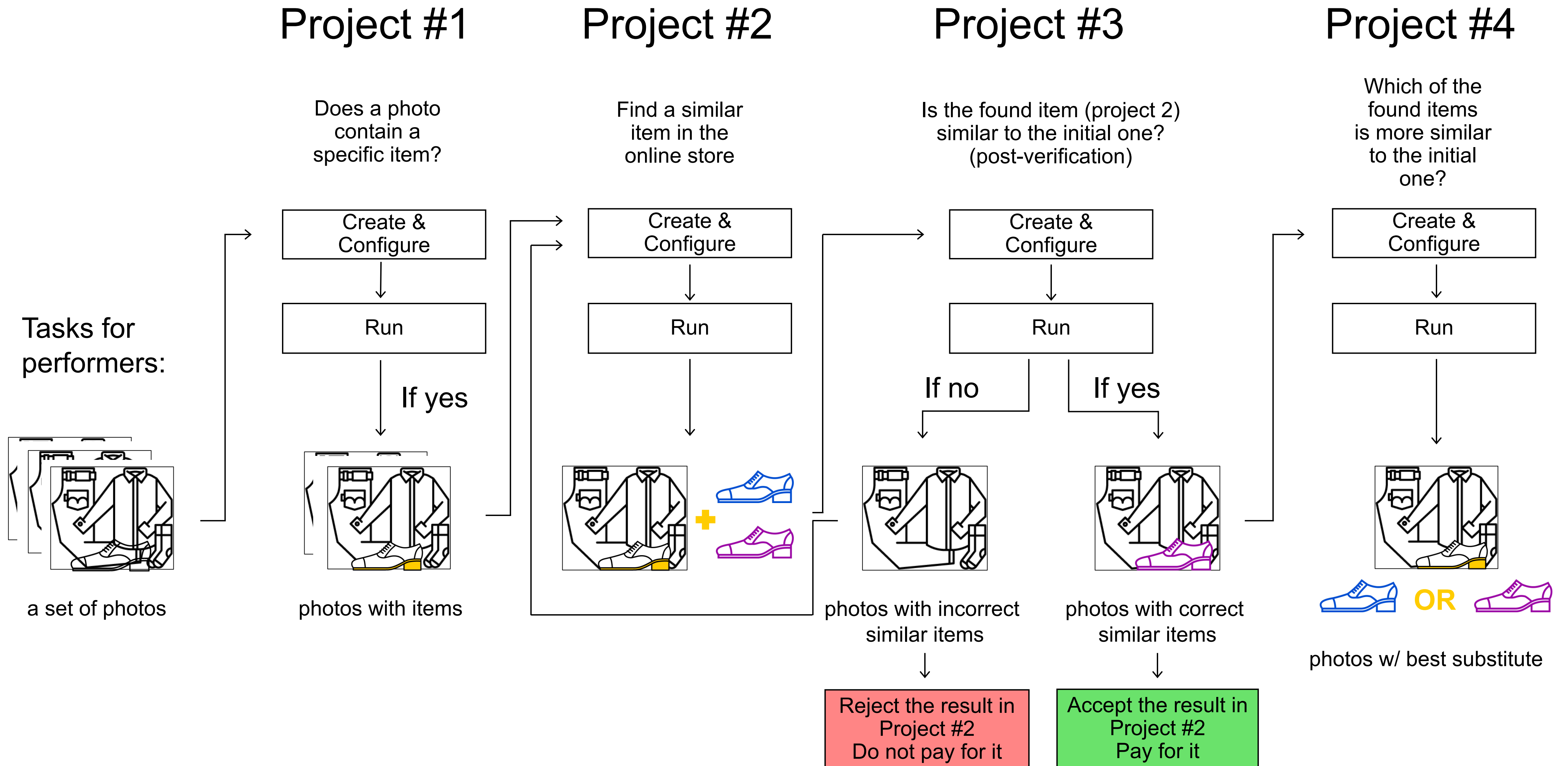
- The account of this requester has money**
- › So, you will run your tasks on real crowd performers!

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



You can divide work within a participant group

Project #1

Project #2

Project #3

Project #4

Does a photo contain a specific item?

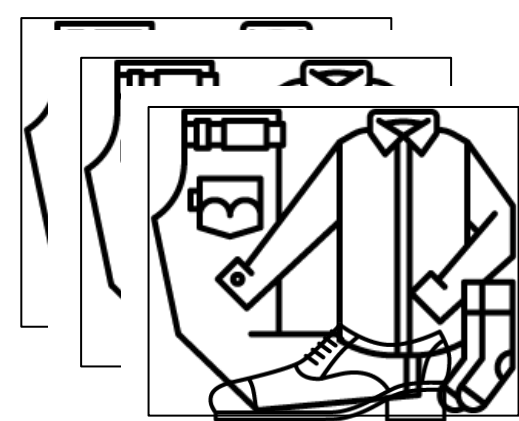
Find a similar item in the online store

Is the found item (project 2) similar to the initial one? (post-verification)

Which of the found items is more similar to the initial one?

Can do in parallel

Tasks for performers:

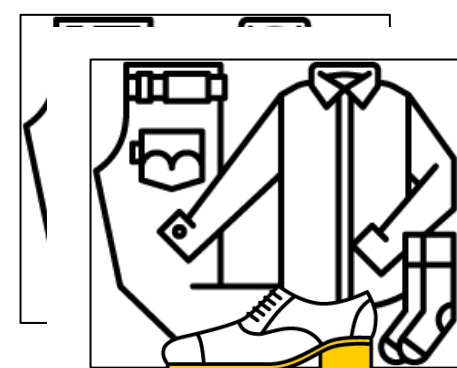


a set of photos

Create & Configure

Run

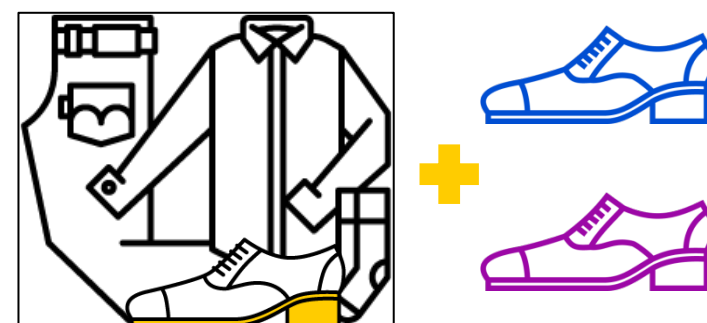
If yes



photos with items

Create & Configure

Run



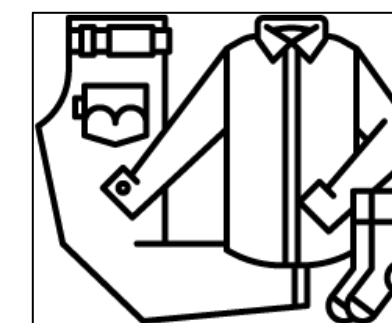
photos with incorrect similar items

Reject the result in Project #2
Do not pay for it

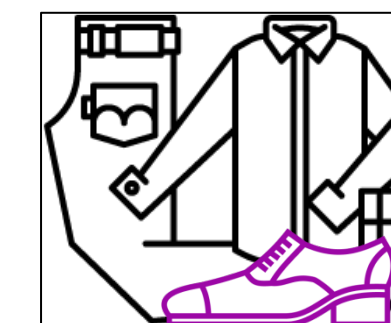
Create & Configure

Run

If no



If yes

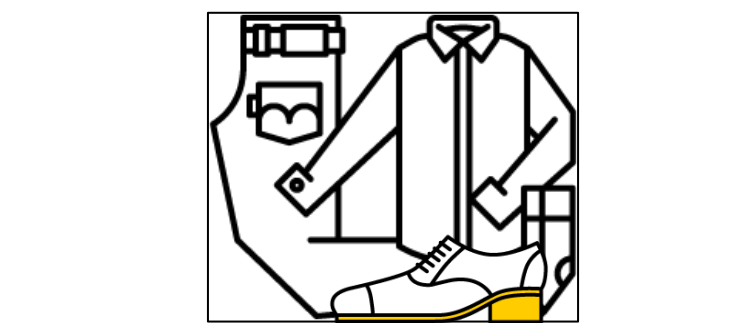


photos with correct similar items

Accept the result in Project #2
Pay for it

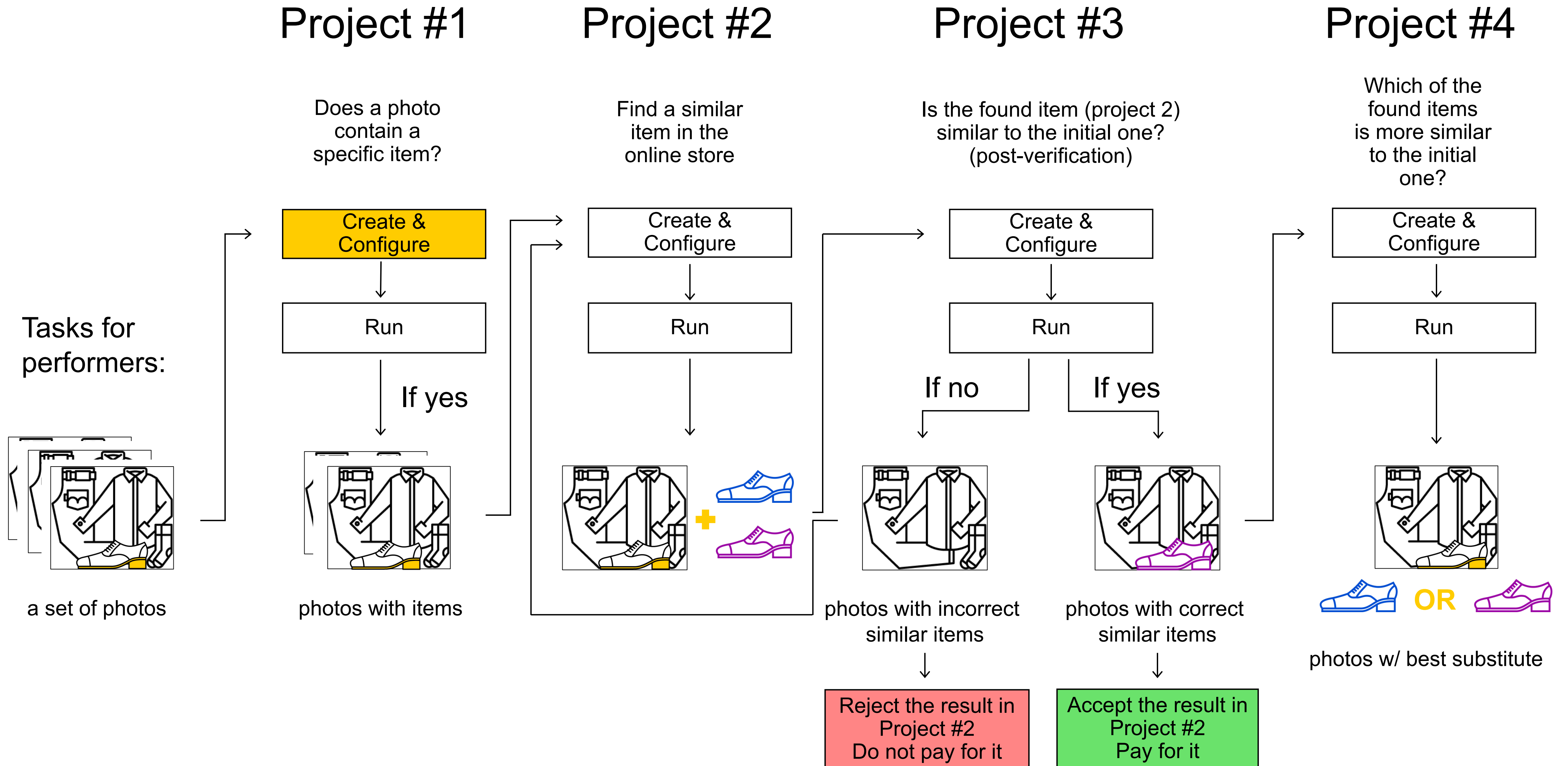
Create & Configure

Run

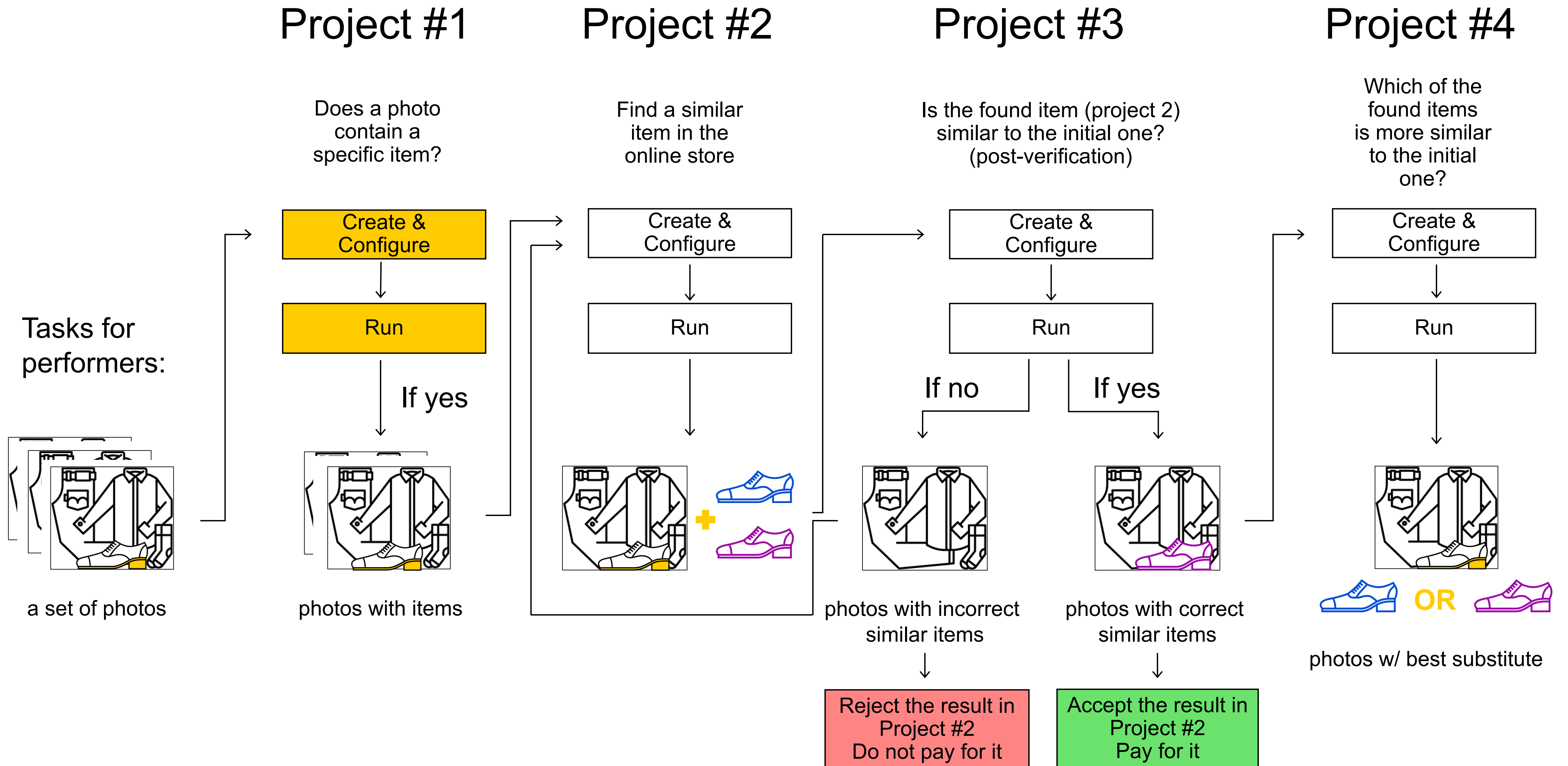


photos w/ best substitute

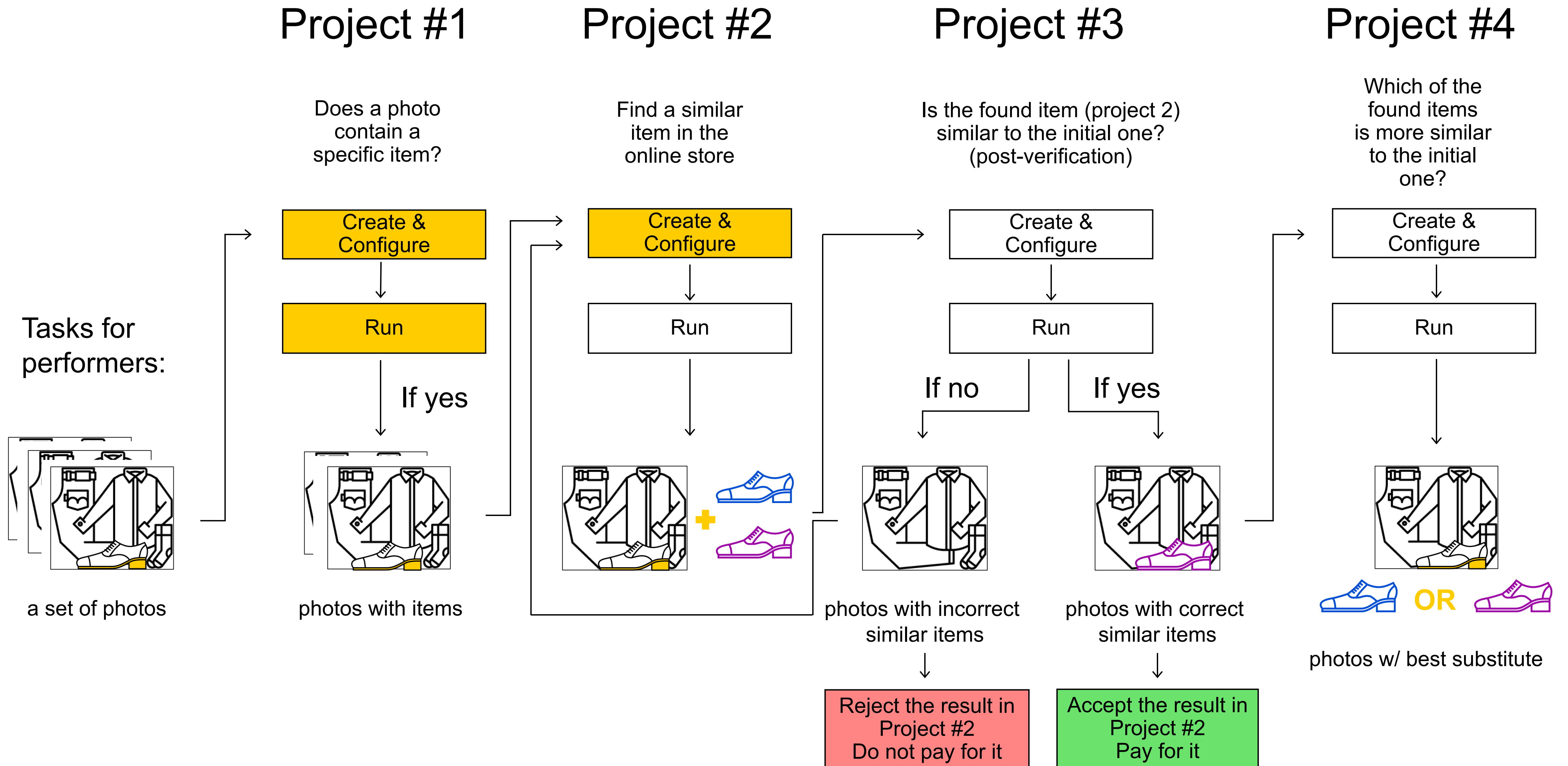
Step #1



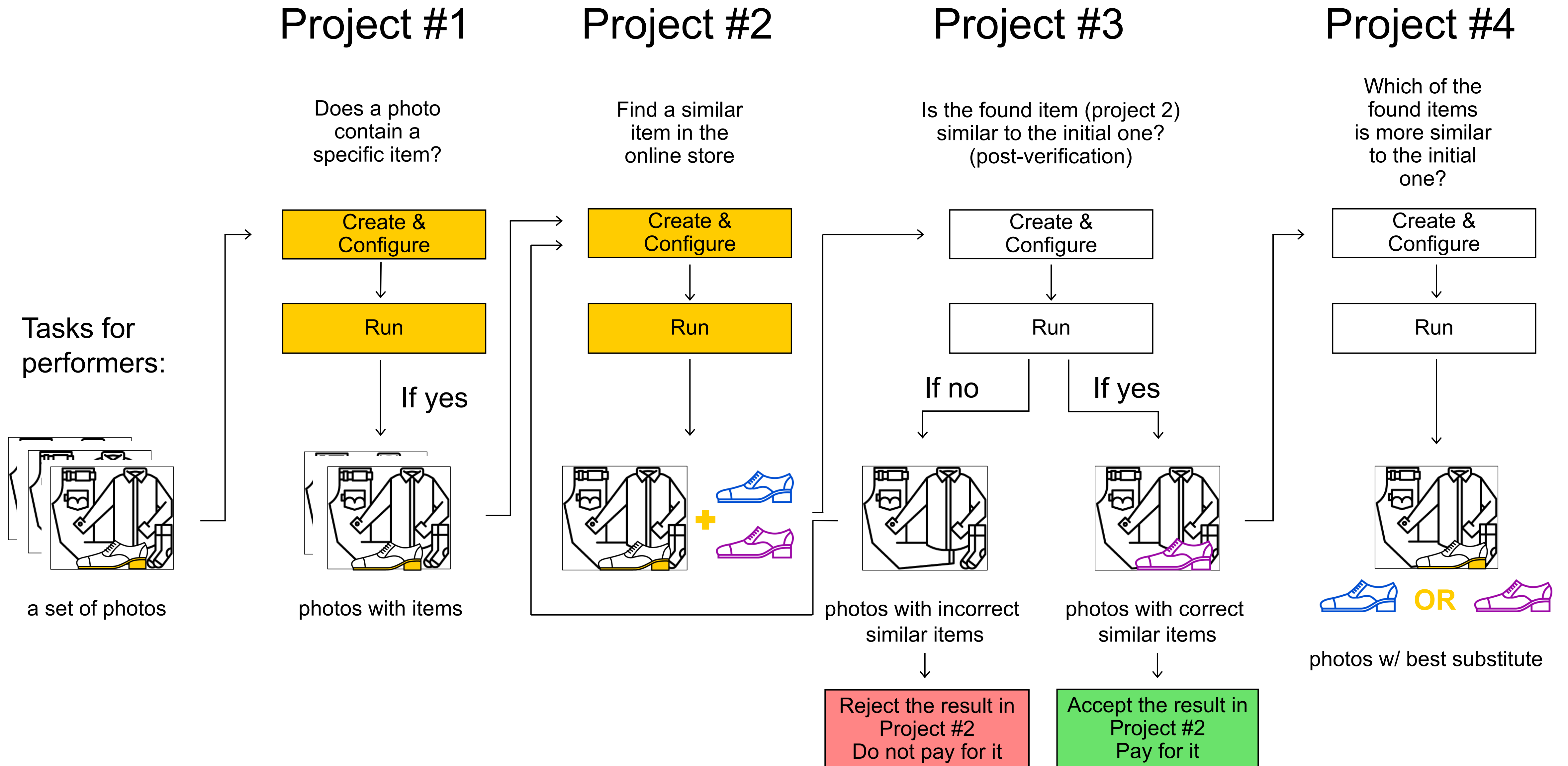
Step #2



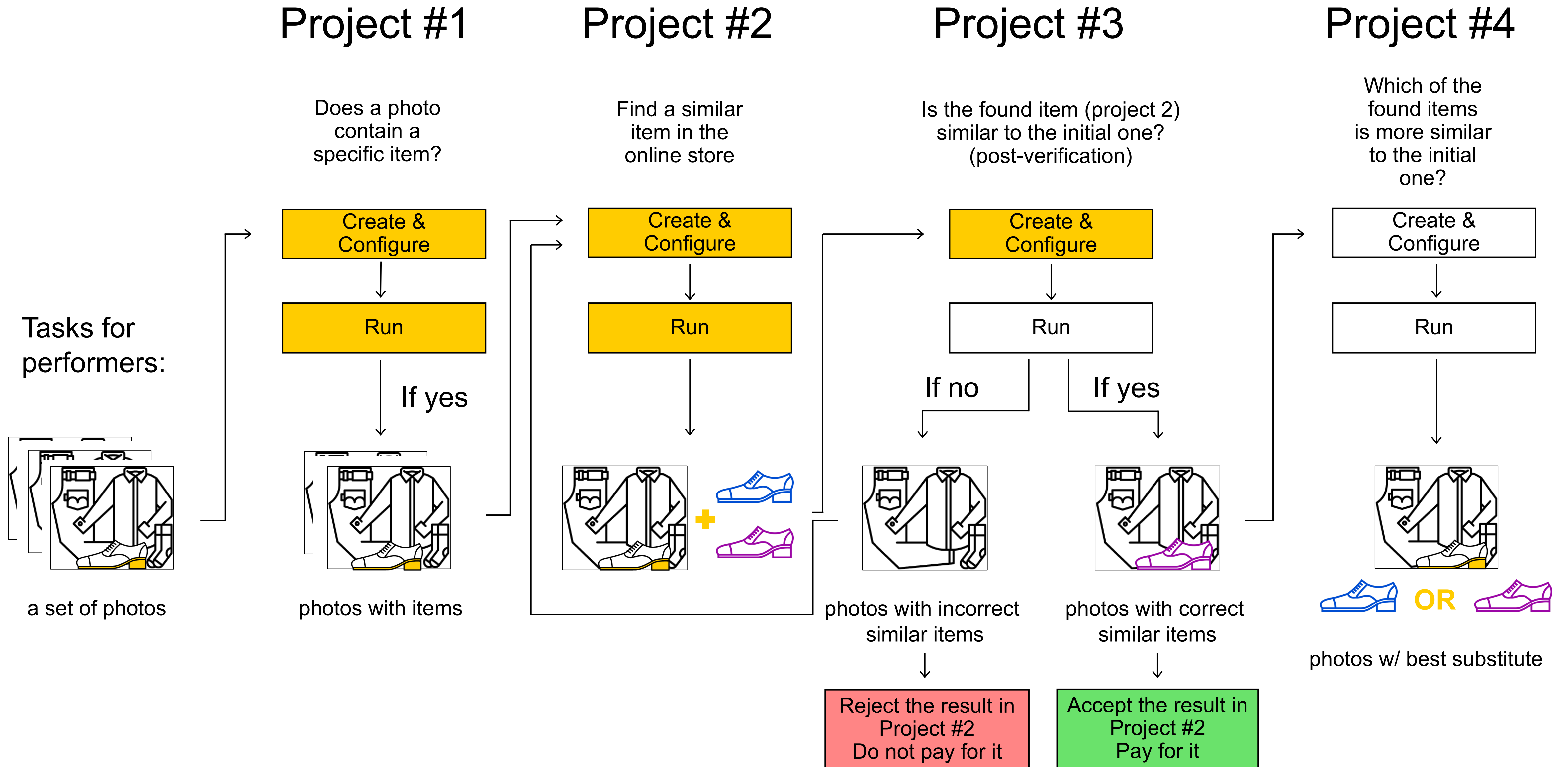
Step #3



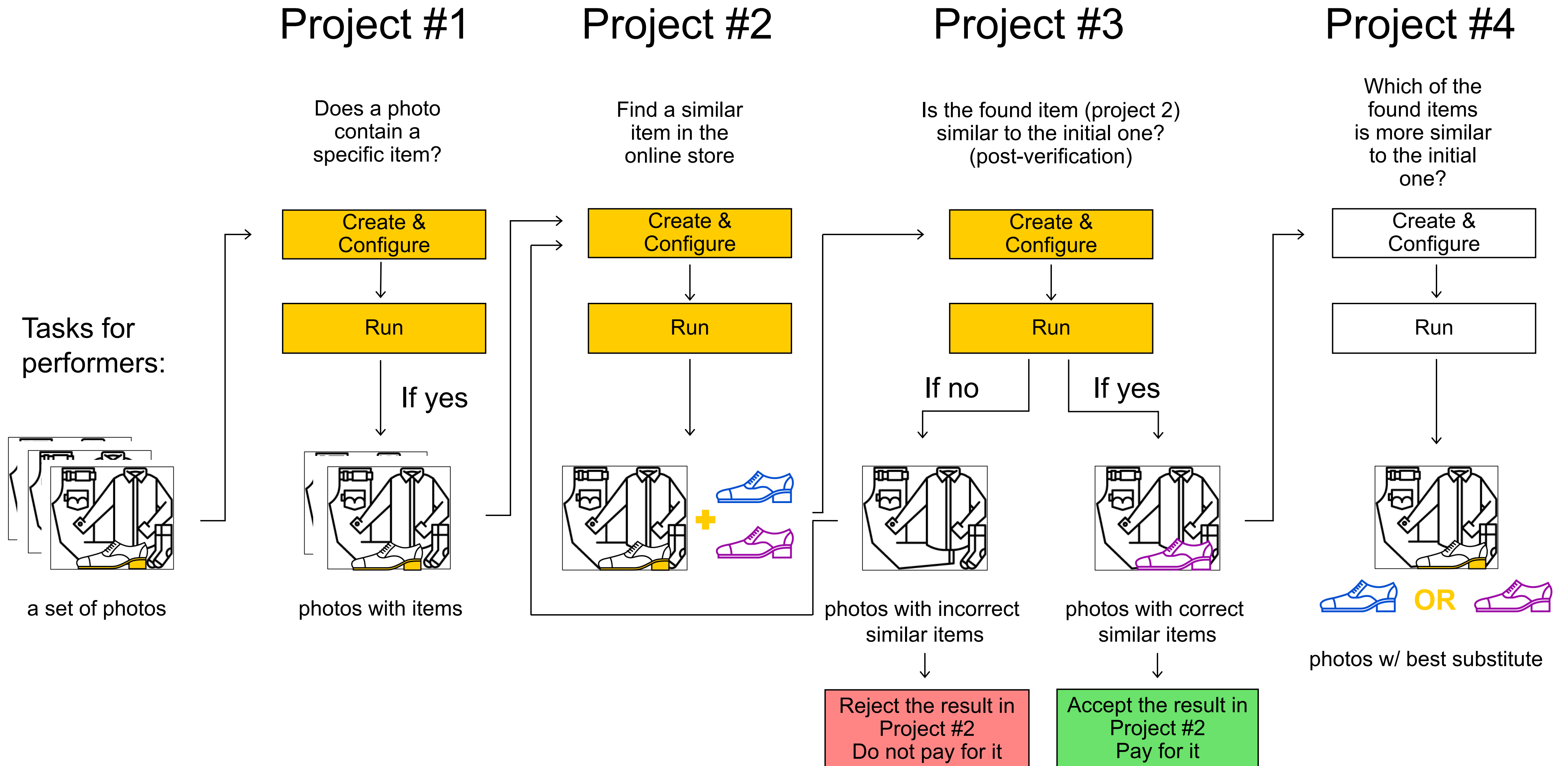
Step #4



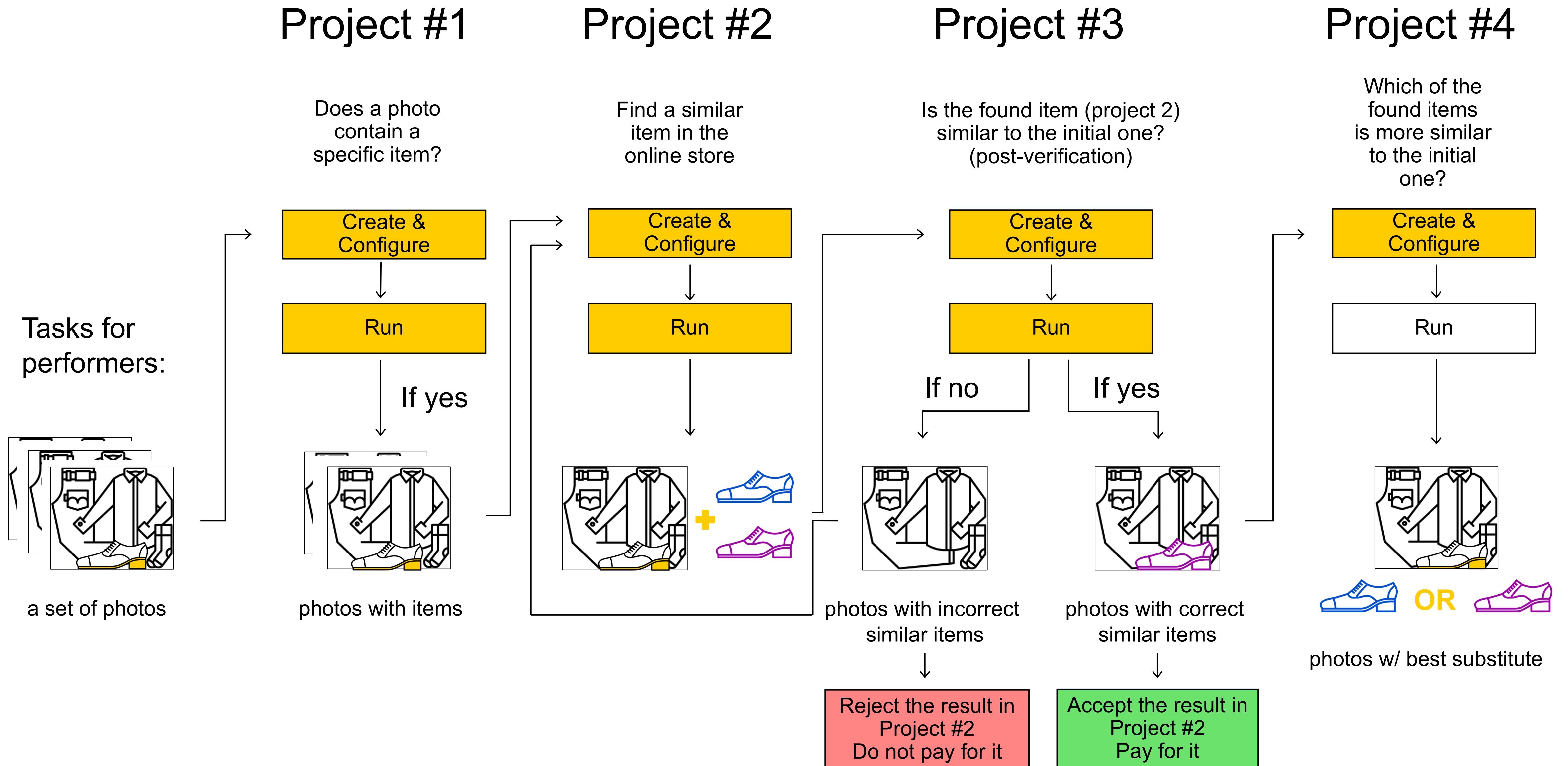
Step #5



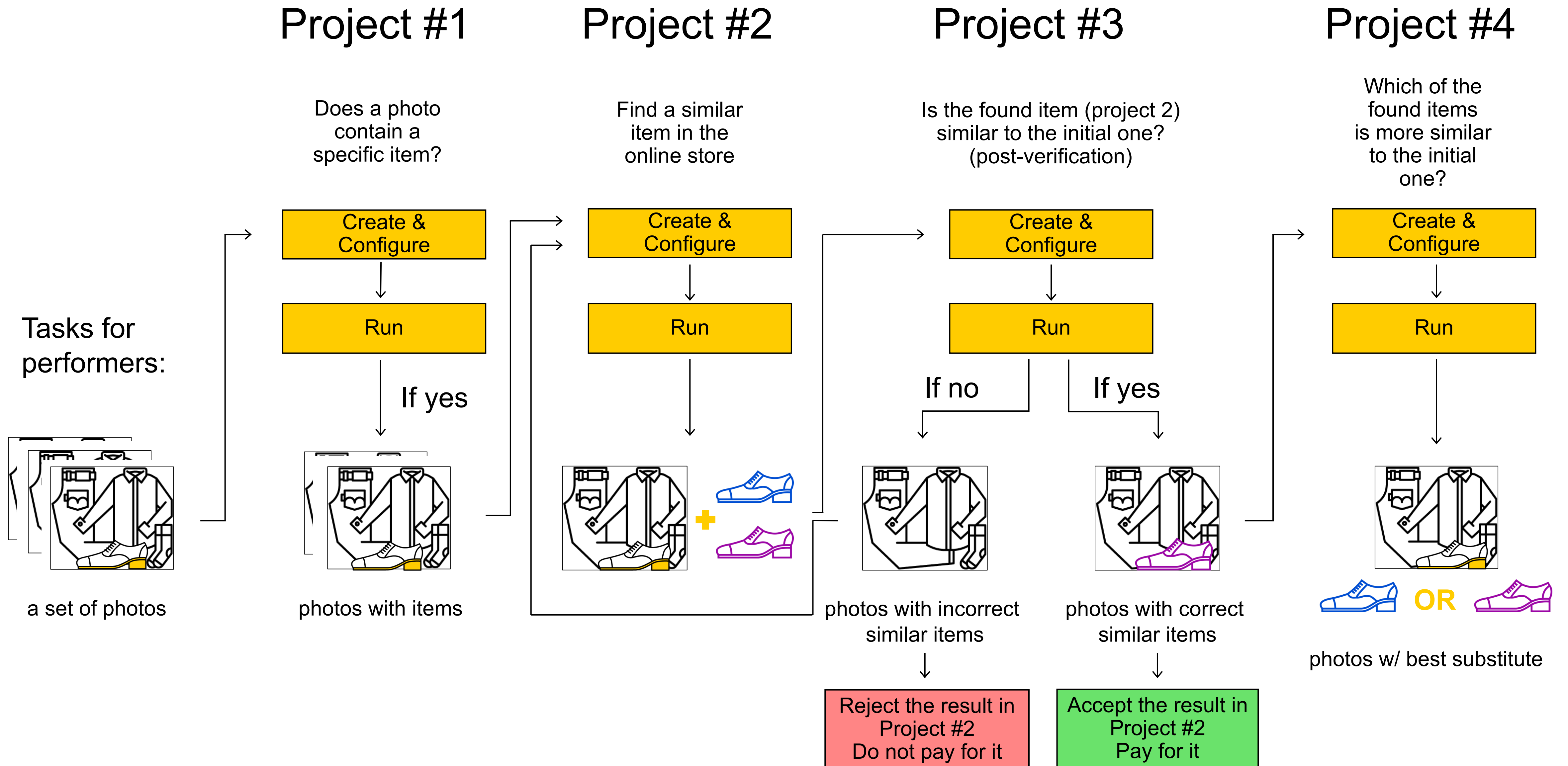
Step #6



Step #7



Step #8



Practice pt 1 starts

You create 4 crowd projects

You connect them in a pipeline

You run them on real performers

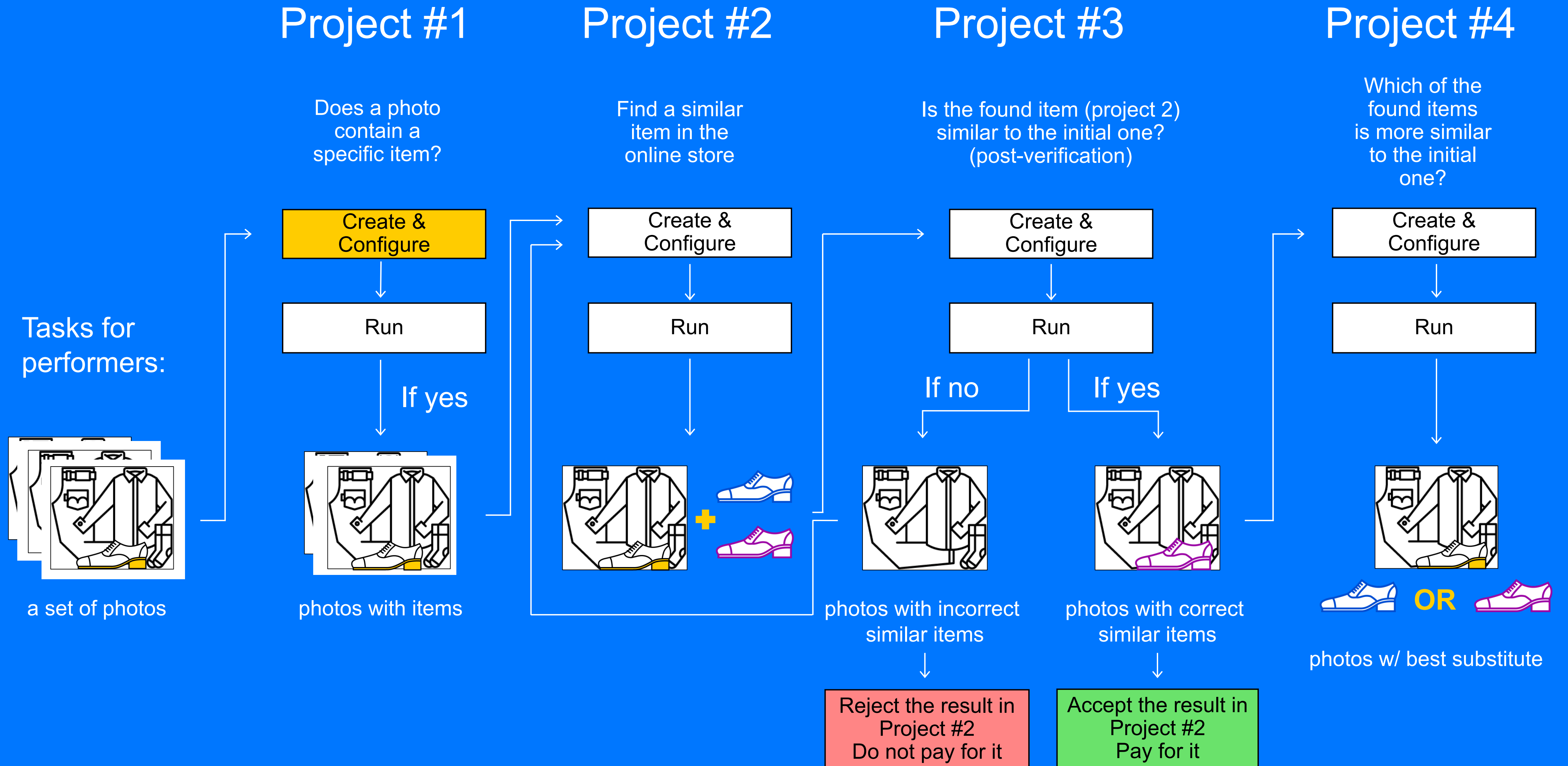
You have:

- › The instruction with all details
- › The dataset with photos

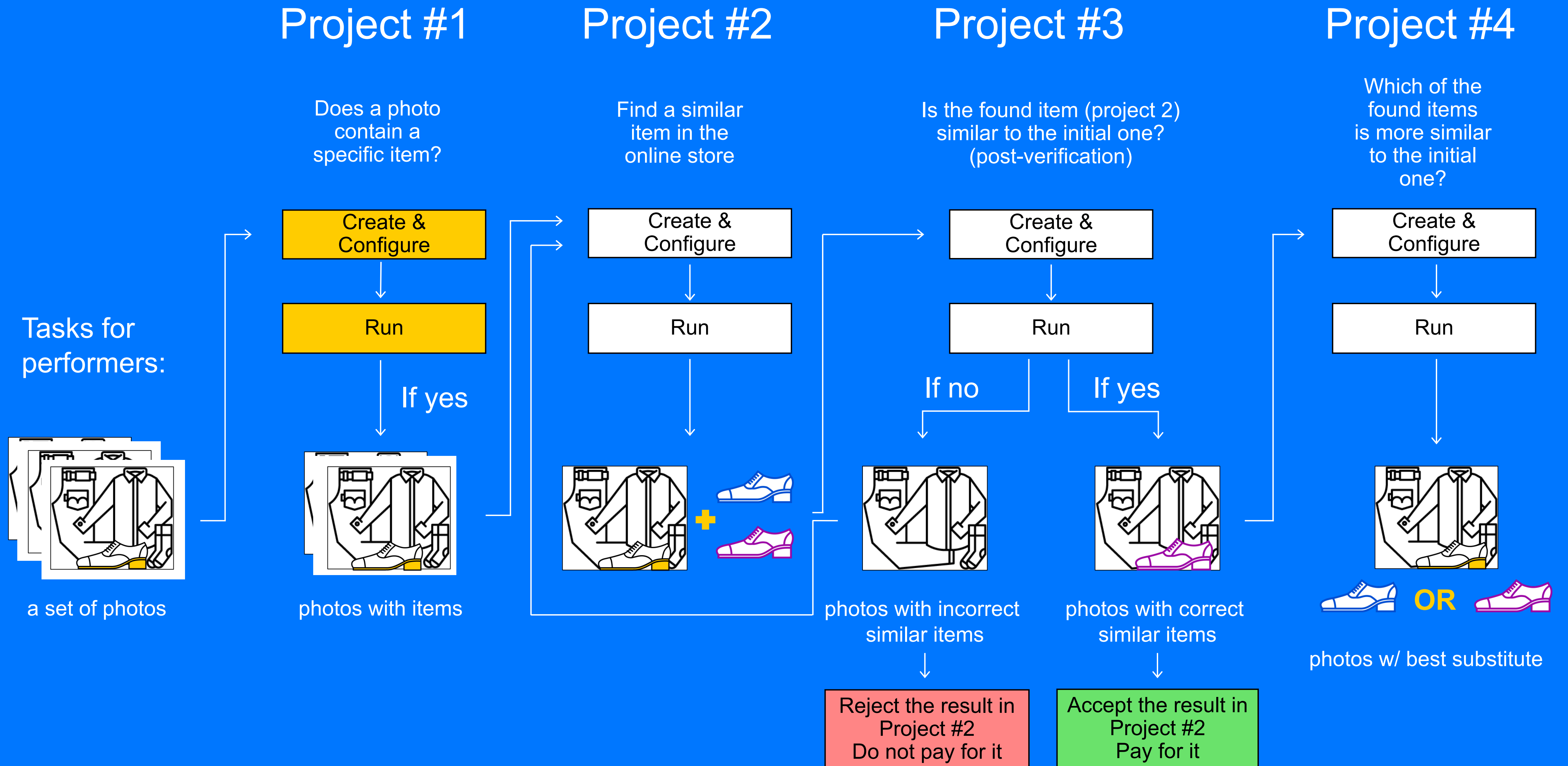
Tutorial chat with materials:



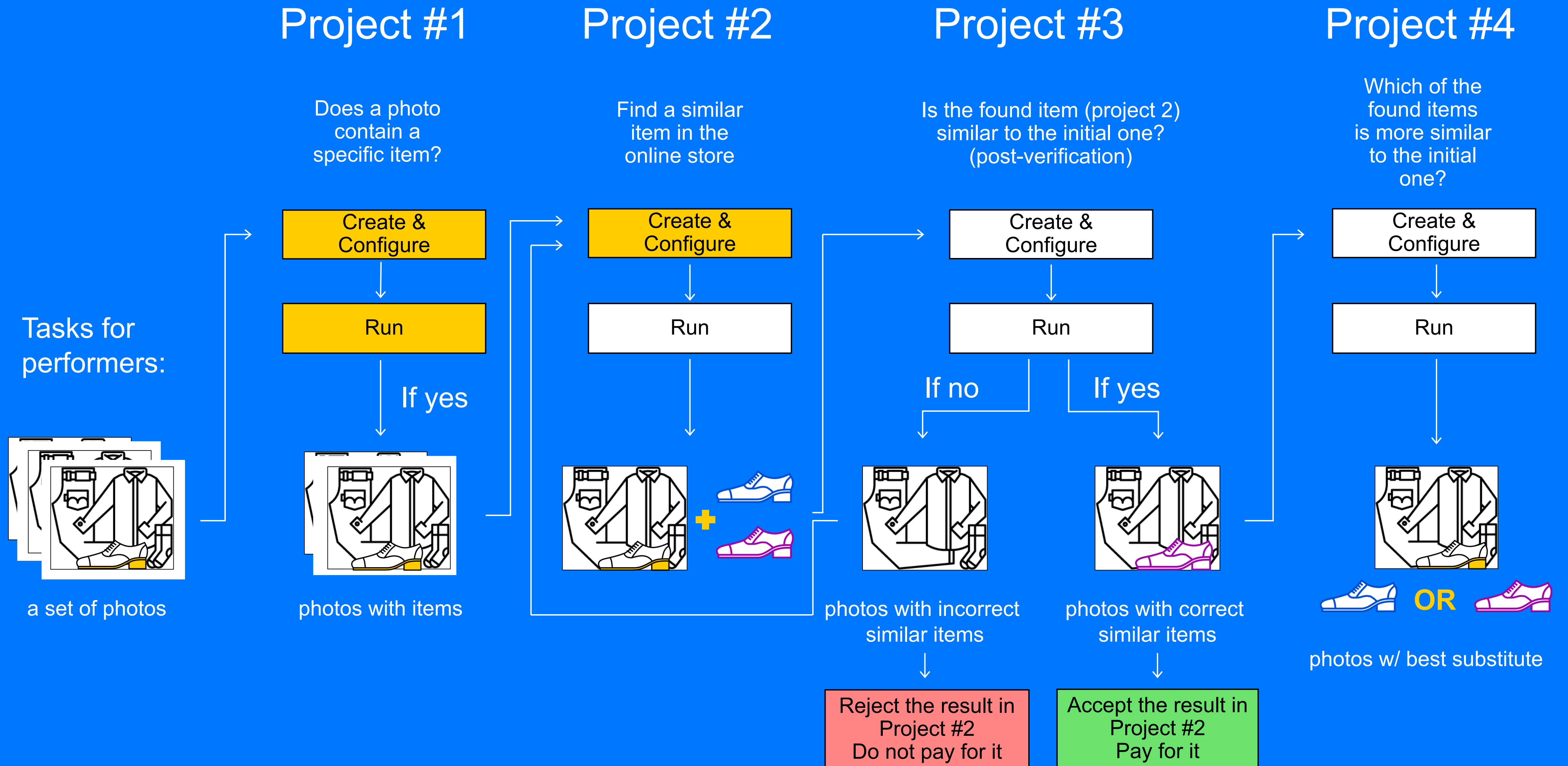
Most of us are at this step



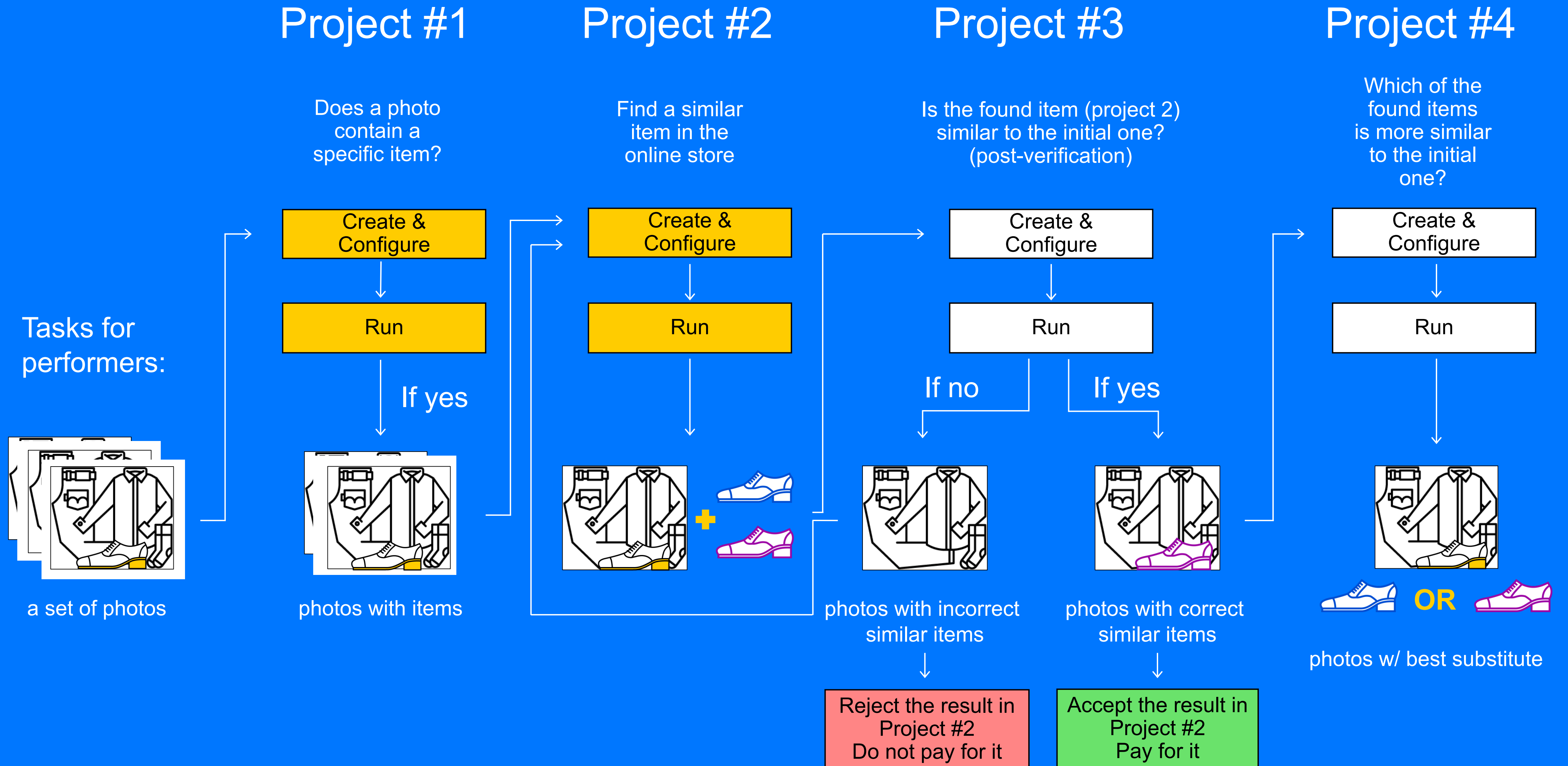
Most of us are at this step



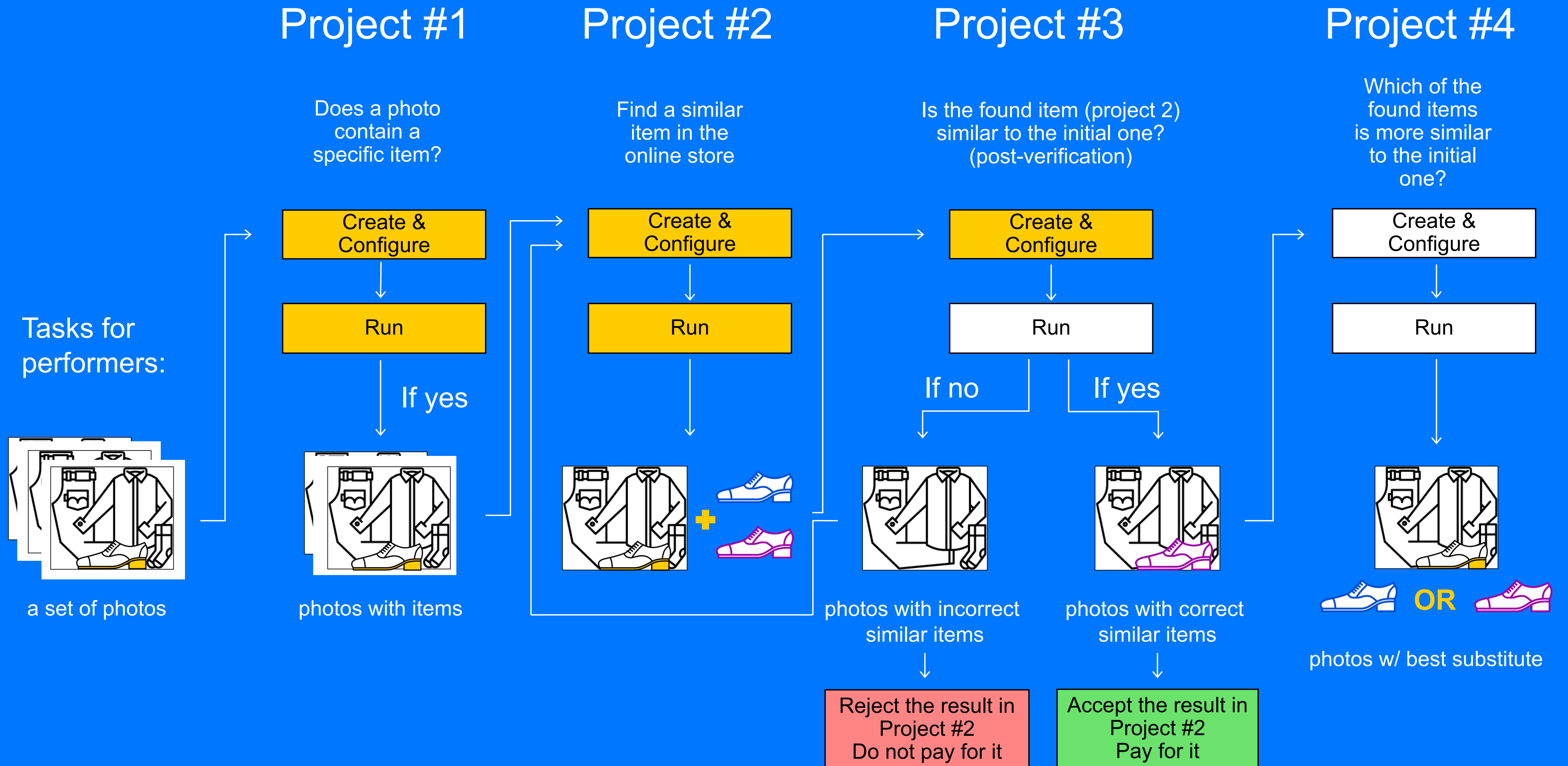
Most of us are at this step



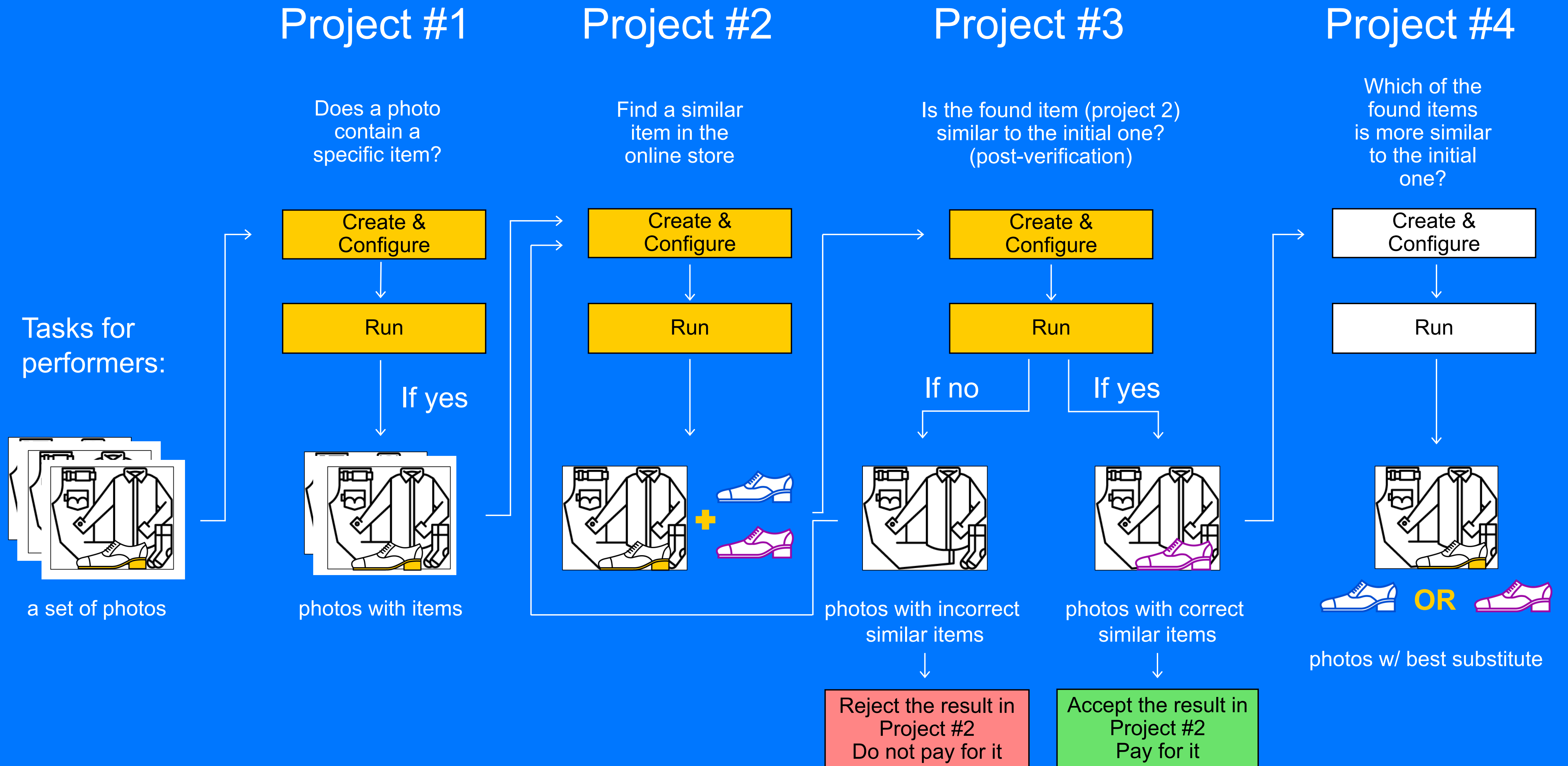
Most of us are at this step



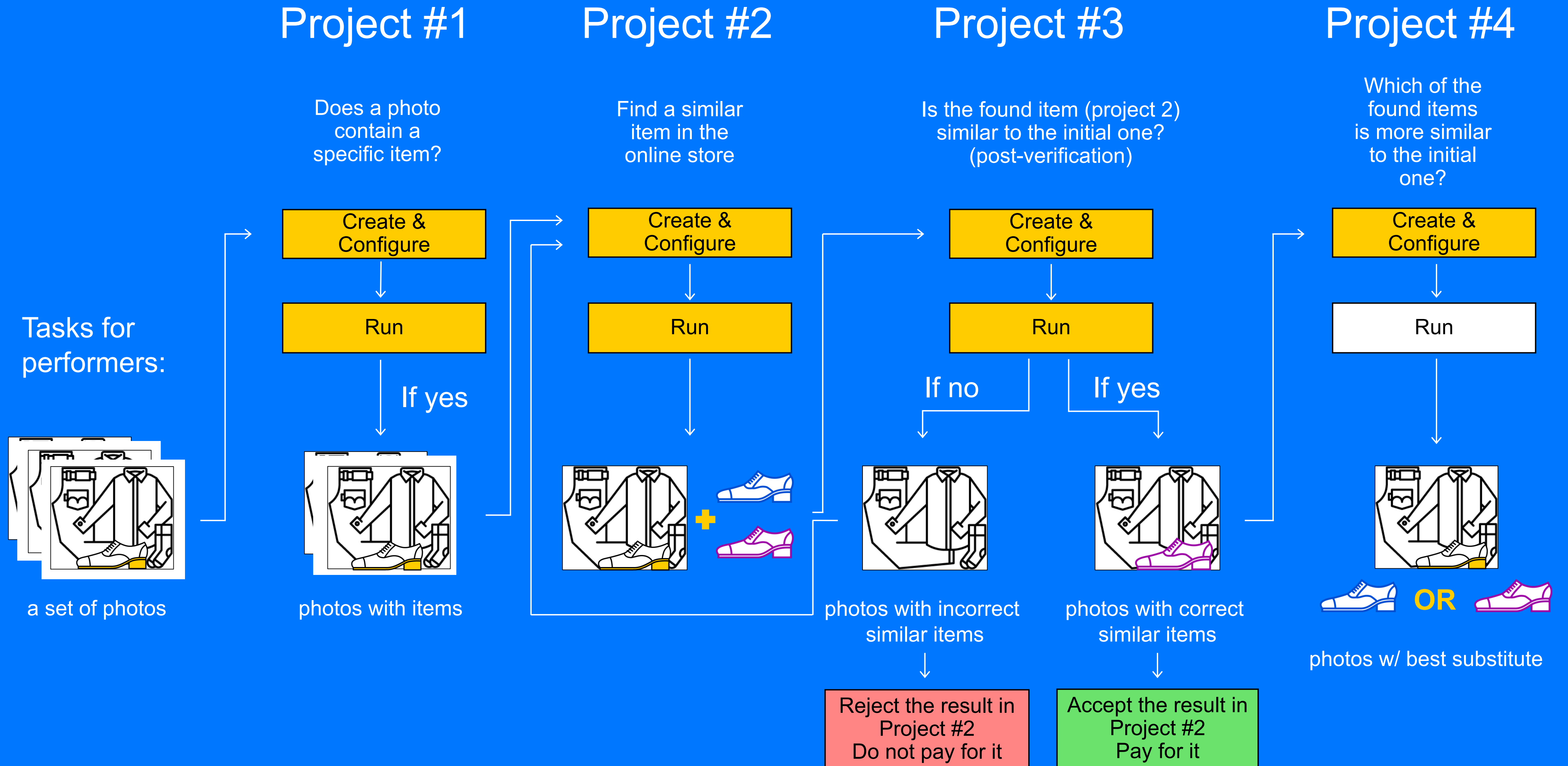
Most of us are at this step



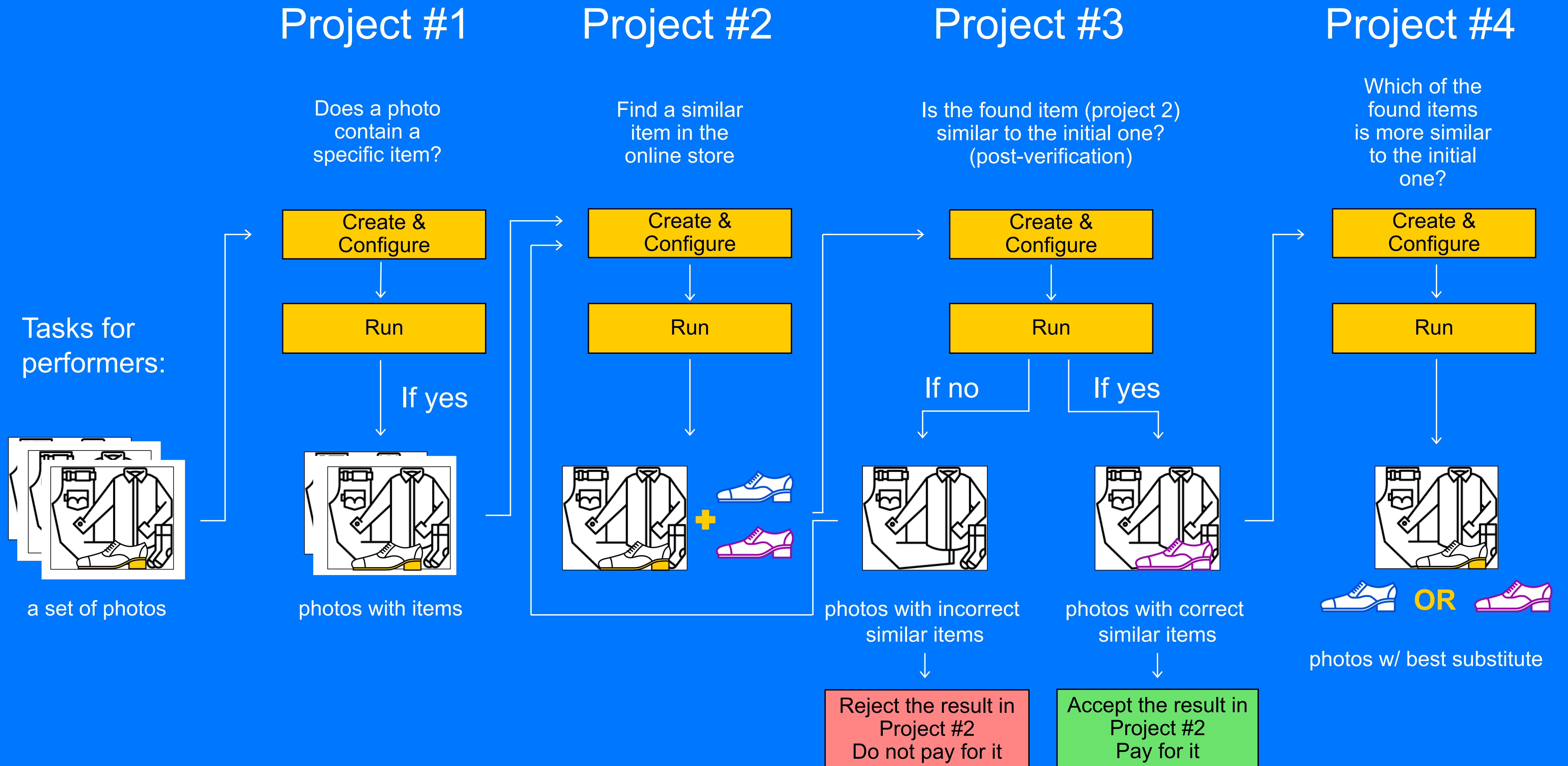
Most of us are at this step



Most of us are at this step



Most of us are at this step



Yandex

**Thank you!
Questions?**

Daria Baidakova

Project Manager



dbaidakova@yandex-team.ru



<https://research.yandex.com/tutorials/crowd/kdd-2019>