Fighting against Marketplace Scams using Community-driven AI

Sinan Ozdemir, Founder/CTO of Shiba

FEATURE STORE SUMMIT 2022

Organized by HOPSWORKS

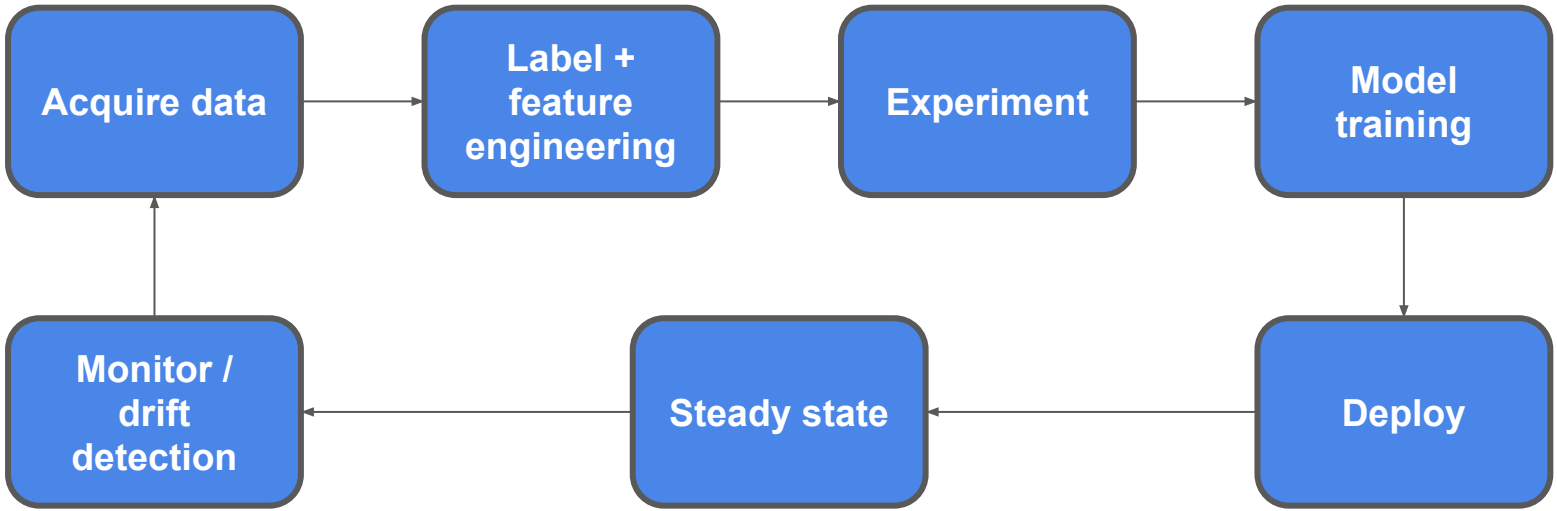
Hi, I'm Sinan Ozdemir

I'm an author, founder, mathematician, ML engineer, teacher, and art collector/lover

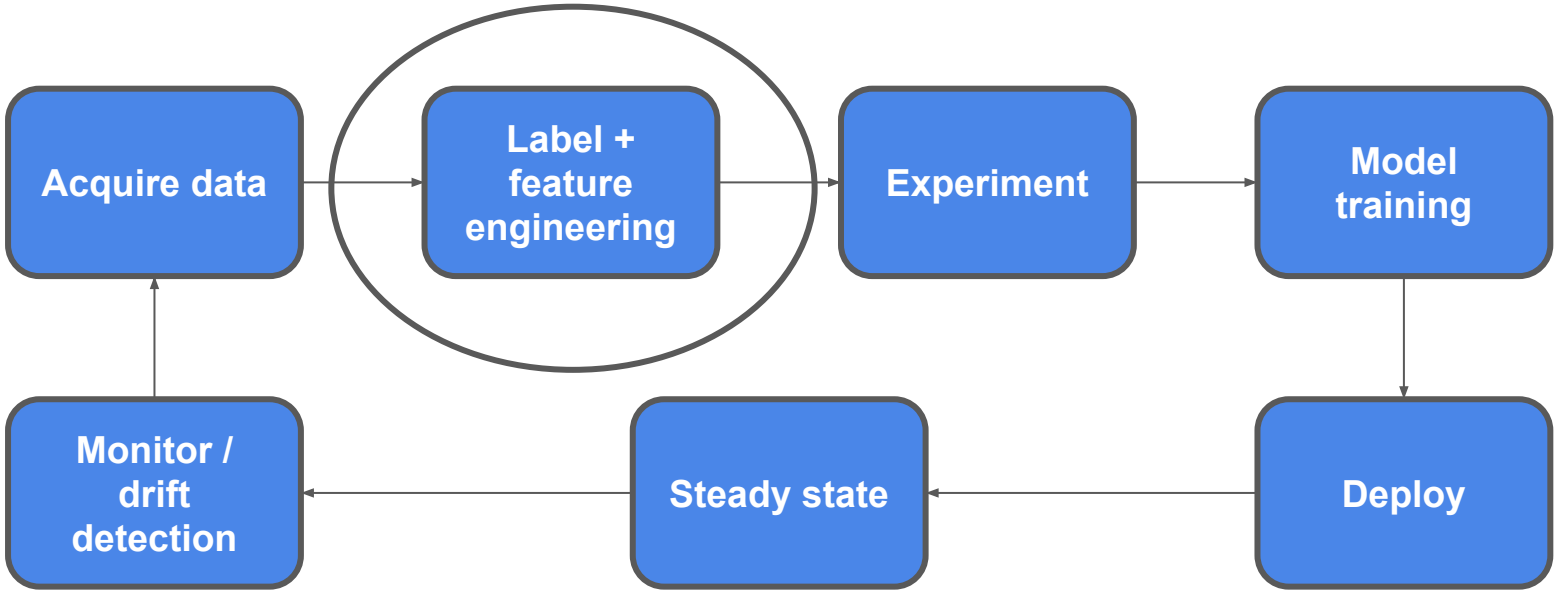
I started collecting art & NFTs during COVID after I sold my last startup, Kylie.ai and **I love feature engineering**

My cat, Euclid, waiting to get an autograph on my latest book

Modern ML Engineering



Zooming in on feature engineering



Feature understanding

Identifying levels of data + exploratory data analysis

Feature Structuring

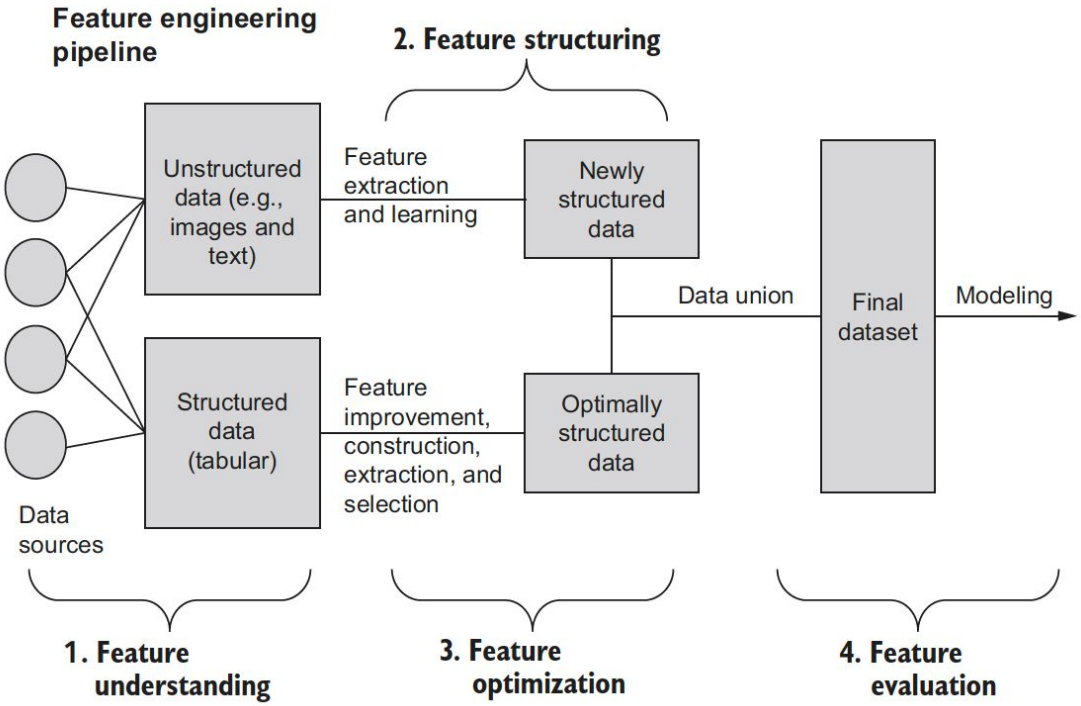
Converting unstructured data into something machine readable

Feature Optimization

Feature selection, improvement, construction, extraction

Feature Evaluation

Testing feature strength via model training



Source: The Feature Engineering Bookcamp - Sinan Ozdemir

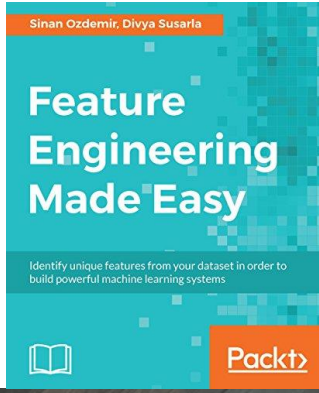
Why Feature Engineering?

A majority of DS/DE/ML time is spent working with data

The biggest boosts in ML performance can often be attributed to feature engineering

Bias in ML pipelines can be best mitigated in the preprocessing phase of model training

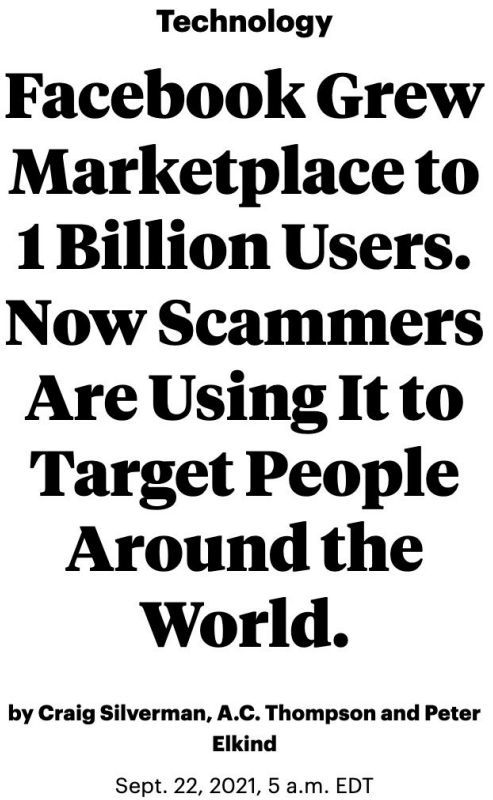
Feature engineering is often how we turn unusable data into **useful and explainable data**



Humans are no strangers to fraud

We all have a story of being, or know someone who has been, tricked, duped, or otherwise bamboozled

Marketplaces, both physical and digital, are fraught with fraud



Technology

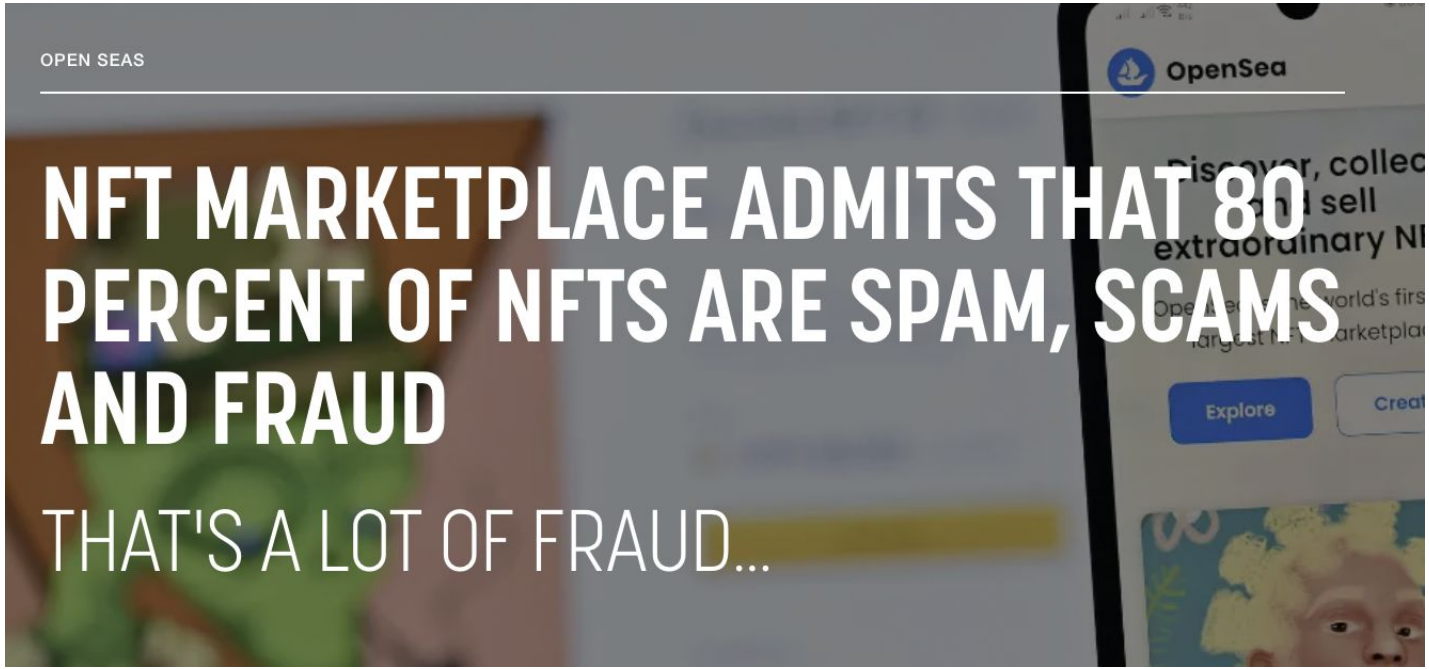
Facebook Grew Marketplace to 1 Billion Users. Now Scammers Are Using It to Target People Around the World.

by Craig Silverman, A.C. Thompson and Peter Elkind

Sept. 22, 2021, 5 a.m. EDT

Source: propublica.org/article/facebook-grew-marketplace-to-1-billion-users-now-scammers-are-using-it-to-target-people-around-the-world

Digital marketplaces especially full with fraud



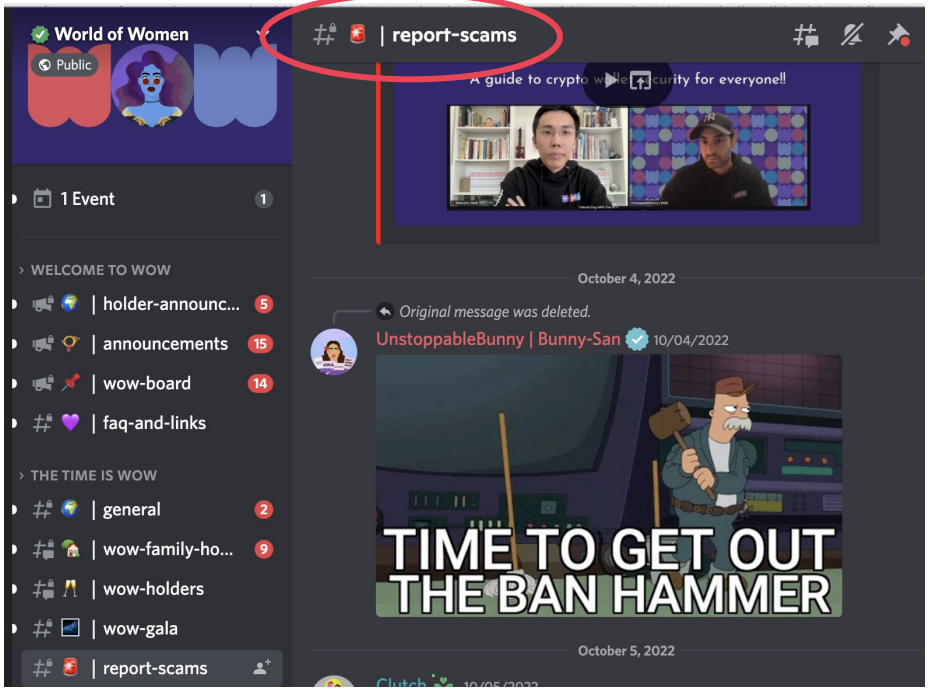
Source: https://futurism.com/the-byte/nft-marketplace-fraud-scams

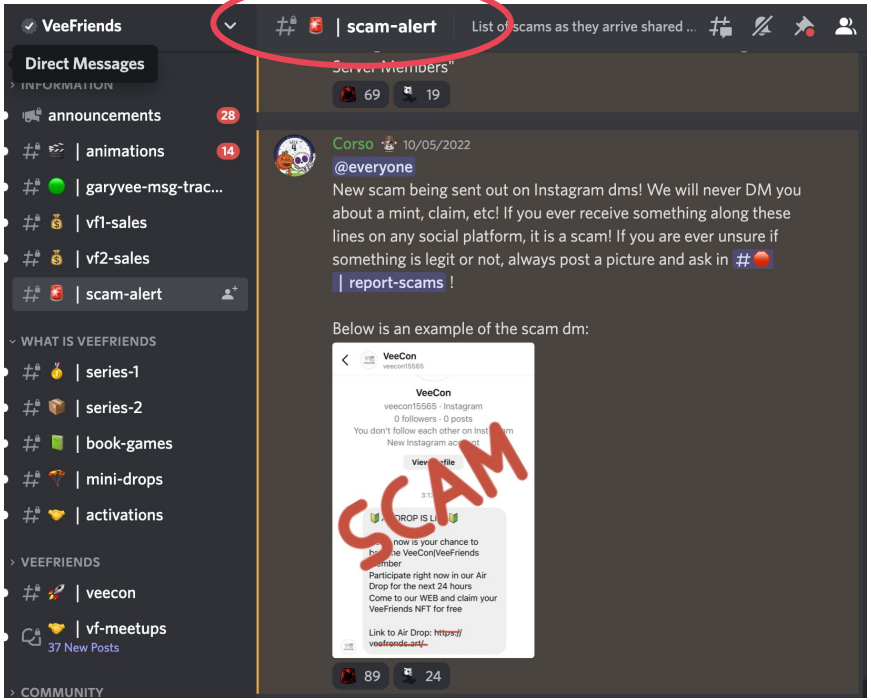
Physical vs digital scams

- Physical marketplace scams (mostly in the secondary market) range from mislabeling items online to being robbed at gunpoint

- Digital marketplace scams (both secondary and people pretending to be primary sources of assets) are generally less physically threatening but **easier to conduct at scale**

NFT communities have spaces to report scams by necessity





Community-driven vigilance

- Discord communities are proactive to alert fellow NFT lovers about potential scams

- They will negatively react and message about people DMing them with scams

- **Communities act as their own sheriff and therefore can provide data to fight scams at scale**

Spam vs Scam

- **Scam** - intentionally fraudulent messages generally designed to steal money/assets. These are more dangerous

- **Spam** - Messages (often referred to as junk) designed to get your attention. Not always scams; it could be someone trying to get attention on their legitimate project through dubious/annoying means

Case study: I got targeted a few days ago with a scam

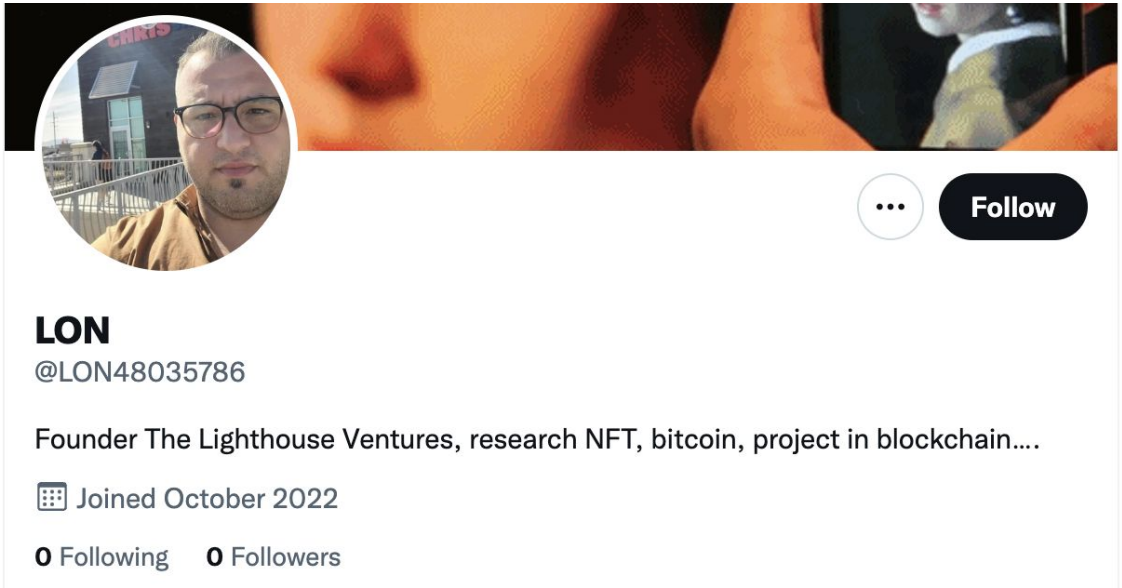
I checked the discord to see if it was already reported. It was

So how do you know?

A few dead giveaways:

User is clearly fake. They have 0 following and followers and only joined twitter in October (at most 8 days ago at the time)

Website/twitter handle is not associated with the project

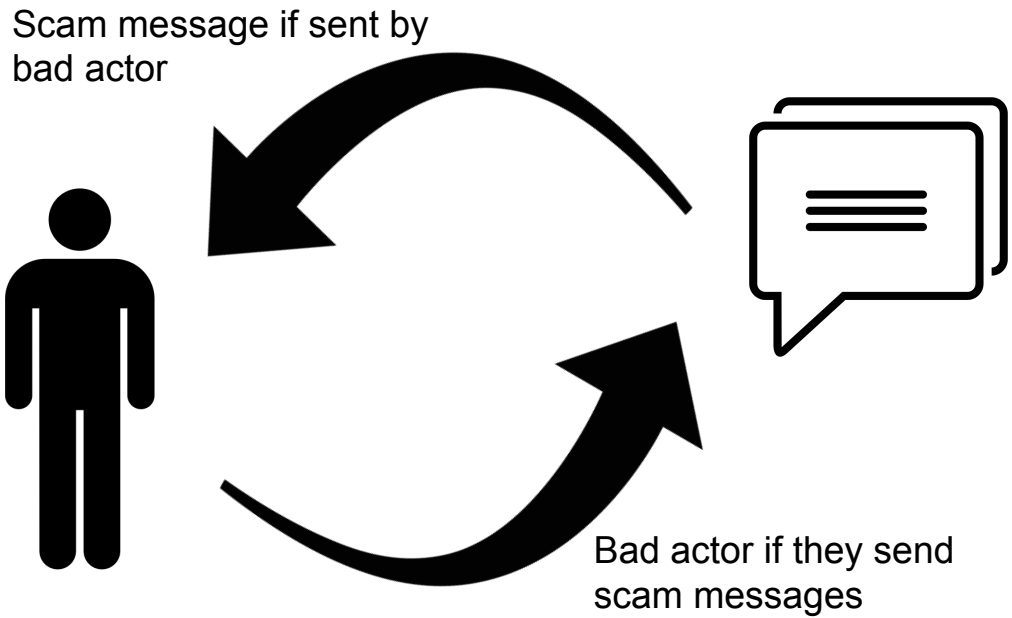


Two entities at play: messages / actors

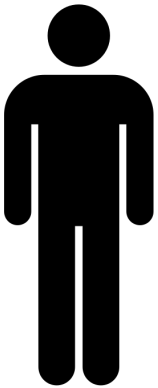
- You have **scam messages** written by **bad actors**

- Each have a different set of features to maintain but are related to each other

Easy to get caught in an endless catch-22



Define features that are independent but related



# of likely scam messages

# of total messages

Account age, # followers, # friends, etc

Number of people mentioned

Use of common scam phrases eg “You’ve been whitelisted” (in the NFT world)

Count of negative reactions in public space

Claiming to be the admin of a project

Multiple feature-drivers

Number of people mentioned

Use of known common scam phrases eg “You’ve been whitelisted” (in the NFT world)

`Bag of Words derived feature`

Count of negative reactions in public space

`Dictionary of negative emojis`

Claiming to be the admin of a project

`Semantic model derived metric (e.g. BERT/GPT)`

So how do feature stores fit in?

Feature Store

Create, define, and iterate on features quickly to discover strong features

Centralized feature location to make experiments/reporting easier

Share features across the organization (message analytics + user analytics)

Monitor features over time for drift and train/test skew

Training/testing skew

Hindsight is 20/20

Historical user and message data have already been reacted to / reported but this isn't true for a message that was sent just 2 minutes ago

For this reason, we rely more on features that are time-agnostic (account age, message content) and less so on time-reliant features (# reactions, # retweets)

Scam Drift

Scams evolve but tend to stick with similar tactics for a while

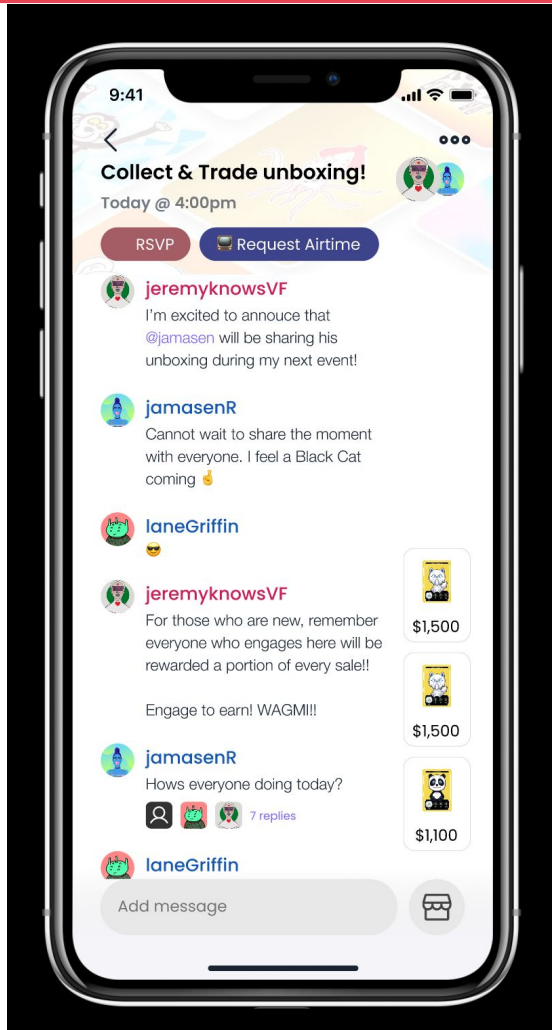
By keeping track of scam reporting channels, we can know about new scams quickly and the new ways/domains/phrases that scammers are using

Yesterday

Feature Store @ Shiba

To build a safe marketplace for physical and digital goods, we need to build an effective **community-driven AI** to drive out scammers and spammy behavior

Community-driven in that many of our features/labels for detecting bad actors/scams are derived from the collective mind of the marketplace users

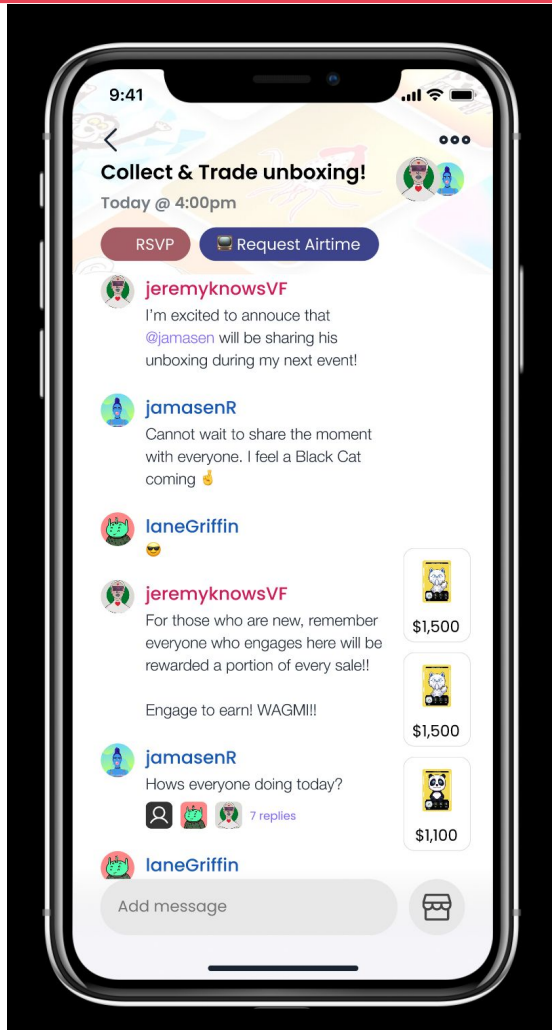


Feature Store @ Shiba

Because we spent time crafting useful features and we rely on modern Transformer-based semantic models we get the best of both worlds:

1. **Speed + performance** of pre-trained language models

2. The ability to explain to someone why their message/account was marked as potential scam and **educate people on what to look out for**

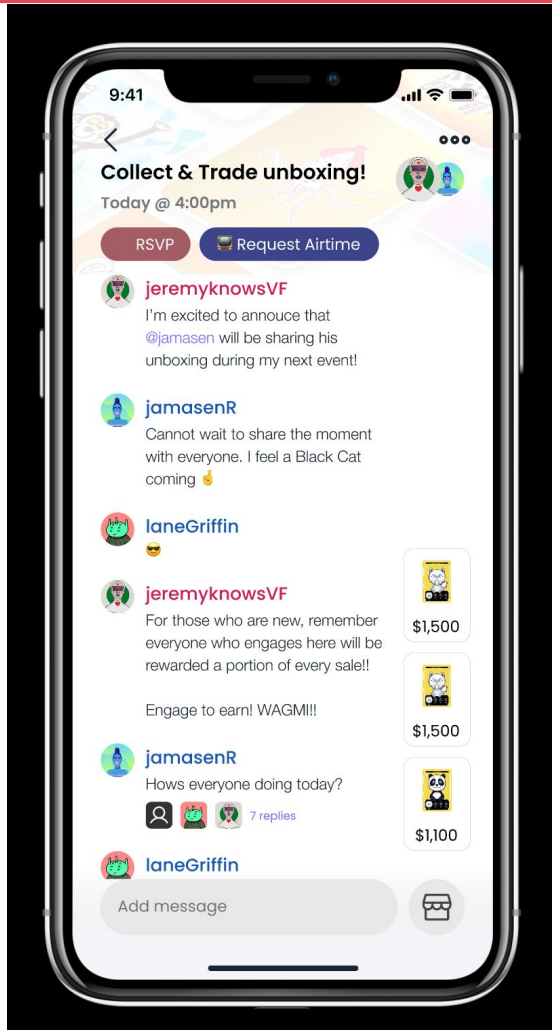


We deserve to shop in peace

I hope we can all agree that whether we want to buy the latest hot NFT project, sell a couch, or buy a piece of art, **no one wants to be scammed**

What you can do

- Be vigilant
- Know what to look for
- See something, say something / Report bad actors



Thank you

Sinan Ozdemir, Founder of Shiba.art

Twitter: @prof_oz

Linkedin: in/sinan-ozdemir