



Spreaker★
From  **iHeart**

Join The Podcast Revolution

Diego Ercoli

Lucidi originali concessi da Emanuele Rampichini - Engineering Manager



Spreaker is how you podcast.

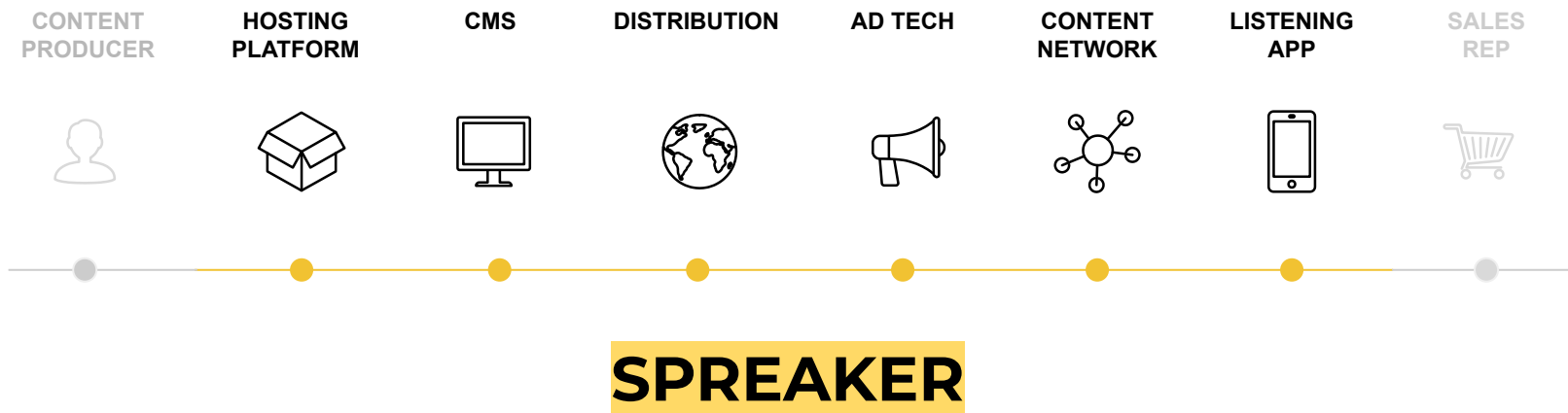
Whether you're producing your passion project or growing a podcast business it has never been easier. Spreaker is the one-stop shop for podcast hosting, creation, distribution and monetization.

Start Podcasting



Within the last 11 years we have moved from being a **podcast hosting and distribution platform** to expanding and innovating **how content producers monetize and hone their craft.**

The Podcast Business Ecosystem



What sets us apart?

- Effortless Publishing
- Enterprise Level Reliability
- Dynamic ad insertion and management
- **Extensive Analytics and Reporting**
- Distribution far and Wide

Spreaker in numbers

>100 k

Podcast producers

>300 M

Downloads per month

4

Continents



Terms & Conditions

Payments are active.

Your payment account has been [verified](#).

\$28.50

Projected monthly payment (\$10.00 minimum).

Next payment of \$14.55 (earned until October 31, 2021) will be paid on or before December 31, 2021.

\$12.81

Paid on November 30, 2021 to your paypal account: theslangpodcast@gmail.com.

Ad Exchange Activation



Don't waste any monetization opportunity! When you activate the Ad Exchange, you receive additional revenues from the ads coming from our open marketplace, even if you don't have any running campaign.

Ad Exchange Statistics

Podcasters earn money from downloads

\$11.58

-0% Previous 30 days



SEE MORE



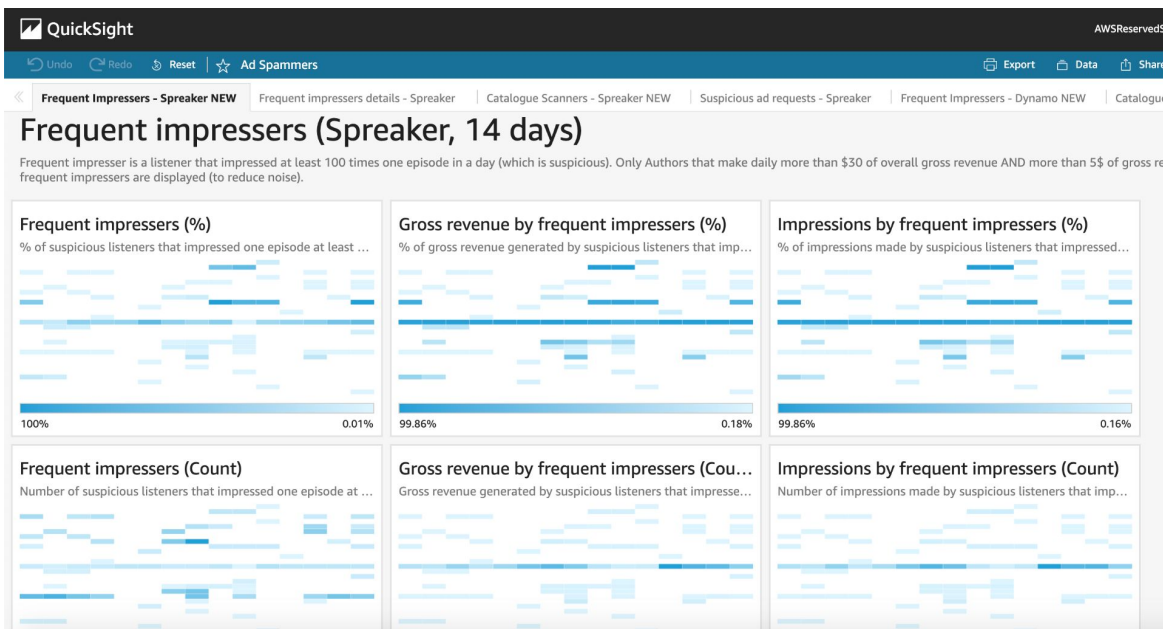
Incentive to game the system!

- If you are able to **generate fake downloads** on your content and not get caught **you can earn an interesting amount of money** :)
- This is **bad for us, for the honest content producers** and **for the industry as a whole**.

How do we deal with that at Spreaker?

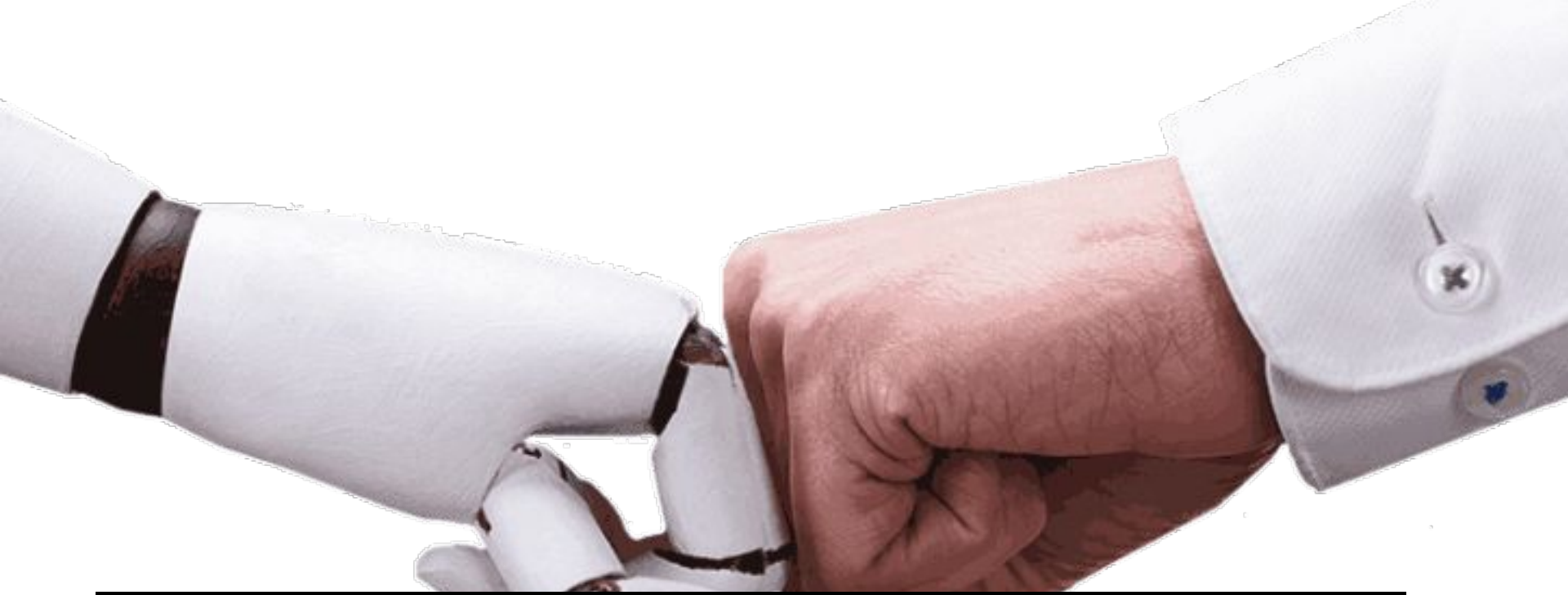
Currently we have a combination of:

- **Data collection** and **aggregation**
- Definition of **suspicious thresholds** based on **experience**
- Data **visualization**
- **Human analysis** and **action**



Is it enough?

- The current approach is **kind of OK at the current scale** but it becomes more and **more unsustainable the more producers we onboard.**



**We want to improve our system
exploring ML techniques**

Project goal

- Explore **algorithms** / **procedures** to automatically discover fake downloads to **identify cheaters**
 - **Training and evaluating ML architectures on a datasets provided by us**

Dataset definition

| Field | Description |
|------------------------|---|
| content_user_id | The id of the podcaster hosting on our platform. |
| content_podcast_id | The id of the podcast owned by the podcaster hosting on our platform. |
| content_episode_id | The id of the episode that is part of the podcast. |
| listener_user_agent | User agent string of the episode download. |
| listener_geolocation | Geolocation down to City from where the download was generated from |
| listener_device | Device family (desktop, mobile etc..) |
| listener_os | Operating system (windows, iOs, Android...) |
| listener_anonymized_ip | Anonymized hash of the ip performing the download. |
| timestamp | Timestamp of the download |

Labelled data

| user_id | is_scammer |
|---------|------------|
| 1 | true |
| 2 | false |
| 3 | false |
| 4 | true |

- We will **provide a set of labelled data coming from supervised human activity**

Expected result

| user_id | is_scammer | p(is_scammer) |
|---------|------------|---------------|
| 1 | true | 0.9 |
| 2 | false | 0.02 |
| 3 | false | 0.2 |
| 4 | true | 0.85 |
| 5 | | 0.7 |
| 6 | | 0.2 |
| 7 | | 0.99 |
| 8 | | 0.12 |

- We want to **test several algorithms** to **correlate feature vectors coming from the initial dataset and the labelled data** and apply it to a larger dataset to **classify users**.

ML Architecture s we want to explore

- Logistic Regression
 - Naive Bayes
 - K-Nearest Neighbors
 - Decision Tree
 - Support Vector Machines
-
- Neural Networks (**BONUS**)