



M.A.G.D.A Event Driven, Serverless and Machine Learning in the Real World

Matteo Rossi

AdTech Lead @ Spreaker from iHeart



Spreaker is a podcast platform

- Within **Spreaker**, podcasts can be hosted, distributed, monetized, discovered, and listened to
- Part of iHeartMedia family since 2020







- Spreaker is a podcast platform
- Within Spreaker, podcasts can be hosted, distributed, monetized, discovered, and listened to
- Part of iHeartMedia family since 2020

Data from 2022 per month

350-400 Millions IAB Certified Downloads

~25 PB of audio delivered through Cloudfront

~30K different ads delivered through the platform

Why are we here?

How serverless and event driven architectures merged with AI and machine learning principles can fix your problems in the real world!









Listener



Share user info User Agent, IP Podcast creator



Targeting

Share content info Podcast category **Advertiser**



Sell the ads based on *targeting*

Rules

I don't want tobacco ads I don't sell tobacco ads







• Broken market rules

Deliver uncategorized or miscategorized ads

• Missed ad revenue opportunities Publishers avoid to enable monetization or want to leave



caterina 12:42 PM @magda we got a complaint because on these 2 podcasts https://www.spreaker.com/show/ and https://www.spreaker.com/show/ and probably other of this user https://admin.spreaker.com/user/ there were ads from tobacco but on both of them the category IAB9-9 is blocked. The ads were listened on 07/30 and 07/31. Can you please check if the ad is miscategorized?



- Collect ad's info manually
- Reactive approach: listen to hundred of ads
- Fix categorization: reach out external Ad Servers

Reports received

multiple reports per day during 2020 US election

Lead time

~3 hours to find the ad ? time to be blocked by the SSP

> **Operational cost** 2-3 person X 3 hours

Proactive approach

Impossible to apply



The solution: M.A.G.D.A.

Machine Augmented Guard for Dynamic Advertising

- Collect ad's info automatically
- Proactive approach: scan ads' inventory on a daily basis
- Fix categorization: in real time inside Spreaker platform
- Named as the person in charge of the Ad operation flow



@magda we got a complaint because on these 2 podcasts
https://www.spreaker.com/show/_______ and
https://www.spreaker.com/show/______ and probably other of this user
https://admin.spreaker.com/user/______ there were ads from_____ tobacco but on
both of them the category IAB9-9 is blocked. The ads were listened on 07/30 and
07/31. Can you please check if the ad is miscategorized?

The challenge

- No impact on the listen experience of the final user
- Serverless event driven architecture to share info between the platform and M.A.G.D.A



Ads sources

The challenge

- Stitching API encodes all the ads to be delivered with the same audio settings of the media
- The output of the Ad Encoder is
 - the ad file encoded
 - a json file with the ad's metadata



Transcription phase

- Collect AWS transcribe output through EventBridge/SNS/SQS
- Merge ad's info (Ad encoder + Transcribe)
- Viewer and search engine through Quicksight



C ~ 2

Lower ad text search Enter value	Lower advertiser search Enter value	Lower uri search Enter value	Ad language	Provider All	
Transcription day between 2023/03/01 • 2023/03/31	IAB category assigned All	Exclude IAB category returned Enter value	IAB category inferred IAB11-4 💌		
Ads list					

Transcription day	provider	advertiser	lineitem	ad_uri	IAB cat. returned	ad_iab	IAB cat.inferred	ad_text
Mar 14, 2023	example	example	example	https://www.example.com	IAB1	IAB11-4	IAB11-4	Lorem ipsum dolor sit amet, consectetur adipi
Mar 14, 2023	example	example	example 1	https://www.example-1.com	IAB2	IAB11-4	IAB11-4	Lorem ipsum dolor sit amet, consectetur adipi

Why EB + SNS + SQS?

- Ad Encoder is multiregion
- M.A.G.D.A. is single region
- Transcribe supports single region input/output
- Use SNS to spread events cross region



Enrich ad's info automatically

AI and Machine learning

- Automatically infer ad's category
- Save outcome in a shared DynamoDB



Classify categories service

 The ad transcription is transformed by a NLP model in a vector (https://huggingface.co/sentence-transformers/paraphrasemultilingual-mpnet-base-v2)

• The classifier category model takes the vector and infers if the ad is for instance political and the probability associated



Classify categories service

- Public NLP model well known and used
- Custom classifier categories model to cover the specific use case
- Time and manual work to collect input to train the classifier categories model

Illegal Content	FALSE -	sheriff also villanueva has gone rogue under investigation for looking the other way when violent deputy gr ignoring leeful subportes just like trung on june 7th vote no on rogue orp also villanueva ad paid for by o villanueva to sheriff 2022 sponsored by smart justice california action fund committee, major funding from action fund not authorized by a candidate or committee controlled by candidate.
Adoption	TRUE -	I found out i was pregnant and i doh't know what to do. I was looking for options. Ihen I found adoptions for from the heart advocates for you, they't explain all your options, help with housing, financial and medical i and empower you to make the right choice for you. I doh't feel alone anymon. I chose open adoption the fi deughter grower qui, i know i made the right choice for both of us to learn more visit adoptions how.
Sexuality	TRUE -	(ingling, jingling, ingling, oh, whoa, you never know what the night might bring. It's the number one libido is women's sexual health, with only one labled bace a week or as needed, you can turn your gorffs into king you can be great? order yours today at chings brands dot com, that's c h i n g a brands dot com, always b good good fines.
Martial Arts	FALSE -	htm? If i get hurt, which gomea want the dog? which gomea take grand to her appointment? which gomea tak which gomea organize the dinner at school? which gomea give me a tash? have, um, and ask pouned?, if y harmed? the out more of a sc dot ou dot a z 8.
Human Resources	FALSE -	wood no malter where you work all university, we provide the uniforms and services you need to work hard after shift, night after night, season after season, we always deliver because you always deliver, discover a wildions.
Crossover	TRUE 🔻	yeah, Neen is no true joy in confirming, but not everyone has what it takes to be efflerent, introducing a or who have what it takes to treak free. The all new infinit qo 55 with a dramatic style that defess expectation infinit qo 55 keal free way, bit free world catch up, viai infinity stewart to learn more about the infinit qo 50
News	TRUE -	Bings are changing so repidly, you need to know what's going on. N. I'm sharon, good. I a. from seven nee the pressing questions interest rates, home prices, your health, the externe weather, proceries, hises in p queensiand's rental crisis? know what's happening? know what matters? no. the news seven news queen source every night at six clicos.
Resume Writing/Advice	TRUE -	tadies, tadies, tadies, iactes sometimes you entry that you're not good enough, troat me, nobody in almost a beaking decade haning my suri aucked out of my body doing sometiming that i didn't low. Simaly, to the heaking decade, i decided uses going to go for it, and reve ended up building the the of my deam, a tim- to tecture i realised that radical confidence is being afraid and doing it, anywas, i wrote this book for you wil you need to go from feeling such and hashinated to doing anything you set your and to.
Donation/Charity	TRUE -	hey guys, Im greg I and every single community needs hences, anyone who registers as an organ or tasu- just one donor can save eight lives and enhance the lives for 75 others be a hero and register as a donor

Classify categories service

- Runs in Python to support Pytorch and Catboost
- Deployed with a custom Docker container to support size of dependencies and models (Lambda package limited to 250MB)

Docker image

FROM public.ecr.aws/lambda/python:3.8 RUN yum install unzip -y && yum clean all # Install the function's dependencies using file requirements.txt # from your project folder. COPY ad-quality/src/functions/classifier-categories/requirements.txt . RUN pip3 install -r requirements.txt --target "\${LAMBDA TASK ROOT}" ARG SPREAKER_ENV RUN curl "https://vn-pdn-\${SPREAKER_ENV}sentence-transformer-model.zip" > sentence-transformer-model.zip && \ unzip sentence-transformer-model.zip -d assets && \ rm sentence-transformer-model.zip # Copy function code COPY ad-quality/src/functions/classifier-categories \${LAMBDA_TASK_ROOT} # Set the CMD to your handler (could also be done as a parameter override outside of the CMD ["handler.handler"]

Serverless framework configuration

assifierCatego	ories:
name:	<pre>\${self:custom.appPrefix}-classifierCategories</pre>
image:	<pre>\${self:custom.spreaker.\${opt:stage}.accountId}.dkr.ecr.</pre>
us-east-1.amaz	<pre>conaws.com/vn-pdn-ad-quality-classifier-categories:\$</pre>
<pre>{param:commit\</pre>	/ersion}
architecture:	x86_64
memorySize:	6144
timeout:	<pre>\${self:custom.classifierTimeout}</pre>
disableLogs:	true

Proactive approach: scan ads' inventory

The search engine

- Dedicated team to find out categorization issues on a daily basis through QuickSight
- Simplified workflow given by the ML inferences

	Ad language	Provider	
	en-GB, en-US	▼ All	
eturned	IAB category inferred		
	IAB11-4	•	

IAB cat. returned	ad_iab_ca	IAB cat.inferred	ad_text	
null	IAB11-4	IAB11-4	american needs reliability roads to keep supplies r	moving ene
IAB11-2	IAB11-4	IAB11-4	amer american needs reliability roads to keep supplies	moving 1e
null	IAB11-4	IAB11-4	energy to power every moment and we need that now. hundreds of new infrastructure projects are	reliability. awaiting 1e
			government approval projects that will supply tho jobs and improve american lives, but new red tape white house could create years of backlog or wor	usands of a from the se project

cancelations. americans can't wait any longer to build back better. we need progress now for better roads, more affordable energy and a reliable supply chain. help break the gridlock. call the white house 2024561111. tell president biden we need him to cut the government red tape and get america moving again to to d 4561111 paid for by the

american highway users alliance.

Proactive approach: fix ads' inventory

The manual categorization

- Manual categorization in the Spreaker Admin application
- Ads metadata database in DynamoDB to share information

Fix categorization: in real time inside Spreaker platform

The internal blocking process

- We merge IAB categories from external ad sources and from M.A.G.D.A
- We discard the ad in case the request wants to block one of the IAB categories
- The block works only for ads already appeared in our platform at least once



The results

BEFORE M.A.G.D.A

Proactive approach

Impossible to apply

Reports received

multiple reports per day during 2020 US election

Lead time

~3 hours to find the ad ? time to be blocked by the SSP

Operational cost

2-3 person X 3 hours

WITH M.A.G.D.A

Proactive approach

~30 minutes per day

Reports received

1 report during 2022 US elections period

Lead time

~5 minutes to find and block internally ? time to be blocked by the SSP

Operational cost

1 person X 5 minutes

Thank you for listening

Matteo Rossi matteo.rossi@spreaker.com







GRAZIE