

The background features abstract organic shapes in teal and orange. A teal shape is in the top-left corner, and a large orange shape is on the right side, partially overlapping a smaller orange circle.

TRINITY COLLEGE SENIOR SEMINAR TALK

TAMING COMPLEXITY IN SOFTWARE

Real-life Stories and Lessons

HYUNSU CHO, CLASS OF 2015

ABOUT ME

2015

Graduate from
Trinity College
Enter Ph.D. @
**University of
Washington**

2017

Summer internship
@ **AWS**
Start managing
open source
project (XGBoost)

2018

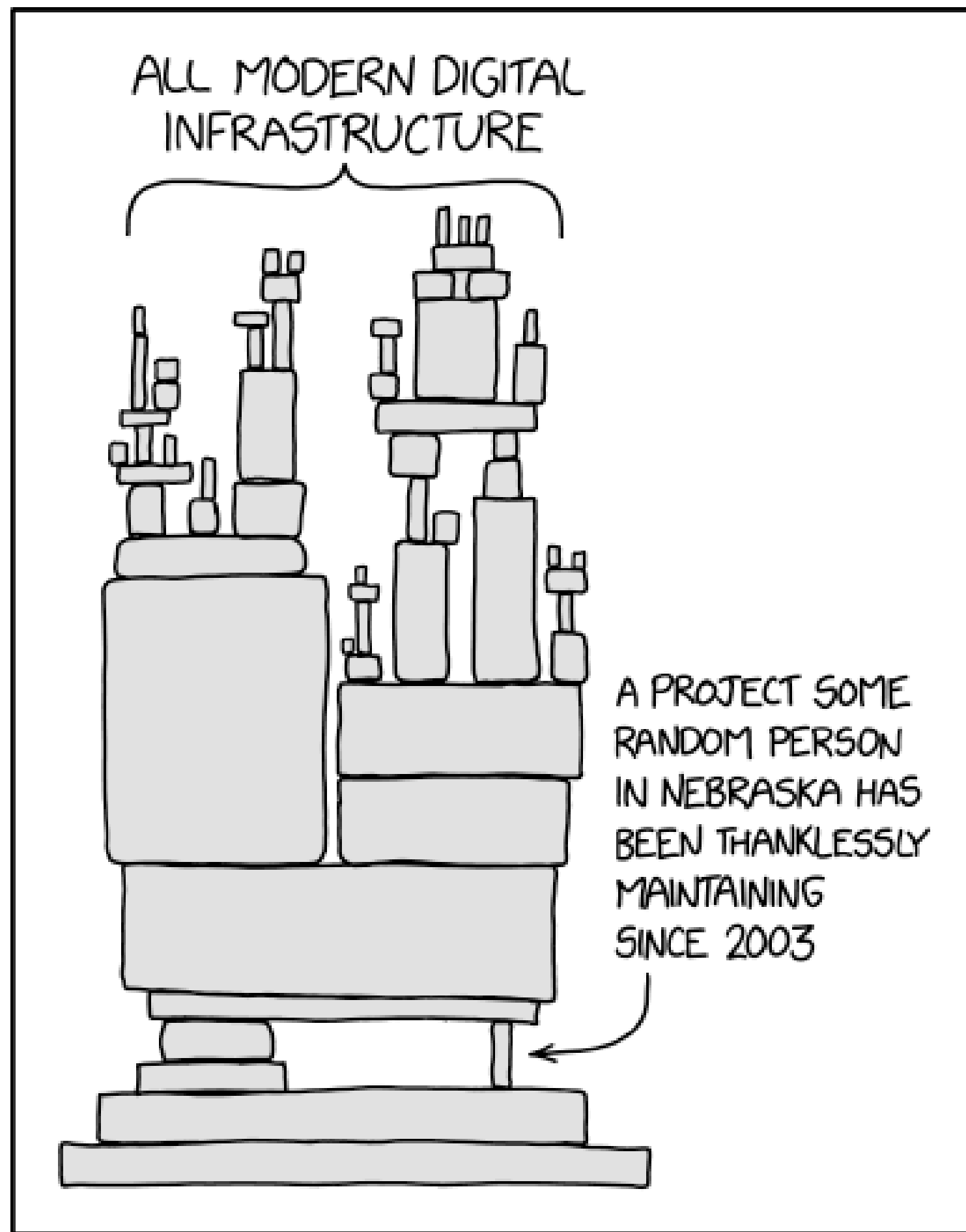
Receive M.S. from
Univ. of Wash.
Join **AWS** as
Applied Scientist



2020

Join **NVIDIA** as
Senior Systems
Software
Engineer





<https://xkcd.com/2347/> "Dependency"

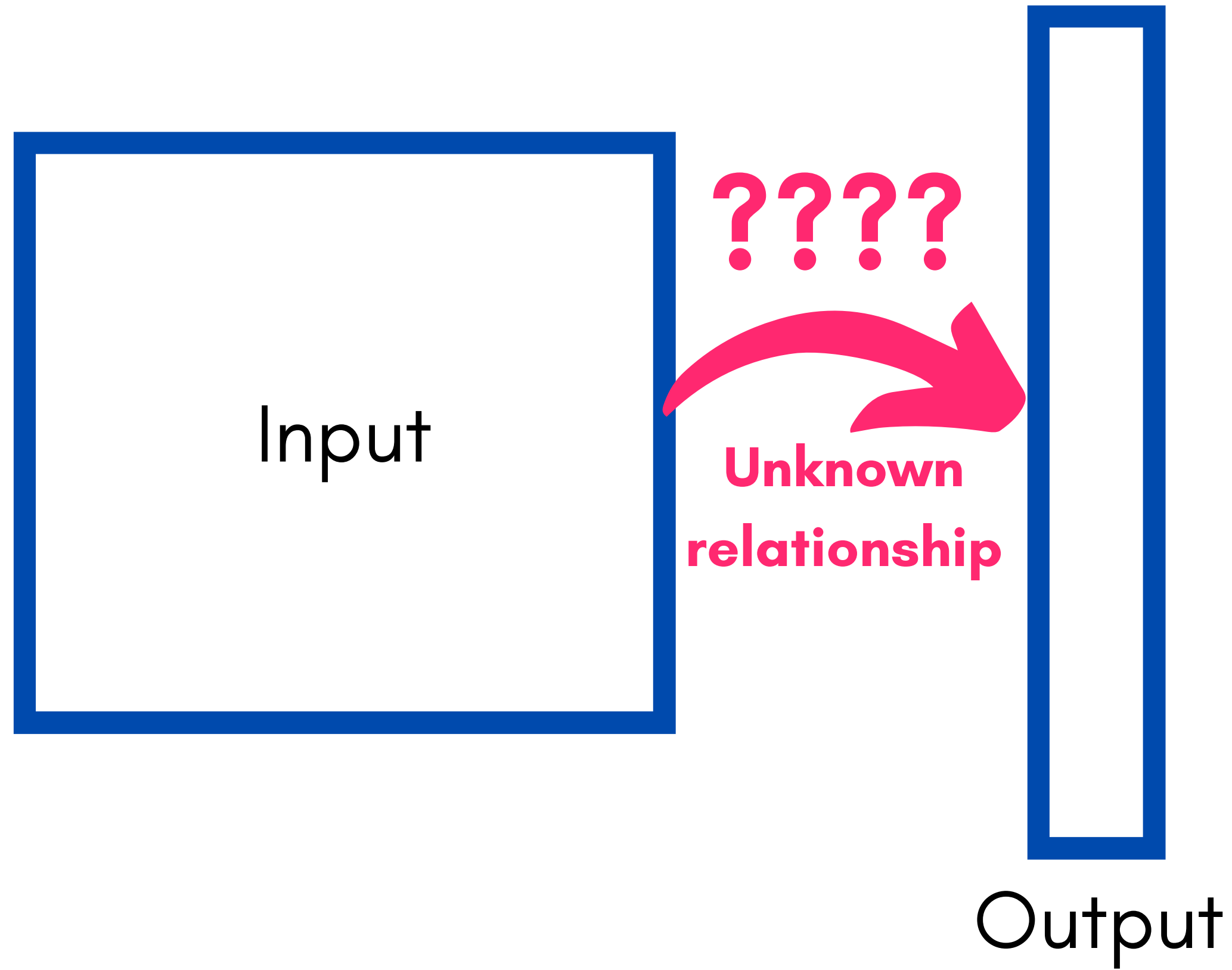
TODAY'S TOPIC



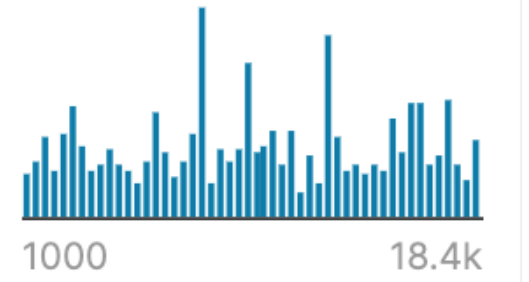
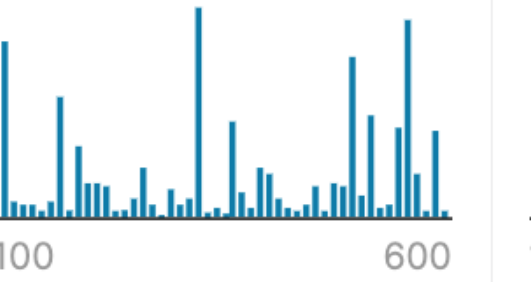
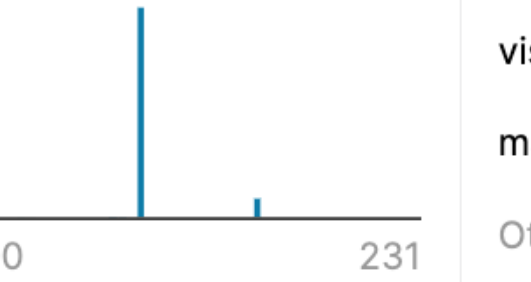
How to deal with software complexity

The background of the slide features a close-up photograph of a person's hand holding a pen, poised to write on a notepad. A small, rectangular sticky note is attached to the notepad, displaying the text 'Ruslan' and 'M - KAZAR' in a handwritten style. The scene is set against a wooden surface. The slide is decorated with large, abstract, organic shapes in orange and red at the top left and a teal shape at the bottom left.

VERY SHORT INTRO TO "AI"

Super hot topic
What I do at my day job

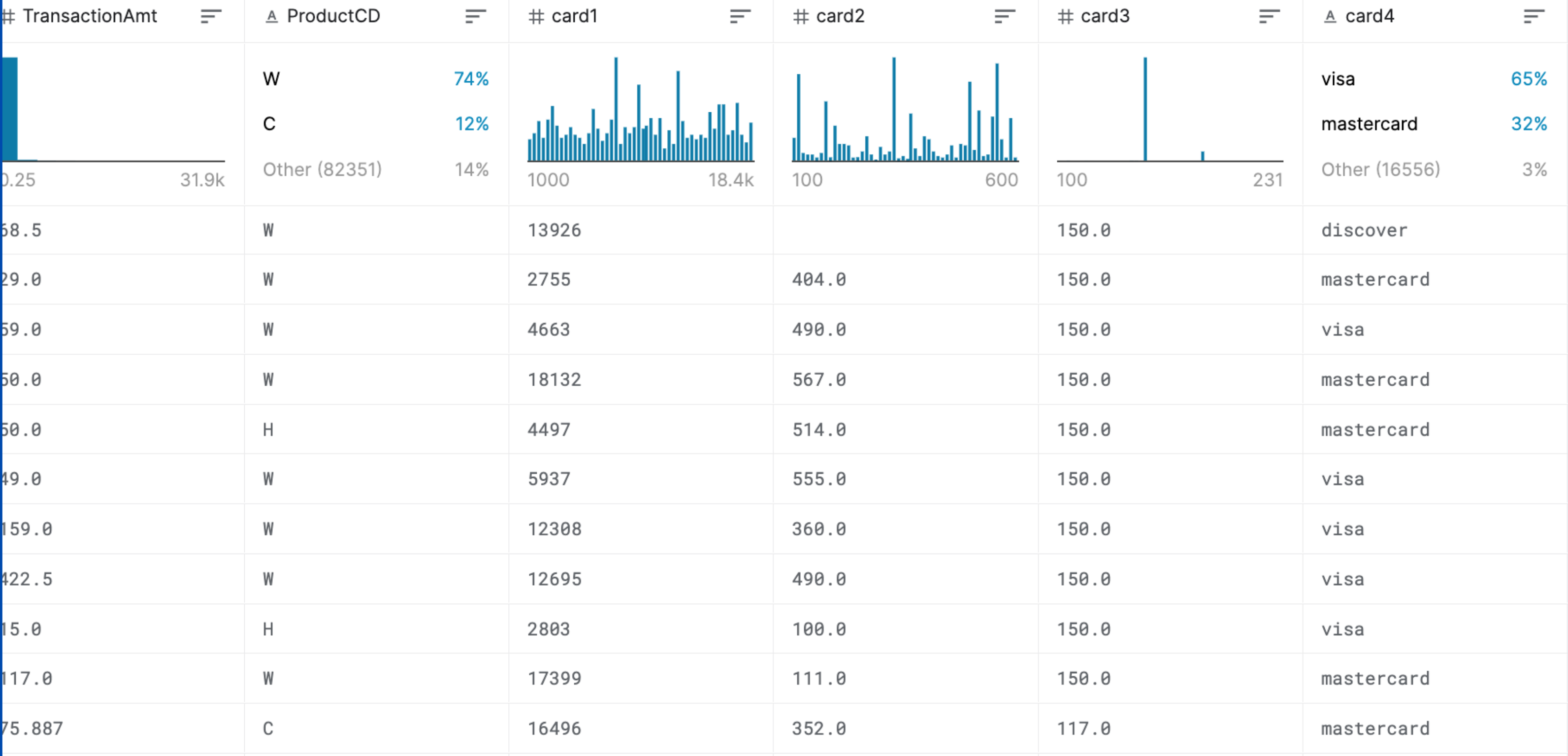
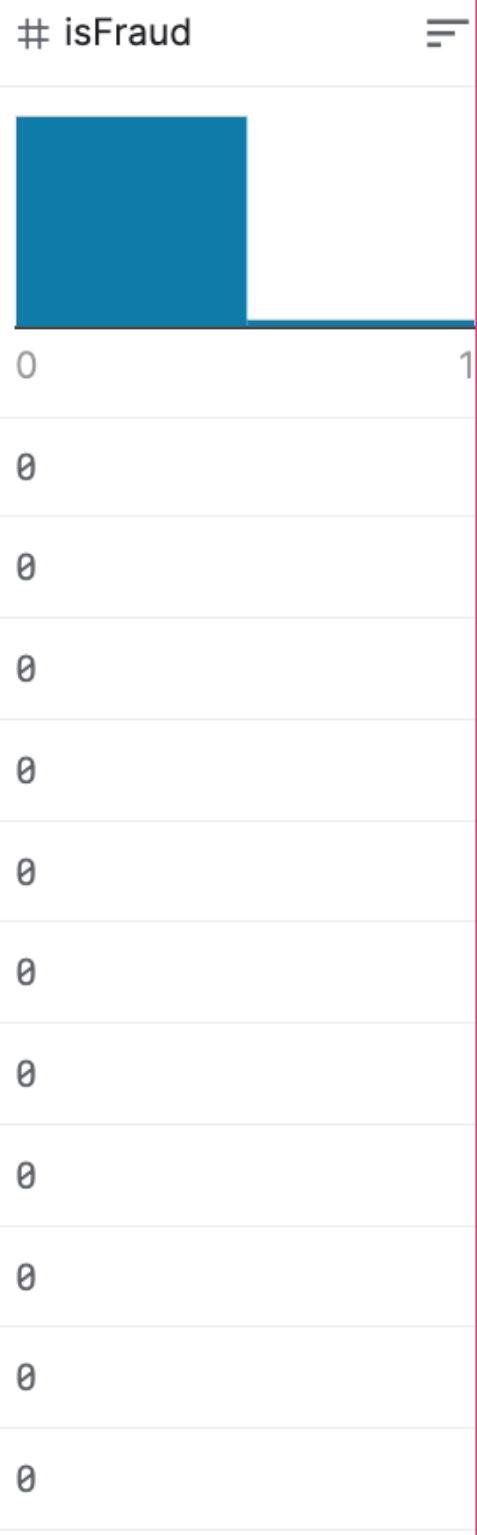


# isFraud	# TransactionAmt	ProductCD	# card1	# card2	# card3	card4
		<div>W74%</div> <div>C12%</div> <div>Other (82351)14%</div>				<div>visa65%</div> <div>mastercard32%</div> <div>Other (16556)3%</div>
0	68.5	W	13926		150.0	discover
0	29.0	W	2755	404.0	150.0	mastercard
0	59.0	W	4663	490.0	150.0	visa
0	50.0	W	18132	567.0	150.0	mastercard
0	50.0	H	4497	514.0	150.0	mastercard
0	49.0	W	5937	555.0	150.0	visa
0	159.0	W	12308	360.0	150.0	visa
0	422.5	W	12695	490.0	150.0	visa
0	15.0	H	2803	100.0	150.0	visa
0	117.0	W	17399	111.0	150.0	mastercard
0	75.887	C	16496	352.0	117.0	mastercard

Example: credit card transaction (some fraudulent)

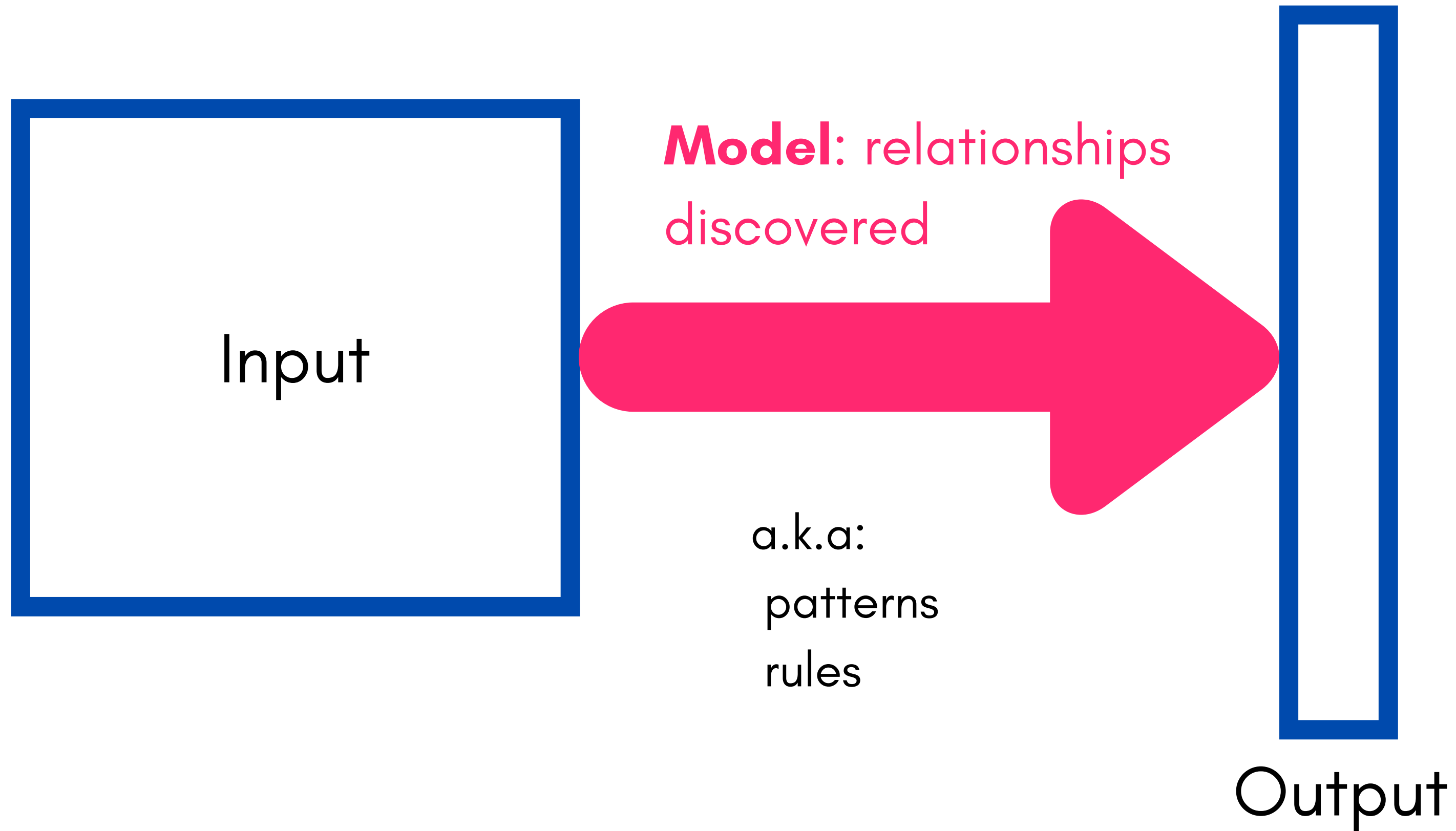
Output: fraud? (Y/N)

Input: attributes of transactions



Example: credit card transaction (some fraudulent)

AI **discovers relationship** between input and output



THE AI WORKFLOW

1 Collect data

Need high-quality data, in great quantity

2 Train a model

Model **discovers all interesting patterns** from the data

3 Test the model

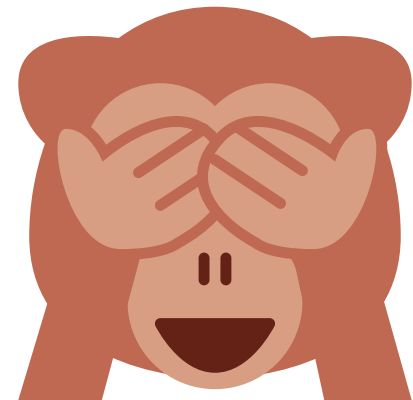
Test the model with similar data and access accuracy. If bad, go back to Step 2.

4 Deploy to production

Search engine, credit card app server, smart-speakers etc.

THE AI WORKFLOW

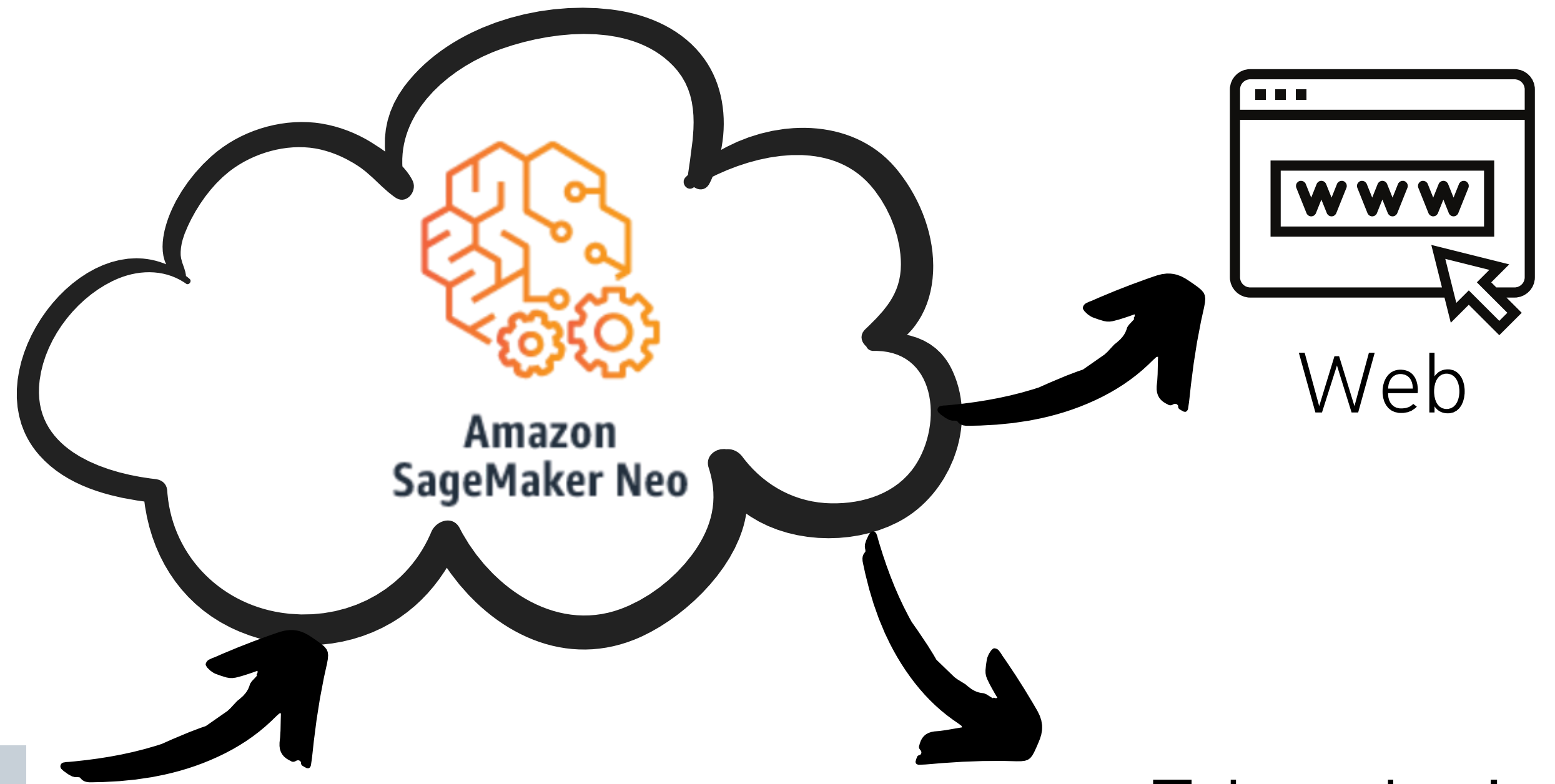
Deployment is HARD!



4

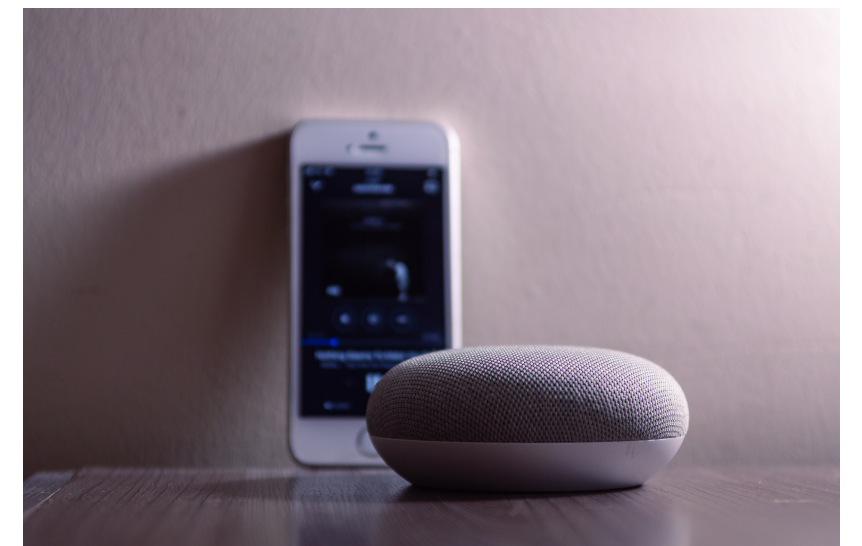
Deploy to production

Search engine, credit card app
server, smart-speakers etc.



Web

Edge device



AI model





TWO LIFE STORIES

My early mistakes, and lessons learned





STORY #1

My internship project goes into production

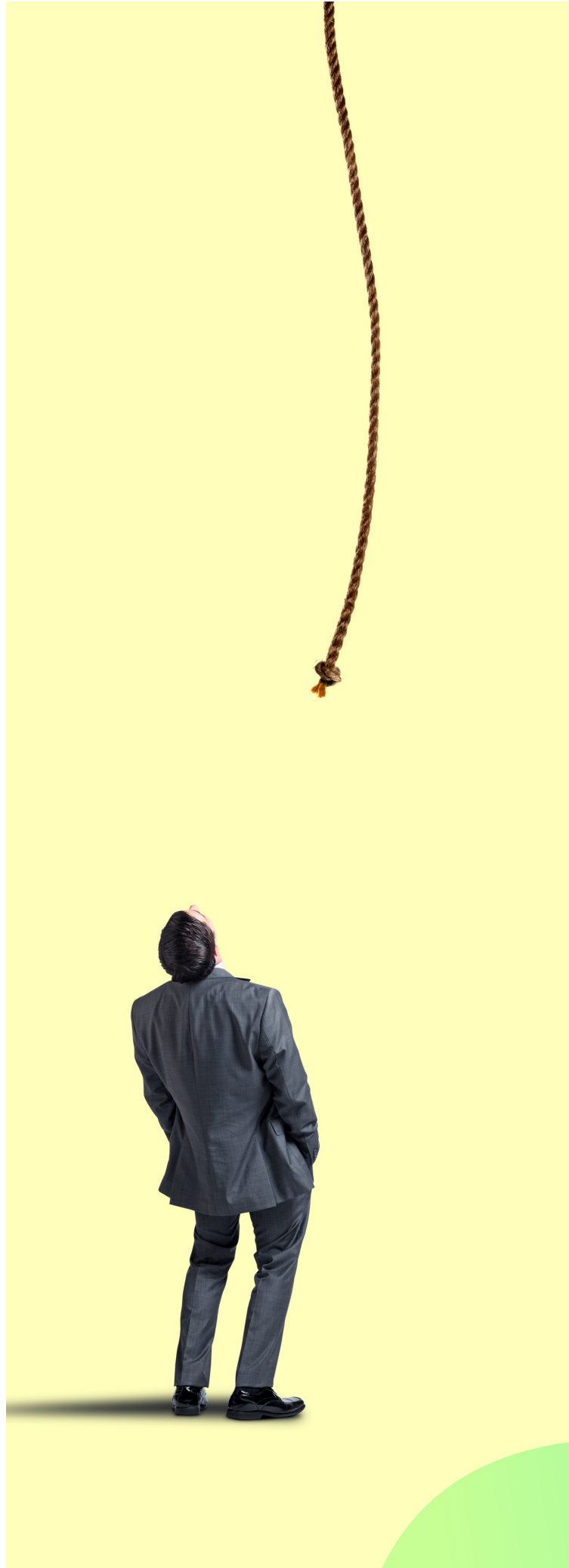
How a "prototype" made it to production and hell broke loose

JAN 2017

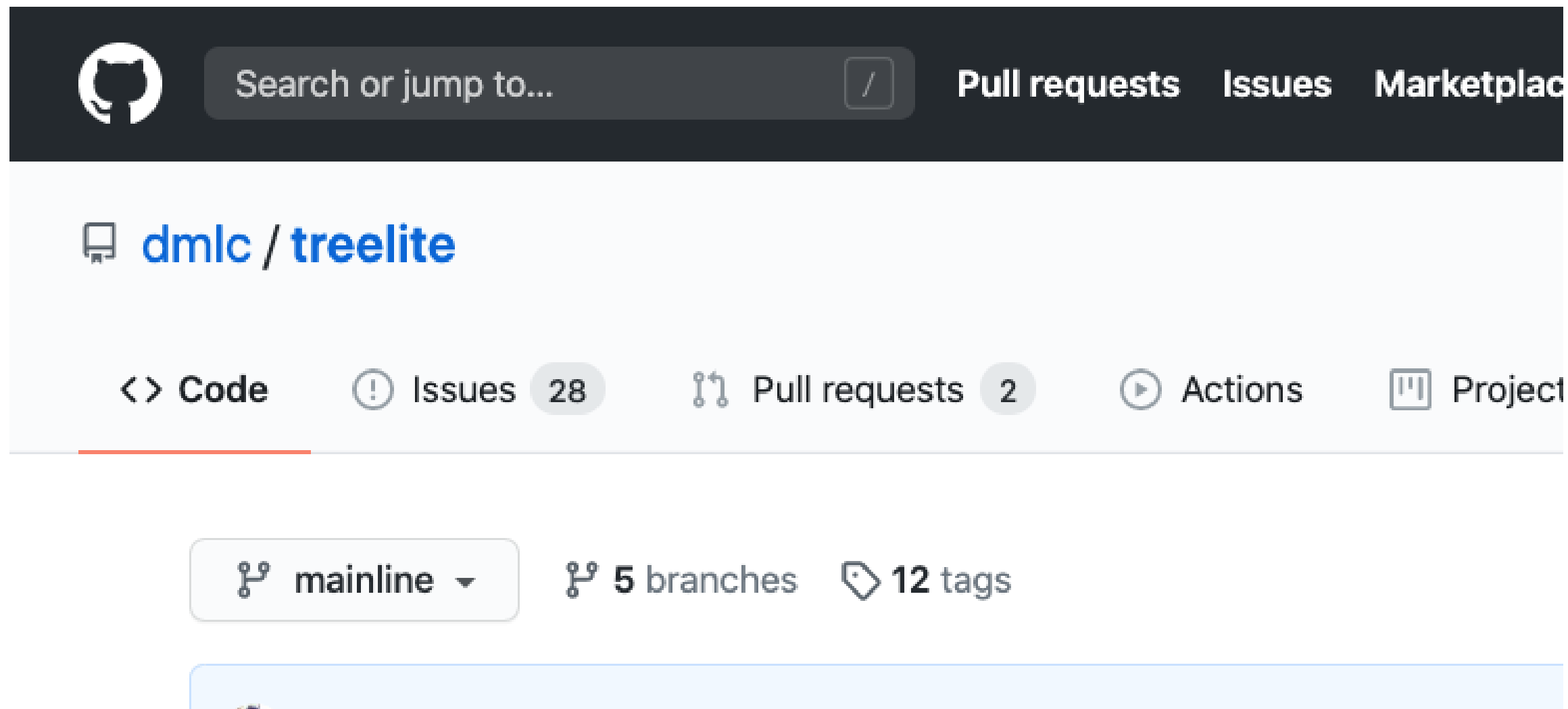


How my Ph.D. ended

JULY 2017



JULY-DEC 2017



Brand new AI tool

30k lines of code from
scratch

Got me a full-time job



Happy Ever After?

...Not so fast!

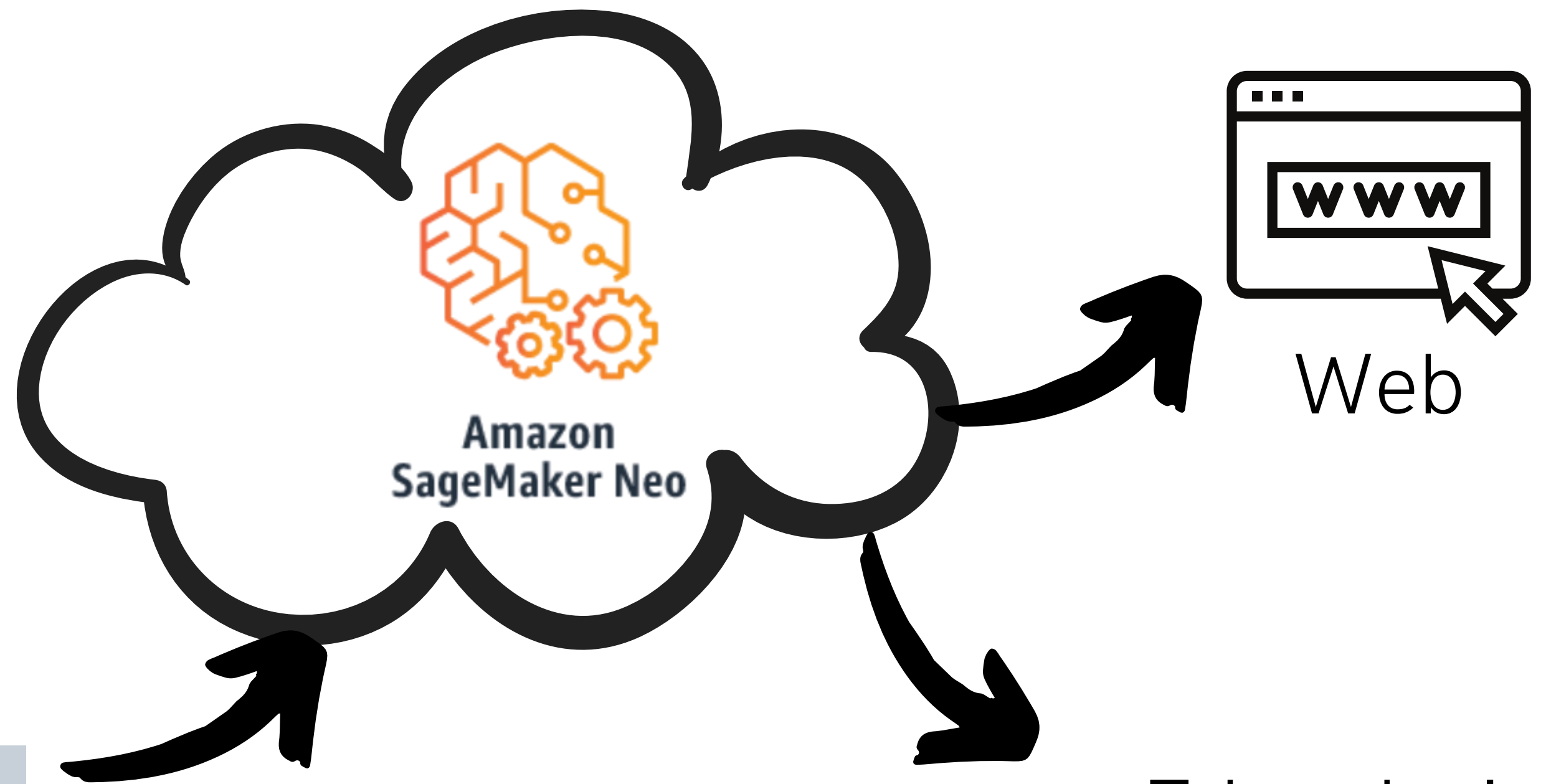
SUMMER 2018 @ AWS: "AI AS SERVICE"

My intern project became part
of a new AWS service!

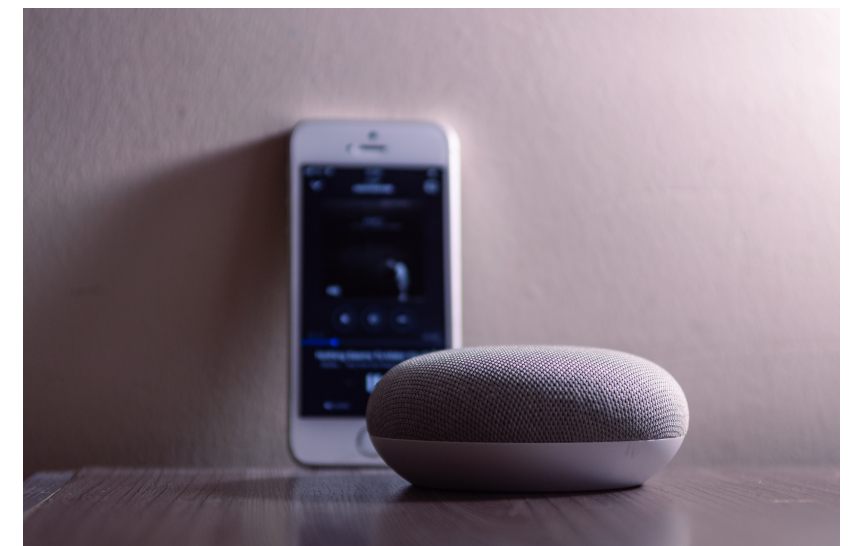
Good: Got promoted

Bad: Saw my prototype crash in
production like a house of cards





Edge device



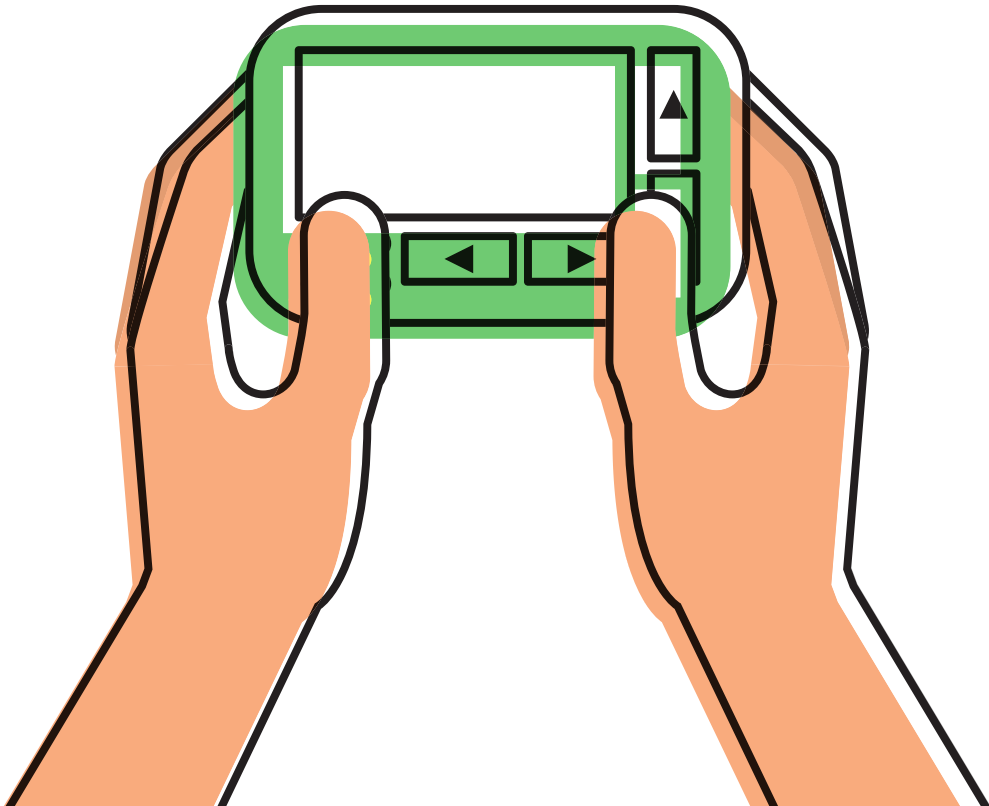
AI model



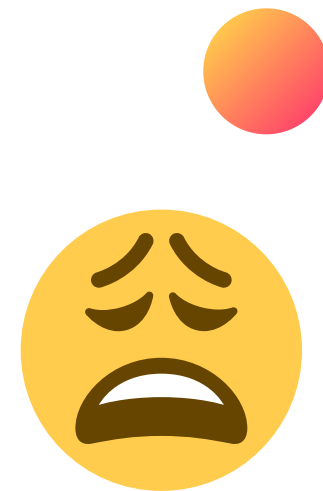
THE ON-CALL EXPERIENCE

Service must be up 24/7

Everyone in the team had
on-call rotation

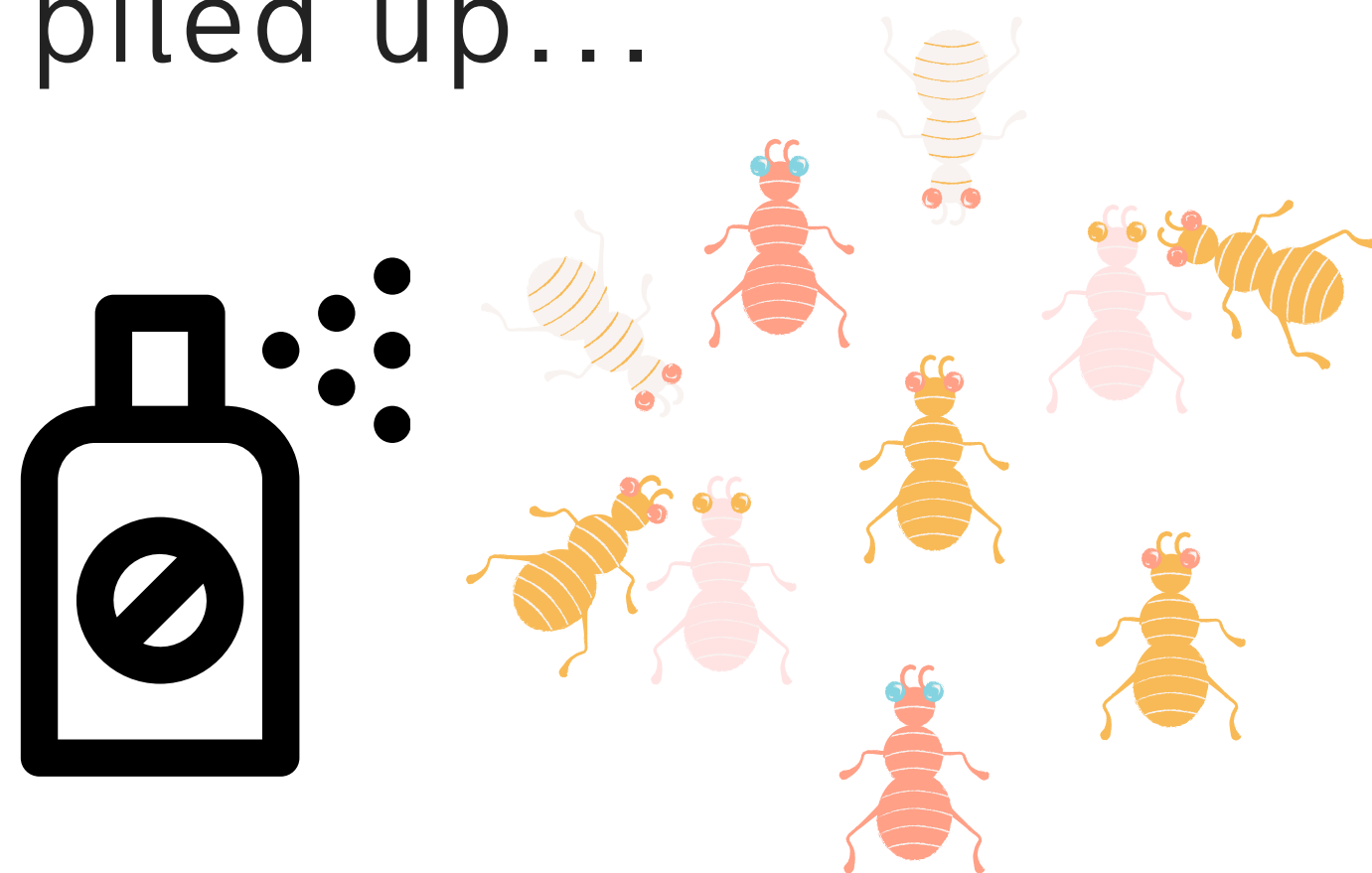


*Pager goes off at 2am
Realize your code
crashed yet another time*



BUGS AFTER BUGS

My code revealed lots of bugs after going into production. Spent several months fixing bugs. Tickets piled up...



BUGS AFTER BUGS

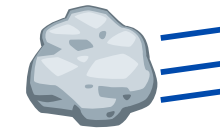
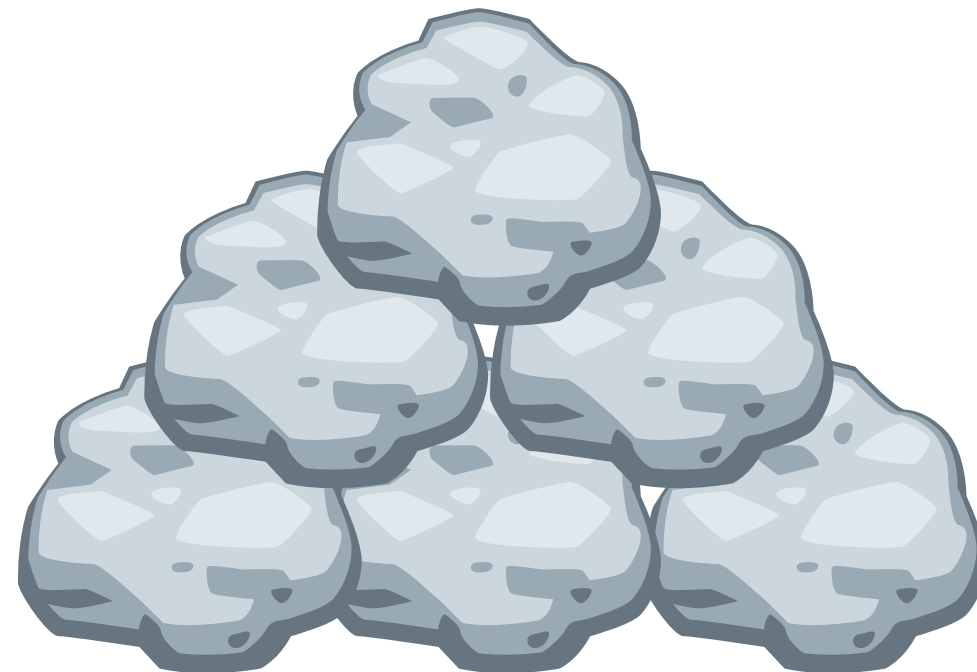
My code revealed lots of bugs
after going into production.
Spent several months fixing
bugs. Tickets piled up...



WHAT WENT WRONG??

I had thrown together bunch of code to get a **prototype** working quickly. Multiple parts were tightly glued together

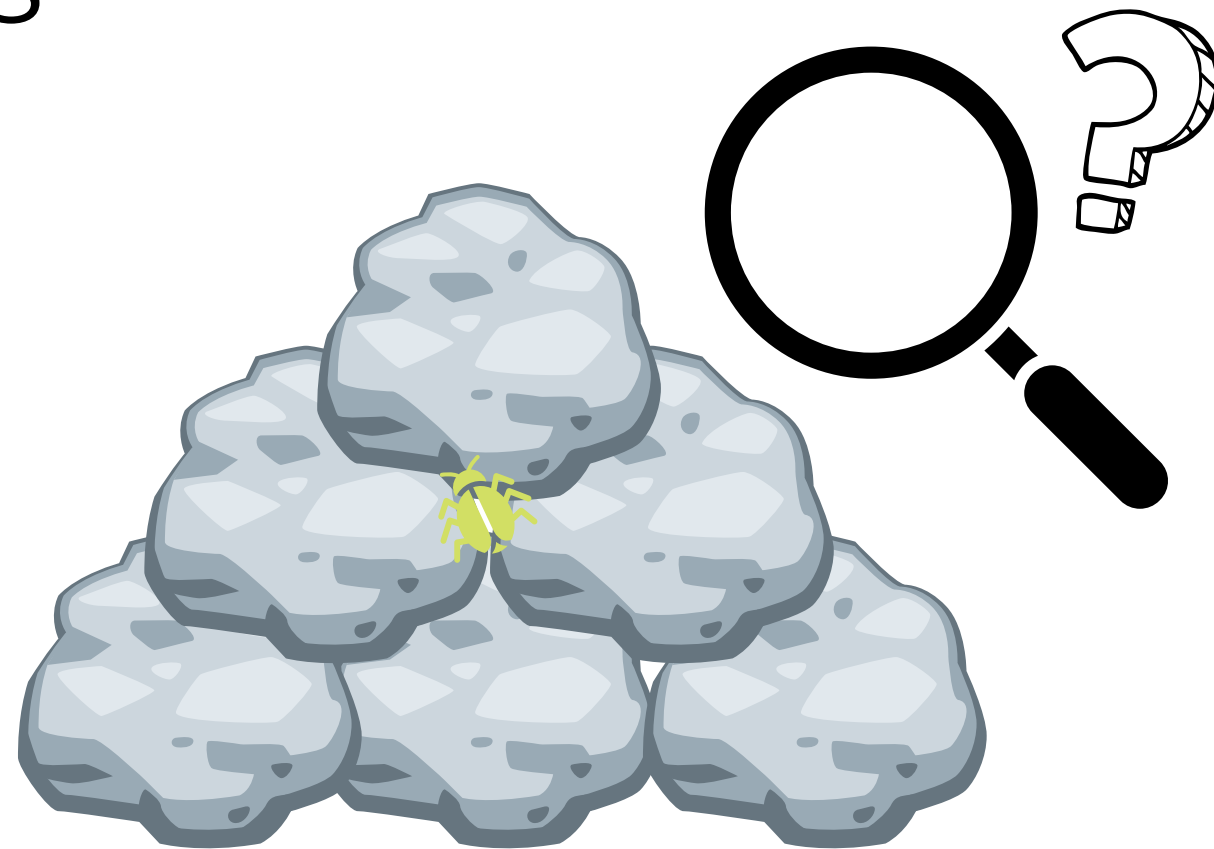
Gonna throw together something quick! Hope it works!



Me (intern)

WHAT WENT WRONG??

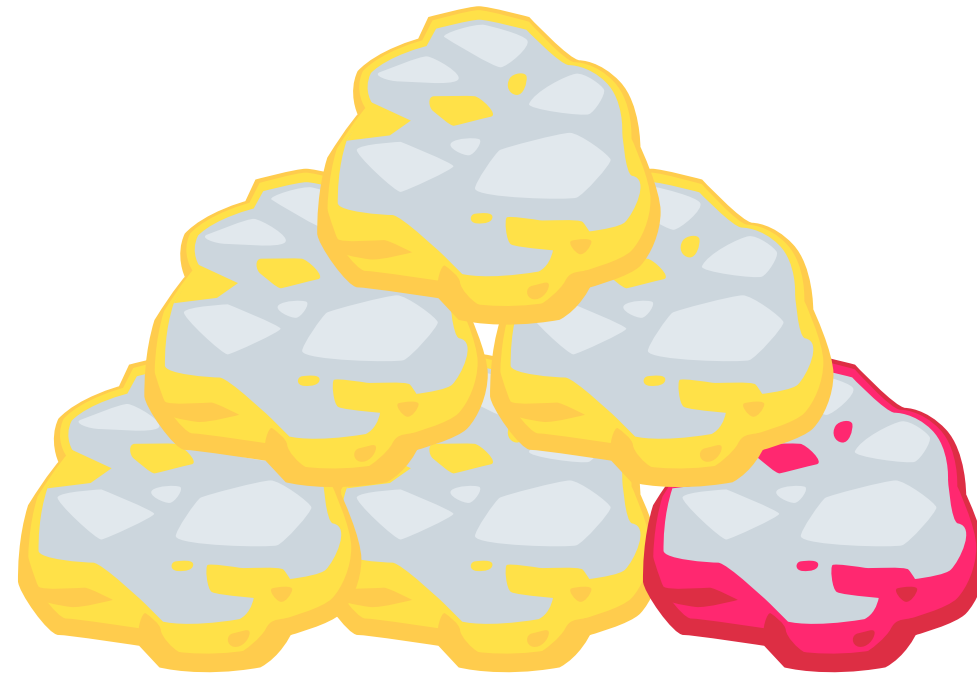
I had thrown together bunch of code to get a **prototype** working quickly. Multiple parts were tightly glued together



Hard to find bug
Hard to know
which part crashed

WHAT WENT WRONG??

I had thrown together bunch of code to get a **prototype** working quickly. Multiple parts were tightly glued together

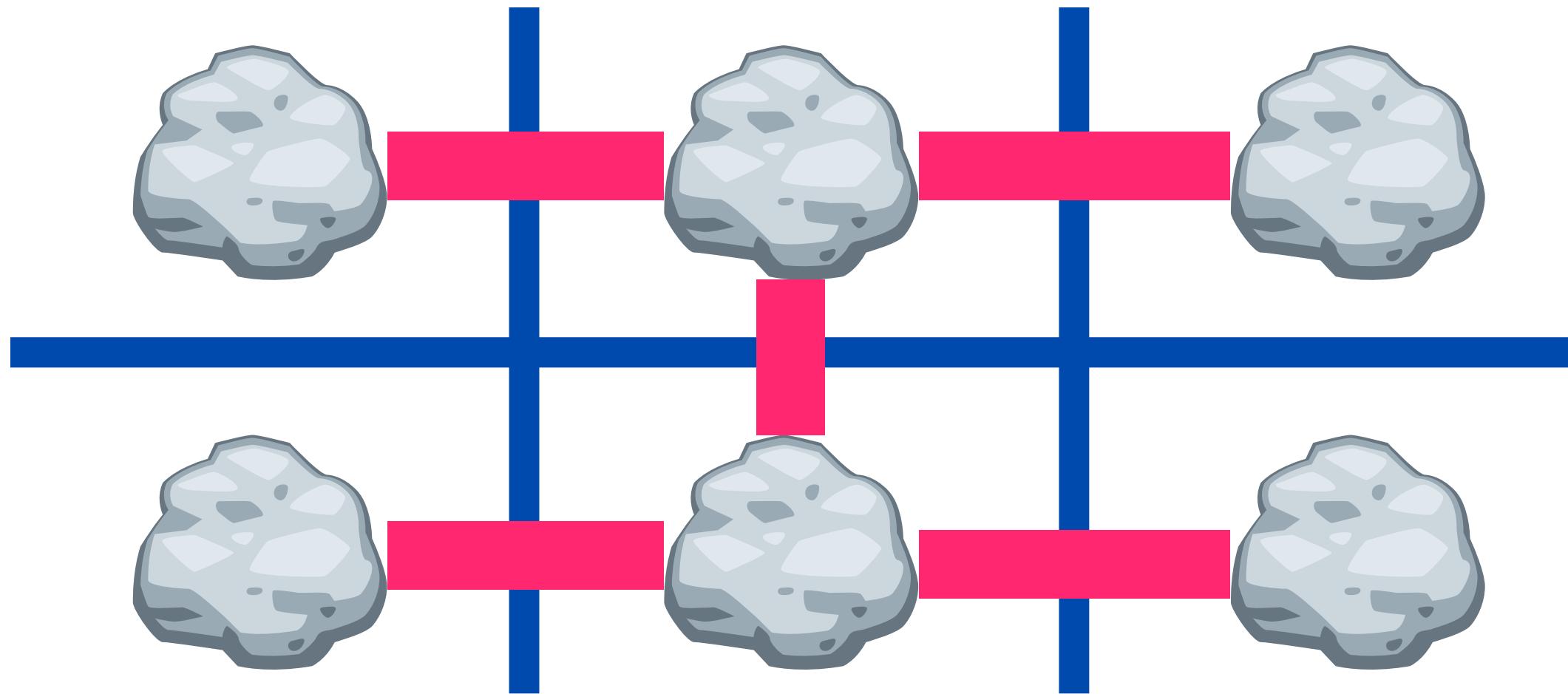


Hard to add a new feature, since you have to change lots of other places ("**big blast radius**")

WHAT WENT WRONG??

If I wrote a **modular, cleanly separated** code,
I would have saved lots of pain later.

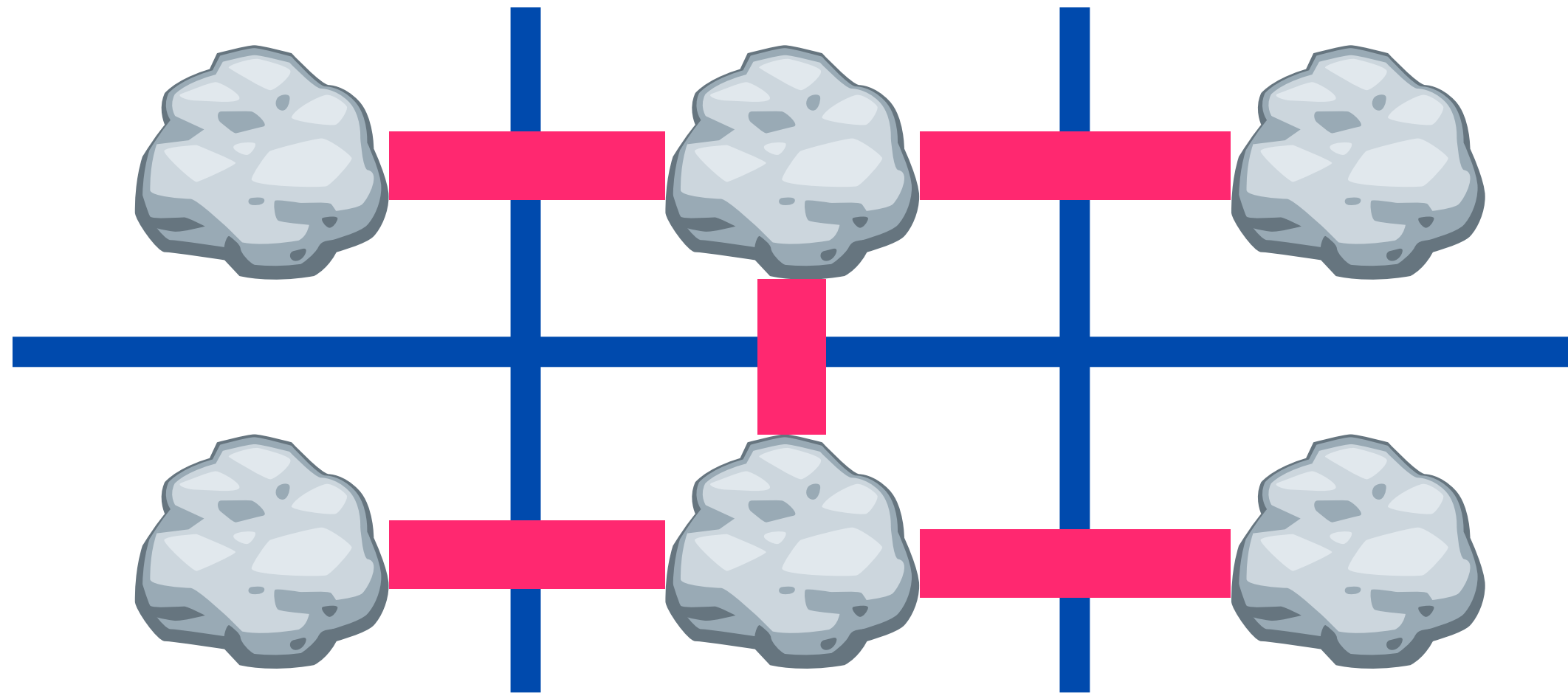
Each module does
only one thing
(separation of
concerns)



WHAT WENT WRONG??

If I wrote a **modular, cleanly separated** code,
I would have saved lots of pain later.

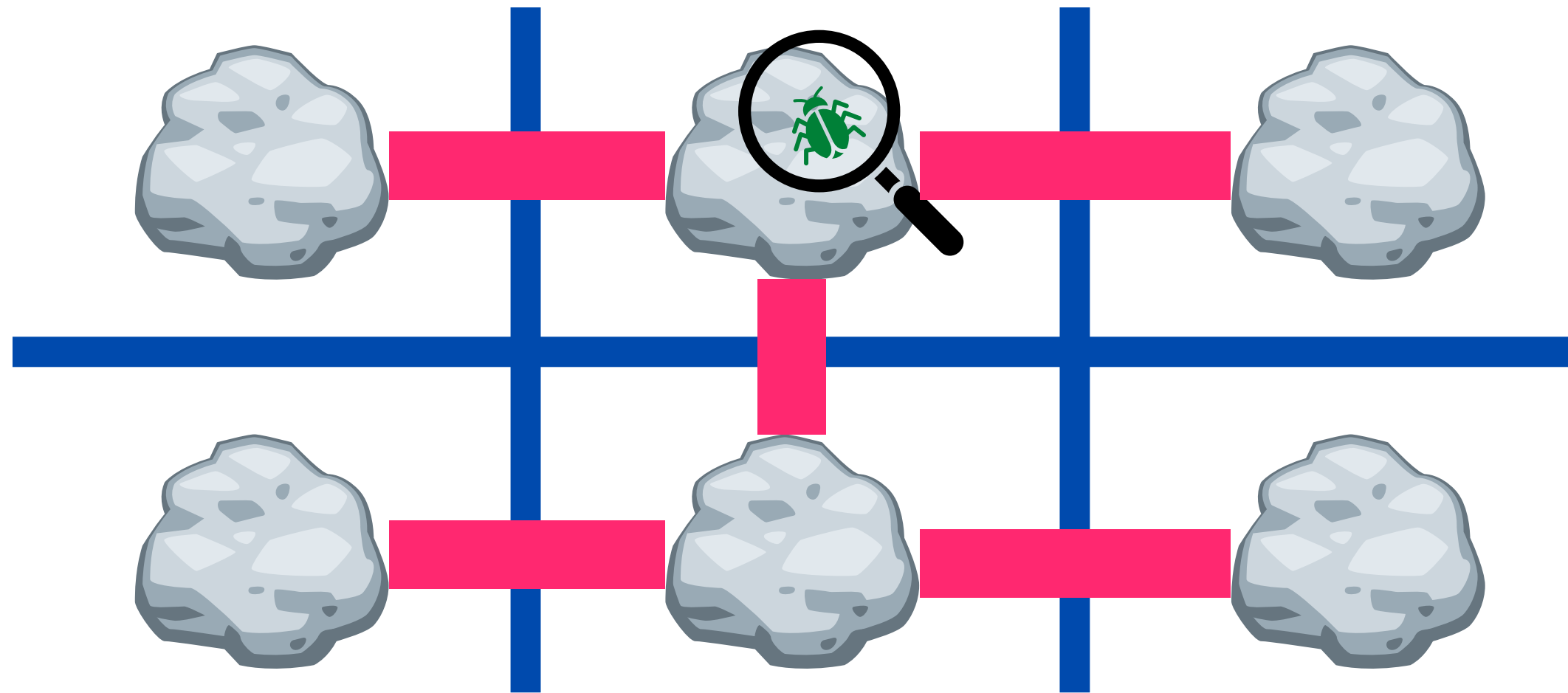
Modules
communicate with
each other only in a
few places
(decoupling)



WHAT WENT WRONG??

If I wrote a **modular, cleanly separated** code,
I would have saved lots of pain later.

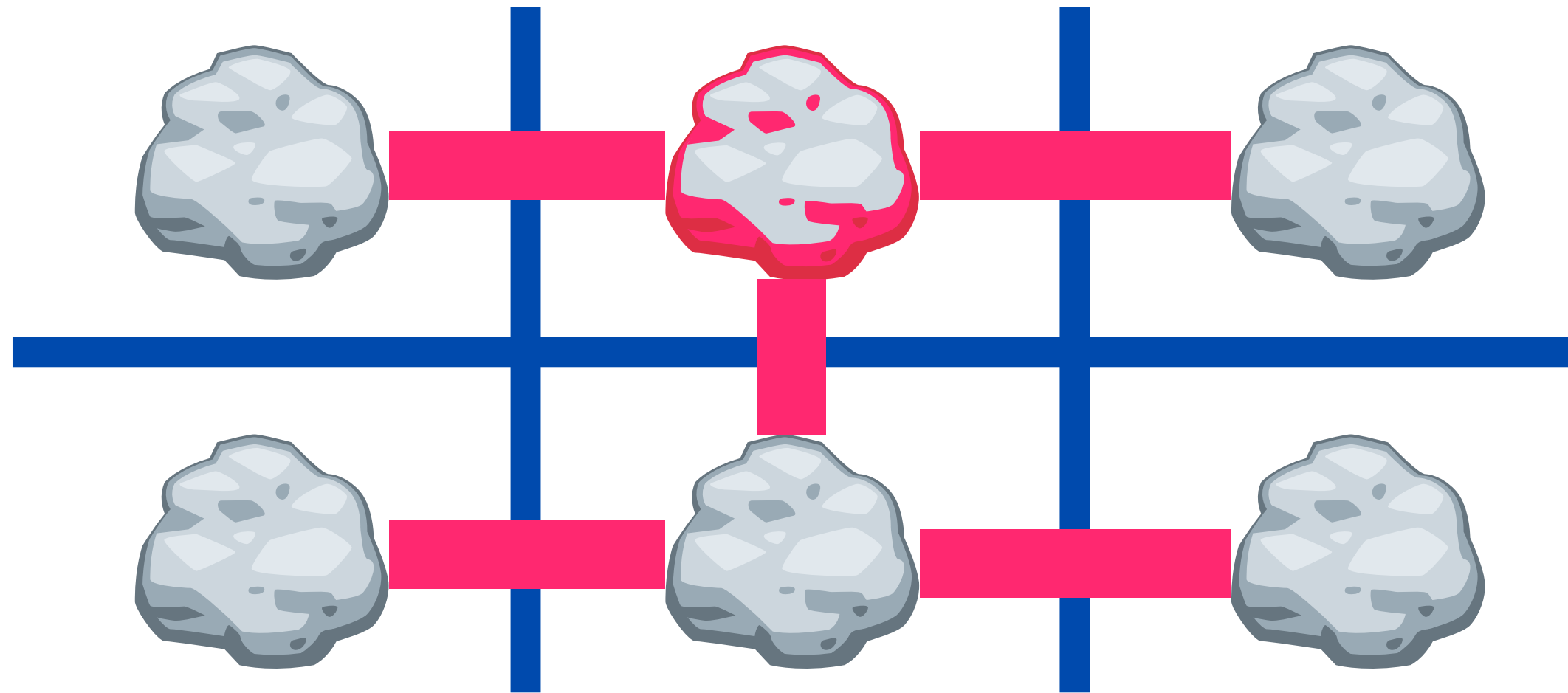
Easy to find which
module crashed
Can **test** each
module one by one



WHAT WENT WRONG??

If I wrote a **modular, cleanly separated** code,
I would have saved lots of pain later.

Easy to **swap out**
one module and
put in another,
while keeping other
modules



WHAT WENT WRONG??

I wrote it 2 yrs ago
but now I have no
idea what it does!



AFTERMATH: MEGA REWRITE (AUGUST 2020)

+4,180 -3,406 ■■■■



Me, this year



STORY #2

The importance of “no”

How I admitted a questionable pull request and ended up hurting the project

THE OPEN SOURCE MOVEMENT

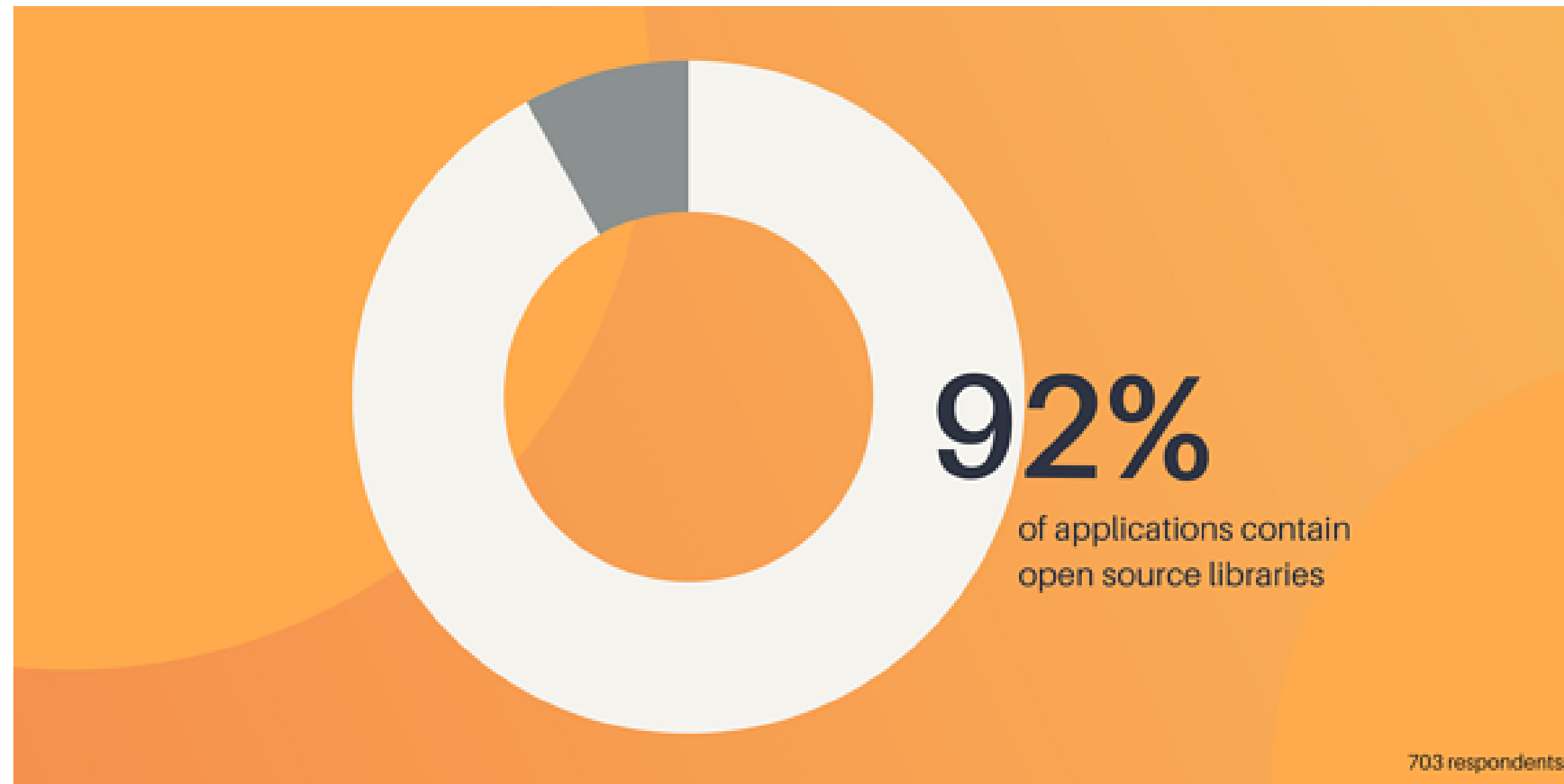
Wikipedia for Code

Code is open; anyone can contribute

Massive community effort

Big impact

Lots of great software developed this way



Tidelift. "Open source is everywhere"

<https://blog.tidelift.com/open-source-is-everywhere-survey-results-part-1>

GitHub

platform for
hosting open
source projects

Search or jump to...

Pull requests

Issues

Marketplace

Explore

dmhc / xgboost

Sponsor

Unwatch

984

Unstar

20k

Fork

7.8k

<> Code

Issues 226

Pull requests 39

Actions

Projects 1

Wiki

Security

Insights

Settings

[CI] Build a Python wheel for aarch64 platform #6253

Merged

hcho3 merged 21 commits into dmhc:master from hcho3:build_aarch64_wheel yesterday

Conversation 10

Commits 21

Checks 24

Files changed 7

+150 -36

Changes from all commits

File filter...

Jump to...

0 / 7 files viewed

Review changes

5 .travis.yml

@@ -15,6 +15,11 @@ jobs:

15- os: linux

16- arch: arm64

17- env: TASK=python_sdist_test

18- os: osx

19- arch: amd64

20- osx_image: xcode10.1

15- os: linux

16- arch: arm64

17- env: TASK=python_sdist_test

18+ - os: linux

19+ - arch: arm64

20+ - env: TASK=python_test

21+ - services:

22+ - docker

23- os: osx

24- arch: amd64

25- osx_image: xcode10.1

37 tests/ci_build/Dockerfile.aarch64

@@ -0,0 +1,37 @@

1+ FROM quay.io/pypa/manylinux2014_aarch64

2+

3+ SHELL ["/bin/bash", "-c"] # Use Bash as shell

4+

5+ # Install all basic requirements

6+ RUN \

7+ yum update -y && \

8+ yum install -y tar unzip wget xz git centos-release-scl-rh yum-utils && \

9+ yum-config-manager --enable centos-scl-rh-testing && \

10+ yum update -y && \

11+ yum install -y devtoolset-7 && \

12+ # Python

13+ wget --no-verbose -O conda.sh https://github.com/conda-forge/miniforge/releases/download/4.8.2-1/Miniforge3-4.8.2-1-Linux-aarch64.sh && \

14+ bash conda.sh -b -p /opt/miniforge-python

15+

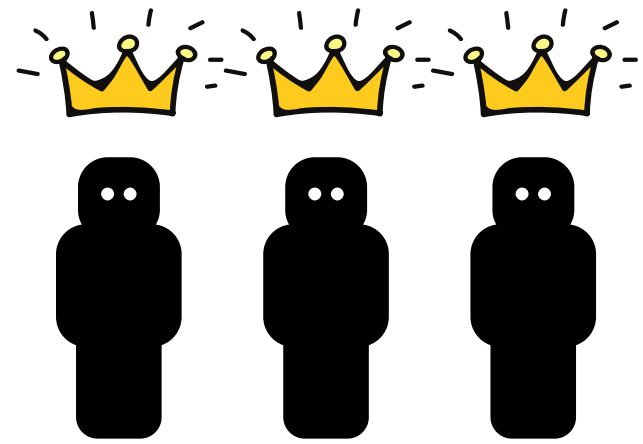
16+ ENV PATH=/opt/miniforge-python/bin:\$PATH

17+ ENV CC=/opt/conda/bin/gcc && CXX=/opt/conda/bin/g++ &&

platform for
hosting open
source projects



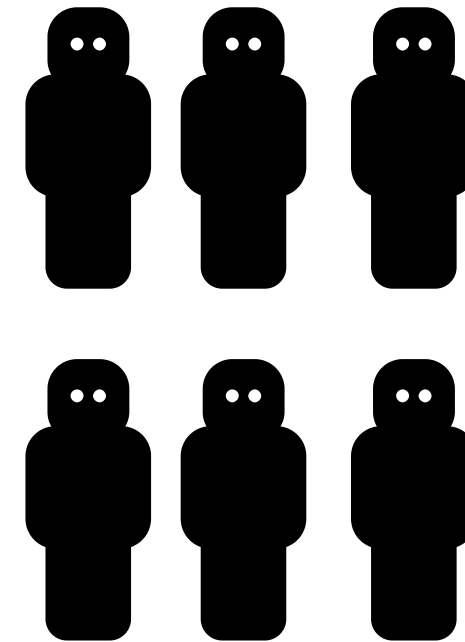
Approve or reject pull requests



Maintainers



Pull request
(proposed changes)



Contributors

Vets contributions for
the long-term benefit
of the project



 **dmlc / xgboost**

 Sponsor

 Watch

984

 Star

20k

 Fork

7.8k



Tianqi Chen

Original creator

Currently Assistant Prof. @
Carnegie Mellon Univ.

Popular AI package

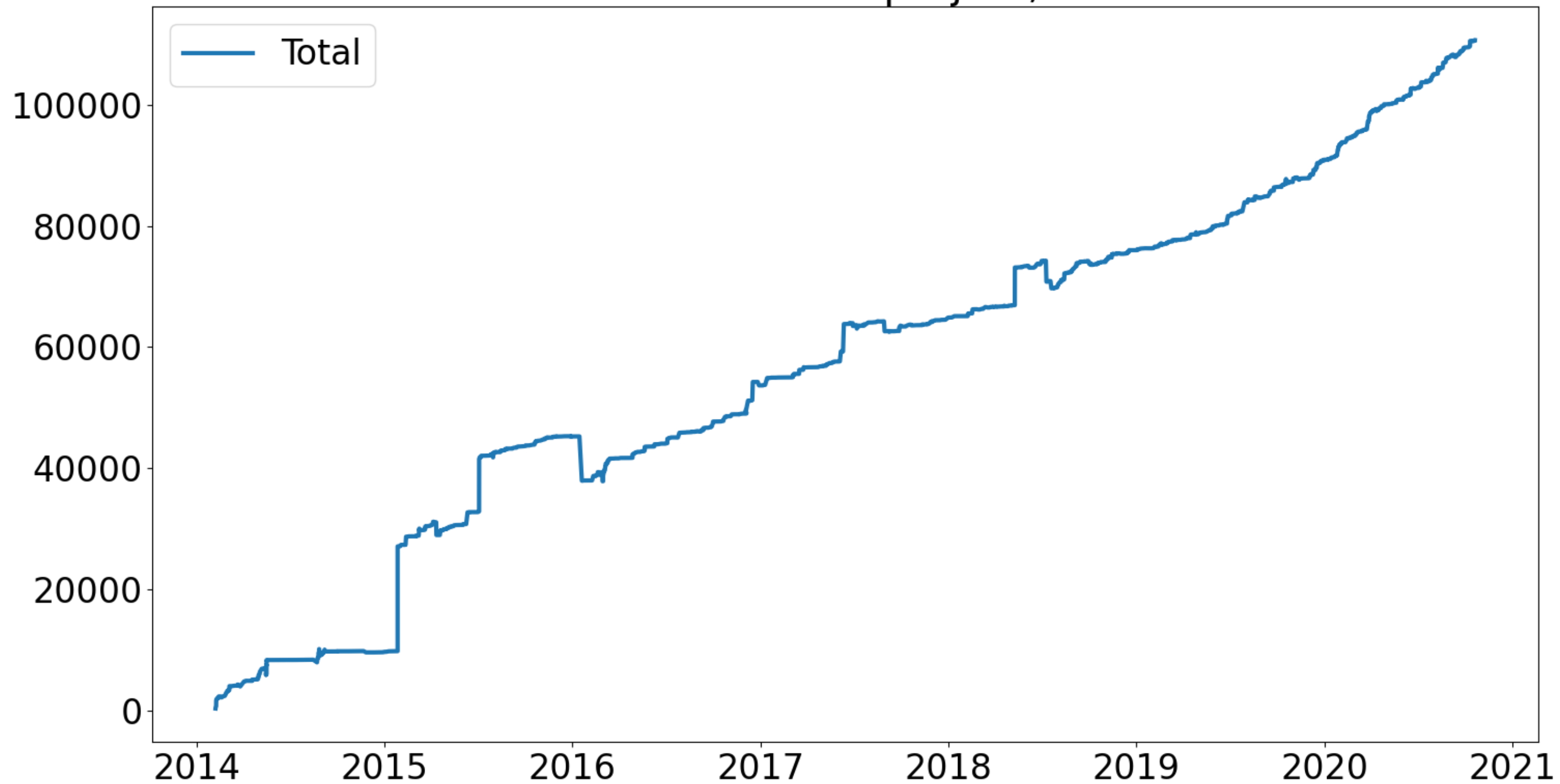
20k+ stars

100k downloads/day

I was appointed as a
maintainer since **Nov 2017**

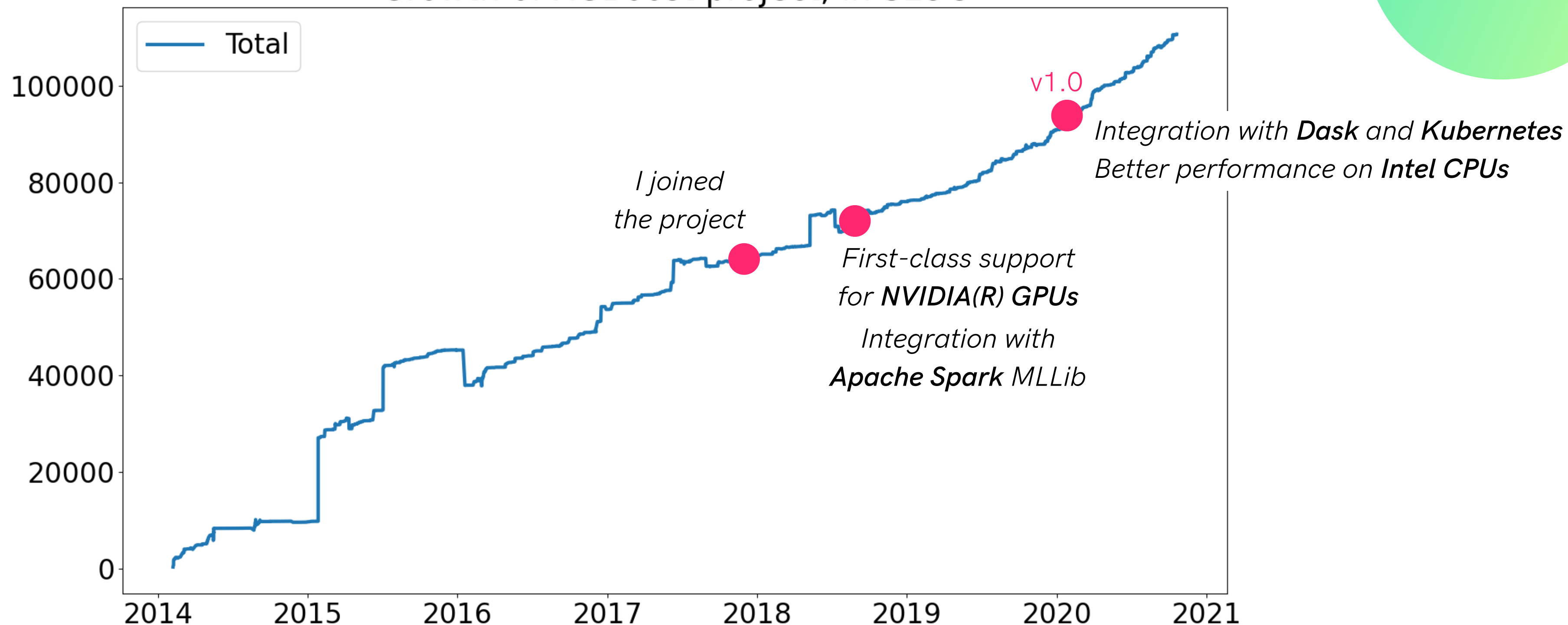
GROWING PAINS

Growth of XGBoost project, in SLOC

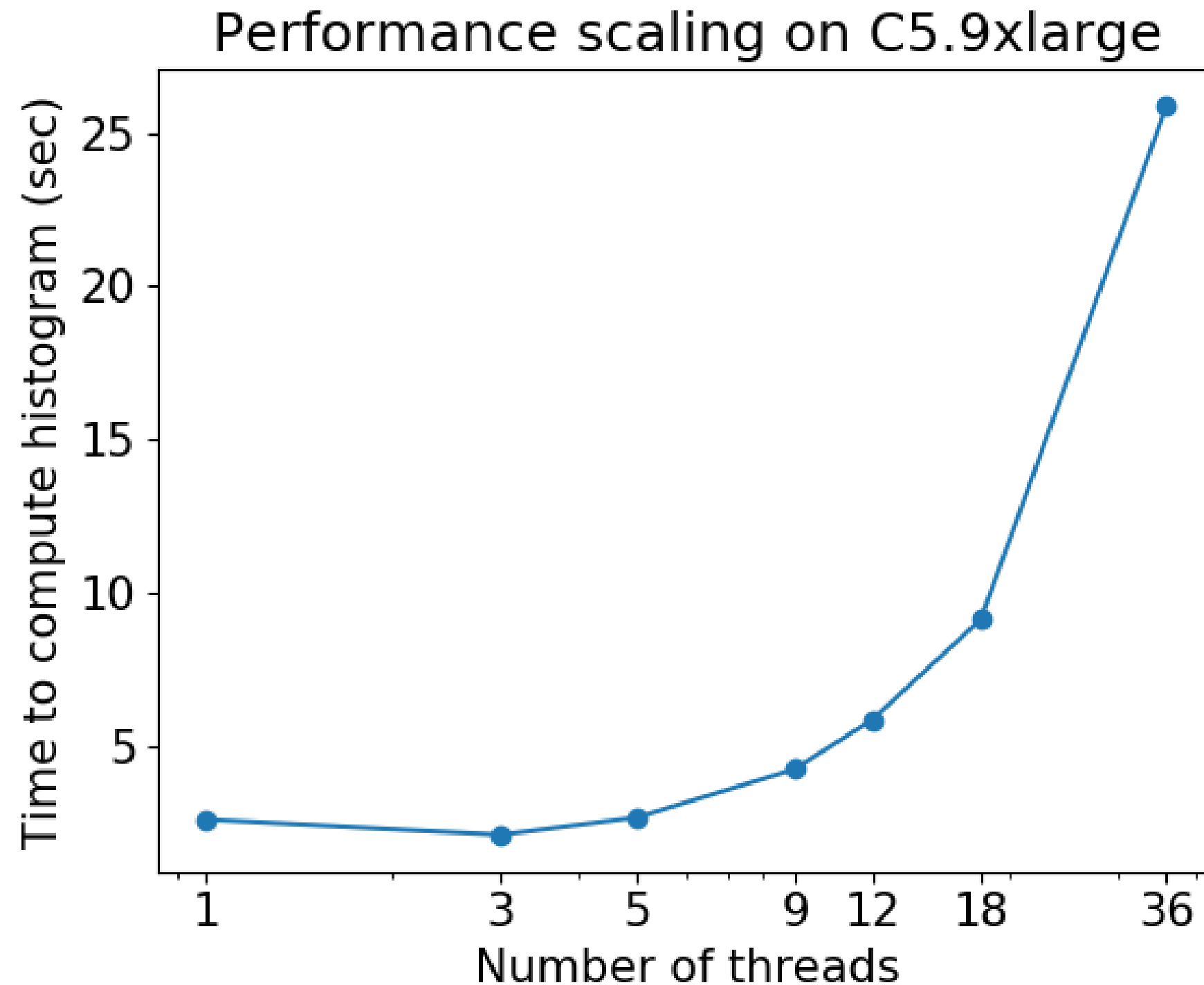


GROWING PAINS

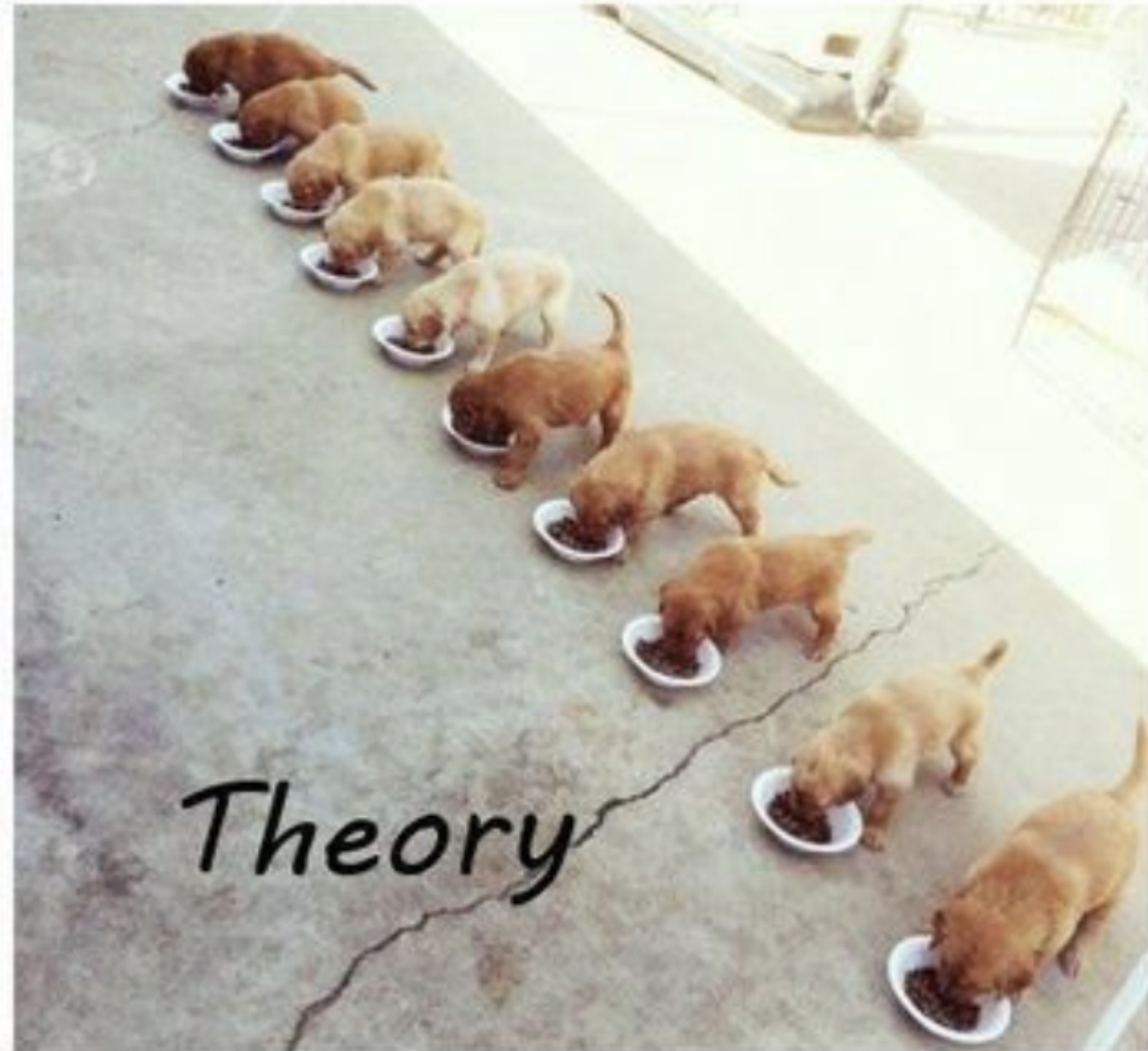
Growth of XGBoost project, in SLOC



OCT. 2018: PERFORMANCE BUG DISCOVERED



Multithreaded programming



JUNE 2019: SOMEONE SENT US A FIX, BUT...



	Abalone	Letters	Mortgage	Higgs1m	Airline	MSRank
BEFORE	10.0	144.6	53.4	123.5	133.2	332.2
AFTER	0.7	10.3	18.3	15.5	55.3	99.1
Difference, times	14.3	14.0	2.9	8.0	2.4	3.4

+1,341 -817 

JUNE 2019: SOMEONE SENT US A FIX, BUT...

It's hard to
understand
their code

BUT they
worked SO
HARD!!

"No" would be too
uncomfortable. I
wanna be nice



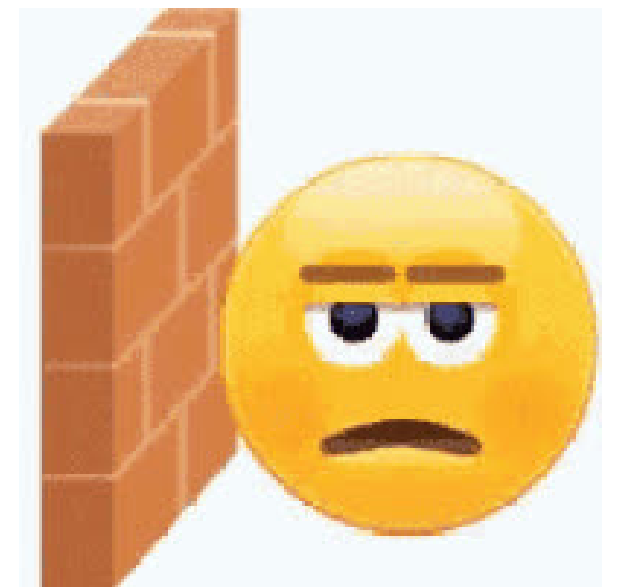


**SEVERAL
MONTHS
LATER...**

NOV 2019: AT IMPASSE

“Those like `CreateTasksForXXX` have to go. I can't allocate time to do a full clean up for CPU Hist. I might prefer maintainability over performance, after banging my head on the monitor, I started to think about reverting the CPU optimization. WDYT?”

<https://github.com/dmlc/xgboost/pull/5005>



NOV 2019: AT IMPASSE

“Let’s revert for now and we’ll try to incorporate some performance improvements in a later date. The state of the current 'hist' code is such that most developers hardly understand it at all and thus slowing down further progress.”

<https://github.com/dmlc/xgboost/pull/5005>



NOV 2019: AT IMPASSE

“After [the pull request], the 'hist' updater became quite unwieldy and currently few developers understand it. In particular, CreateTasksForXXX functions have created multiple layers of indirection, making it unclear which objects are being computed/modified.”

<https://github.com/dmlc/xgboost/pull/5008>

AFTERMATH

Slight awkward moment when the contributor found out we reverted their contribution

Fortunately they understood our decision and even submitted a replacement pull request that was much more legible.

As a result, we were able to start making further improvements



LESSON LEARNED

Code legibility and clarity are just as important as performance.

Open source projects should consider long-term impact of the new proposed changes

Sometimes you have to say "no"

STORY SO FAR

Software is awesome
We ask software lots of things
Complexity will come and bite you
BUT... there is a way out



GOOD HABITS GO A LONG WAY

I will teach you good habits you
can start practicing today

Some of the habits will **help your
senior project!**

Avoid pitfalls and mistakes that'd derail your project
Complete your project with high-quality code and a
peace of mind

#-1. START EARLY

Do not procrastinate!

Earlier you start, you can rescue yourself more easily out of mistakes and troubles.

Rushing for deadlines will make you sloppy

#0. EXPECT TROUBLE

Fact: all software have **bugs**. So does your code.

Prepare for **possible delays**.

“Schedule padding” / “Underpromise, Overdeliver”

#1. START SMALL, ADD SMALL

Adding lots of code in a short time period is a recipe for disaster

Add a little bit of code (a function or a class) at a time.

Then:

- Test the function/class by itself (later discussed)
- Let it rest a bit and review (*)
- Save it to a version control system (next slide)
- Repeat

(*) It's best to do code review in a pair or a group. Multiple eyes can uncover lots of issues

#2. VERSION CONTROL

Use a version control system, e.g. **Git**. Also back up your code to **GitHub**

Git helps you in multiple ways:

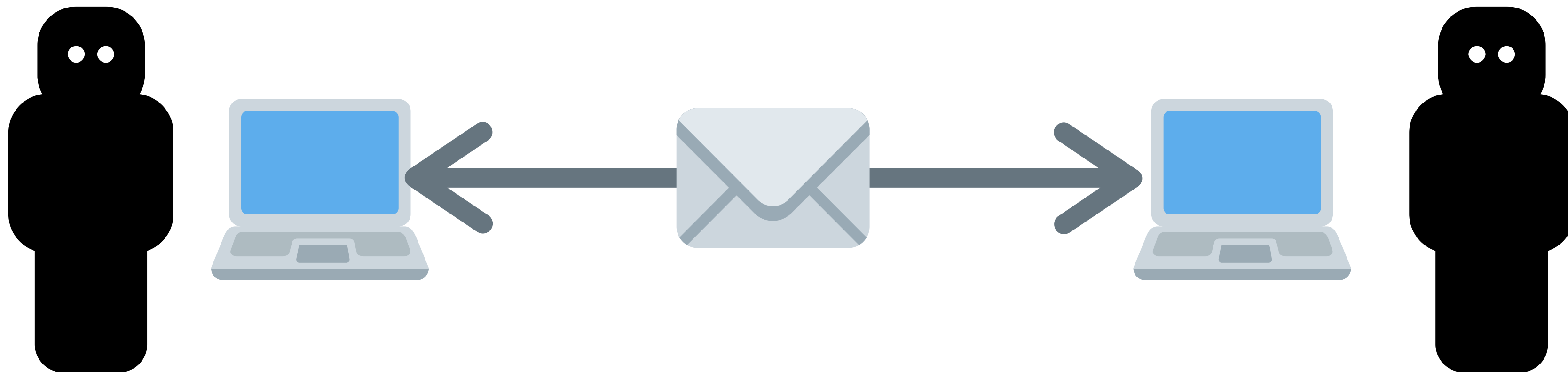
- Collaboration
- Documentation of changes over time / work log
- Ability to roll back changes

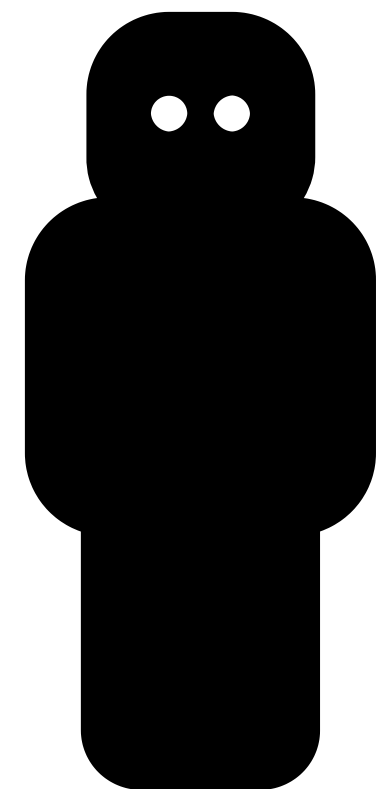
Collaboration



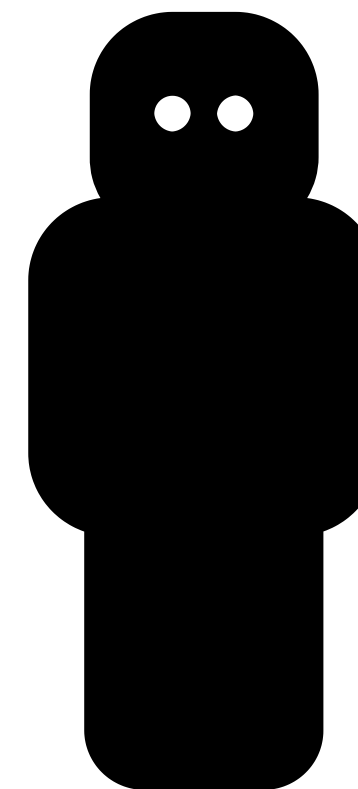
Meeting in a lab may not always be desirable

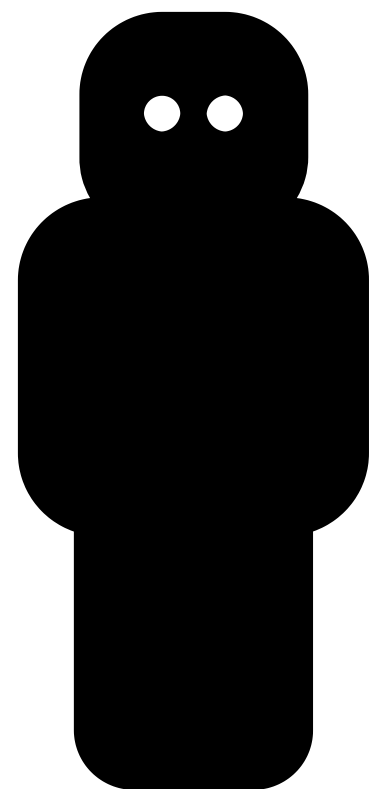
- Lab might be full
- Hard to find time that work for 2+ people
- Social distancing requirement



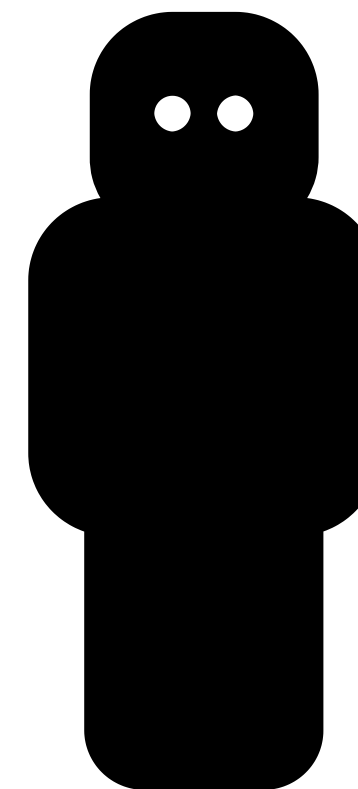


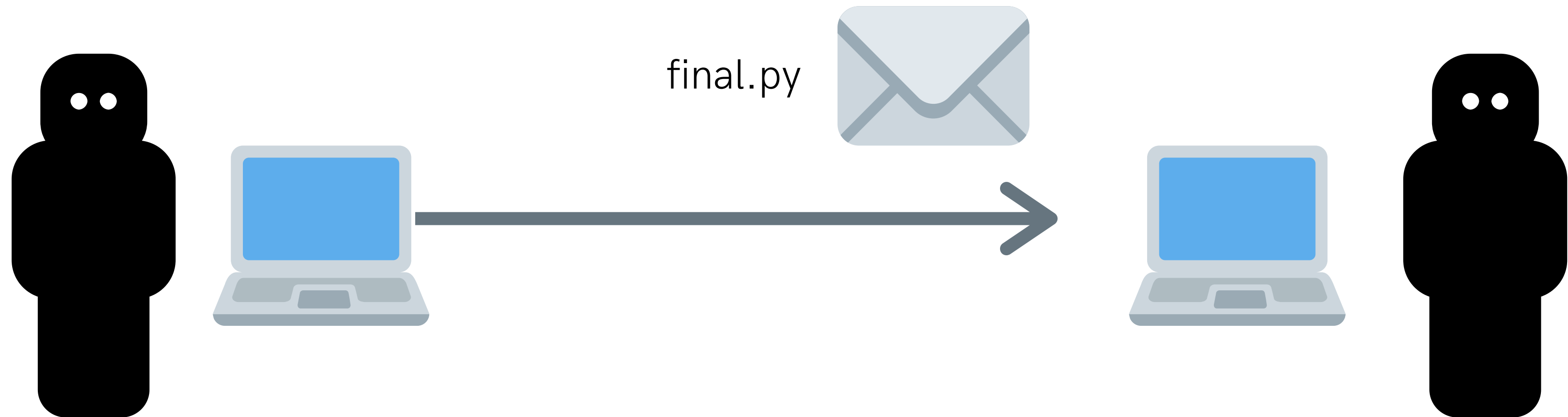
draft.py

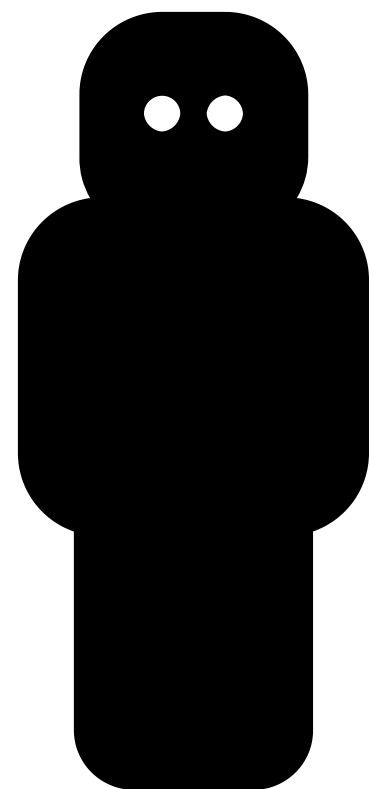




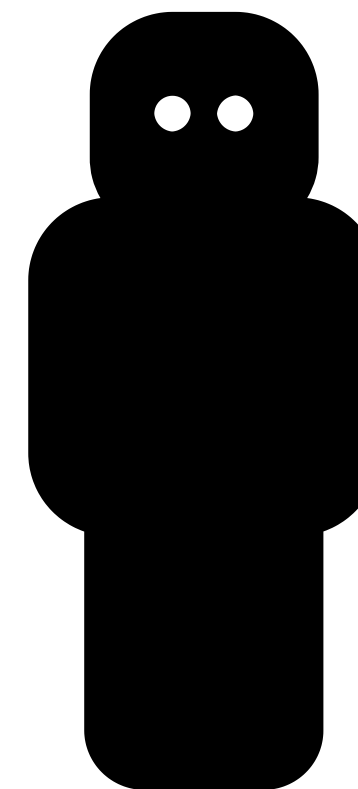
revised.py

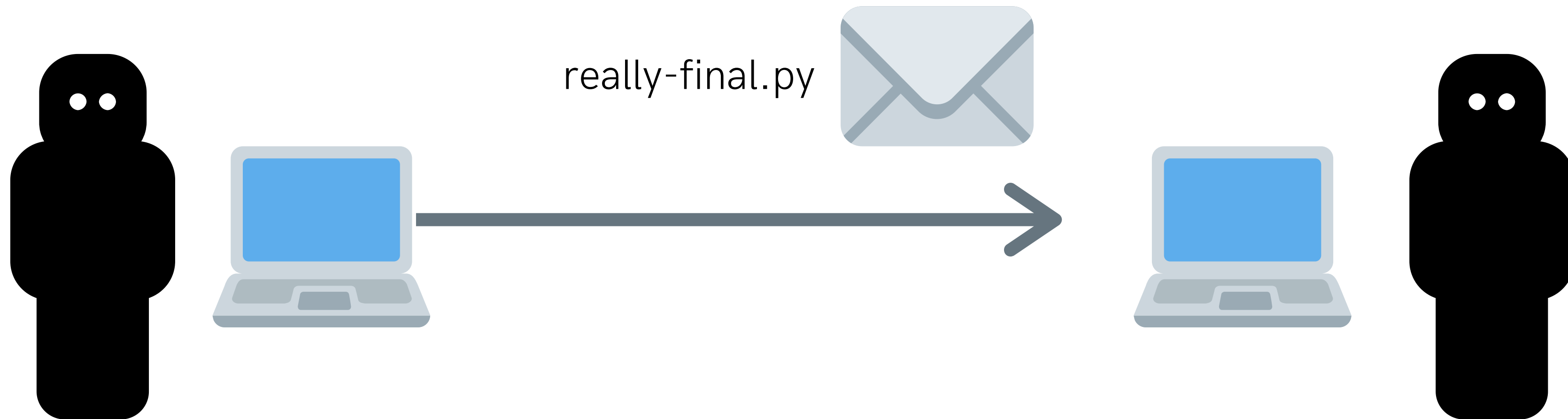


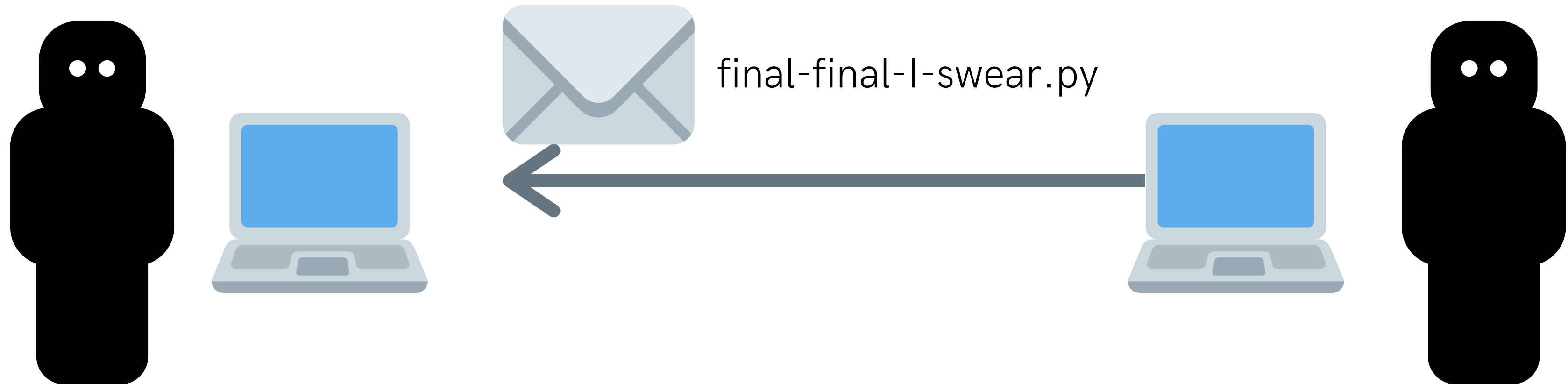


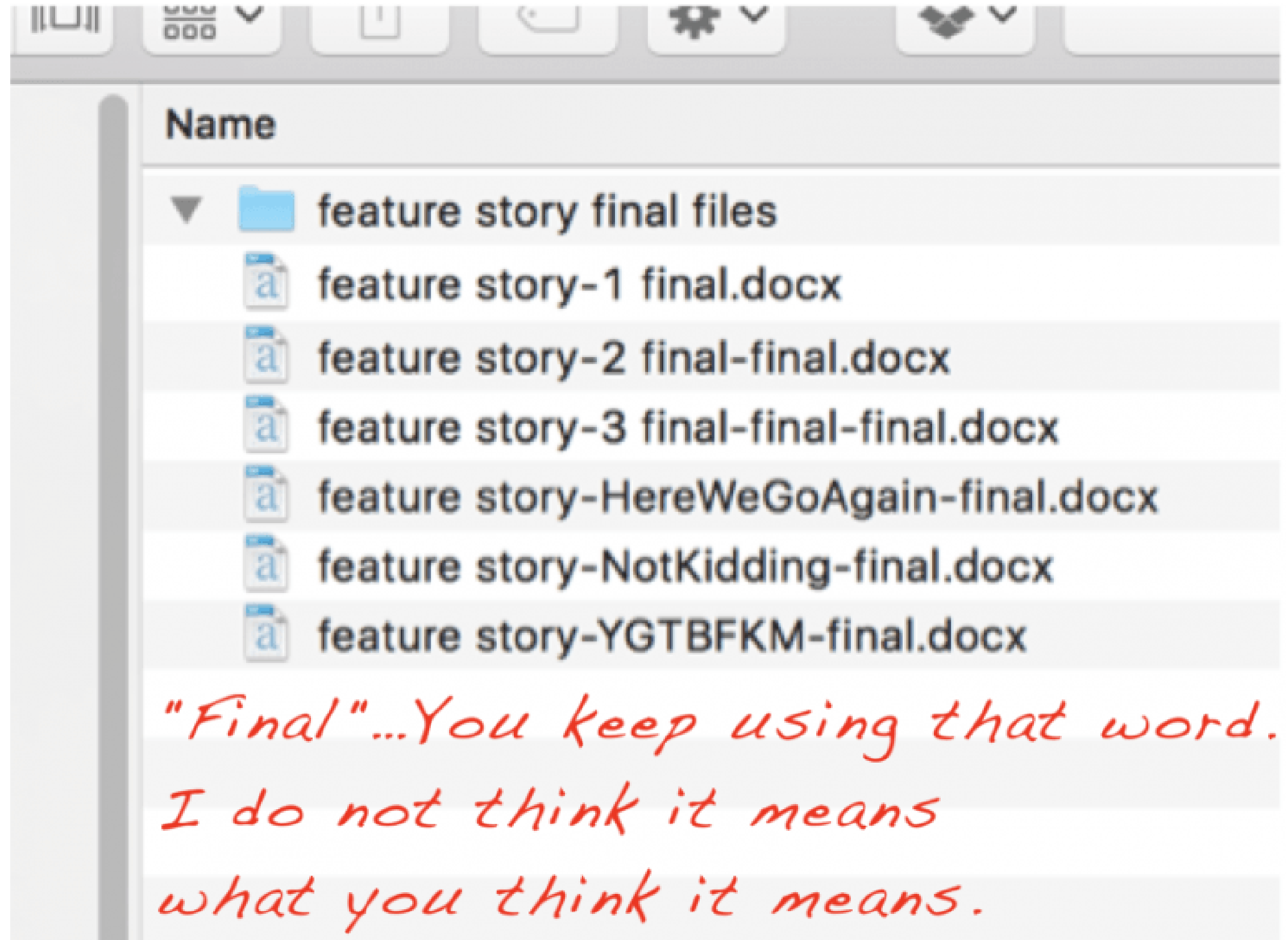


final2.py



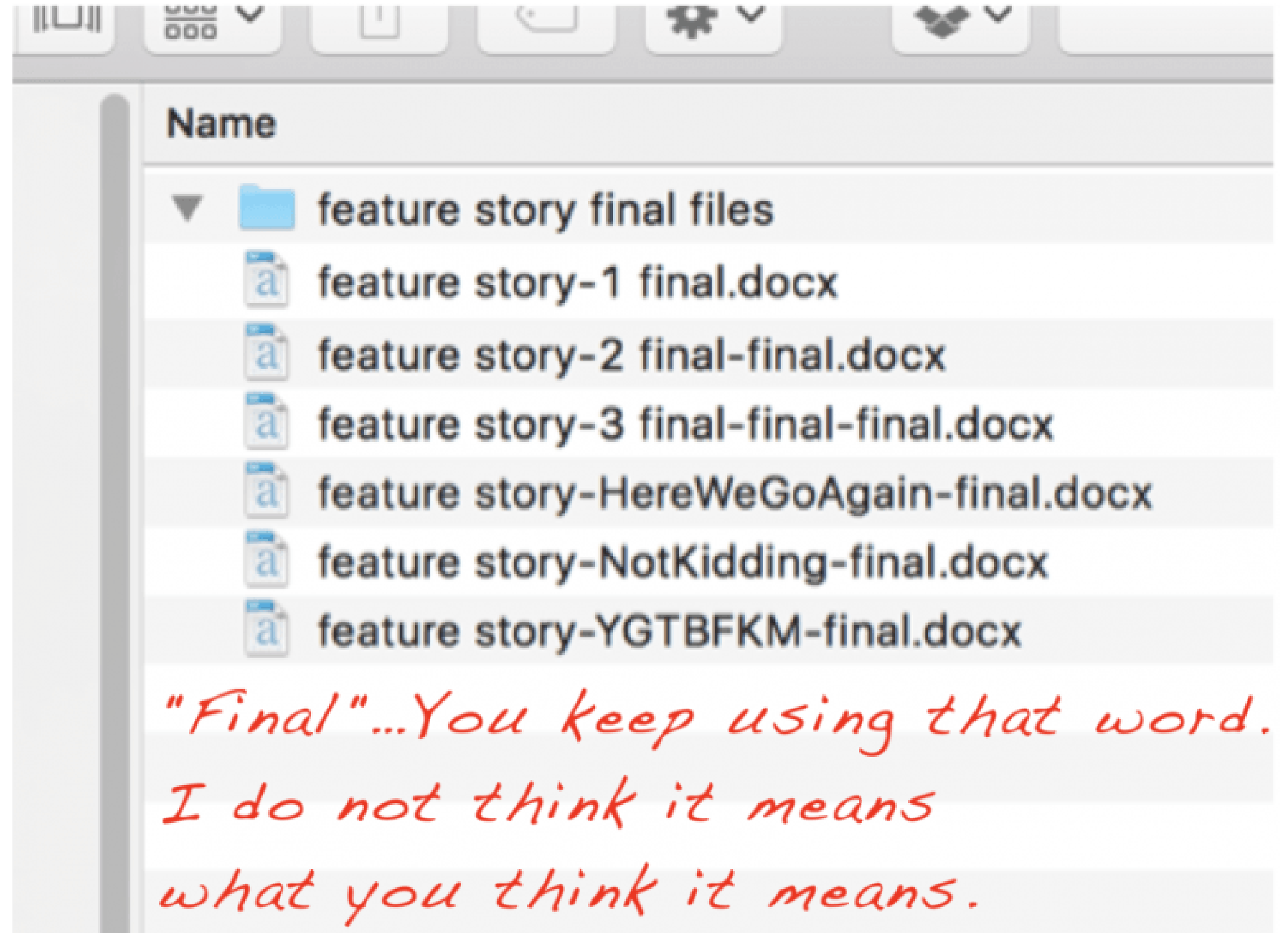






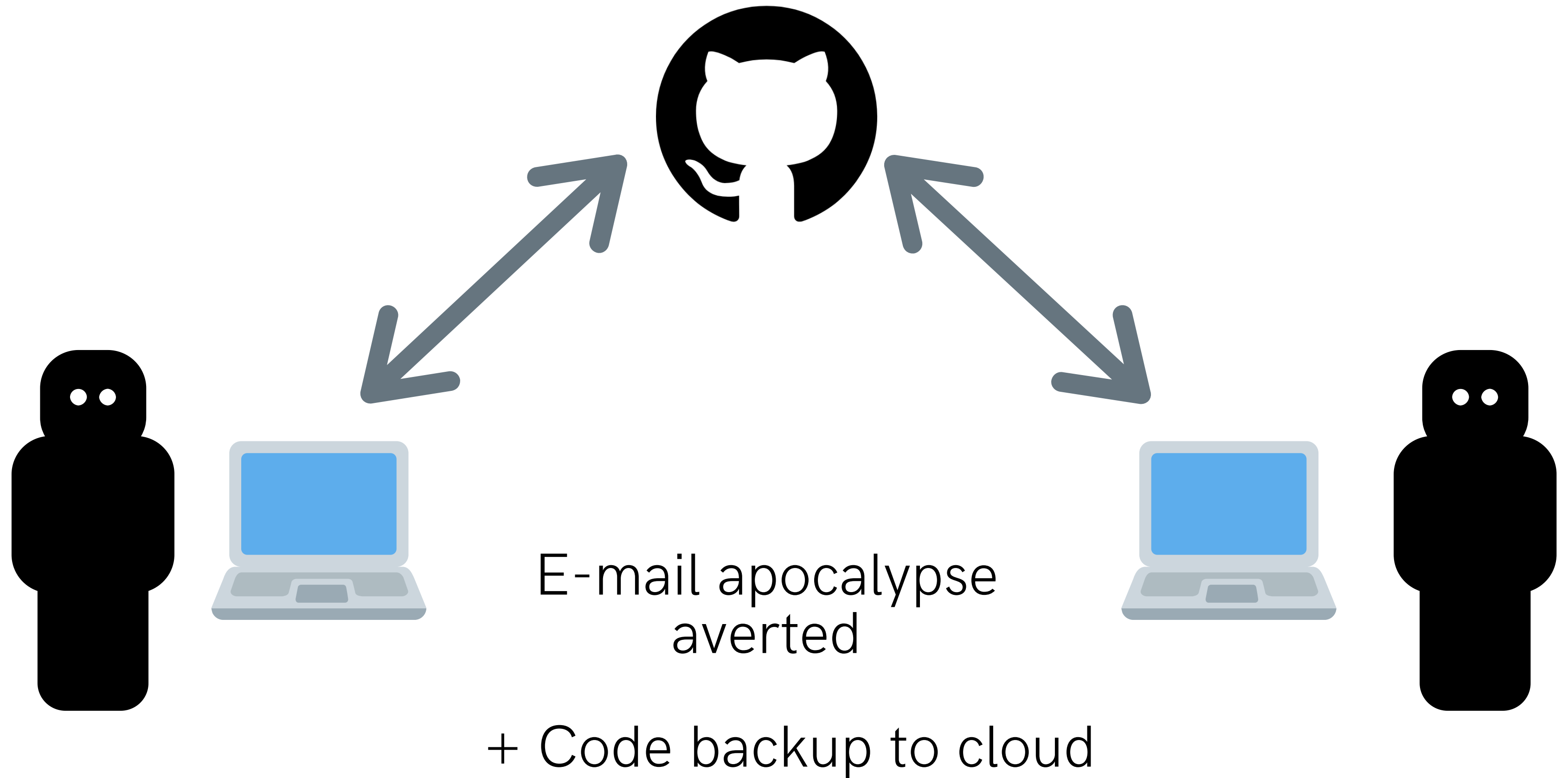
<https://doctorfreelance.com/file-naming-conventions/>

**Too many
e-mails
Too many
files
Not
trackable
Might submit
wrong version (!!)**



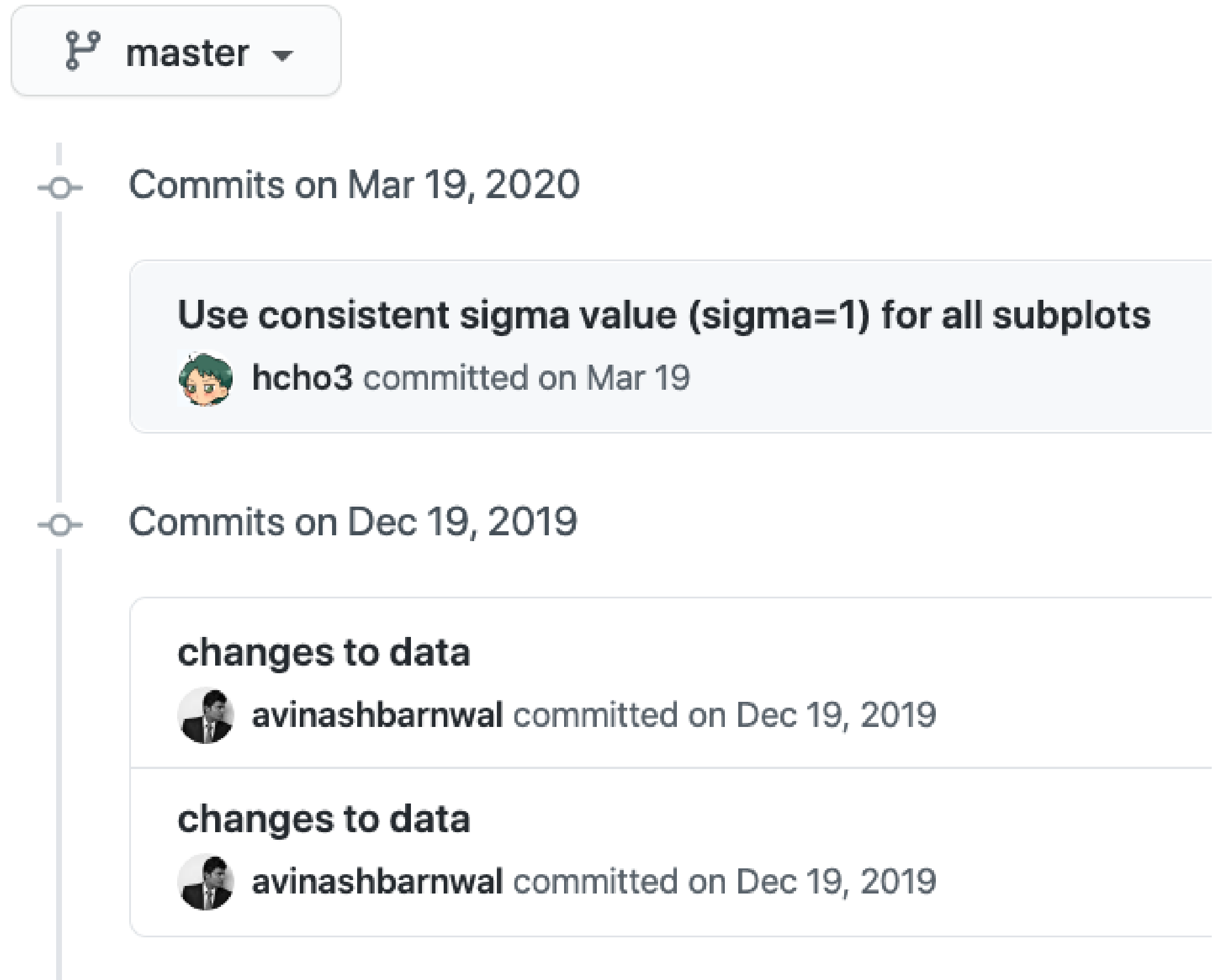
<https://doctorfreelance.com/file-naming-conventions/>

Collaboration: Git(Hub) to the Rescue



Collaboration: Git(Hub) to the Rescue

Track changes,
without duplicating files



Collaboration: Git(Hub) to the Rescue

Code reviews and pull requests

Use consistent sigma value (sigma=1) for all subplots #7

Merged

avinashbarnwal merged 1 commit into avinashbarnwal:master from unknown repository on Mar 19

Conversation 0

Commits 1

Checks 0

Files changed 2

Changes from all commits

File filter...

Jump to...

0 / 2 files viewed

Review changes

16 AFT/R/combined_assignment.R

@@ -249,10 +249,10 @@ distribution.list = list(gaussian=list(uncensored_l=loss(type="uncensored",t.low

249

left_h=hessian(type="lef

249

left_h=hessian(type="lef

250

right_h=hessian(type="ri

250

right_h=hessian(type="ri

251

interval_h=hessian(type=

251

interval_h=hessian(type=

252

-

252

+

253

-

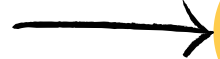
253

+

Collaboration: Git(Hub) to the Rescue

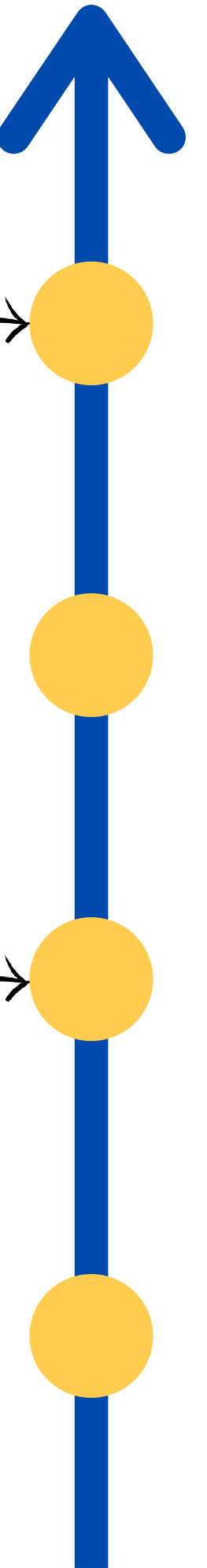
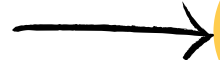
Git history

My program crashes here....



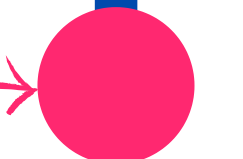
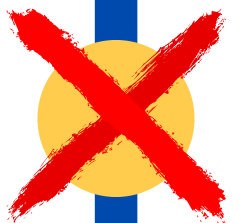
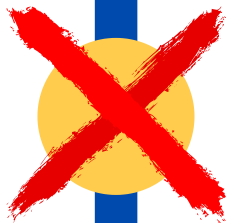
Roll back changes
as necessary

... but not here



Collaboration: Git(Hub) to the Rescue

Git history



Roll back changes
as necessary

Roll back to this version



Collaboration: Git(Hub) to the Rescue

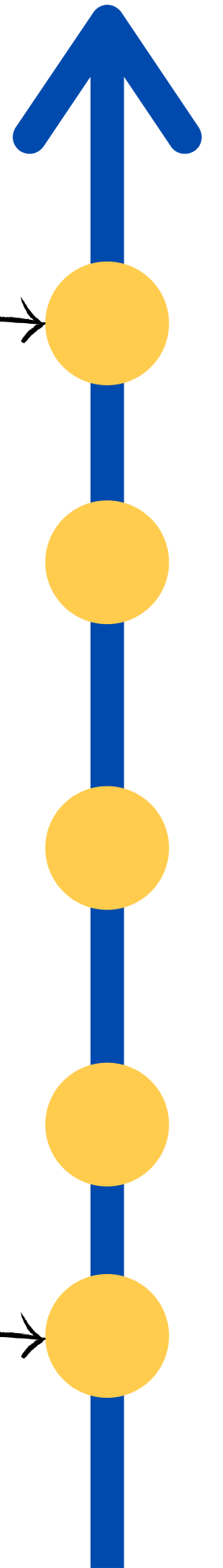
Git history

Find out which
change (commit)
introduce a bug
(git bisect)

My program crashes here....

Somewhere a bug was introduced

... but not here

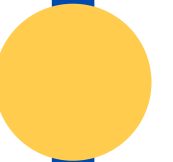
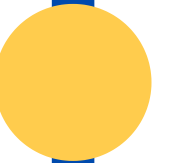


Collaboration: Git(Hub) to the Rescue

Git history



My program crashes here....

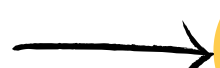


Find out which
change (commit)
introduce a bug
(git bisect)

Exact commit that introduced the bug



... but not here



Collaboration: Git(Hub) to the Rescue

Sign up at [GitHub.com](https://github.com) and learn Git today! It's free!

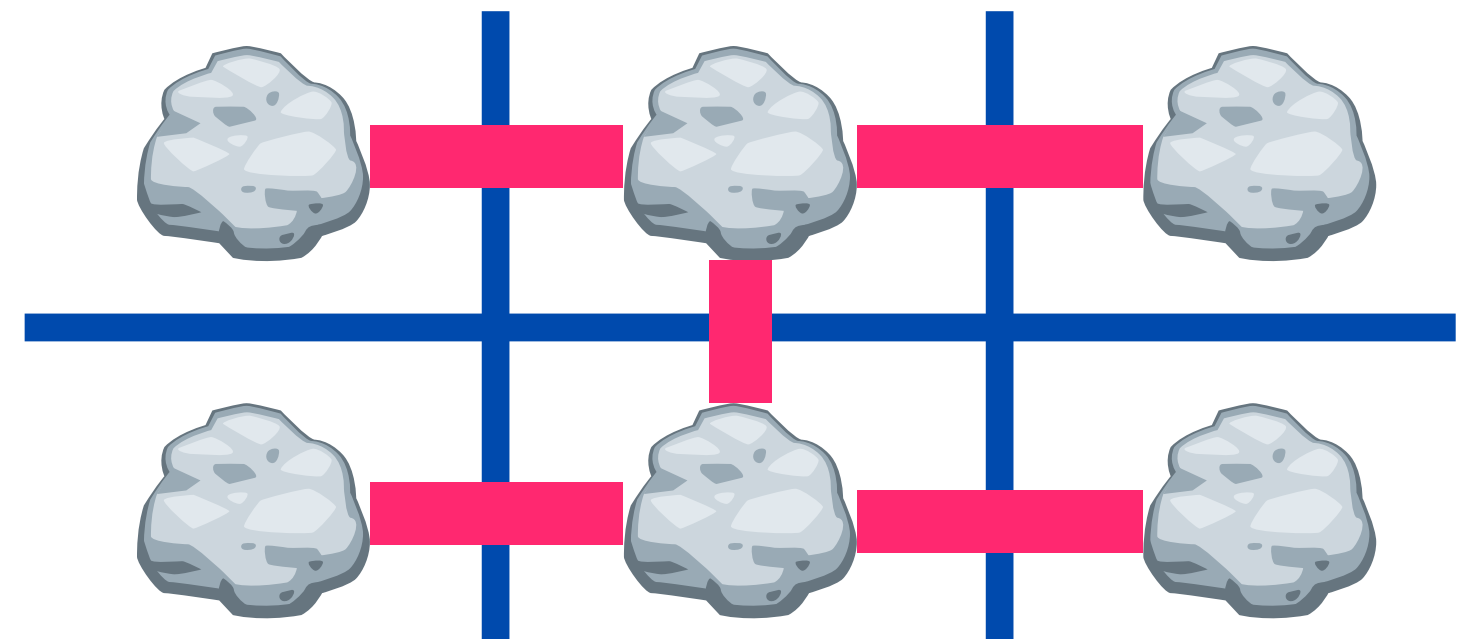
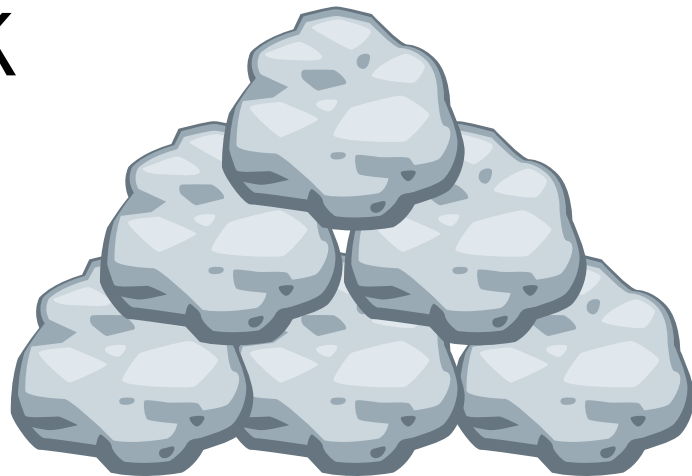
This will pay off immensely

Note: Make sure to use **private repository** to store your senior project

#3. MODULAR DESIGN

Organize your code into multiple **modules**

- Quite difficult to get it right first time
- Advice #1 definitely helps
- Get better with experience, e.g.
 - Open source
 - Internship
 - Work



#4. UNIT TESTING

Add small bit of code at a time, and then test it

- **Unit test**: Given an input, does a function produce a correct output?
 - If not, your code has a bug; fix it!

#4. UNIT TESTING

```
def area_of_rectangle(width, height):  
    return width * height
```

```
def test_rectangle():  
    assert area_of_rectangle(1, 4) == 4  
    assert area_of_rectangle(4, 1) == 4  
    assert area_of_rectangle(2, 2) == 4  
    assert area_of_rectangle(2.5, 1.6) == 4
```

#4. UNIT TESTING

Add small bit of code at a time, and then test it

- **Mocking**: Create fake parts that interact with a given module

Menu 1

Menu 2

Menu 3

Placeholder for Menu 1

#4. UNIT TESTING

Add small bit of code at a time, and then test it

- **Mocking**: Create fake parts that interact with a given module

Menu 1

Menu 2

Menu 3

Placeholder for Menu 2

#4. UNIT TESTING

Add small bit of code at a time, and then test it

- **Mocking**: Create fake parts that interact with a given module

Menu 1

Menu 2

Menu 3

Placeholder for Menu 3

#4. UNIT TESTING

Add small bit of code at a time, and then test it

- Tools
 - Pytest (Python)
 - Jest (JavaScript)
 - Google Tests (C++)
 - JUnit (Java)

CONCLUSION

Good habits let you avoid worst pitfalls of software complexity.

CONCLUSION

Good habits let you avoid worst pitfalls of software complexity.

Experience with real-world software trains your sense, to smell "bad code" from miles away

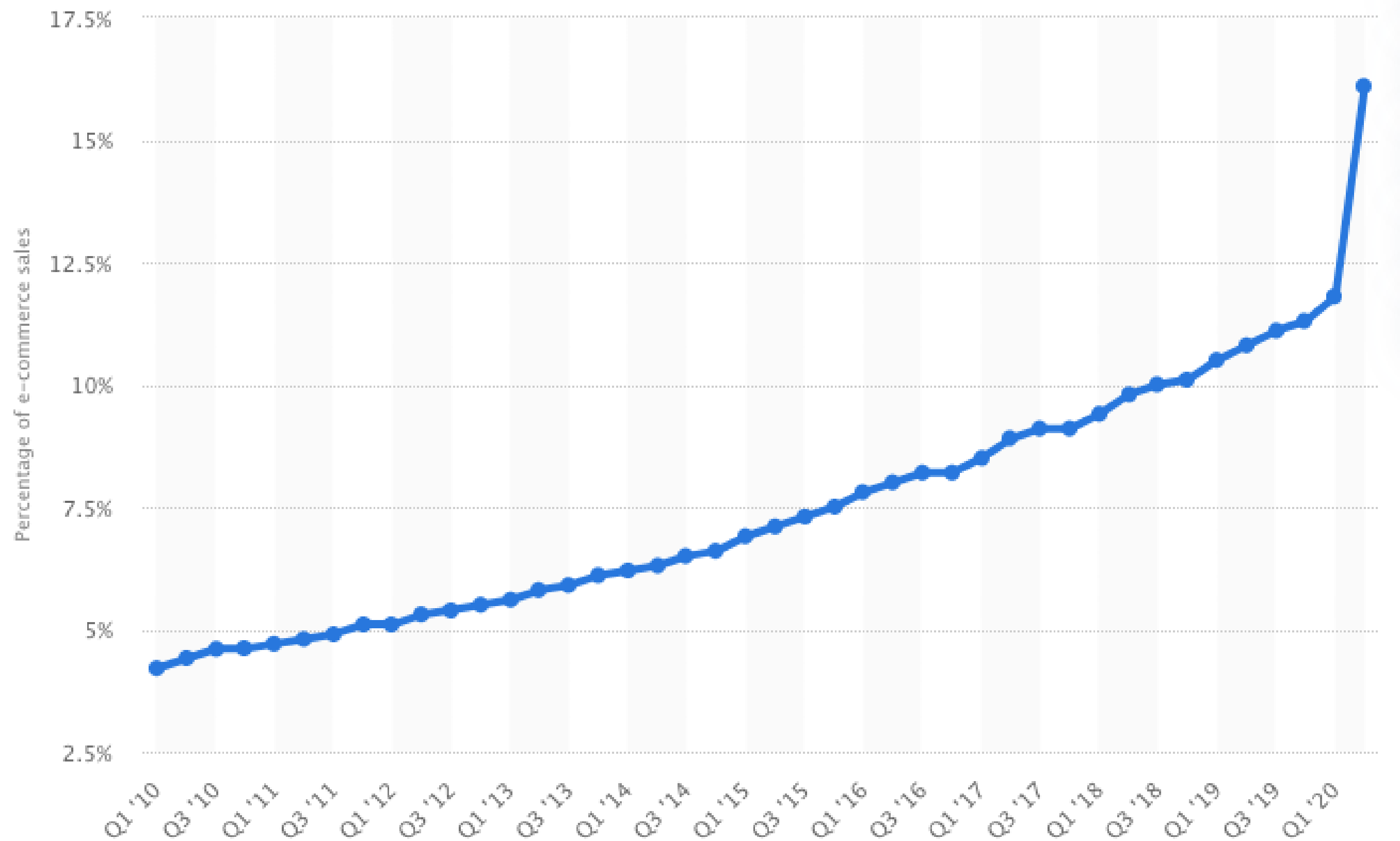
- Consider getting an **internship** / contributing to an **open-source project**
- You should avoid mistakes, but mistakes are often fastest way for you to learn

FYI: We are hiring interns

YEAR 2021: DIFFICULT TIME, ALSO WITH OPPORTUNITIES

Economy still under effect of the pandemic

Keep in mind of the opportunities



Quarterly share of e-commerce sales of total U.S. retail sales

<https://www.statista.com/statistics/187439/share-of-e-commerce-sales-in-total-us-retail-sales-in-2010/>



Small businesses (*) comprise 99.5% of firms and employ 52% of Americans (**). They took the brunt of the current recession. Many are building up IT infrastructure and establishing online presence.

(*) < 500 full-time employees

(**) As of 2019. "Business Employment Dynamics," U.S. Bureau of Labor Statistics

<https://www.bls.gov/bdm/bdmfirmsize.htm>

Small businesses (*) comprise 99.5% of firms and employ 52% of Americans (**). They took the brunt of the current recession. Many are building up IT infrastructure and establishing online presence.

=> Software is expanding its reach in the national economy. The pandemic will eventually pass, but the trend will continue.

(*) < 500 full-time employees

(**) As of 2019. "Business Employment Dynamics," U.S. Bureau of Labor Statistics

<https://www.bls.gov/bdm/bdmfirmsize.htm>

Small businesses (*) comprise 99.5% of firms and employ 52% of Americans (**). They took the brunt of the current recession. Many are building up IT infrastructure and establishing online presence.

=> Software is expanding its reach in the national economy. The pandemic will eventually pass, but the trend will continue.

=> We need more software developers.

(*) < 500 full-time employees

(**) As of 2019. "Business Employment Dynamics," U.S. Bureau of Labor Statistics

<https://www.bls.gov/bdm/bdmfirmsize.htm>

Advice

- Be a good self-learner
- Be a good listener
- Be coachable
- Be willing to wear multiple hats (if needed)
- Do not box yourself. Step outside of the comfort zone.
- Above all, **pursue excellence in coding**. The more experience you get, the better.



Feel free to reach out to me at
phcho@nvidia.com

