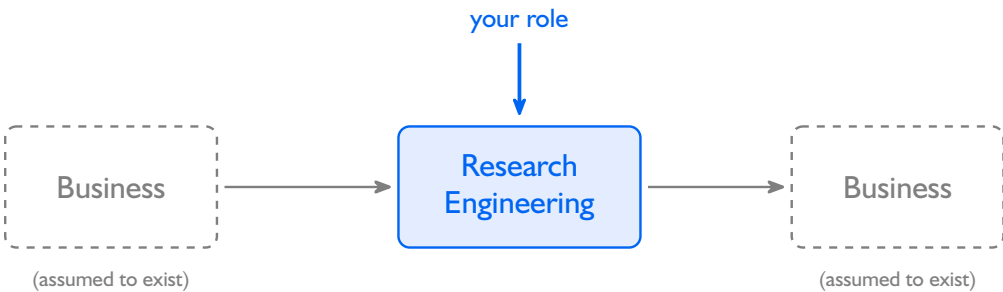
CV/ML Industry

Andrey Stotskiy



29 October 2026

Scope


your role
Business
Research
Engineering
Business
(assumed to exist)
(assumed to exist)

Typical ? CV/ML Work Pipeline

1



read technical specification

Typical ? CV/ML Work Pipeline

1



read technical specification

2



download train & test data

Typical ? CV/ML Work Pipeline

1



read technical specification

2



download train & test data

3



open jupyter on your laptop

Typical ? CV/ML Work Pipeline

1



read technical specification

2



download train & test data

3



open jupyter on your laptop

4

```
from torch import magic
# вжух 🐱 🪄 ✨
magic.solve_problem()
```

write code, get SoTA

Typical ? CV/ML Work Pipeline

1



read technical specification

2



download train & test data

3



open jupyter on your laptop

4

```
from torch import magic
# вжух 🐱 🪄 ✨
magic.solve_problem()
```

write code, get SoTA

5





send `model.pt` to boss

Typical ? CV/ML Work Pipeline

1



read technical specification

2



download train & test data

3



open jupyter on your laptop

4

```
from torch import magic
# вжух 🐱 🪄 ✨
magic.solve_problem()
```

write code, get SoTA

5



send `model.pt` to boss

6



get paid


My job here is done.

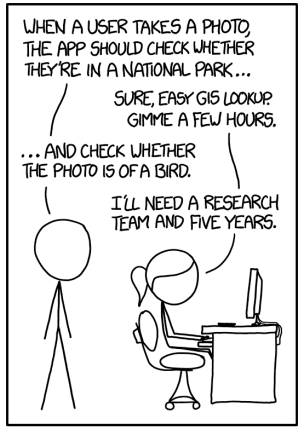
Identify the Task



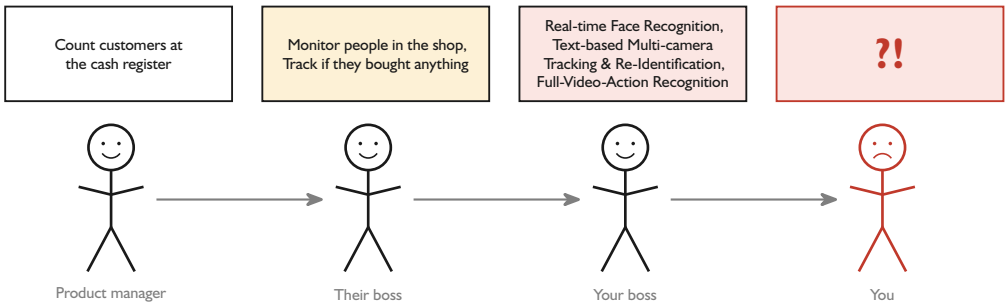
Identify the Task



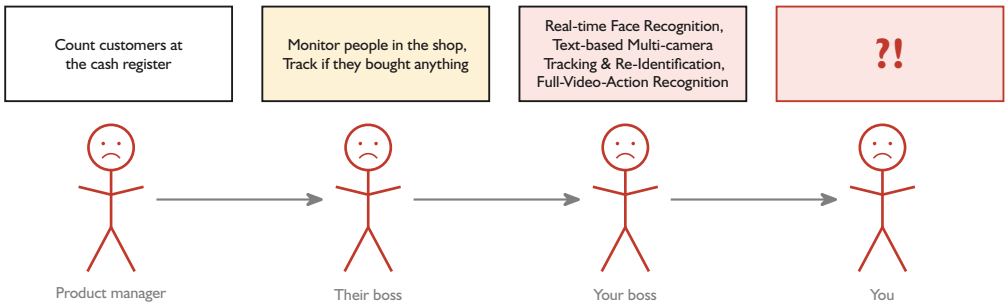
xkcd 1425: Tasks



Identify the Task

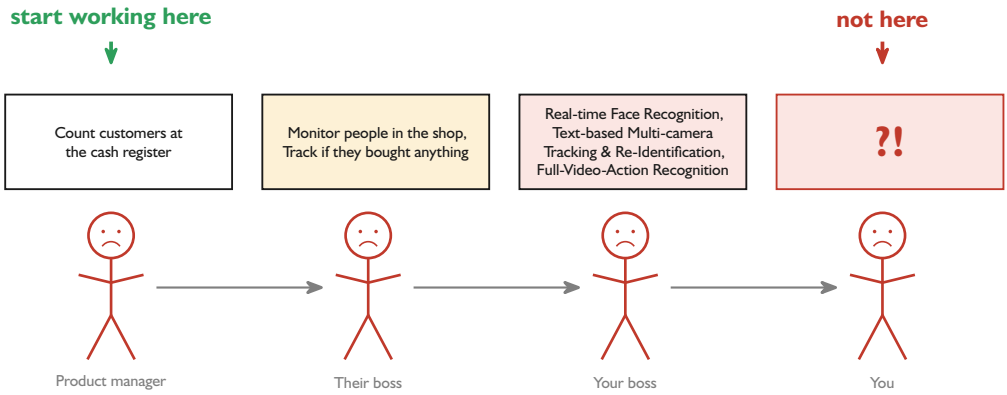


Identify the Task

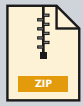


everyone is sad

Identify the Task



Prepare the Data

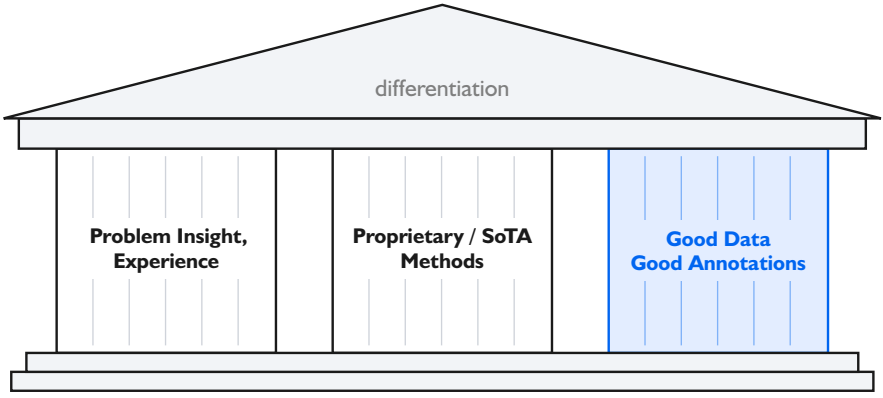


download train & test data

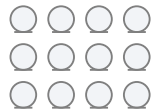
Prepare the Data



Why would anyone pay *you*?



Who labels your data?



Crowd platforms

MTurk, Toloka,
Яндекс.Задания



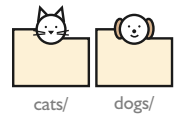
In-house team

CVAT, etc
(self-hosted)



Domain experts

Doctors,
Engineers,
etc



You

Any Image Viewer,
One Free Weekend

low cost, low quality

high quality, high cost

Practice Data Hoarding



source image: `cam07/2024-03-11.jpg`

detector: `plates-v3.2, conf 0.91`

crop box: `(412, 230, 188, 52)`

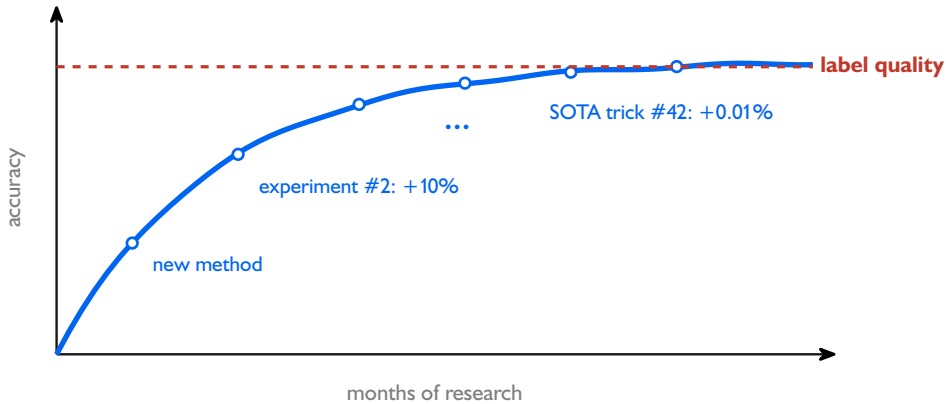
labeling tool: `CVAT, task 118`

labeled by: `annotator #7`

labeled at: `2024-03-12 14:02`

label: `AP31[FT]V7918`

Limiting factor: Annotation quality



Train the Model

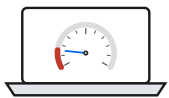


open jupyter on your laptop

Train the Model

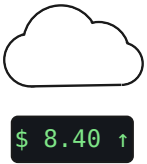


Where to train?



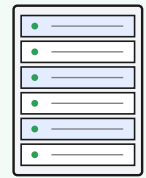
Laptop / Colab

for debugging only



Rent in the cloud

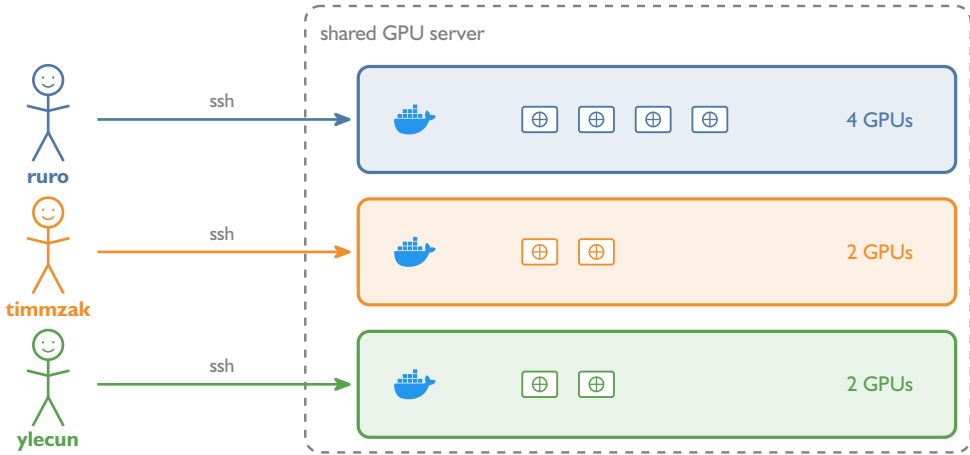
too expensive long term



Buy your own servers

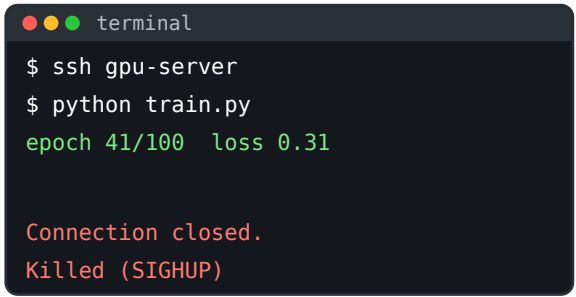
min ammortized costs,
max control & convenience

Sharing Compute

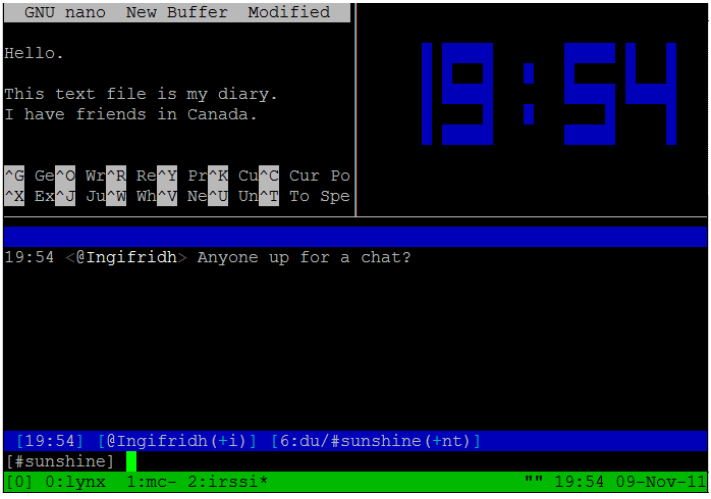


slurm, singularity, docker (with the nvidia/cuda image)

Long-running Training



tmux

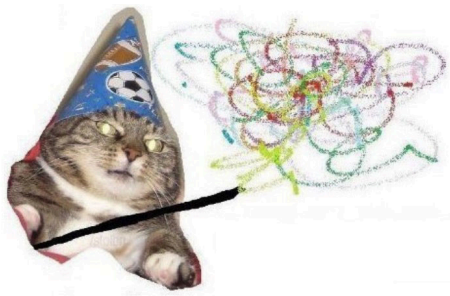

GNU nano New Buffer Modified
Hello.
This text file is my diary.
I have friends in Canada.
^G Ge ^O Wr ^R Re ^Y Pr ^K Cu ^C Cur Po
^X Ex ^J Ju ^W Wh ^V Ne ^U Un ^T To Spe
19:54
19:54 <@Ingifridh> Anyone up for a chat?
[19:54] [@Ingifridh(+i)] [6:du/#sunshine(+nt)]
[#sunshine]
[0] 0:lynx 1:mc- 2:irssi* "" 19:54 09-Nov-11

Write “Industrial Grade” Code

```
from torch import magic
# вжух 🐱 🪄 ✨
magic.solve_problem()
```

write code, get SoTA

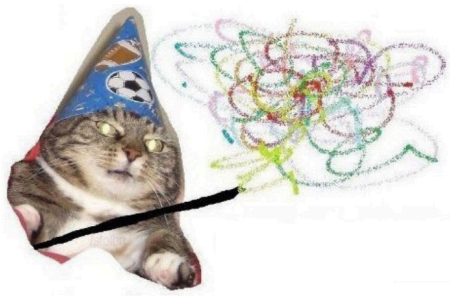
Write “Industrial Grade” Code



```
from torch import magic
# вжух 😺 🪄 ✨
magic.solve_problem()
```

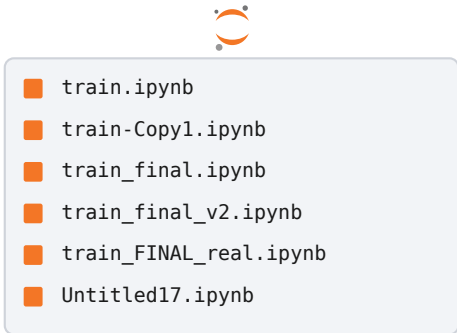
write code, get SoTA

Write “Industrial Grade” Code

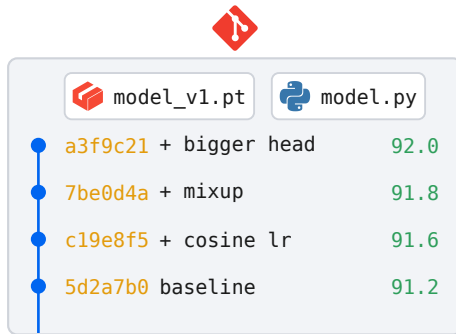




Research Code Versioning



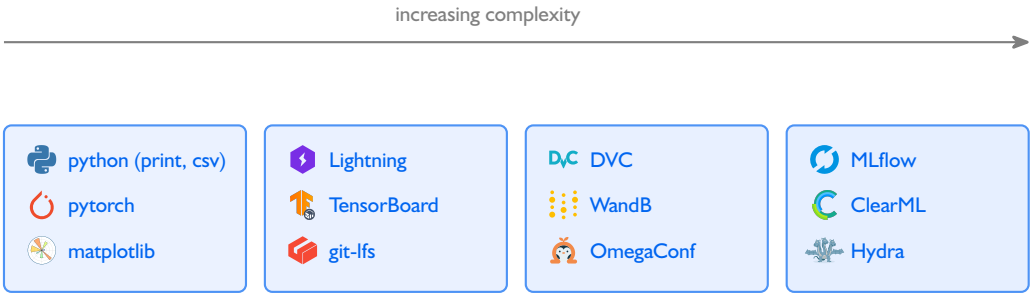
which one gave 92%?



git checkout a3f9c21

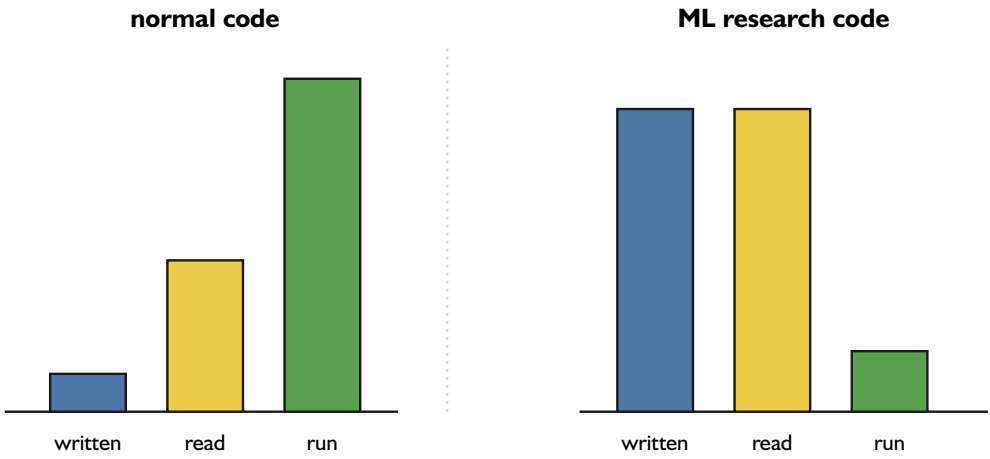
git, git-lfs

Machine Learning Libraries and Frameworks



Try to use whatever is already used in your company.
Only move to the more complex solutions, if you actually need it.

ML Research Code is not Code



Config Parsing is Evil

config.yaml + config_loader.py + …

```
aug:
  - {type: HFlip, p: 0.5}
  - {type: Crop, size: 224}

model:
  source: timm
  arch: resnet18
  ...

# N more settings
```

```
def build_aug_from_config(cfg): ...

def build_model_from_config(cfg): ...

# N! code paths, 0 tests
```

=

train.py + …

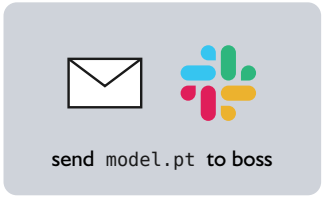
```
aug = A.Compose([
    A.HorizontalFlip(p=0.5),
    A.RandomCrop(224, 224),
]

class MyModel(Sequential):
    self.backbone = timm.create_model(...)
    ...

# only code that is actually used
```

better in 90% of cases

Prepare for Production



Prepare for Production



Prepare for Production

Done! Model attached: model.pt

Great! Where do I click?

Can it run on our cameras?

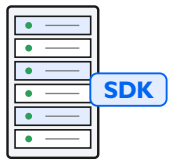
Where is the demo website?

Where does inference run?



Your servers

SaaS



Their servers

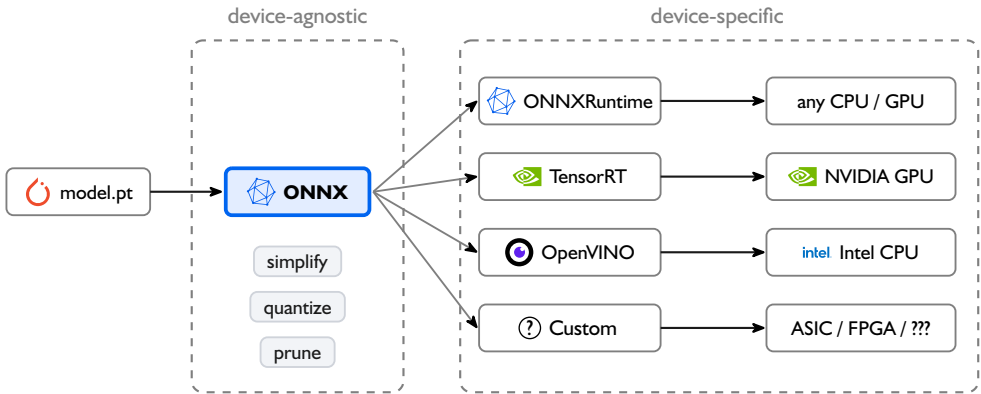
on-premise



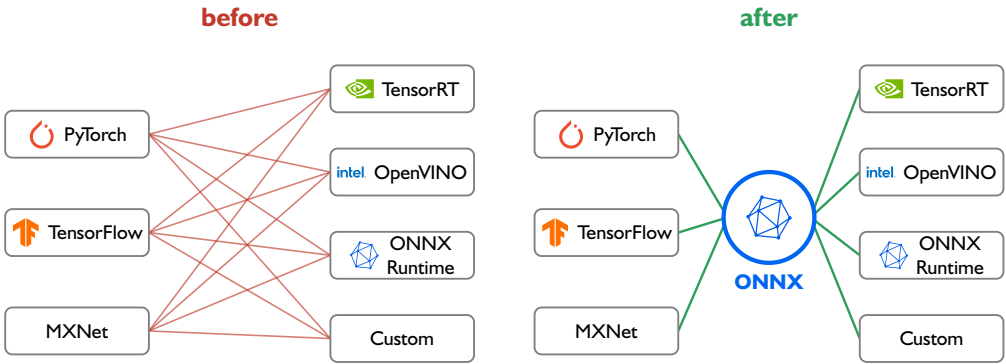
Their devices

on-device

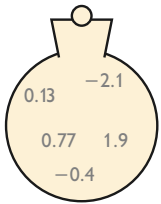
Inference Frameworks



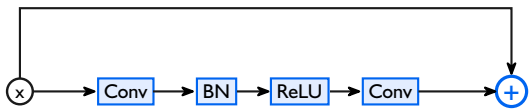
ONNX (Open Neural Network Exchange)



Model Weights vs Full Model Definition



model.pt



model.onnx
(+ model.data)

For real this time?



Cash Money



Cash Money

