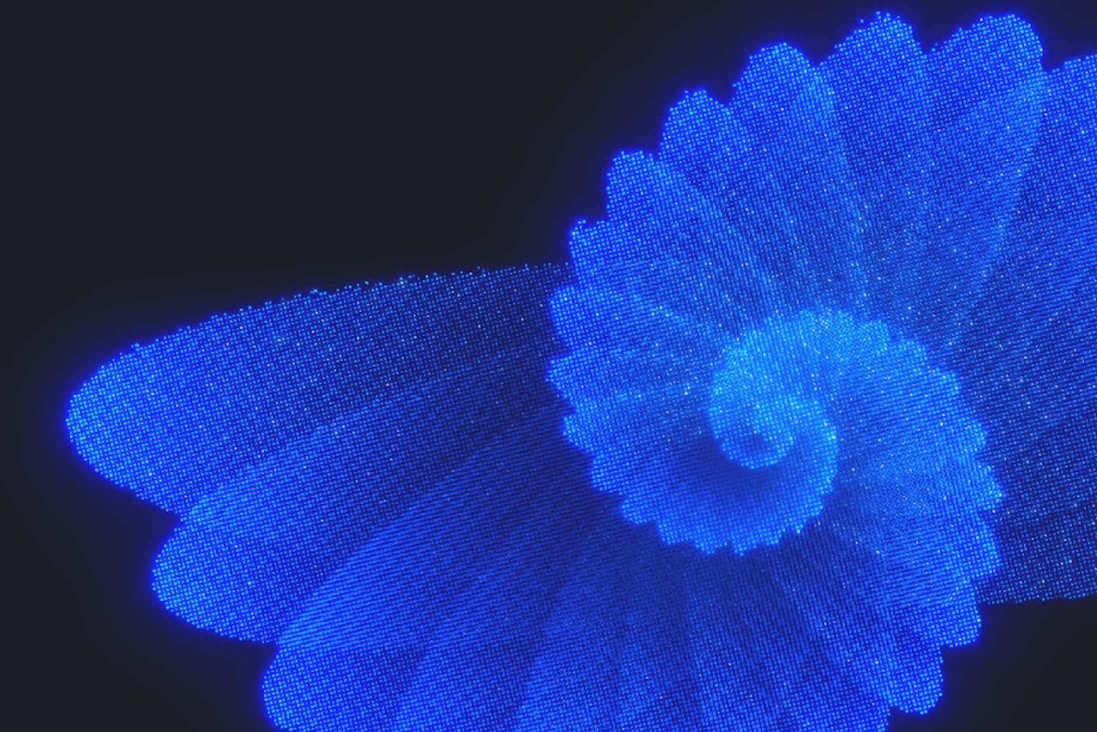
Toloka

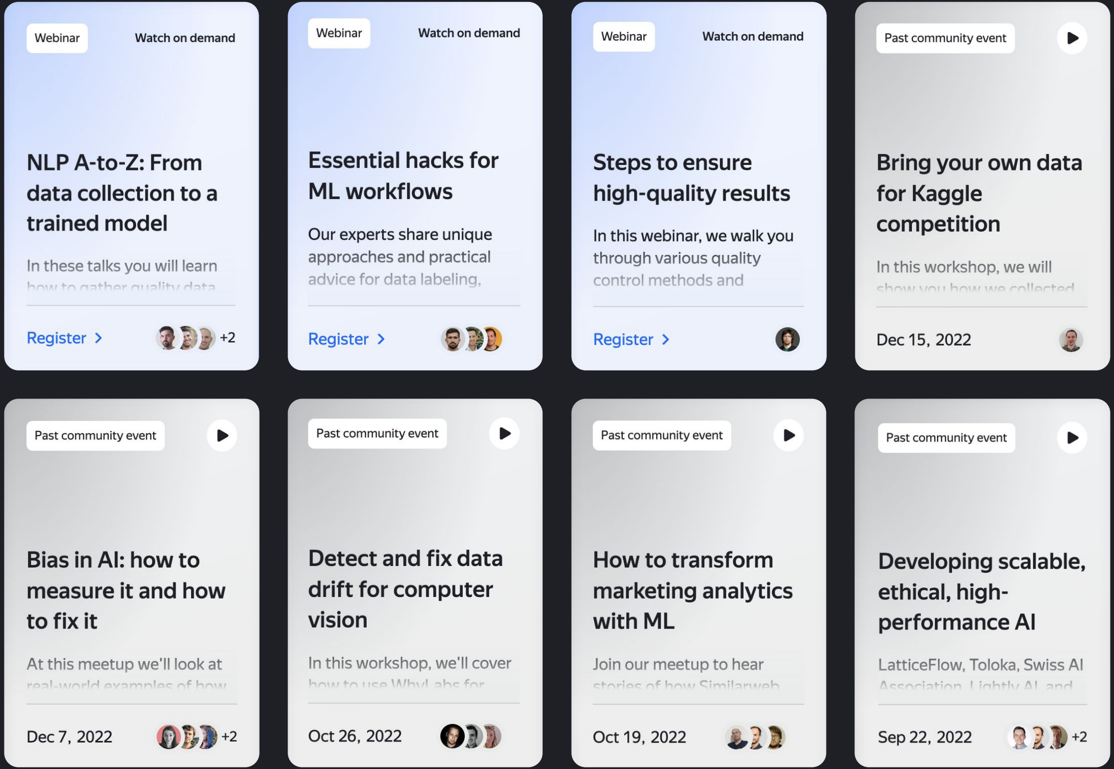
Taboola AliExpress Booking.com

Data-Driven AI meetup

Content moderation
with humans and ML



Explore data-driven world with us



Webinar
Watch on demand
NLP A-to-Z: From data collection to a trained model
In these talks you will learn
how to gather quality data
Register >
+2

Webinar
Watch on demand
Essential hacks for ML workflows
Our experts share unique approaches and practical advice for data labeling,
Register >

Webinar
Watch on demand
Steps to ensure high-quality results
In this webinar, we walk you through various quality control methods and
Register >

Past community event
Bring your own data for Kaggle competition
In this workshop, we will
show you how we collected
Dec 15, 2022

Past community event
Bias in AI: how to measure it and how to fix it
At this meetup we'll look at
real-world examples of how
Dec 7, 2022
+2

Past community event
Detect and fix data drift for computer vision
In this workshop, we'll cover
how to use WhyLabs for
Oct 26, 2022

Past community event
How to transform marketing analytics with ML
Join our meetup to hear
stories of how Similarweb
Oct 19, 2022

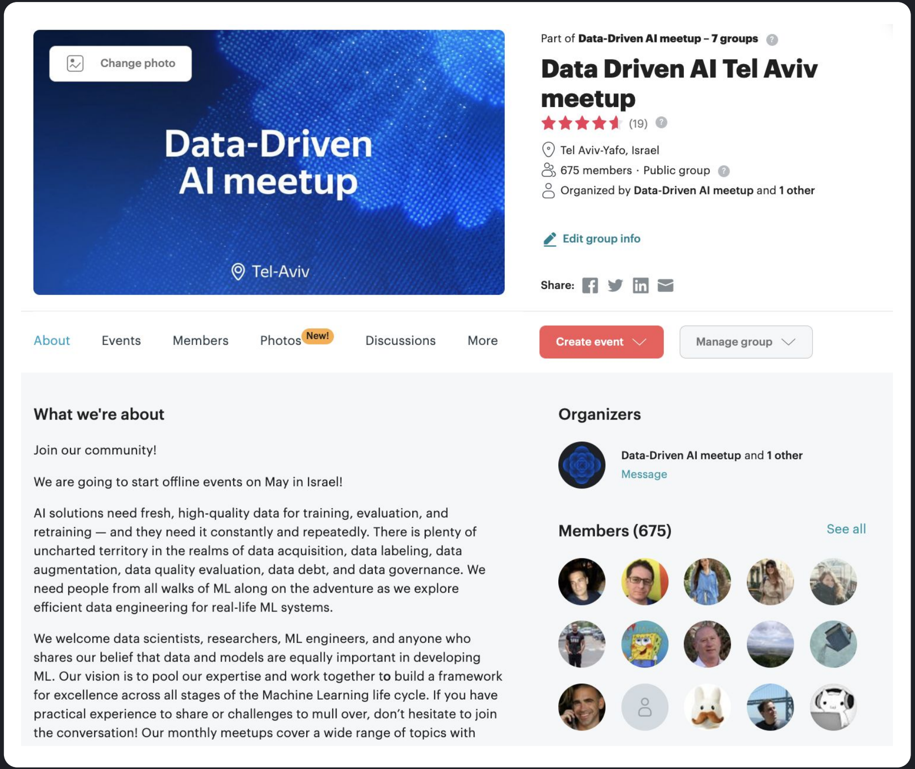
Past community event
Developing scalable, ethical, high-performance AI
LatticeFlow, Toloka, Swiss AI
Association, Lightly AI, and
Sep 22, 2022
+2

Enter your email
Join

https://toloka.ai/community/

slack

Explore data-driven events in Berlin


Change photo
Data-Driven AI meetup
Tel-Aviv
Edit group info
Share:
About
Events
Members
Photos New!
Discussions
More
Create event
Manage group
See all
Message

Part of **Data-Driven AI meetup – 7 groups**

Data Driven AI Tel Aviv meetup

★★★★☆ (19)

Tel Aviv-Yafo, Israel

675 members · Public group

Organized by **Data-Driven AI meetup** and **1 other**

What we're about

Join our community!

We are going to start offline events on May in Israel!

AI solutions need fresh, high-quality data for training, evaluation, and retraining — and they need it constantly and repeatedly. There is plenty of uncharted territory in the realms of data acquisition, data labeling, data augmentation, data quality evaluation, data debt, and data governance. We need people from all walks of ML along on the adventure as we explore efficient data engineering for real-life ML systems.

We welcome data scientists, researchers, ML engineers, and anyone who shares our belief that data and models are equally important in developing ML. Our vision is to pool our expertise and work together to build a framework for excellence across all stages of the Machine Learning life cycle. If you have practical experience to share or challenges to mull over, don't hesitate to join the conversation! Our monthly meetups cover a wide range of topics with

Organizers

Data-Driven AI meetup and **1 other**

Members (675)

meetup

https://www.meetup.com/meetup-group-data-driven/

Agenda

18:30 – 19:00 — Item moderation and data multi-labeling for e-commerce, Elena Gruntova, Product Director, AliExpress

19:00 – 19:30 — AI content moderation: the truth is out there Gal Cohen, Product Manager, Taboola

19:30 – 19:45 — Break + networking

19:40 – 20:30 — Panel discussion about moderation with Aliexpress, Booking.com, Taboola and Toloka

Ask questions = receive answers and gifts from our community

Powering AI. Empowering people

Olga Megorskaya

CEO

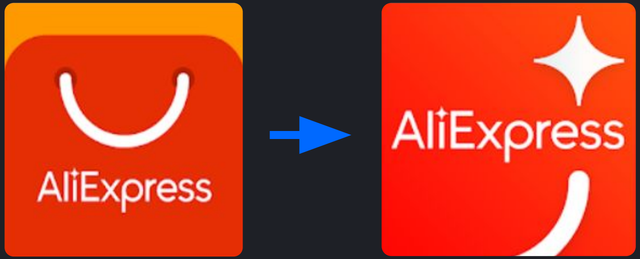
AliExpress

Item moderation and data multi-labeling for multinational e-commerce

Elena Gruntova, CPO at AliExpress CIS

AliExpress CIS

- Launched in 2010 as a B2B, later B2C marketplace by Alibaba Group
- In Sept 2019 Aliexpress CIS countries became a JV, Alibaba Group & local investors. Product & development localization



Previously CIS main Aliexpress app was powered by Alibaba. From May 23d 2022 AliExpress CIS got new official app, and it's locally developed.

FY 2021

GMV

306 bln ₽ +46% growth

total GMV
excluding services

110 bln ₽ +124% growth

local business GMV
excluding services

The assortment

211 mln the total number of SPU — +52% growth

16,5 mln the number of local SPU — +184% growth

3,2 bln the total number of SKU — +48% growth

The sellers

400K the total number of sellers — +69% growth

102,5K the number of local sellers — +193% growth

The audience

28,7 mln the number of buyers

80,2 mln the number of unique users for the year

35 mln MAU

12,5 mln DAU

Reasons to localize item moderation

Local legal regulations

- Devices for secretly obtaining information
- Goods without certification
- Weapons
- Prohibited substances
- Other banned products

Cultural differences

- Health-related matters
- Body measurement
- Open body demonstration limits
- Culturally or politically charged images

Classification

OK

Product without display restrictions. These products can be sold and displayed everywhere

Restricted

The product is restricted to display. Can only be shown upon direct request

Adult

Adult products related to sex. We separate it into a separate category since we mean a separate product logic for these products (blair, age confirmation, display only on requests)

Prohibited

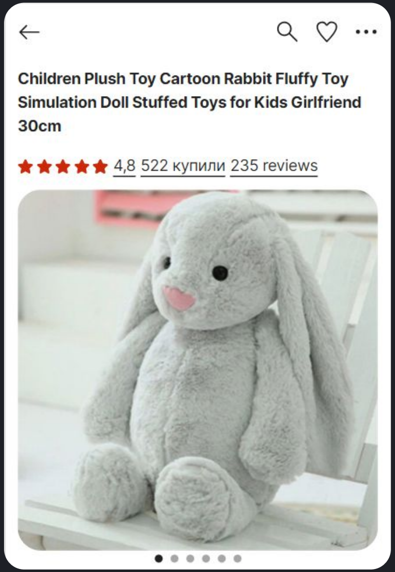
Products that are prohibited for sale according to the laws of the Russian Federation & CIS or company policy. Can't be shown anywhere and sold at all

Suspect to Fraud

Products with unreasonably high or unreasonably low price, or product without clear description

Classification: examples

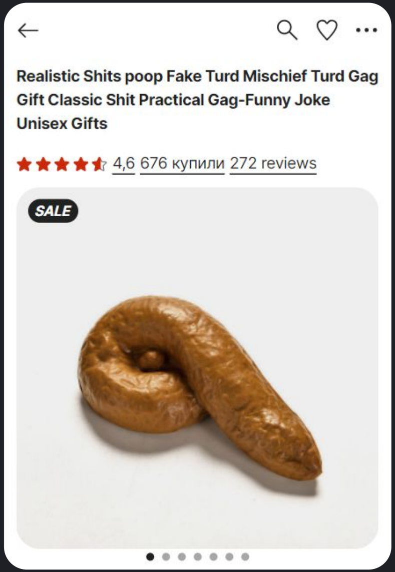
✓



Children Plush Toy Cartoon Rabbit Fluffy Toy Simulation Doll Stuffed Toys for Kids Girlfriend 30cm

★★★★★ 4,8 522 купили 235 reviews

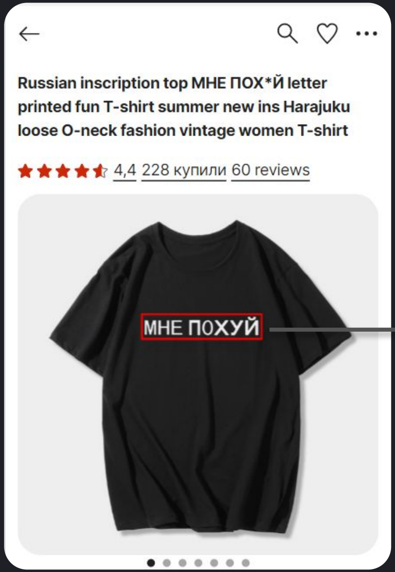
!



Realistic Shits poop Fake Turd Mischief Turd Gag Gift Classic Shit Practical Gag-Funny Joke Unisex Gifts

★★★★☆ 4,6 676 купили 272 reviews

!

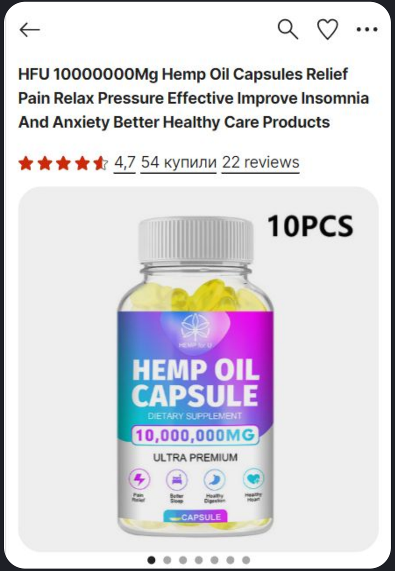


Russian inscription top МНЕ ПОХ*Й letter printed fun T-shirt summer new ins Harajuku loose O-neck fashion vintage women T-shirt

★★★★☆ 4,4 228 купили 60 reviews

I don’t give a shit

Classification: examples

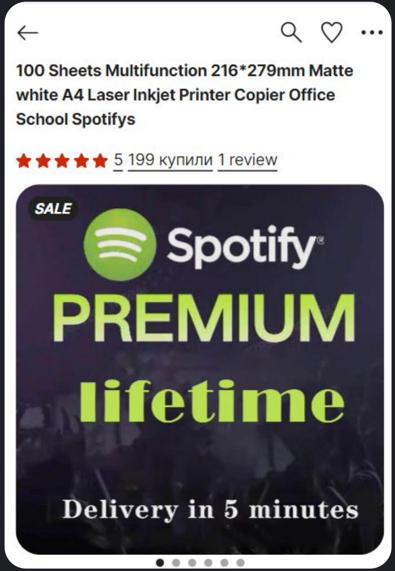


HFU 10000000Mg Hemp Oil Capsules Relief Pain Relax Pressure Effective Improve Insomnia And Anxiety Better Healthy Care Products

★★★★★ 4,7 54 купили 22 reviews

10PCS
HEMP OIL CAPSULE
DIETARY SUPPLEMENT
10,000,000MG
ULTRA PREMIUM
CAPSULE

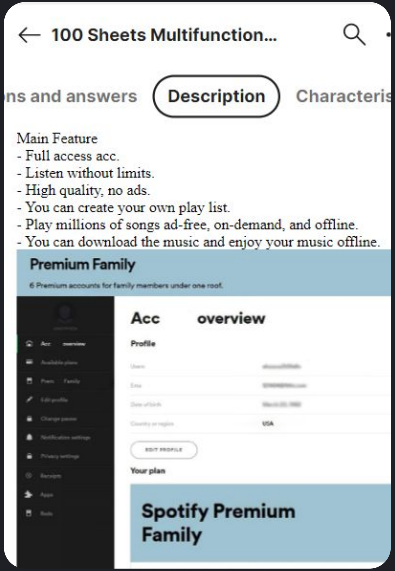
Product description specifies illegal substances



100 Sheets Multifunction 216*279mm Matte white A4 Laser Inkjet Printer Copier Office School Spotifys

★★★★★ 5 199 купили 1 review

SALE
Spotify
PREMIUM
lifetime
Delivery in 5 minutes



100 Sheets Multifunction...
ns and answers
Description
Characteris

Main Feature
- Full access acc.
- Listen without limits.
- High quality, no ads.
- You can create your own play list.
- Play millions of songs ad-free, on-demand, and offline.
- You can download the music and enjoy your music offline.

Premium Family
6 Premium accounts for family members under one roof.
Acc overview
Profile
USA
EDIT PROFILE
Your plan
Spotify Premium Family

Title states it's a legal product but Q&A shows otherwise.
And subscription can be sold by the service itself only

Items prohibited for sale

✓ Criminal activity

✓ Prohibited weapons

✓ Smoking related items

✓ Radical politicsl

✓ Prescribed drugs

✓ Alcohol

✓ Not for sale items, test items

✓ Contains dangerous substances (toxic, radiation hazard…)

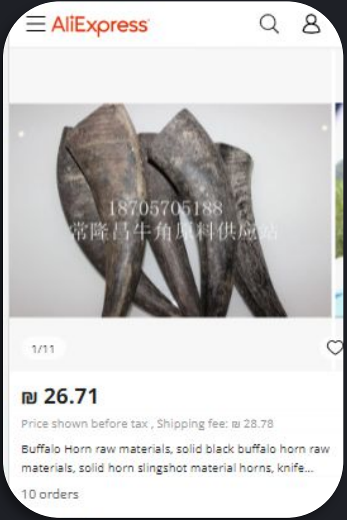
✓ Porno, sexual violence

✓ Animals, plants, illegal hunting and fishing

✓ Certified services

✓ IP

✓ Illegal drugs

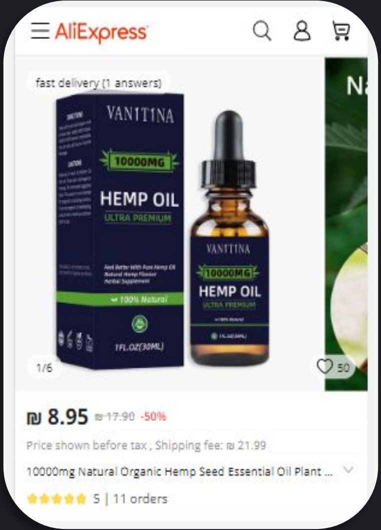

AliExpress
18705705188
常隆昌牛角原料供应站
1/11

₪ 26.71

Price shown before tax , Shipping fee: ₪ 28.78

Buffalo Horn raw materials, solid black buffalo horn raw materials, solid horn slingshot material horns, knife...

10 orders

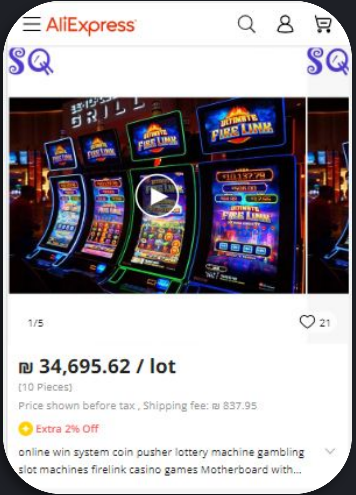

AliExpress
fast delivery (1 answers)
VANITINA
10000MG
HEMP OIL
ULTRA PREMIUM
1FL.OZ(30ML)
1/6
50

₪ 8.95 ~~₪ 17.90~~ -50%

Price shown before tax , Shipping fee: ₪ 21.99

10000mg Natural Organic Hemp Seed Essential Oil Plant ...

★★★★★ 5 | 11 orders


AliExpress
SQ
FIRE LINK
1/5
21

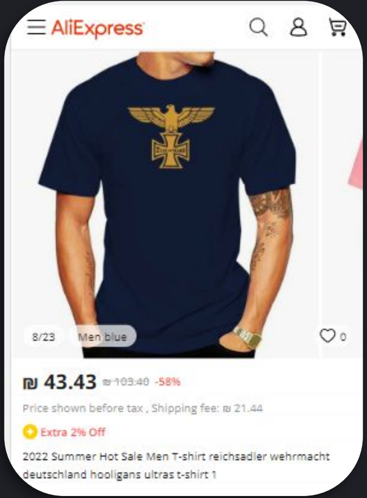
₪ 34,695.62 / lot

(10 Pieces)

Price shown before tax , Shipping fee: ₪ 837.95

Extra 2% Off

online win system coin pusher lottery machine gambling slot machines firelink casino games Motherboard with...


AliExpress
8/23
Men blue
0

₪ 43.43 ~~₪ 103.40~~ -58%

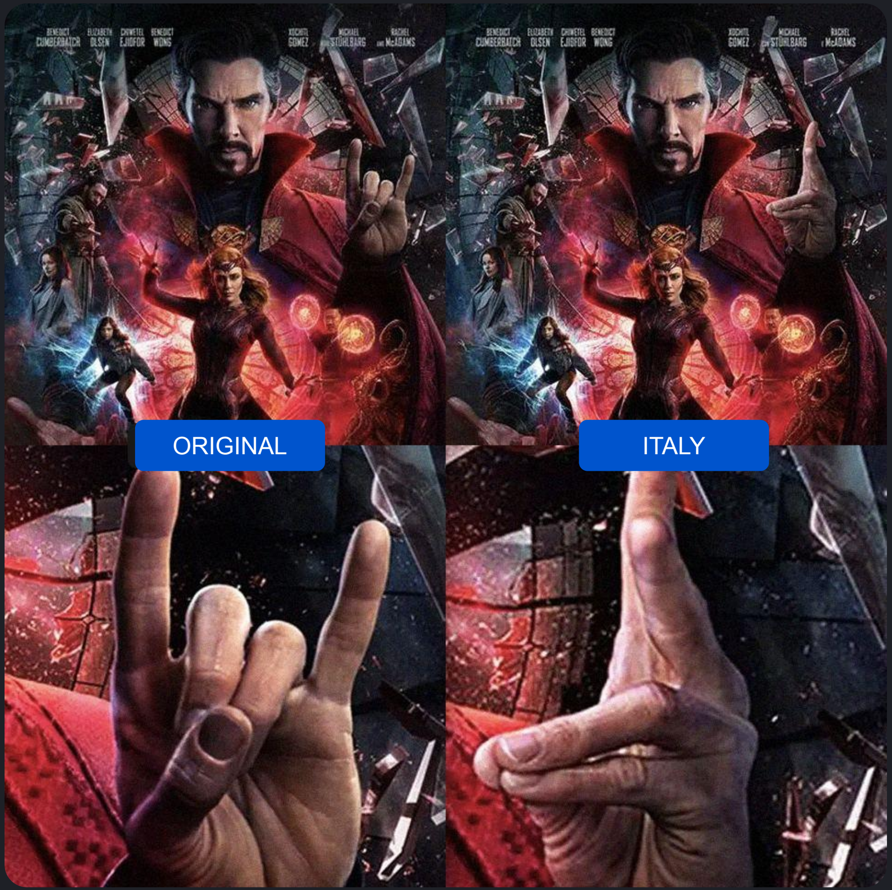
Price shown before tax , Shipping fee: ₪ 21.44

Extra 2% Off

2022 Summer Hot Sale Men T-shirt reichsadler wehrmacht deutschland hooligans ultras t-shirt 1

Cultural differences

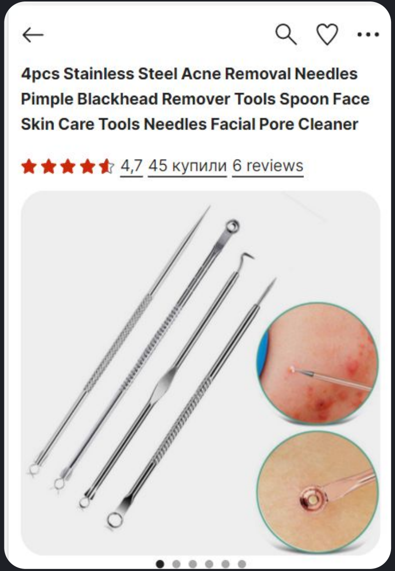
- Integrated and maintained system of socially acquired values, beliefs, and rules of conduct
- Distinguishable from one societal group to another
- Which impact the range of accepted behaviors



* Jackson A., Guerra N.S. (2011) Cultural Difference. In: Goldstein S., Naglieri J.A. (eds) Encyclopedia of Child Behavior and Development. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-79061-9_752

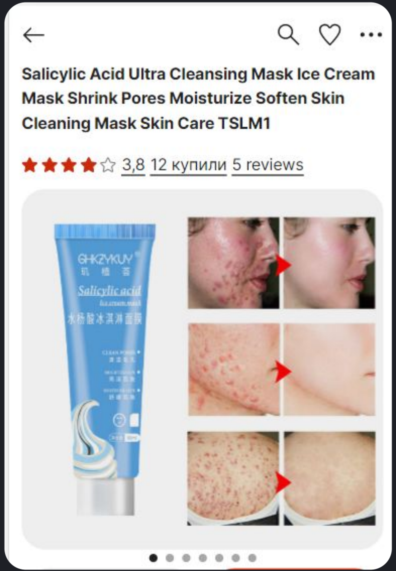
Cultural differences: main topics

Medicine



4pcs Stainless Steel Acne Removal Needles Pimple Blackhead Remover Tools Spoon Face Skin Care Tools Needles Facial Pore Cleaner

4,7 45 купили 6 reviews

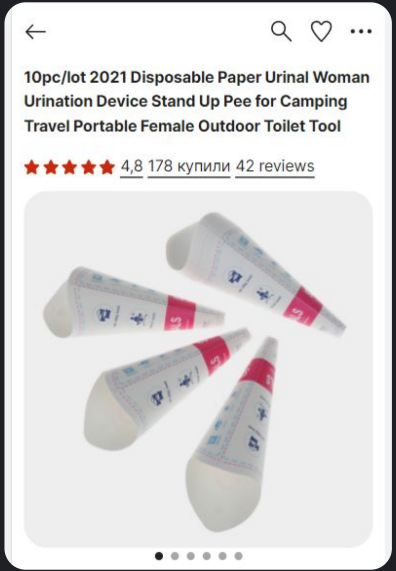


Salicylic Acid Ultra Cleansing Mask Ice Cream Mask Shrink Pores Moisturize Soften Skin Cleaning Mask Skin Care TSLM1

3,8 12 купили 5 reviews

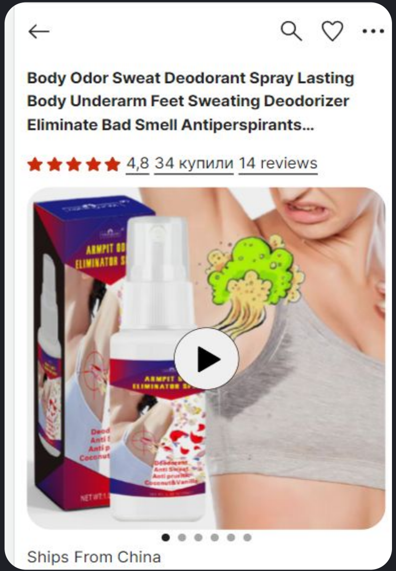
🇨🇳 ✓ !

Personal hygiene



10pc/lot 2021 Disposable Paper Urinal Woman Urination Device Stand Up Pee for Camping Travel Portable Female Outdoor Toilet Tool

4,8 178 купили 42 reviews



Body Odor Sweat Deodorant Spray Lasting Body Underarm Feet Sweating Deodorizer Eliminate Bad Smell Antiperspirants...

4,8 34 купили 14 reviews

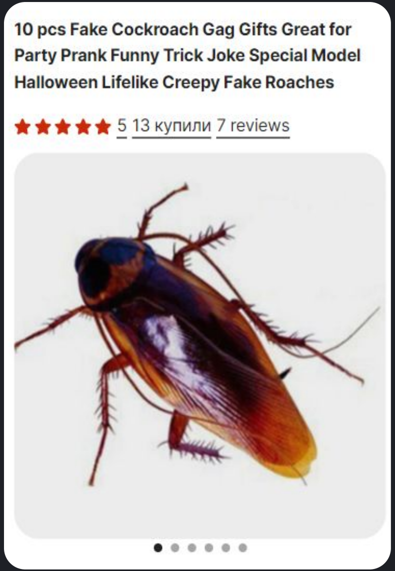
Ships From China

🇨🇳 ✓ !

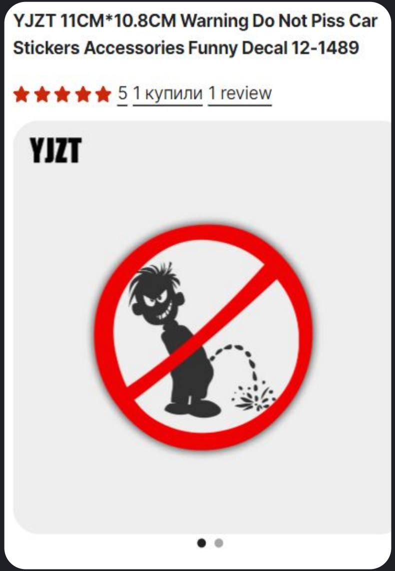
Toloka AliExpress

Cultural differences: more examples

Jokes


10 pcs Fake Cockroach Gag Gifts Great for Party Prank Funny Trick Joke Special Model Halloween Lifelike Creepy Fake Roaches

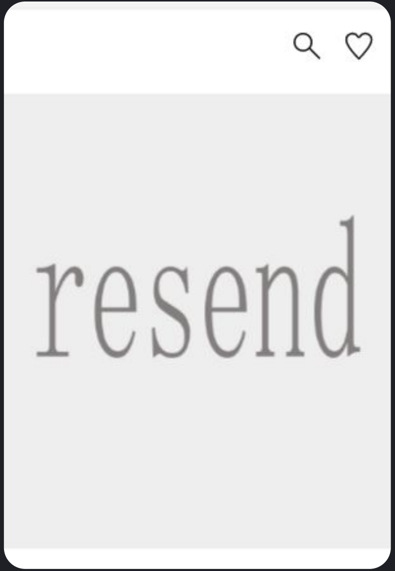
★★★★★ 5 13 купили 7 reviews

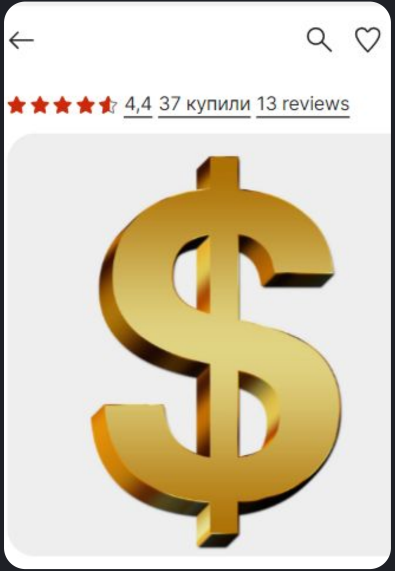

YJZT 11CM*10.8CM Warning Do Not Piss Car Stickers Accessories Funny Decal 12-1489

★★★★★ 5 1 купили 1 review

🇨🇳 ✓ !

Suspicion

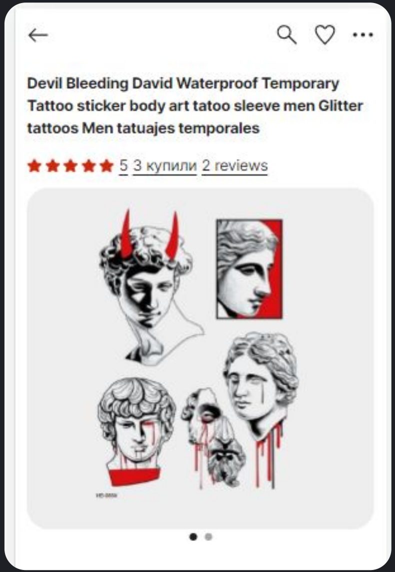



★★★★☆ 4,4 37 купили 13 reviews

🇨🇳 ✓

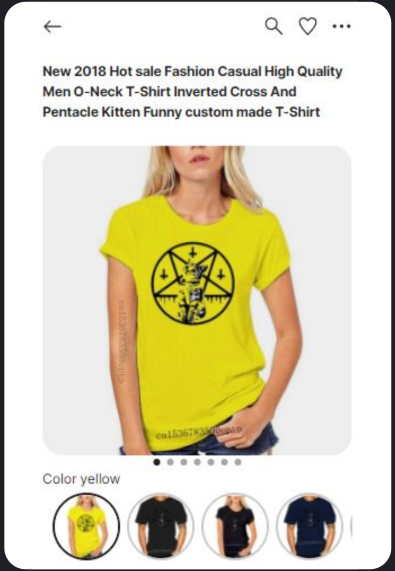
Cultural differences: cultural context

Cultural context



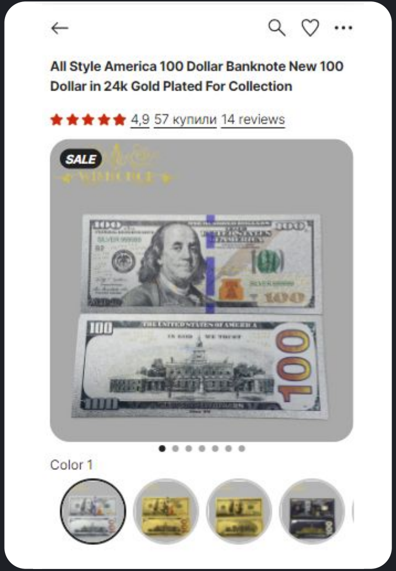
Devil Bleeding David Waterproof Temporary Tattoo sticker body art tatoo sleeve men Glitter tattoos Men tatuajes temporales

★★★★★ 5 3 купили 2 reviews



New 2018 Hot sale Fashion Casual High Quality Men O-Neck T-Shirt Inverted Cross And Pentacle Kitten Funny custom made T-Shirt

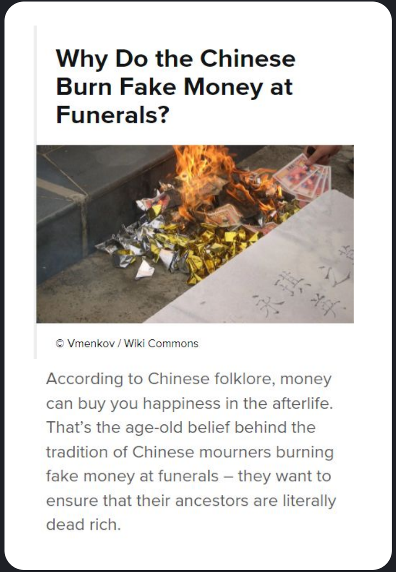
Color yellow



All Style America 100 Dollar Banknote New 100 Dollar in 24k Gold Plated For Collection

★★★★★ 4,9 57 купили 14 reviews

Color 1



Why Do the Chinese Burn Fake Money at Funerals?

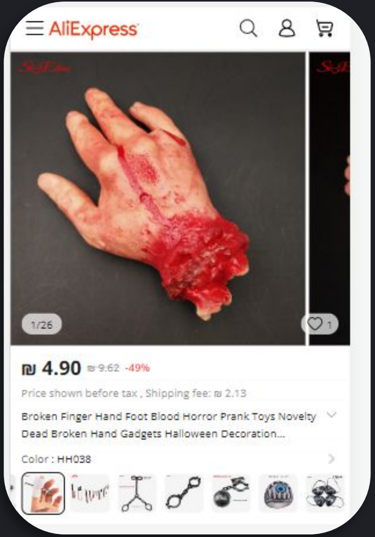
© Vmenkov / Wiki Commons

According to Chinese folklore, money can buy you happiness in the afterlife. That's the age-old belief behind the tradition of Chinese mourners burning fake money at funerals – they want to ensure that their ancestors are literally dead rich.

✓ !

Items we hide: only direct link or query

- ✓ Disgusting
- ✓ Insulting
- ✓ Obscene language
- ✓ Unreasonable attraction of attention to the product
- ✓ Realistic depiction of disease and injury

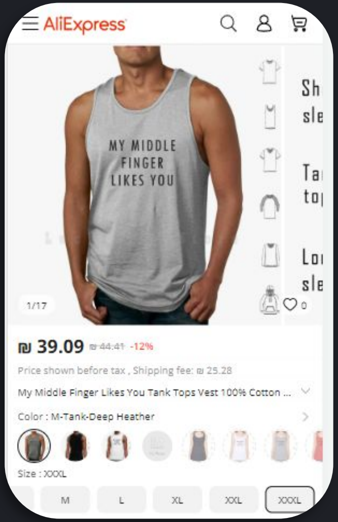

AliExpress
1/26
1

₪ 4.90 ~~₪ 9.62~~ -49%

Price shown before tax , Shipping fee: ₪ 2.13

Broken Finger Hand Foot Blood Horror Prank Toys Novelty Dead Broken Hand Gadgets Halloween Decoration...

Color : HH038


AliExpress
MY MIDDLE FINGER LIKES YOU
1/17
0

₪ 39.09 ~~₪ 44.41~~ -12%

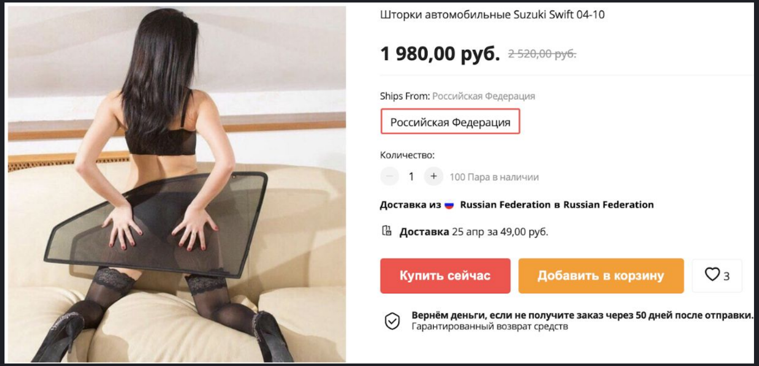
Price shown before tax , Shipping fee: ₪ 25.28

My Middle Finger Likes You Tank Tops Vest 100% Cotton ...

Color : M-Tank-Deep Heather

Size : XXXL

M
L
XL
XXL
XXXL



Шторки автомобильные Suzuki Swift 04-10

1 980,00 руб. ~~2 520,00 руб.~~

Ships From: Российская Федерация

Российская Федерация

Количество:

1 100 Пара в наличии

Доставка из Russian Federation в Russian Federation

Доставка 25 апр за 49,00 руб.

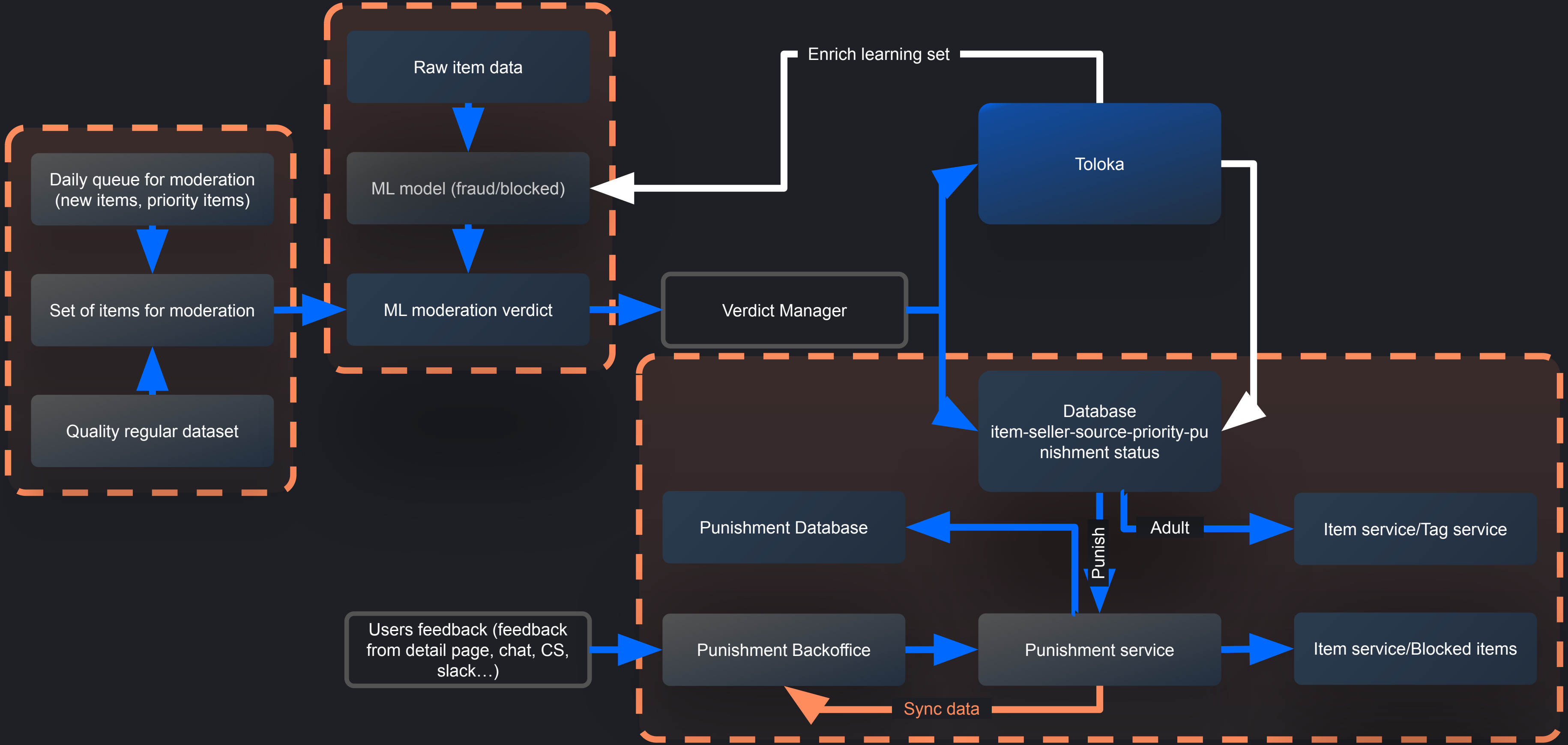
Купить сейчас
Добавить в корзину
3

Вернём деньги, если не получите заказ через 50 дней после отправки.
Гарантированный возврат средств

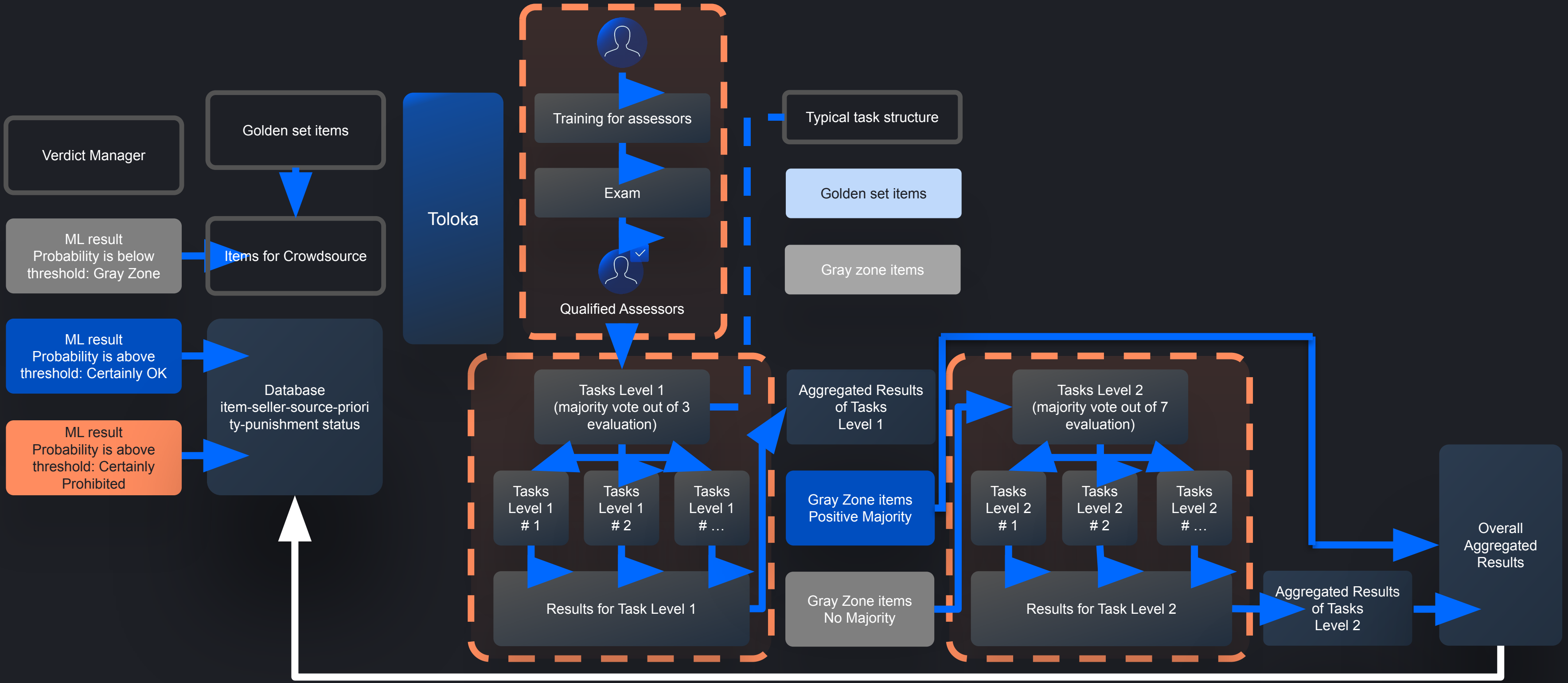
Moderation metrics for product subsets

	Advertising feed	New products	Top selling items	Top GMV items	Items from SERPs for 1k most popular search queries
✓ OK	99.55%	96.76%	97.94%	96.15%	92.87%
! Restricted	0.10%	0.14%	0.57%	1.53%	0.11%
18+ Adult	0.15%	0.58%	1.13%	0.85%	5.93%
✗ Prohibited	0.04%	0.04%	0.11%	1.02%	0.10%
404 Error	0.16%	2.48%	0.25%	0.45%	1.01%

Pipeline for whole project


Raw item data
ML model (fraud/blocked)
ML moderation verdict
Enrich learning set
Toloka
Daily queue for moderation (new items, priority items)
Set of items for moderation
Quality regular dataset
Verdict Manager
Database item-seller-source-priority-punishment status
Punishment Database
Punish
Adult
Item service/Tag service
Users feedback (feedback from detail page, chat, CS, slack…)
Punishment Backoffice
Punishment service
Item service/Blocked items
Sync data

Pipeline for labeling


Verdict Manager

Golden set items

Items for Crowdsource

ML result
Probability is below
threshold: Gray Zone

ML result
Probability is above
threshold: Certainly OK

ML result
Probability is above
threshold: Certainly
Prohibited

Database
item-seller-source-priority-punishment status

Toloka

Training for assessors

Exam

Qualified Assessors

Typical task structure

Golden set items

Gray zone items

Tasks Level 1
(majority vote out of 3
evaluation)

Tasks
Level 1
1

Tasks
Level 1
2

Tasks
Level 1
…

Results for Task Level 1

Aggregated Results
of Tasks
Level 1

Gray Zone items
Positive Majority

Gray Zone items
No Majority

Tasks Level 2
(majority vote out of 7
evaluation)

Tasks
Level 2
1

Tasks
Level 2
2

Tasks
Level 2
…

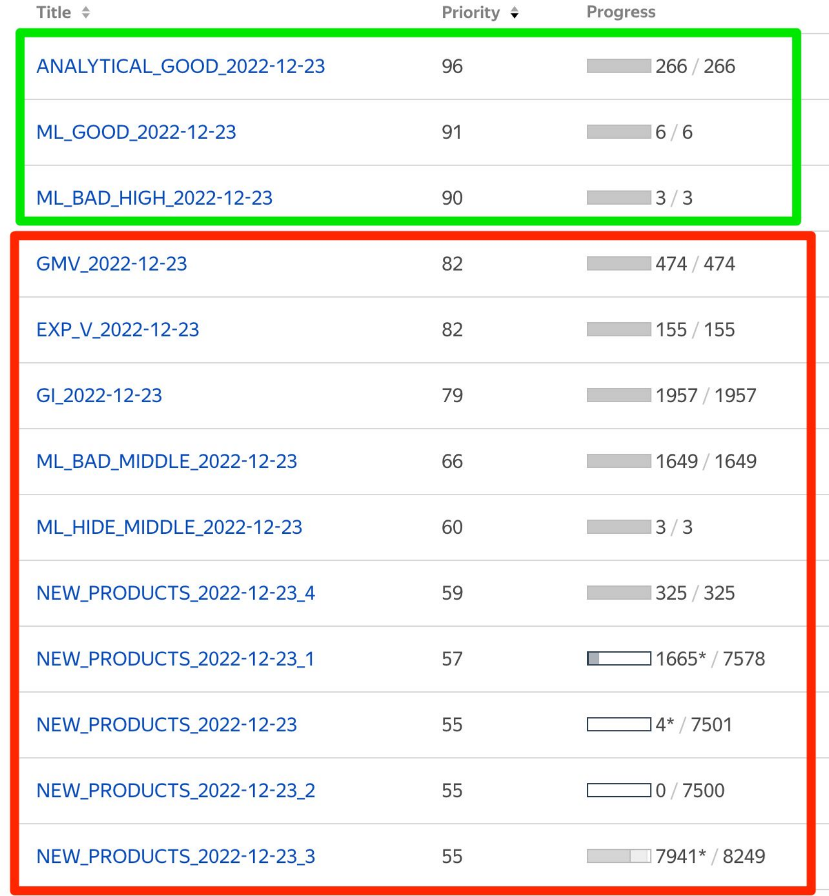
Results for Task Level 2

Aggregated Results
of Tasks
Level 2

Overall
Aggregated
Results

More details on daily queue

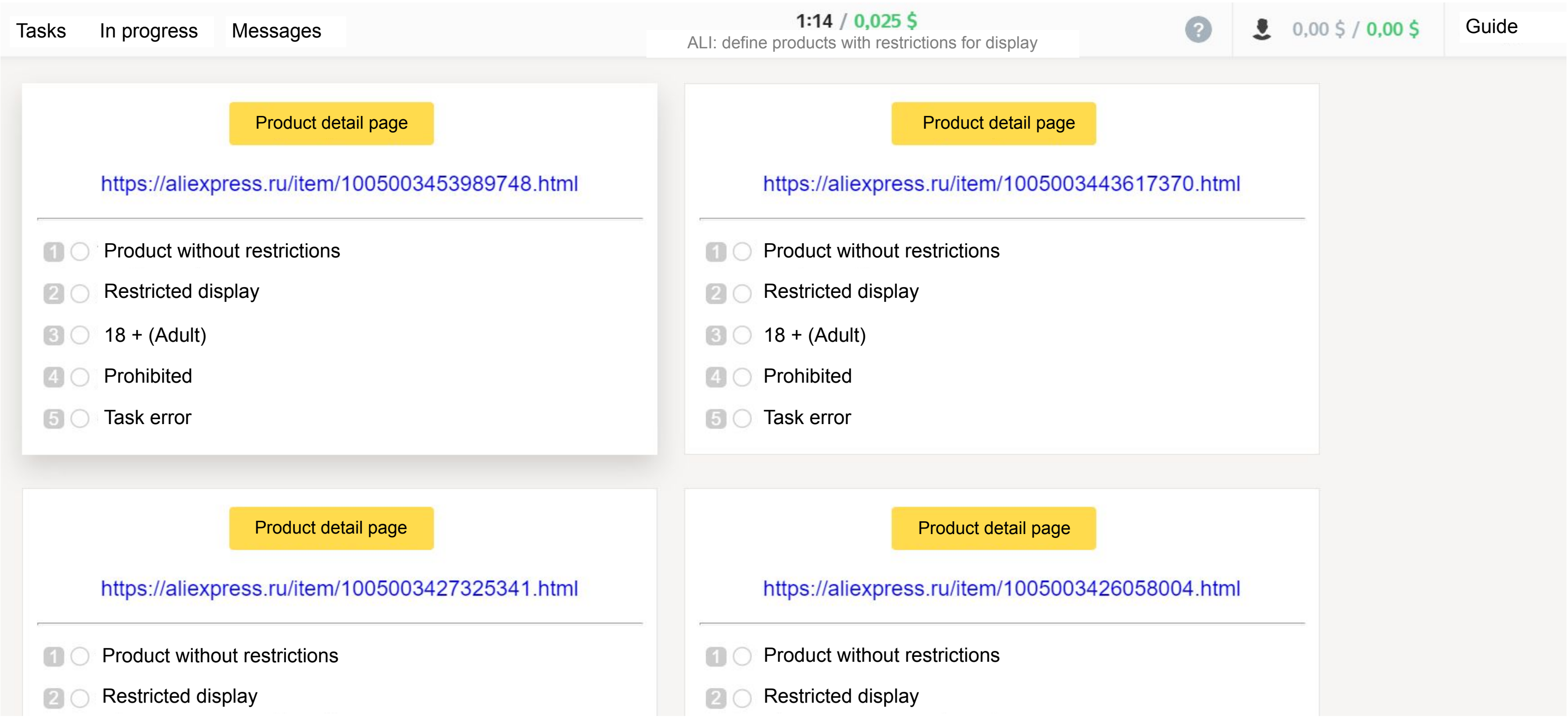
- Product priorities: new products, top items by exposure, top items by GMV, Golden items
- Analytical inputs: customers behaviour, item performance, regular features…
- ML inputs within thresholds



Title	Priority	Progress
ANALYTICAL_GOOD_2022-12-23	96	266 / 266
ML_GOOD_2022-12-23	91	6 / 6
ML_BAD_HIGH_2022-12-23	90	3 / 3
GMV_2022-12-23	82	474 / 474
EXP_V_2022-12-23	82	155 / 155
GI_2022-12-23	79	1957 / 1957
ML_BAD_MIDDLE_2022-12-23	66	1649 / 1649
ML_HIDE_MIDDLE_2022-12-23	60	3 / 3
NEW_PRODUCTS_2022-12-23_4	59	325 / 325
NEW_PRODUCTS_2022-12-23_1	57	1665* / 7578
NEW_PRODUCTS_2022-12-23	55	4* / 7501
NEW_PRODUCTS_2022-12-23_2	55	0 / 7500
NEW_PRODUCTS_2022-12-23_3	55	7941* / 8249

* Jackson A., Guerra N.S. (2011) Cultural Difference. In: Goldstein S., Naglieri J.A. (eds) Encyclopedia of Child Behavior and Development. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-79061-9_752

Specific task from Ali Moderation project in Toloka



ALI: define products with restrictions for display

Product detail page

https://aliexpress.ru/item/1005003453989748.html

- 1 ○ Product without restrictions
- 2 ○ Restricted display
- 3 ○ 18 + (Adult)
- 4 ○ Prohibited
- 5 ○ Task error

Product detail page

https://aliexpress.ru/item/1005003443617370.html

- 1 ○ Product without restrictions
- 2 ○ Restricted display
- 3 ○ 18 + (Adult)
- 4 ○ Prohibited
- 5 ○ Task error

Product detail page

https://aliexpress.ru/item/1005003427325341.html

- 1 ○ Product without restrictions
- 2 ○ Restricted display

Product detail page

https://aliexpress.ru/item/1005003426058004.html

- 1 ○ Product without restrictions
- 2 ○ Restricted display

Toloka metrics at 6 months scale

1500 → 00
participants

200 → 0 000
SKU daily

0,017 → ,01
per SKU

15 sec
Avrg time per one SKU

~92% → .7%
Quality control

66.7%
Training quality

Key learnings

- ✓ General ML classification model aided by human evaluation is a powerful complex when big data tasks are at a stake
- ✓ Using effective and scalable crowdsource mechanics allows for weakly formalized domains evaluation (e.g. cultural differences), and it’s sensitive to current context and recent changes
- ✓ Moreover, these data can be transferred and efficiently performed by ML algorithm: ML can learn from human evaluation and scale results
- ✓ Human verdicts are used to get final conclusions, but also help improve ML performance – it’s a double action trigger

Thank you!

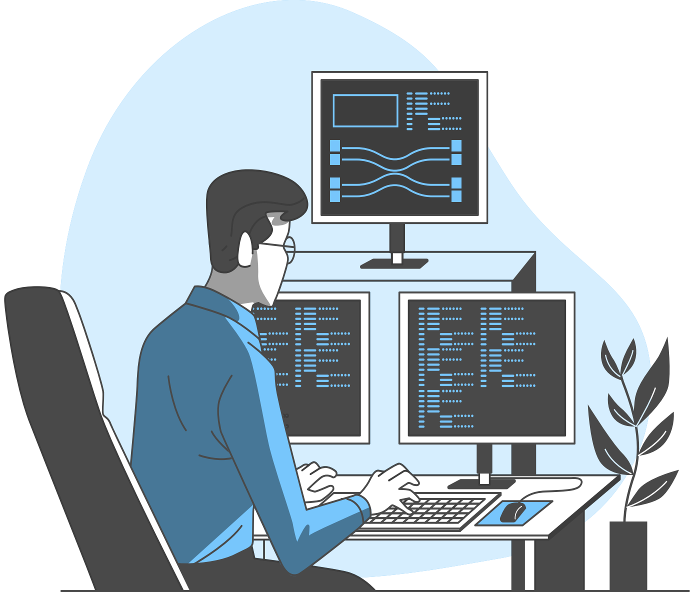
Elena Gruntova
CPO at AliExpress CIS

elena.gruntova@aliexpress.ru

Lena Gruntova
WhatsApp business account

AI Content Moderation

The Truth is Out There

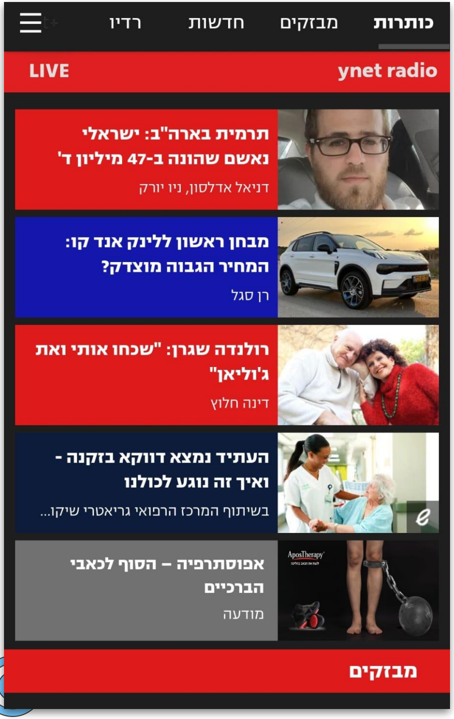


Gal Cohen
Product Manager
Taboola

NEW TEAM MEMBER
WELCOME!

Taboola

Publishers



תרמית בארה"ב: ישראלי נאשם שהונה ב-47 מיליון ד'
דניאל אדלסון, ניו יורק

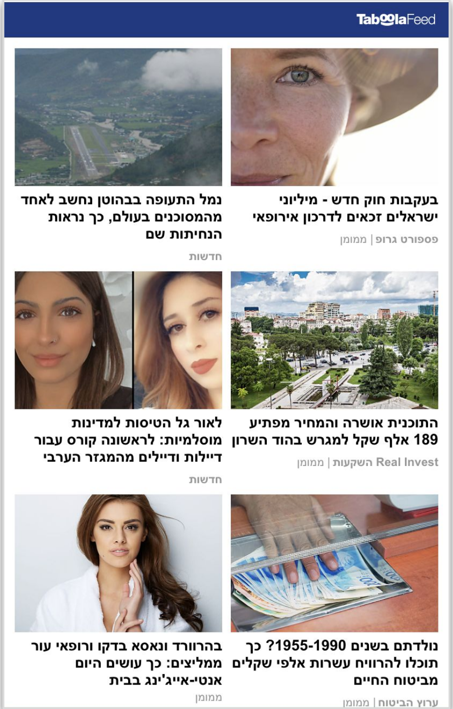
מבחן ראשון ללינק אנד קו: המחיר הגבוה מוצדק?
רן סגל

רולנדה שגרן: "שכחו אותי ואת ג'וליאן"
דינה חלוץ

העתיד נמצא דווקא בזקנה - ואיך זה נוגע לכולנו
בשיתוף המרכז הרפואי גריאטרי שיקו...

אפוסתרפיה – הסוף לכאבי הברכיים
מודעה

Advertisers



בעקבות חוק חדש - מיליוני ישראלים זכאים לדרכון אירופאי
פספורט גרופ | ממומן

נמל התעופה בבהוטן נחשב לאחד מהמסוכנים בעולם, כך נראות הנחיתות שם
חדשות

התוכנית אושרה והמחיר מפתיע 189 אלף שקל למגרש בהוד השרון
Real Invest השקעות | ממומן

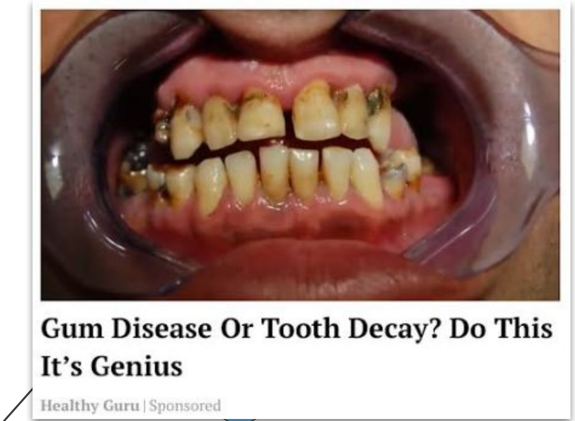
לאור גל הטיסות למדינות מוסלמיות: לראשונה קורס עבור דיילות ודיילים מהמגזר הערבי
חדשות

נולדתם בשנים 1955-1990? כך תוכלו להרוויח עשרות אלפי שקלים מביטוח החיים
ערוץ הביטוח | ממומן

בהרוורד ונאסא בדקו ורופאי עור ממליצים: כך עושים היום אנטי-אייג'ינג בבית
ממומן

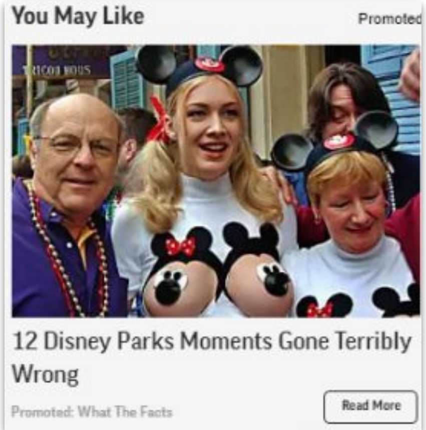
Goal #1:

Block non-compliant content



Gum Disease Or Tooth Decay? Do This It's Genius

Healthy Guru | Sponsored



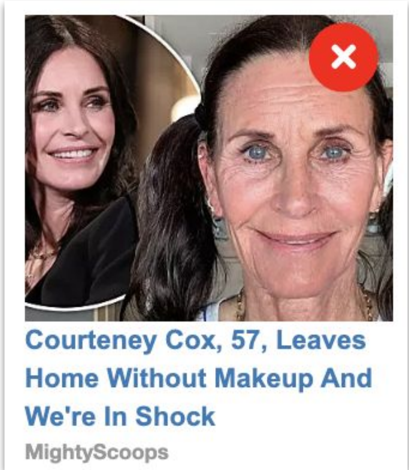
You May Like

Promoted

12 Disney Parks Moments Gone Terribly Wrong

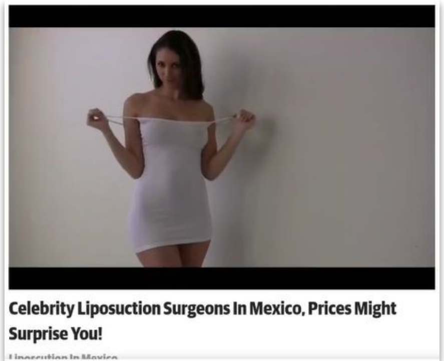
Promoted: What The Facts

Read More



Courteney Cox, 57, Leaves Home Without Makeup And We're In Shock

MightyScoops



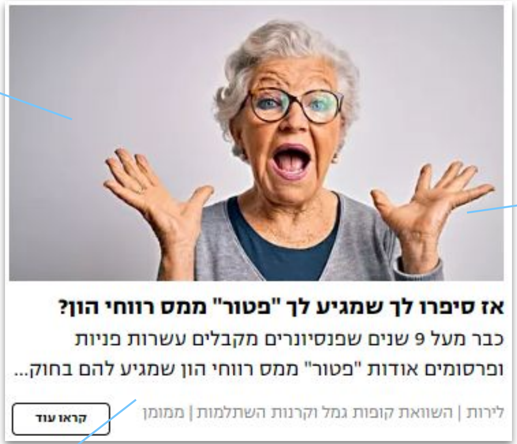
Celebrity Liposuction Surgeons In Mexico, Prices Might Surprise You!

Goal #2:

Label advertisement correctly

Quality level:

High



אז סיפרו לך שמגיע לך "פטור" ממס רווחי הון?

כבר מעל 9 שנים שפנסיונרים מקבלים עשרות פניות ופרסומים אודות "פטור" ממס רווחי הון שמגיע להם בחוק...

לירות | השוואת קופות גמל וקרנות השתלמות | ממומן

Category:

Investments

Advertisement type:

CTA

Here. We.

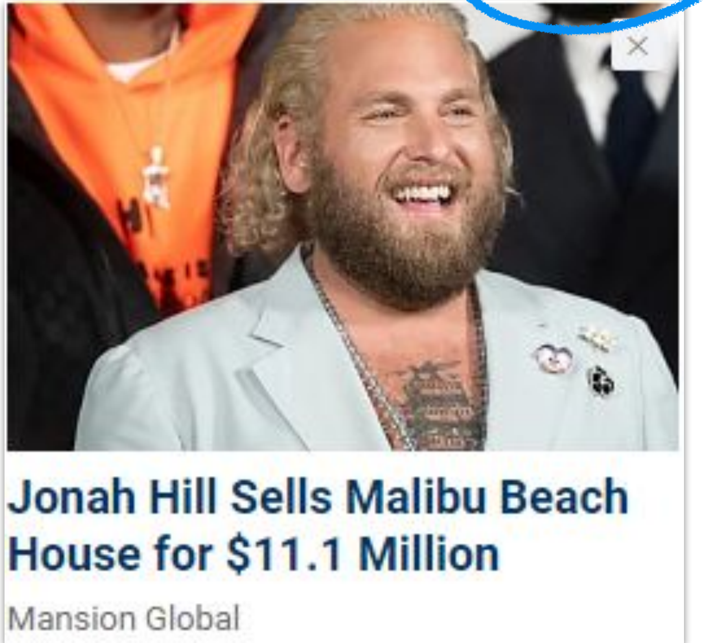
GO!

Content policy rule #1

The category of the ad must reflect its actual essence

Which category is it?

Real Estate or [✓ Gossip]



Jonah Hill Sells Malibu Beach House for $11.1 Million

Mansion Global

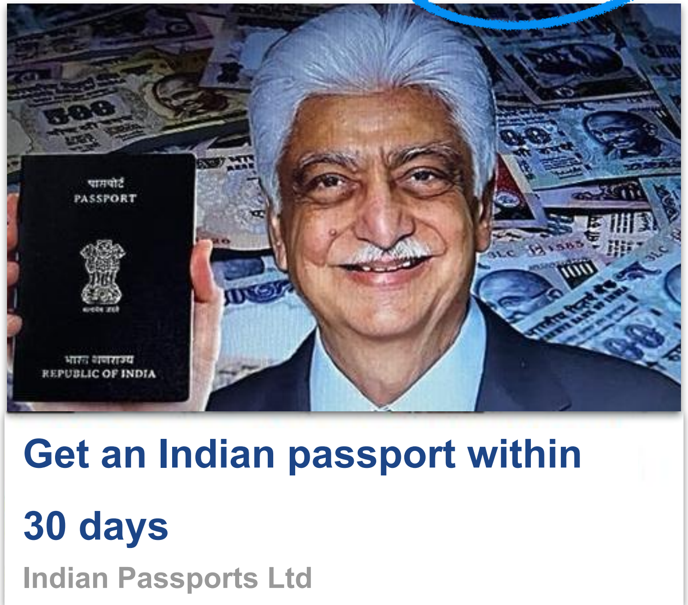
Human review challenge #1

Inconsistency

Content policy rule #2

Celebrities cannot advertise products, unless they are getting paid for it

Approve or Reject?


पासपोर्ट
PASSPORT
भारत गणराज्य
REPUBLIC OF INDIA

Get an Indian passport within 30 days

Indian Passports Ltd

Azim Premji

Human review challenge #2

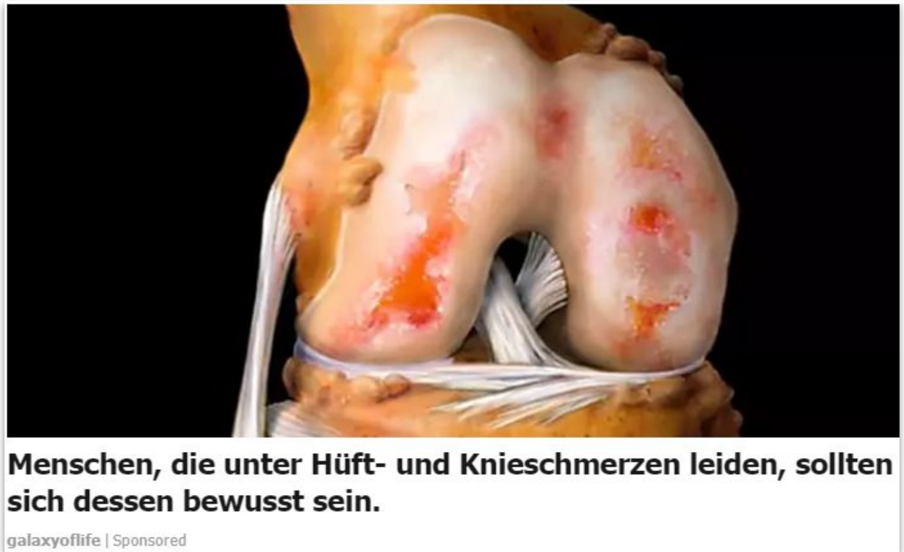
Localization

Content policy rule #3

Taboola does not allow any disturbing imagery

Approve or Reject?

$85k each month



Menschen, die unter Hüft- und Knieschmerzen leiden, sollten sich dessen bewusst sein.

galaxyoflife | Sponsored

Right answer: we need to consult with the business

Human review challenge #3

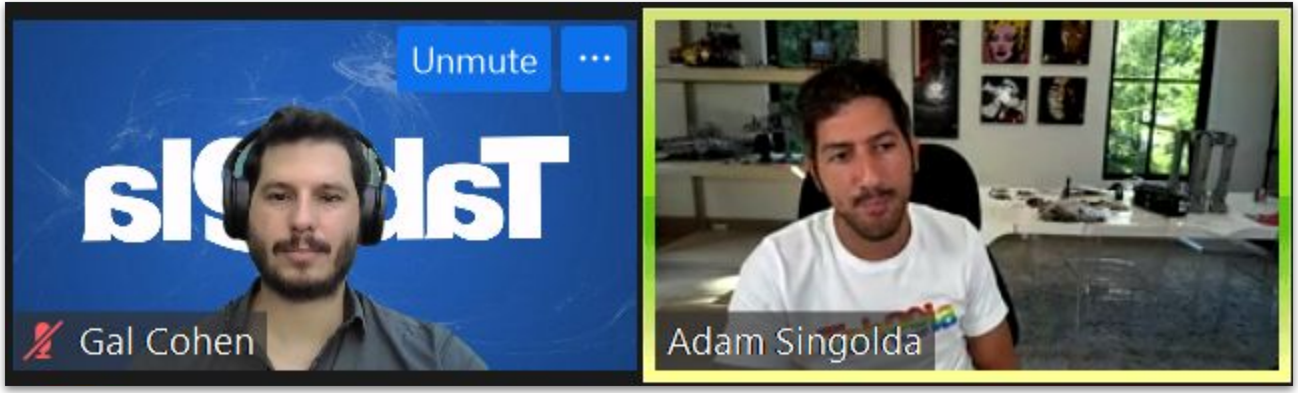
Business dependency

150k decisions

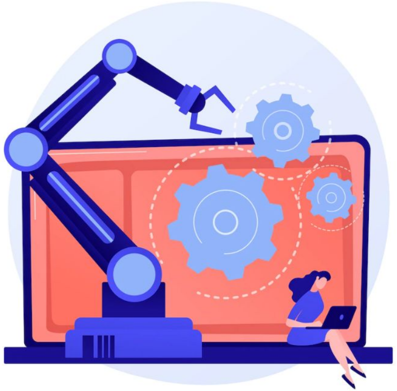
Each day

(say: “*wow!*”)

TLV: 2am **NYC: 7pm**


Unmute
···
Gal Cohen
Adam Singolda

Automations



Predefined rules

Obvious decisions

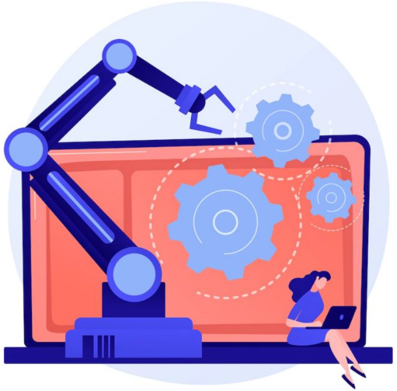
Automation Examples

Yesterday

=

Today

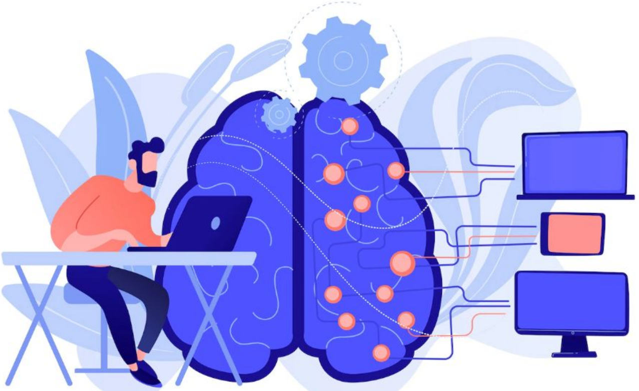
Automations



Predefined rules

Obvious decisions

AI

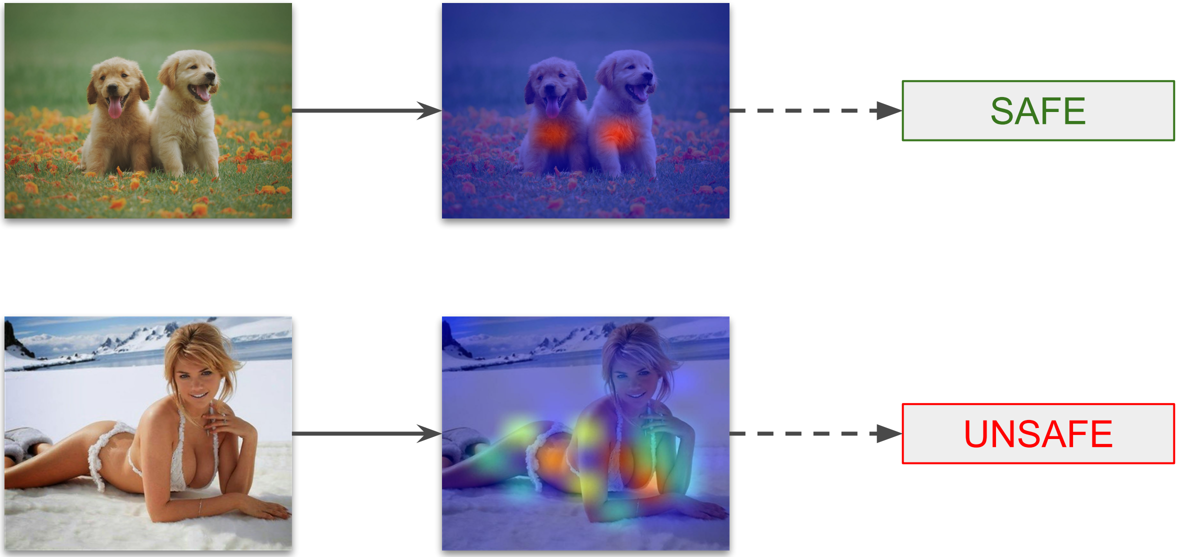


Machine learning

Standard decisions

AI Examples

Image Safety

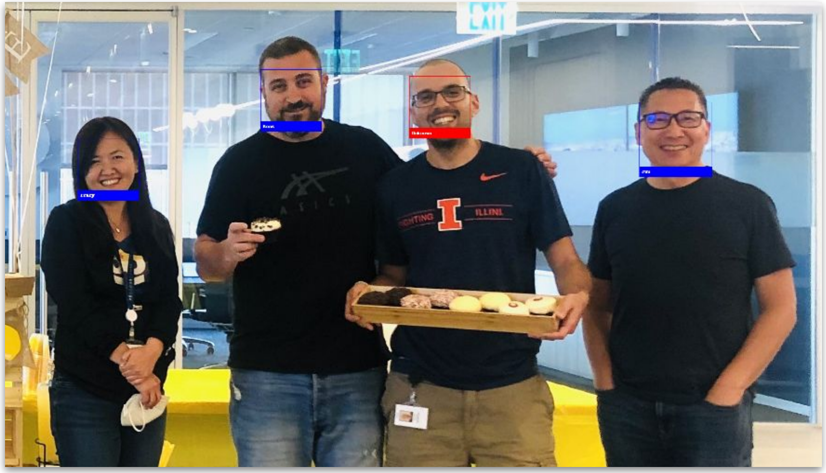


SAFE

UNSAFE

AI Examples

Face Recognition

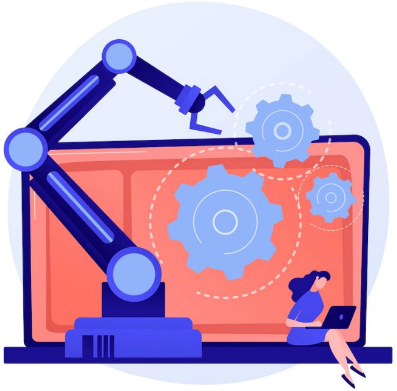

tracy
Boaz
Unknown
Jun

Jun, RnD director @taboola

Tracy, group manager @taboola

Boaz, Algorithm engineer @taboola

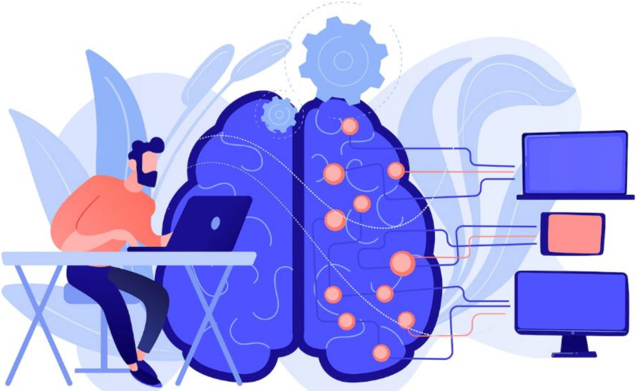
Automations



Predefined rules

Obvious decisions

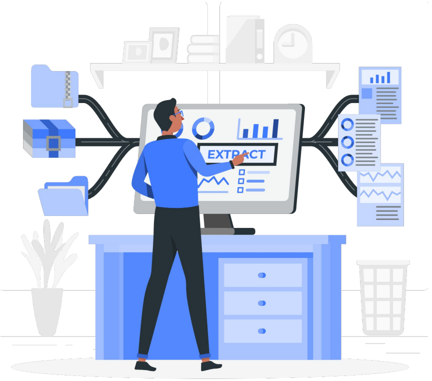
AI



Machine learning

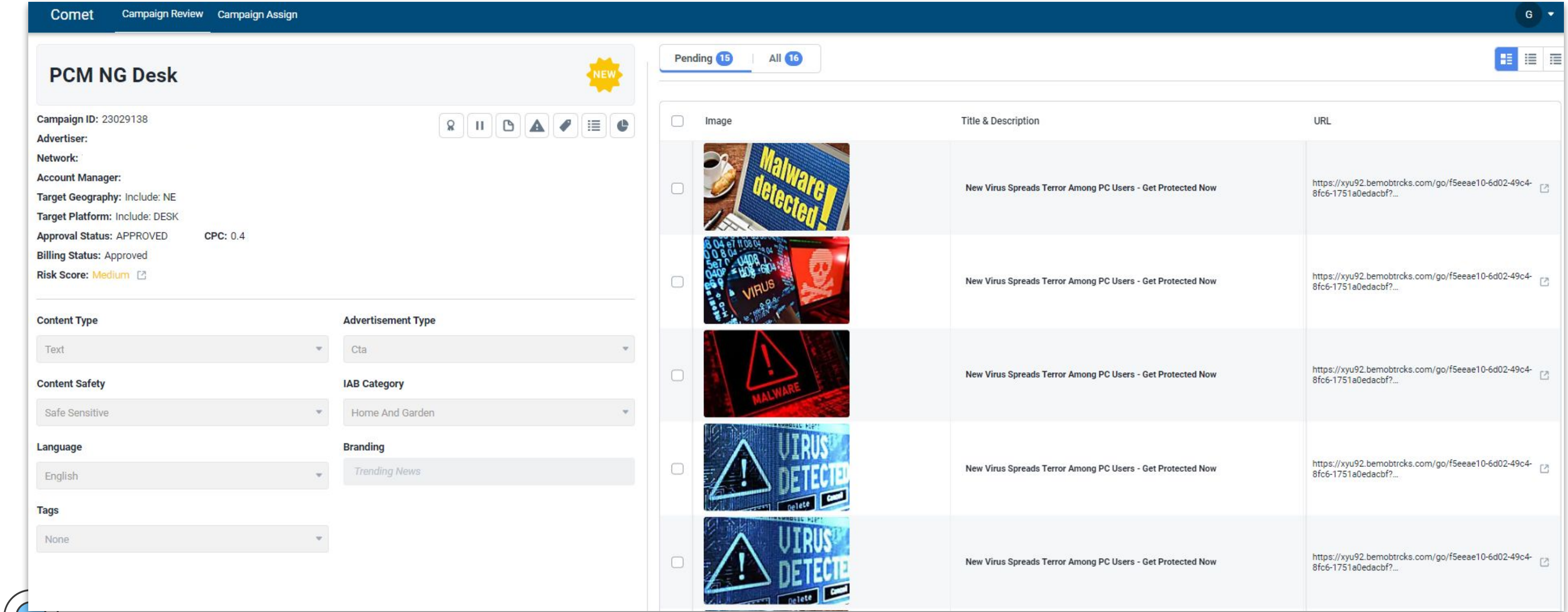
Standard decisions

CRT



Manual review

Hardest decisions


Comet
Campaign Review
Campaign Assign
G
NEW
Pending 15
All 16
Malware detected!
VIRUS
MALWARE
VIRUS DETECTED
Delete
Cancel

PCM NG Desk

Campaign ID: 23029138
Advertiser:
Network:
Account Manager:
Target Geography: Include: NE
Target Platform: Include: DESK
Approval Status: APPROVED **CPC:** 0.4
Billing Status: Approved
Risk Score: Medium

Content Type
Text

Advertisement Type
Cta

Content Safety
Safe Sensitive

IAB Category
Home And Garden

Language
English

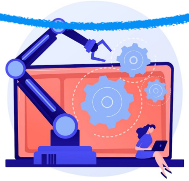
Branding
Trending News

Tags
None

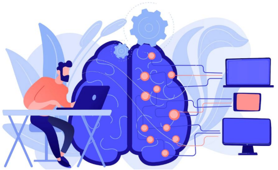
☐	Image	Title & Description	URL
☐		New Virus Spreads Terror Among PC Users - Get Protected Now	https://xyu92.bemobtrcks.com/go/f5eeae10-6d02-49c4-8fc6-1751a0edacbf?...
☐		New Virus Spreads Terror Among PC Users - Get Protected Now	https://xyu92.bemobtrcks.com/go/f5eeae10-6d02-49c4-8fc6-1751a0edacbf?...
☐		New Virus Spreads Terror Among PC Users - Get Protected Now	https://xyu92.bemobtrcks.com/go/f5eeae10-6d02-49c4-8fc6-1751a0edacbf?...
☐		New Virus Spreads Terror Among PC Users - Get Protected Now	https://xyu92.bemobtrcks.com/go/f5eeae10-6d02-49c4-8fc6-1751a0edacbf?...
☐		New Virus Spreads Terror Among PC Users - Get Protected Now	https://xyu92.bemobtrcks.com/go/f5eeae10-6d02-49c4-8fc6-1751a0edacbf?...

Who makes more decisions?

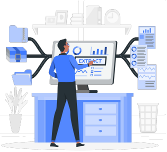
Automations

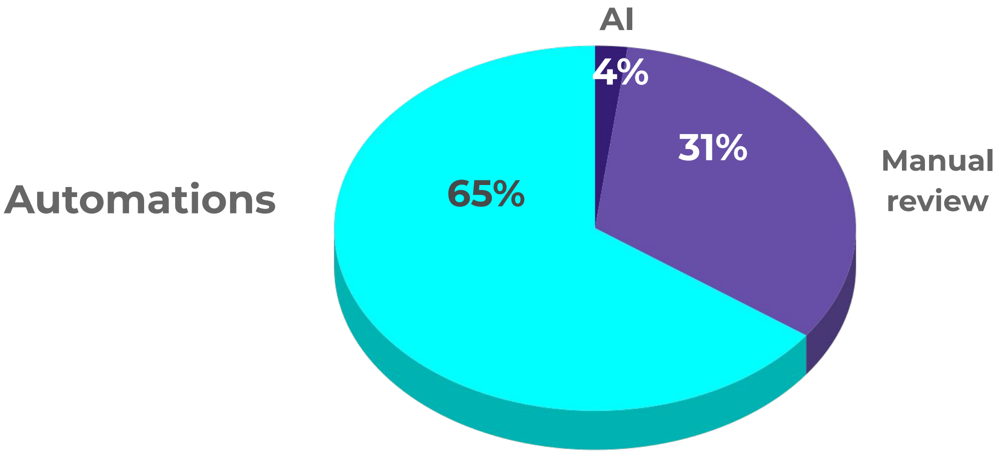


AI



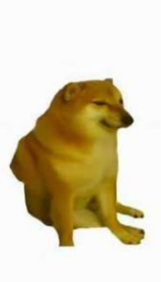
CRT

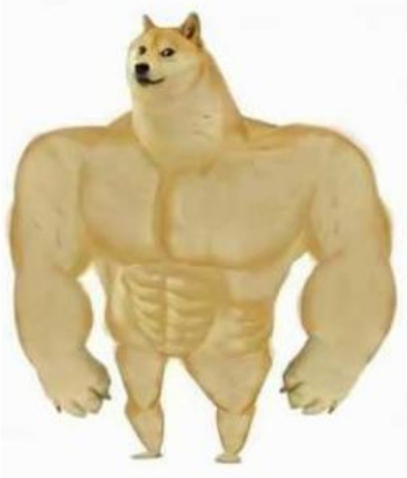




AI

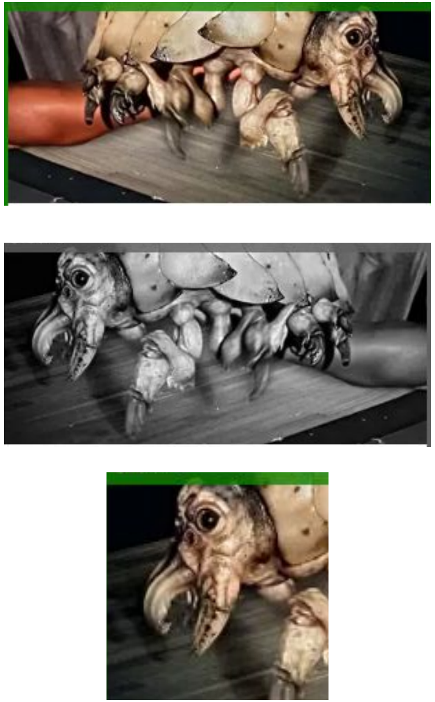
Automations





Yesterday

Today



AI-Powered Automation

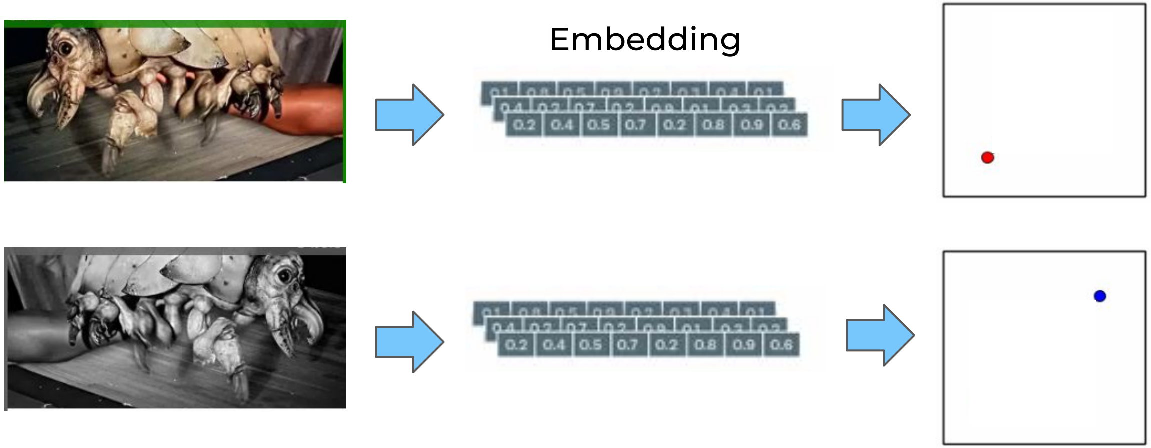
Image Simiarity Model

“Similar!”

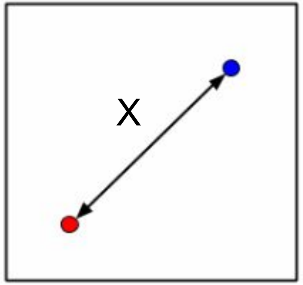
“Not similar!”

Image similarity model

How does it work?


Embedding
0.2
0.4
0.5
0.7
0.2
0.8
0.9
0.6
0.2
0.4
0.5
0.7
0.2
0.8
0.9
0.6

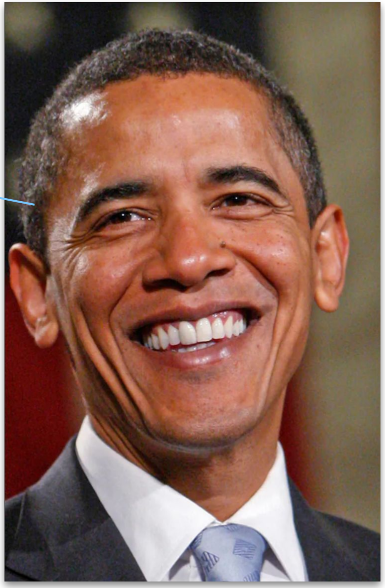
Euclidean distance



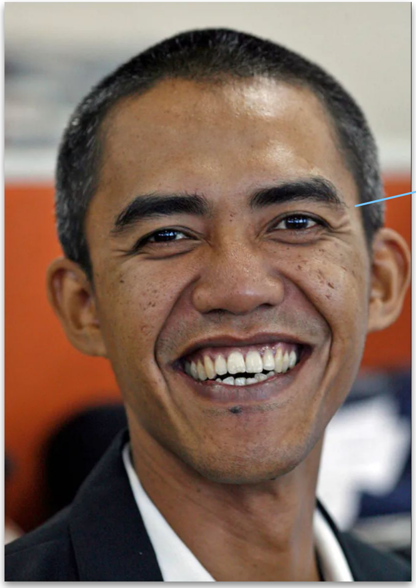
If X < threshold then **SIMILAR**

If X >= threshold then **NOT SIMILAR**

“Similar!”



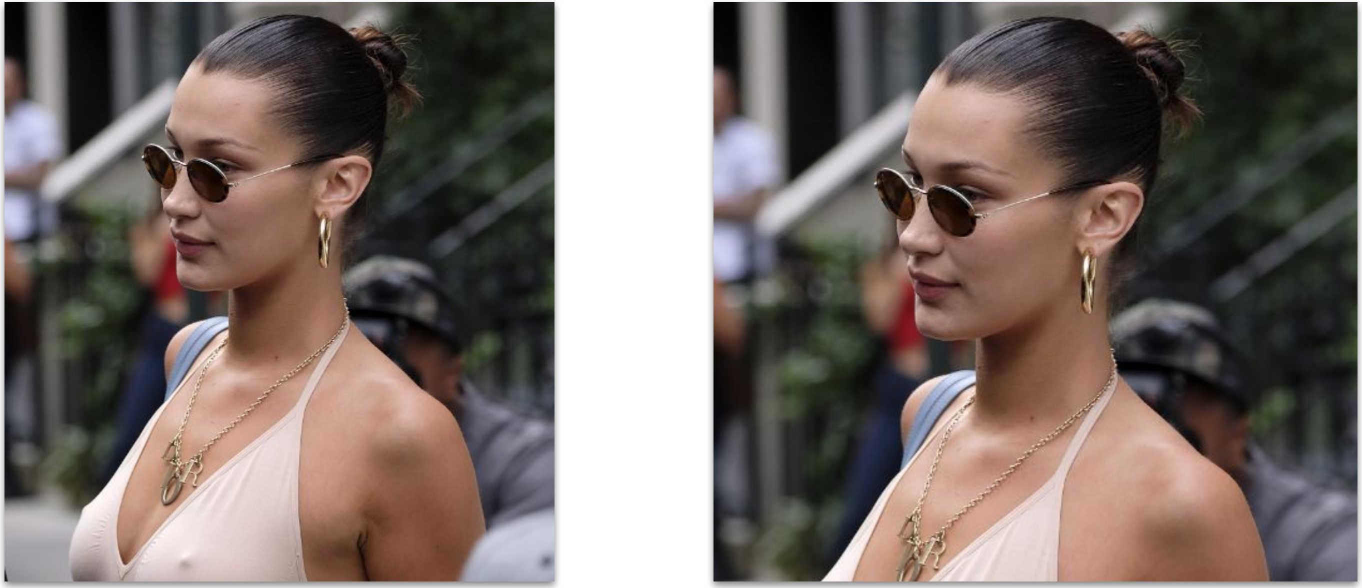
Obama



Not Obama

Similar BUT not the same person

“Similar!”



Similar BUT a different decision

Huston, we have a problem...

What is “similar”?

0.7

0.1

0.2

In other words

Which threshold should we choose?

0.4

0.8

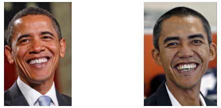
0.9

0.5

0.3

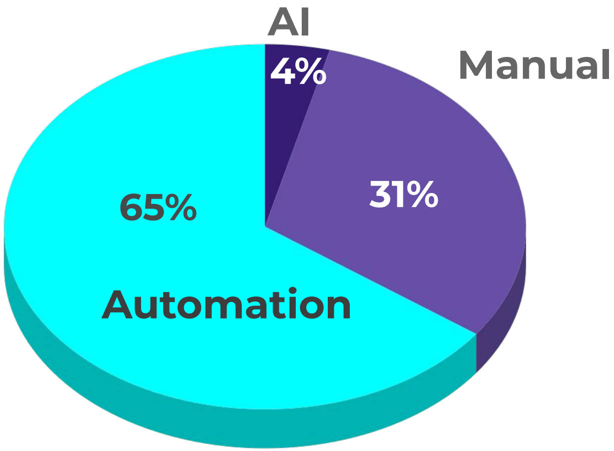
We need

Humans

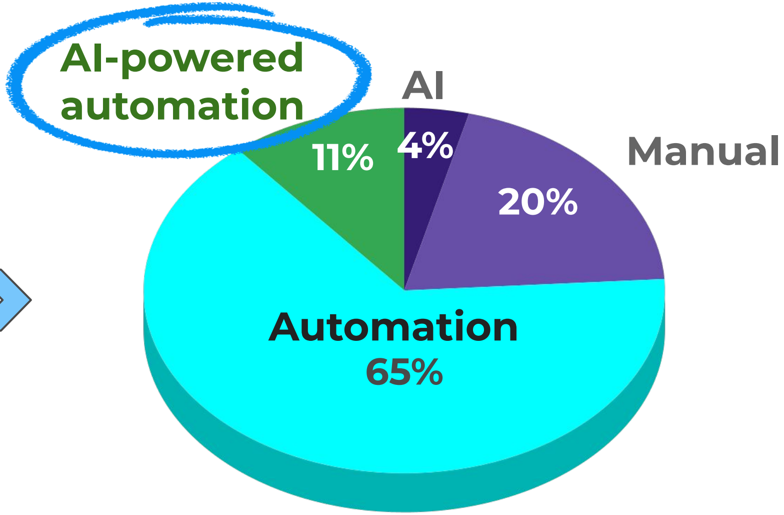
Original + New				**Ground truth** (majority vote)	AI prediction	T1 = **0.8**	T1 = **0.4**	T1 = **0.6**
	Similar	Similar	Not similar	**Similar**	**0.55**	**0.55** < 0.8 ✓ **Similar**	**0.55** > 0.4 ✗ **NOT similar**	**0.55** < 0.6 ✓ **Similar**
	Similar	Not similar	Not similar	**NOT similar**	**0.75**	**0.75** < 0.8 ✗ **Similar**	**0.75** > 0.4 ✓ **NOT similar**	**0.75** > 0.6 ✓ **NOT similar**

Threshold	Precision	Recall	Accuracy
0.8	50%	100%	50%
0.4	Undefined	0%	50%
0.6	100%	100%	100%

Before



After



Take-aways

- Keep an open mind
- Use best practice
- Join Taboola. We are awesome.

Thank You

Gal Cohen

Product Manager at Taboola

Break + networking

Toloka

Panel discussion about Moderation

Toloka

1. What is your content moderation journey?

2. What do you like most about ML for moderation? How do you work around its disadvantages?

3. How does a content moderator’s typical day look like?

Feedback and more ideas for the next event

Toloka

Thanks!