GDC

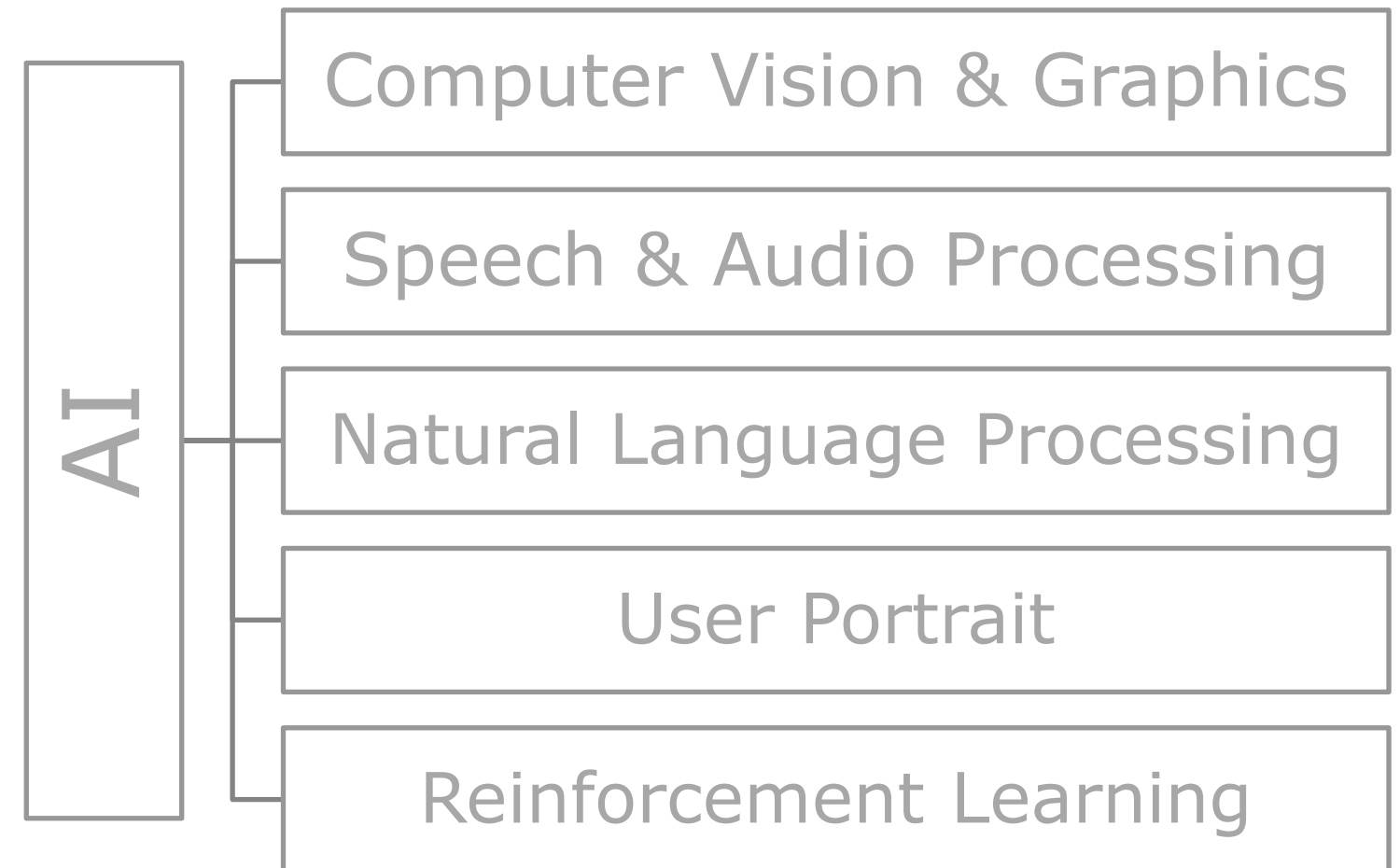# Game Environment Guardian
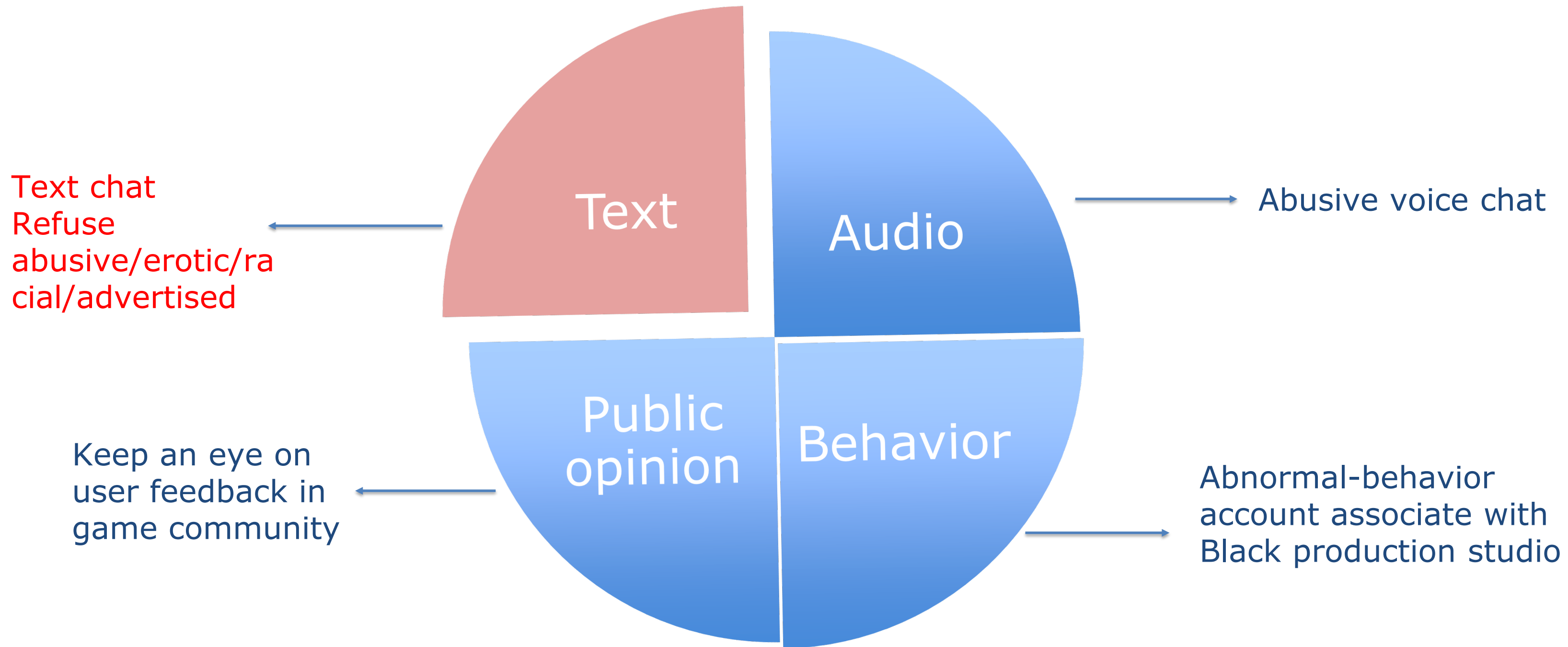
Wuhe Zou (Member: Nan Liang, Jian Xin, Xiangxing Kong, weidong Zhang)
Senior Engineer from AI Lab in NetEase Games

NetEase Games AI Lab

About us

- NetEase Games AI Lab
- Established in late 2017
- Games plus AI

```
        ┌─────────────────────────────────┐
        │ Computer Vision & Graphics      │
        ├─────────────────────────────────┤
        │ Speech & Audio Processing       │
   ┌──┐ ├─────────────────────────────────┤
   │AI│─┤ Natural Language Processing     │
   └──┘ ├─────────────────────────────────┤
        │ User Portrait                   │
        ├─────────────────────────────────┤
        │ Reinforcement Learning          │
        └─────────────────────────────────┘
```

Game Environment Guardian System X AI
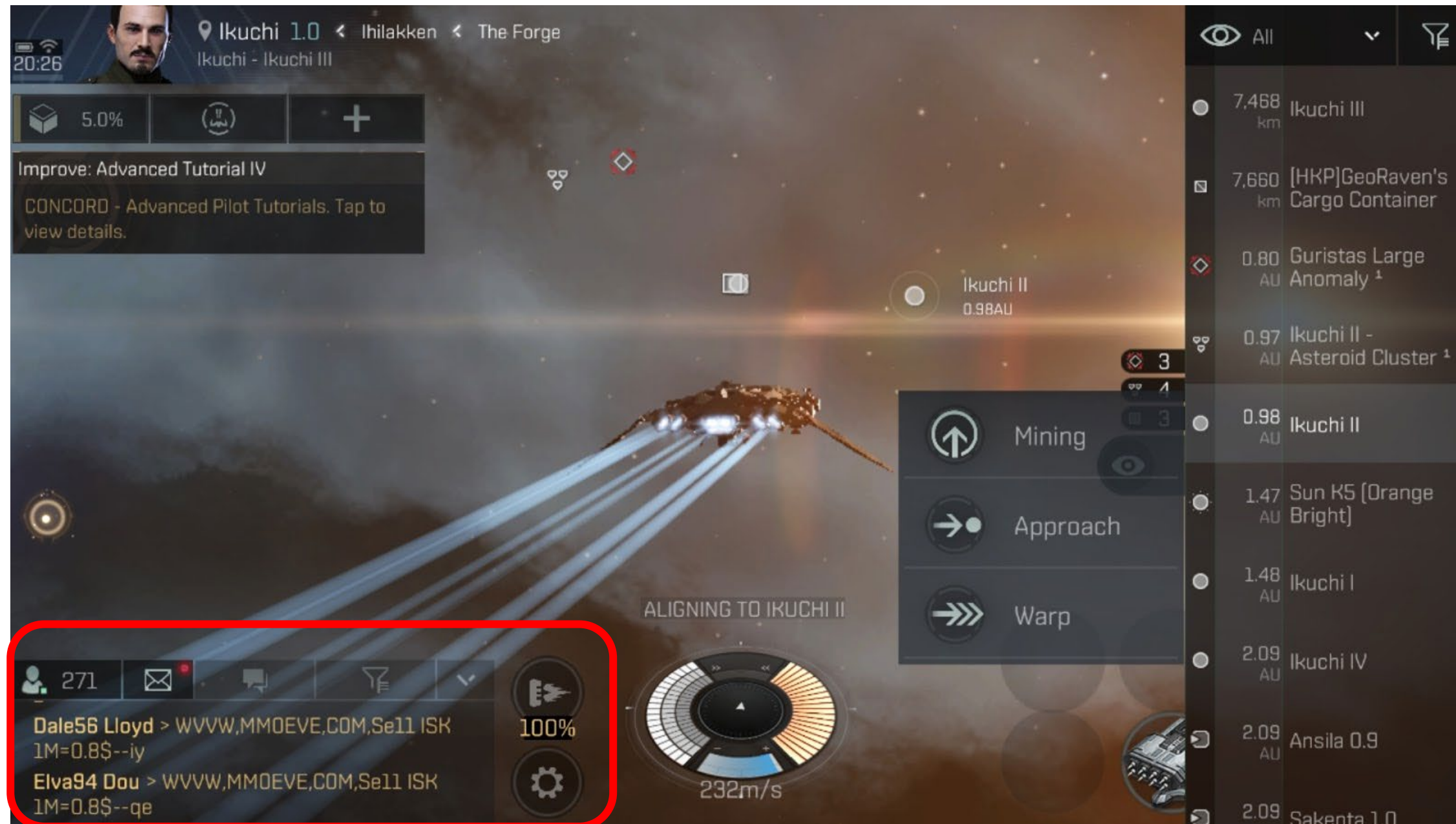
A case in EVE

*The advertising about selling currency in communication channel inside **EVE:echoes***

A case in EVE

Variant words
Homophonic words
…

WVVW , MMOEVE , C0M , Sell ISK 1M=0.8$--iy

W   .   o   random charter

Inappropriate chat categories

Abuse
- racial discrimination
- quarrels
- unfriendly evaluation about other's game performance

Eroticism
- sexual harassment
- sexual suggestion

Abuse and Eroticism may make the community atmosphere **tense** and hurt others feeling

Inappropriate chat categories

Advertising
- game boosting service
- cheat script for sale
- sell in-game currency farmed by script bot

Real-money-trade
- offline trading between players

Unofficial trade may be **fraud** and interfere with the game environment

Traditional Method

Dale56 Lloyd > WVVW,MMOEVE,COM,Sell ISK
1M=0.8$--iy
Elva94 Dou > WVVW,MMOEVE,COM,Sell ISK
1M=0.8$--qe

- **key words**
  - www, wvvw
  - com, c0m

- **regular expression.**
  - [wv]*
  - c[o0]m

Traditional Method

variant of "com"

ccoomm c0m cdm
℃om ℃0m çom ç0m
(om (0m

variant of "www"

vvvw wvvw wwvv
3w 3vv 6v 3v
Щщщ ωωω

Combine of "www" and "com"

ωωω. … .Com
vvvvw. … .℃0m
3w. … . Çom
Щщщ . … .c0m
… … …

Plenty variants
the characters in key words can
be replaced by many other
characters

Labor consuming
hard to exhaustive all combinations
keep updating keywords is labor
consuming
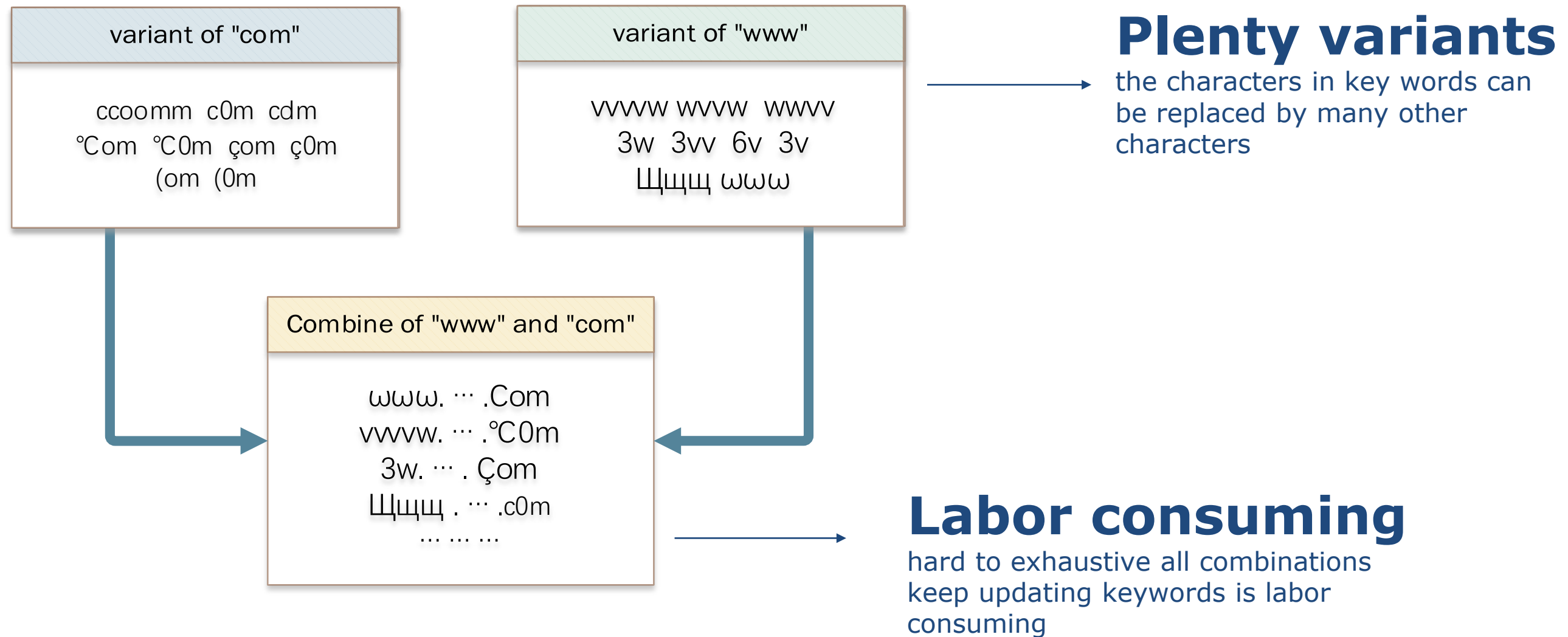
Our machine learning based solution

Input

Preprocess

Normalize

Regex recall

Models

Output
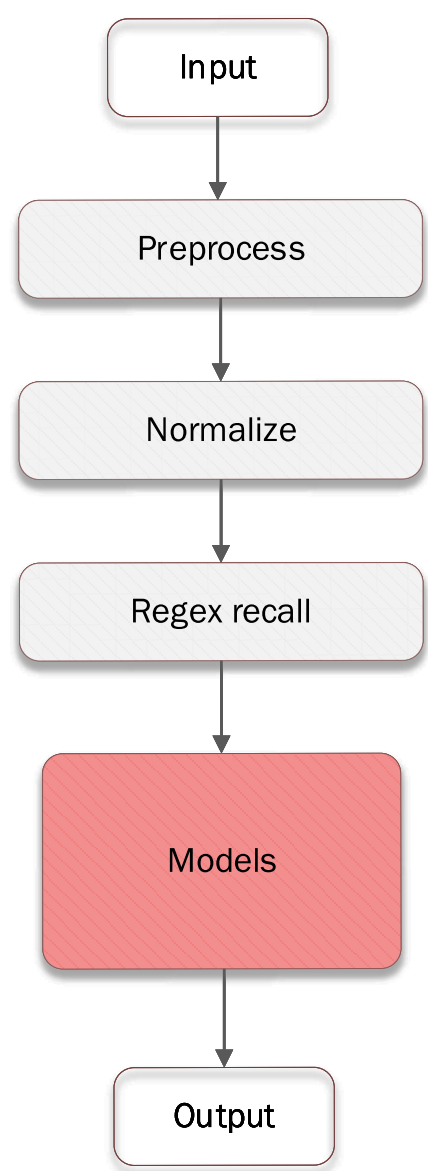
Machine Learning Models
Preprocess
Regex recall
Normalize

emoji
numeric
suitable for fixed phrases
Normalize

$12345 \rightarrow$ **N**

"====="
℃ → **c**
repeated
letters
character
mapping

lightGBM

"coomm"
A → **a**
repeated
uppercase to lowercase
suitable for complex
expression

Transformer

In "preprocess", we removed unconventional
expressions.
In "Normalize", we perform characters mapping to
reduce diversity.
In "regex recall", we use simple regex to roughly search
chats that may contains illegal content.

CNN

efficient and suitable
for most case

BERT

high performance
Sutiable for offline server

Convolutional Neural Network
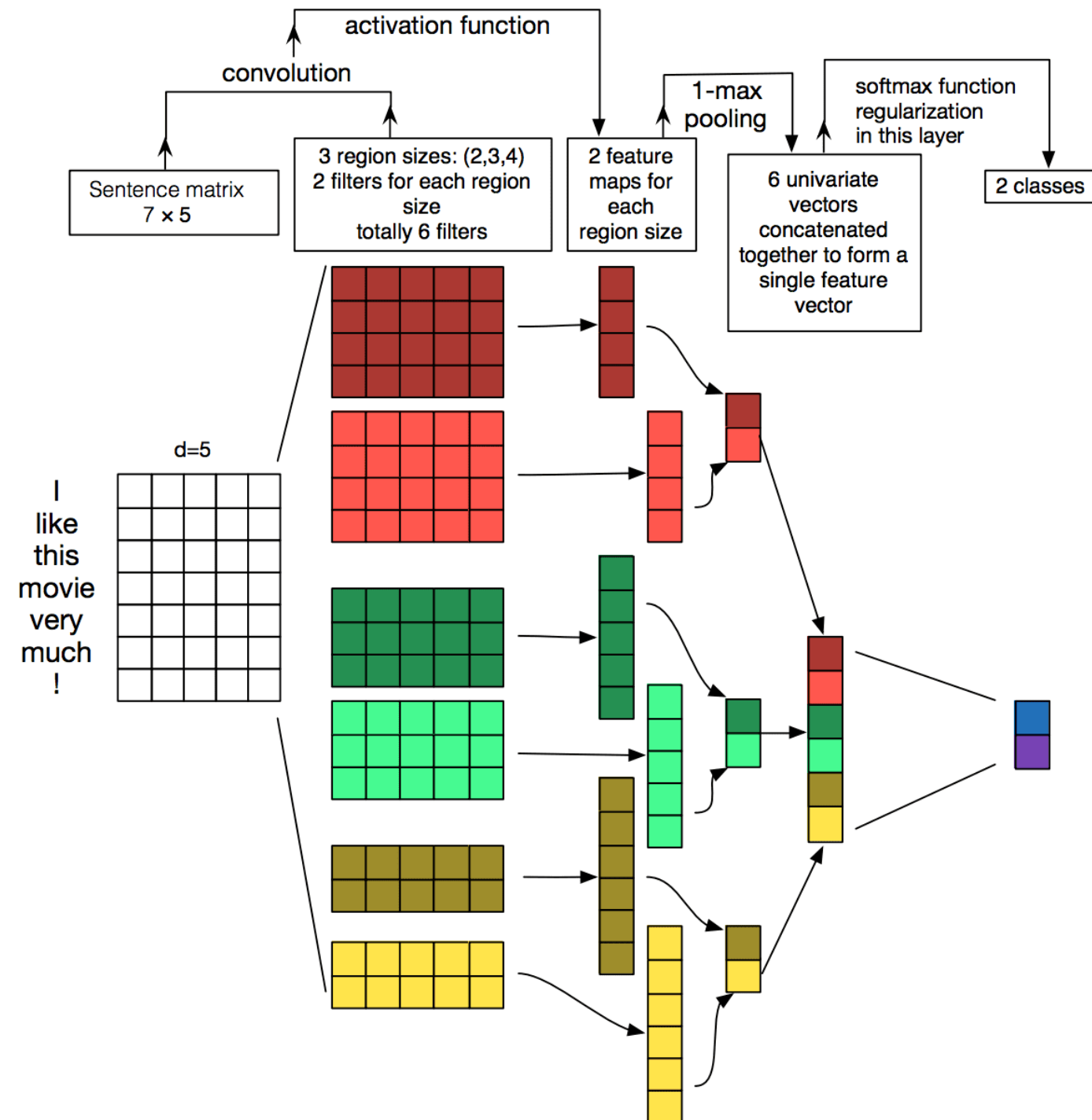
High level

Low level

Sell ... $

size 4   size 2

WVVW , MMOEVE , C0M , Sell ISK 1M=0.8$--iy

In low layer, different size filters learns different features. In high layer, filters learns longer features.

Efficient and reliable

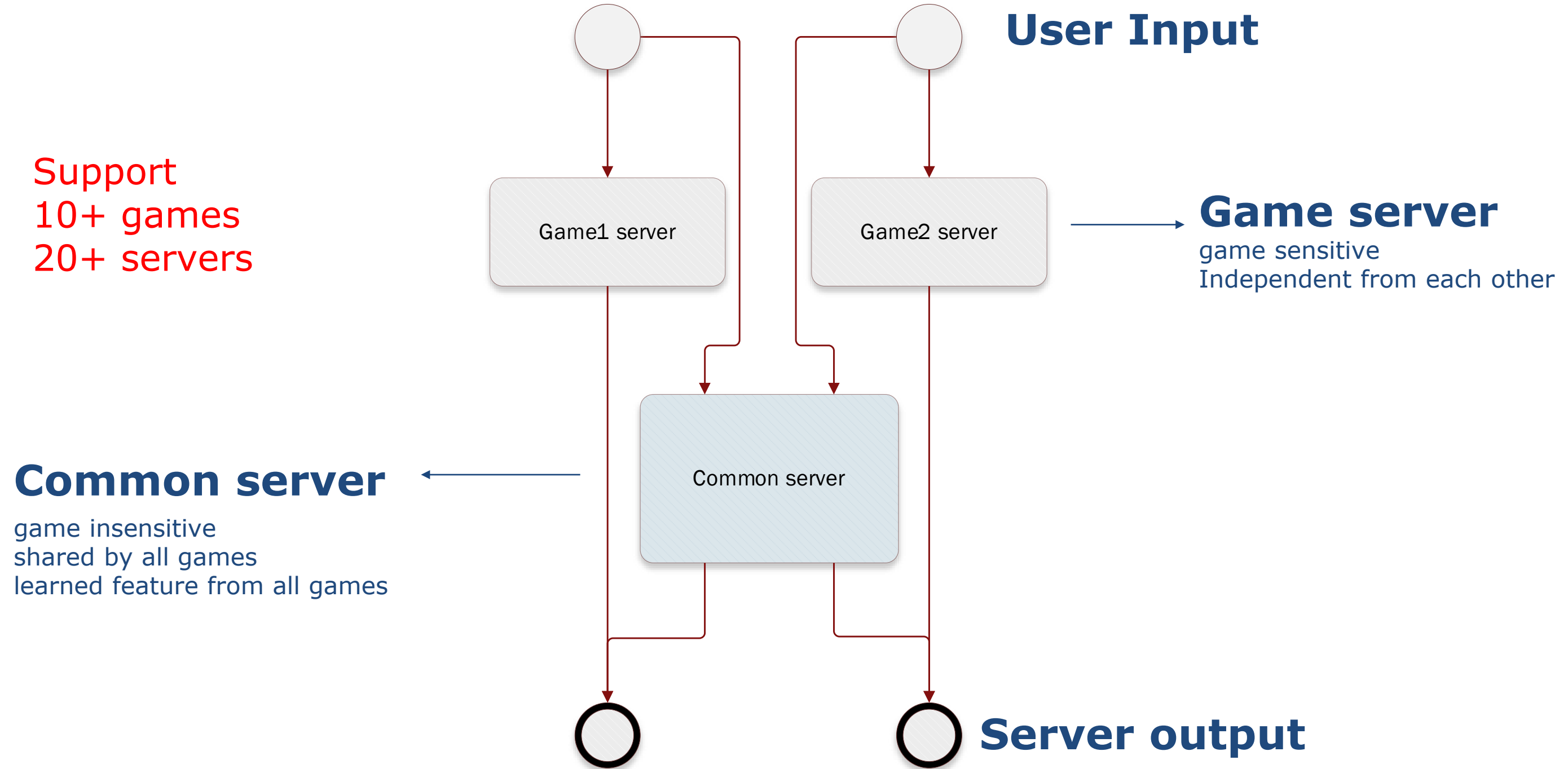Process texts within 100 characters in 7ms

CPU info: Inter E5 2.20GHz

Accuracy > 97%

Content Classification

Support
10+ games
20+ servers

User Input

Game server
game sensitive
Independent from each other

Game1 server

Game2 server

Common server

Common server
game insensitive
shared by all games
learned feature from all games

Server output

Automatic Learning

Iterate wording
- **Frequent retry**
- **Chat script**
- **Community language change**

Automatic Learning

Cache illegal content
expiration time whin 10 min
the num caches is less than 10
cached separately for each player

Our ML based solution
check legality with different models

Chat content → Legality check

illeagl

Cached content

www.mmoeve.com Sell ISK 1M=0.8$

www.mmoeve.com 0.8$ = 1M ISK

ωωω, mmoeve, com Sell ISK

legal

Similarity check

Train Dataset

Compare with cached content
just compare with their own cache

Append to dataset
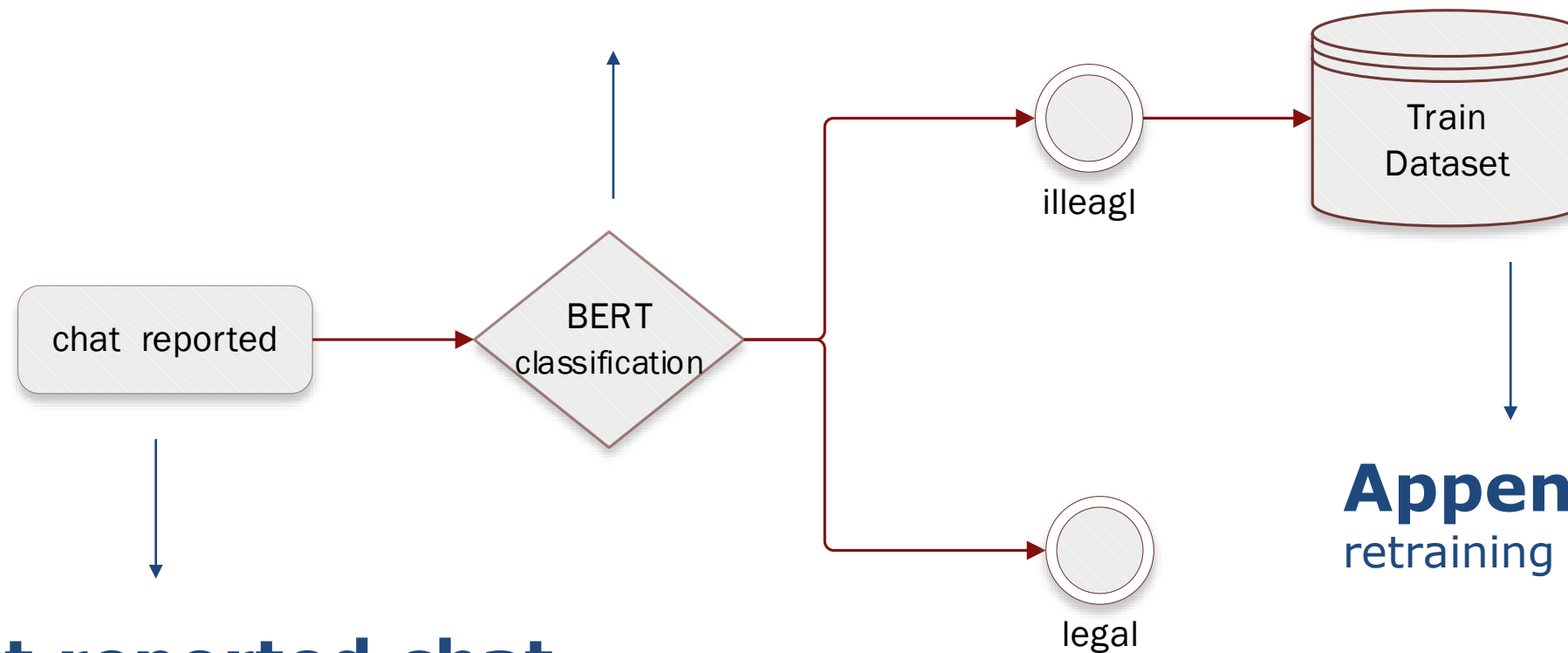training the model to learn new feature

Automatic Learning

True report
or
In a mood

Automatic Learning

Classify with offline model
offline model has better performance
but more resource consume

chat reported → BERT classification

→ illeagl → Train Dataset

→ legal

Append to dataset
retraining online and offline models

Collect reported chat
collect chats and its context reported by players

Text chat
Refuse
abusive/erotic/ra
cial/advertised

Audio

Abusive voice chat

Text

Keep an eye on
user feedback in
game community

Public
opinion

Behavior

Abnormal-behavior
account associate with
Black production studio
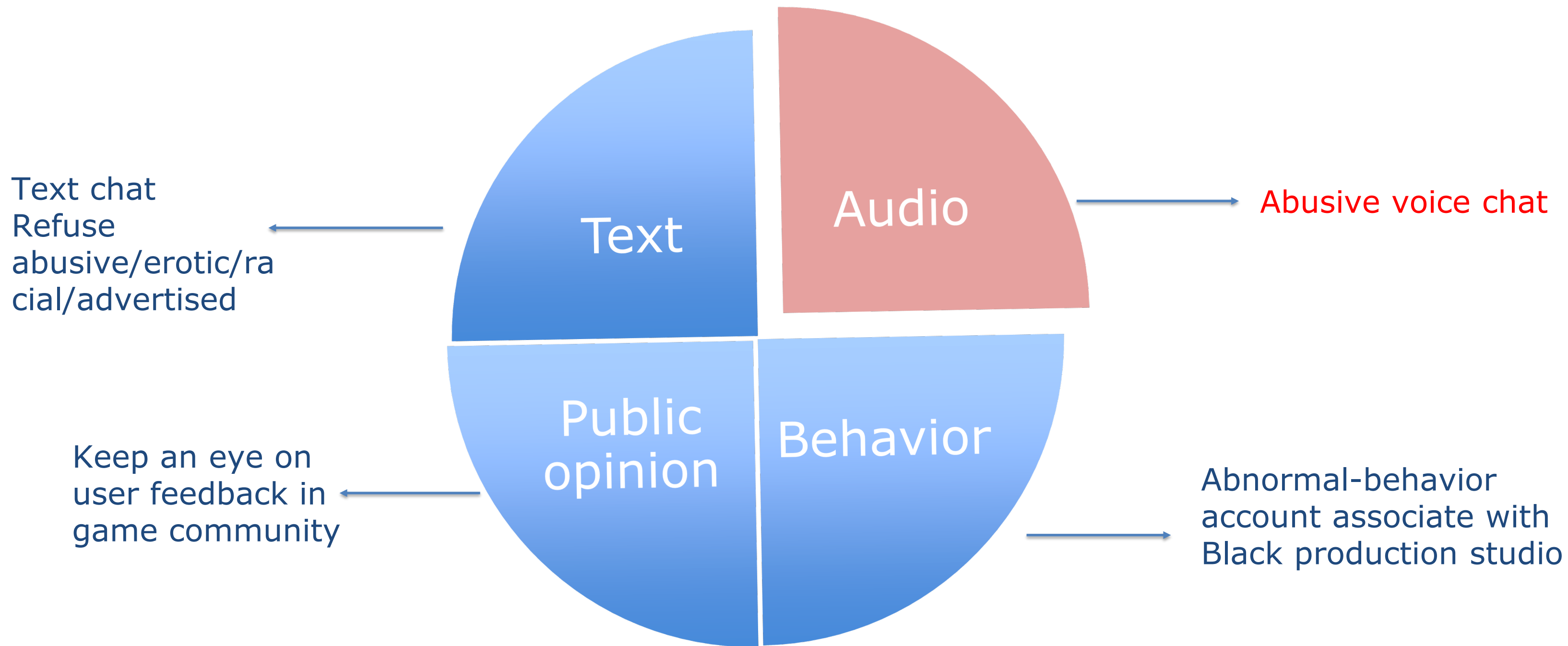
Game Environment Guardian System X AI

Why

Background

- **Voice systems**

    More and more popular in competitive mobile games

    More convenient and faster than typing

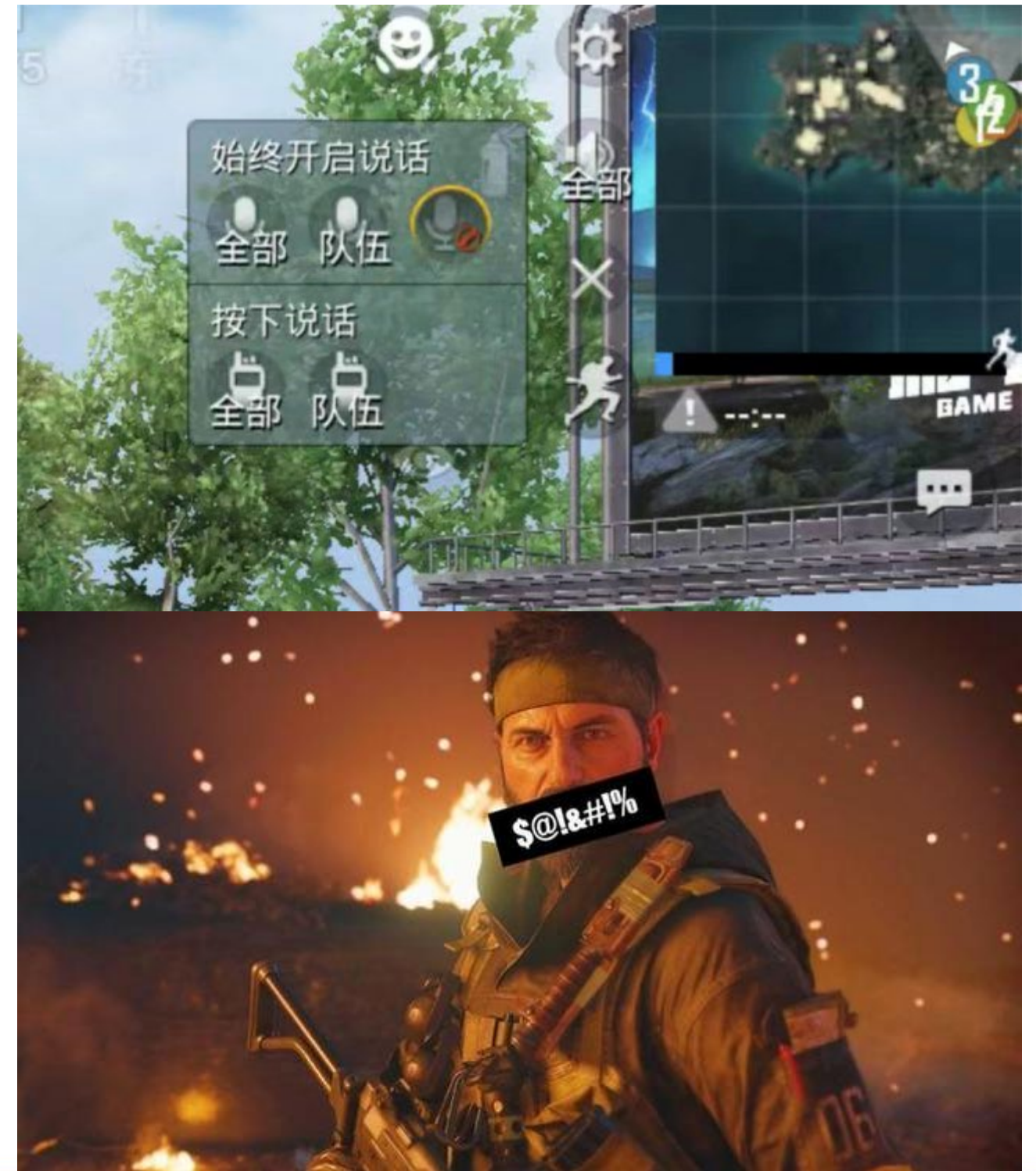- **Abusive language is increasing**

    Vent dissatisfaction

    Swear, curse and insult other players

- **Lead to the loss of users**

    Report released by ADL (Anti-Defamation League)

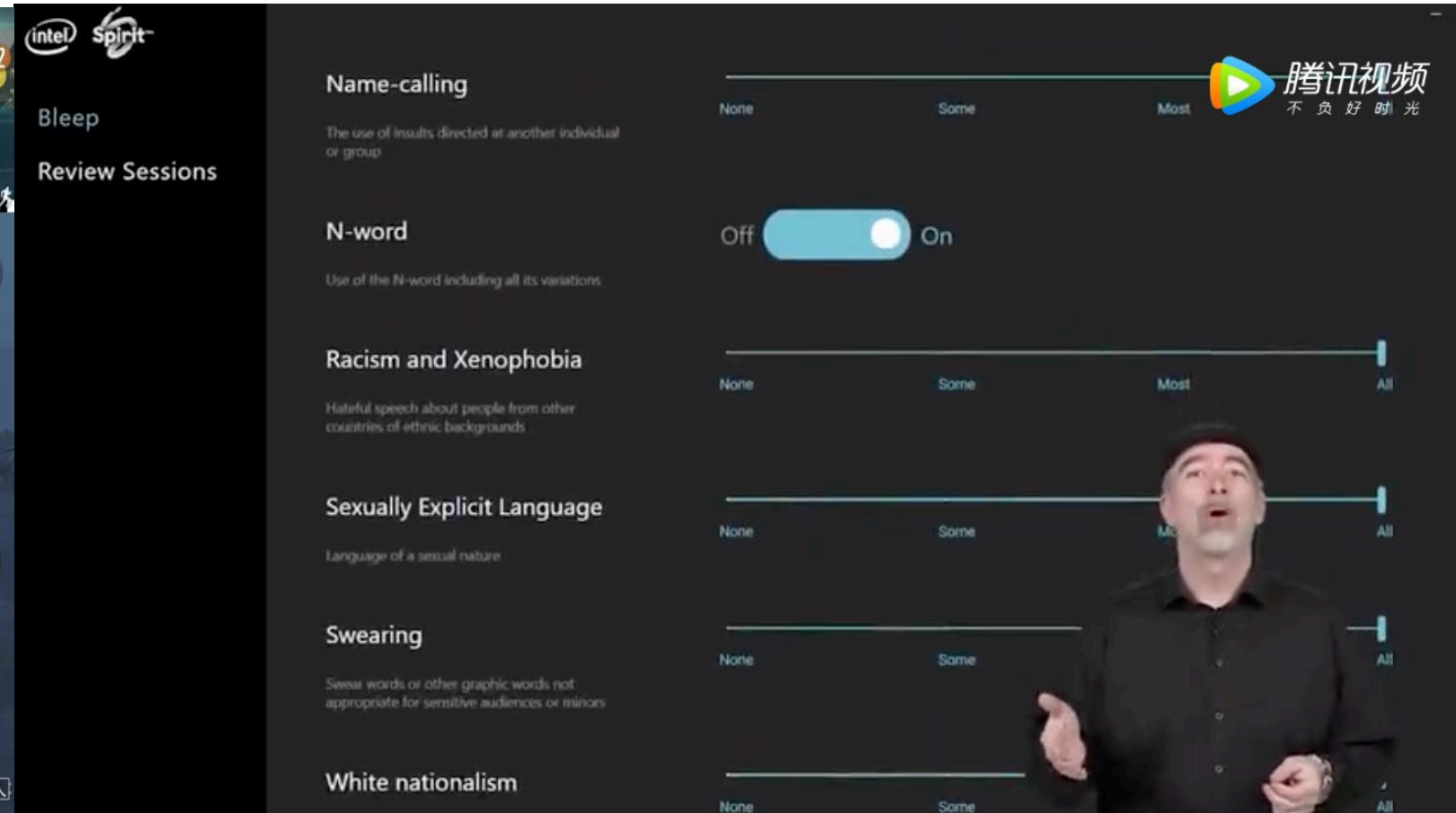   22% of players gave up games because of harassments

Industry Solution

- **Trun off the voice channel**
  - PUBG Mobile
  - Honor of Kings

- **Filter sensitive words**
  - Intel Bleep
  - ToxMod

Challenges

Save me, please. Fuck~ I'm on the second floor. There is a mother fucker lying in ambush.

Cost n seconds to collect context

- **Intuitional method**

Traditional audio detection method

problem → Real-time detection challenge

partially useful information

- **Our solution**
  - Audio segmentation
  - Phoneme Recognition
  - Tradeoff precision and latency

How

Phoneme

Phoneme	Example	Translation
AA	odd	AA D
AE	at	AE T
AH	hut	HH AH T
AO	ought	AO T
AW	cow	K AW
AY	hide	HH AY D
B	be	B IY
CH	cheese	CH IY Z
D	dee	D IY
DH	thee	DH IY
EH	Ed	EH D
ER	hurt	HH ER T
EY	ate	EY T
F	fee	F IY
G	green	G R IY N
HH	he	HH IY
IH	it	IH T
IY	eat	IY T
JH	gee	JH IY
K	key	K IY
L	lee	L IY
M	me	M IY
N	knee	N IY
NG	ping	P IH NG
OW	oat	OW T
OY	toy	T OY
P	pee	P IY
R	read	R IY D
S	sea	S IY
SH	she	SH IY
T	tea	T IY
TH	theta	TH EY T AH
UH	hood	HH UH D
UW	two	T UW
V	vee	V IY
W	we	W IY
Y	yield	Y IY L D
Z	zee	Z IY
ZH	seizure	S IY ZH ER

A set of 39 phonemes

The smallest unit of syllable
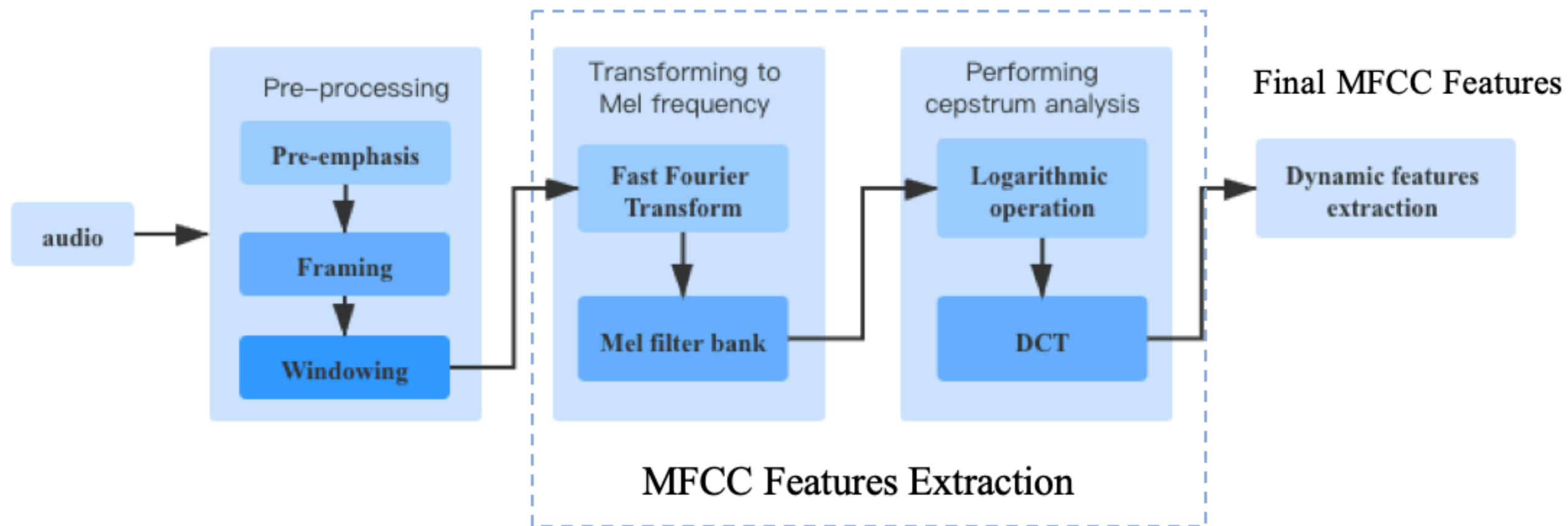
'me' ~ 'M' + 'IY'

from Carnegie Mellon University

Speech Features Extraction

Mel-scale Frequency Cepstral Coefficients (MFCC)

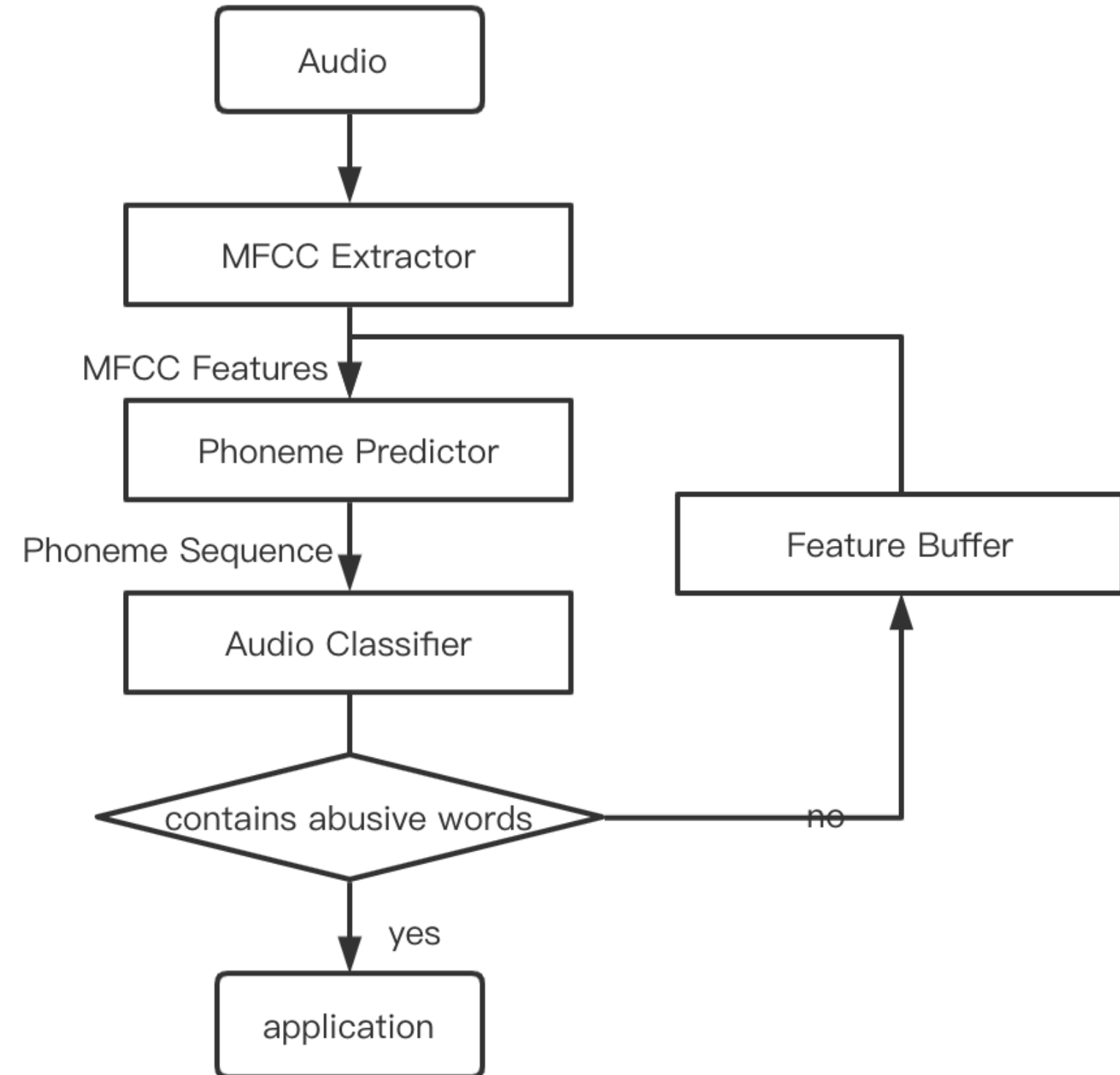# Audio Detection System

- **Target**
  Filter improper audio segments
  Low latency

- **Procedure**
  Segment audio stream
  Extraction speech feature(MFCC)
  Predict phoneme sequence
  Classify abuse or not

Phoneme Predictor

- **Training Data**

  Pre-trained on 1.26 million open source speech samples, fine-tuned on 40k+ game samples.

- **Model**

  A CNN model similar to the ResNet34(Residual Neural Network) structure.

- **Input**
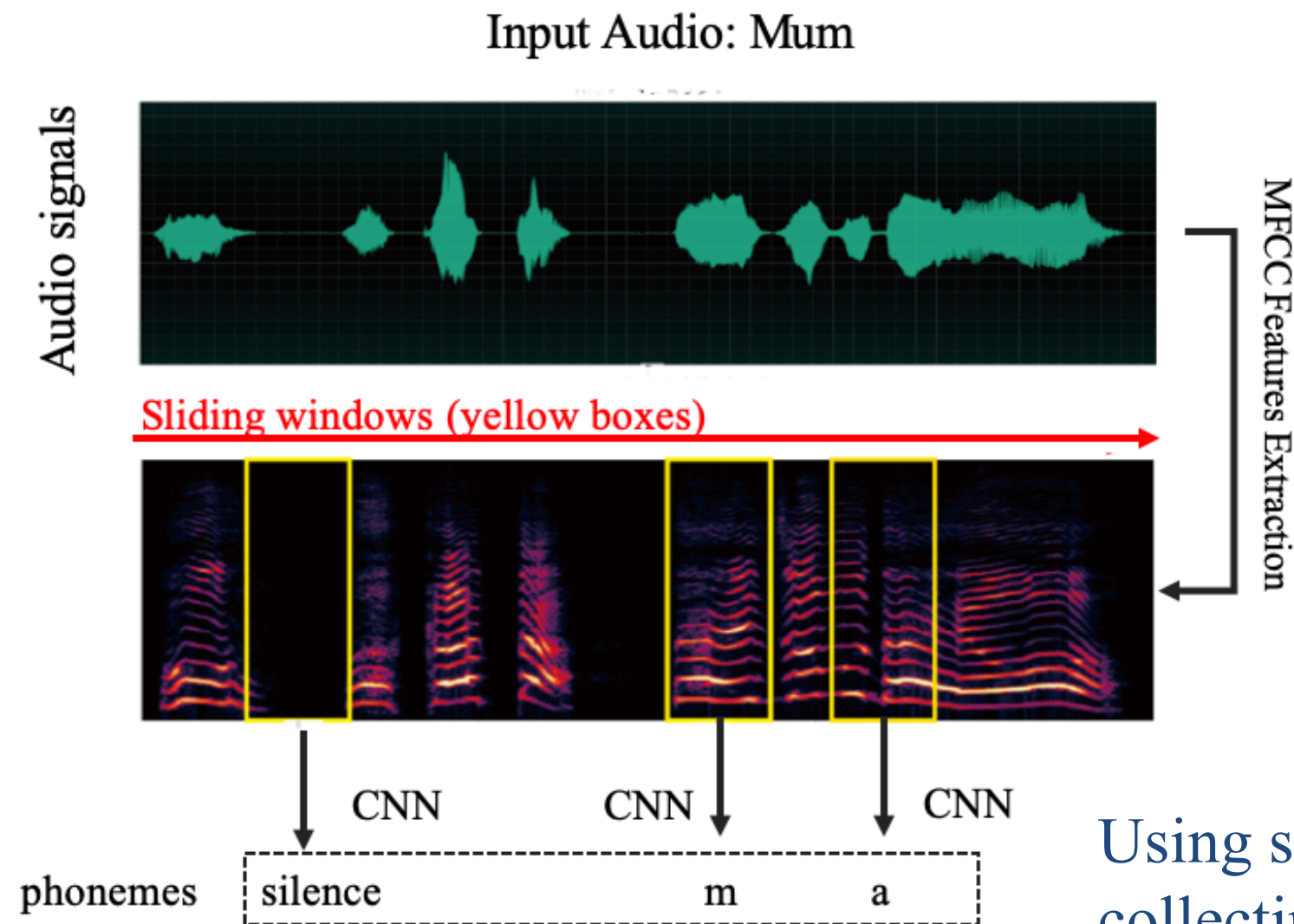
  MFCC features

- **Output**

  C + 1 categories (C is the number of phonemes, and one label indicating silence position)

Phoneme Predictor

Accuracy: 85%
Datasets: aishell2, thchs30,
aidatatang (public available)

Using sliding window strategy on the features, then
collecting all non-silent phonemes into a sequence.
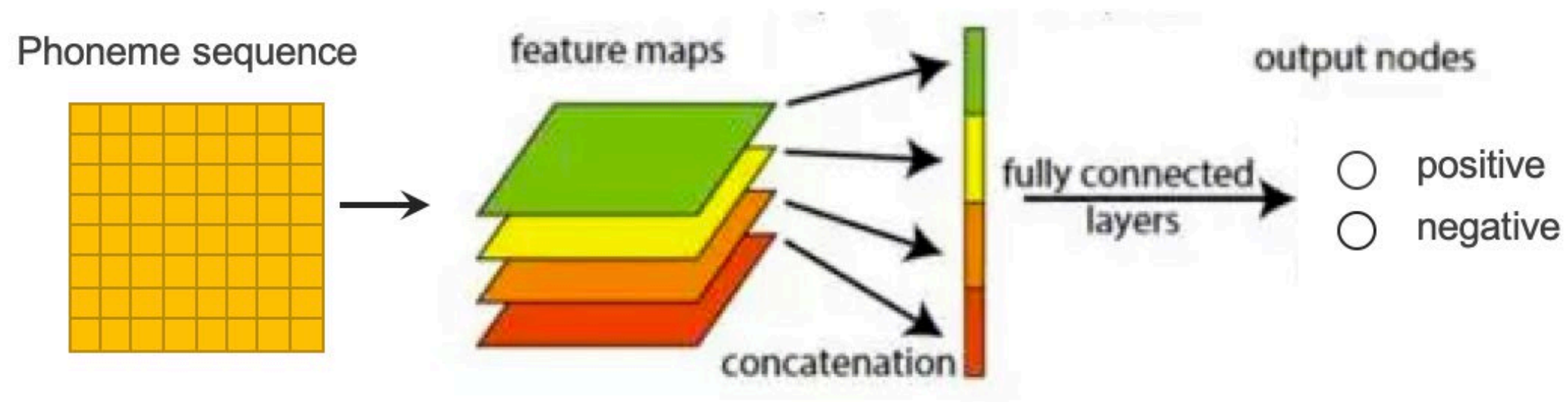
Audio Classifier

- **Training Data**

  Voice data of real players

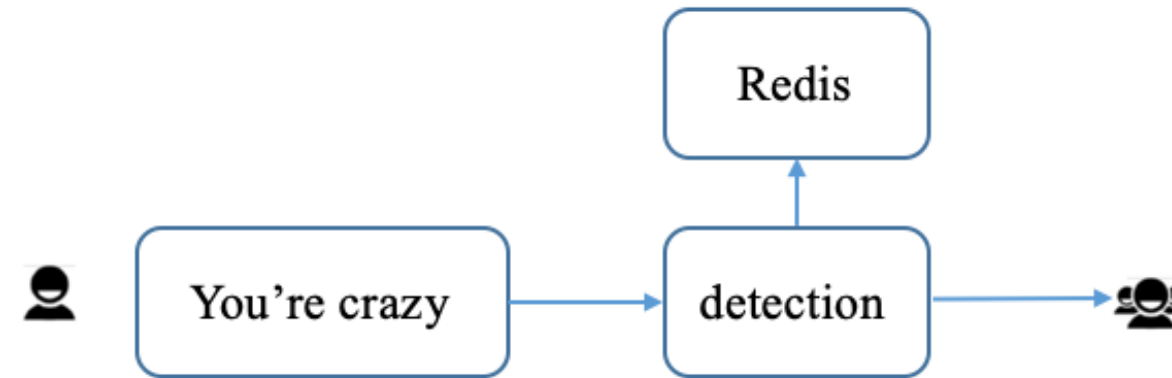 Manually tagged audio segments (5k)

  Segments containing swear words are positive samples, others are negative
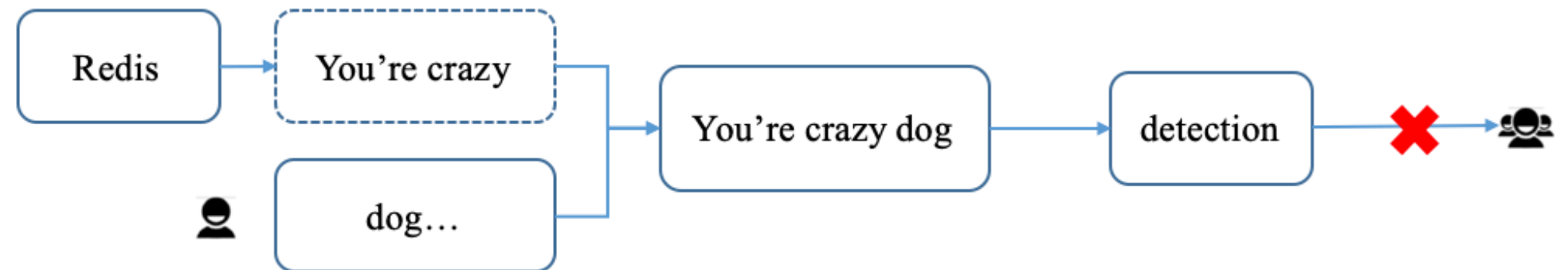
- **Model**

Caching Mechanism

- Alleviate the breakpoint problem caused by audio segmentation.
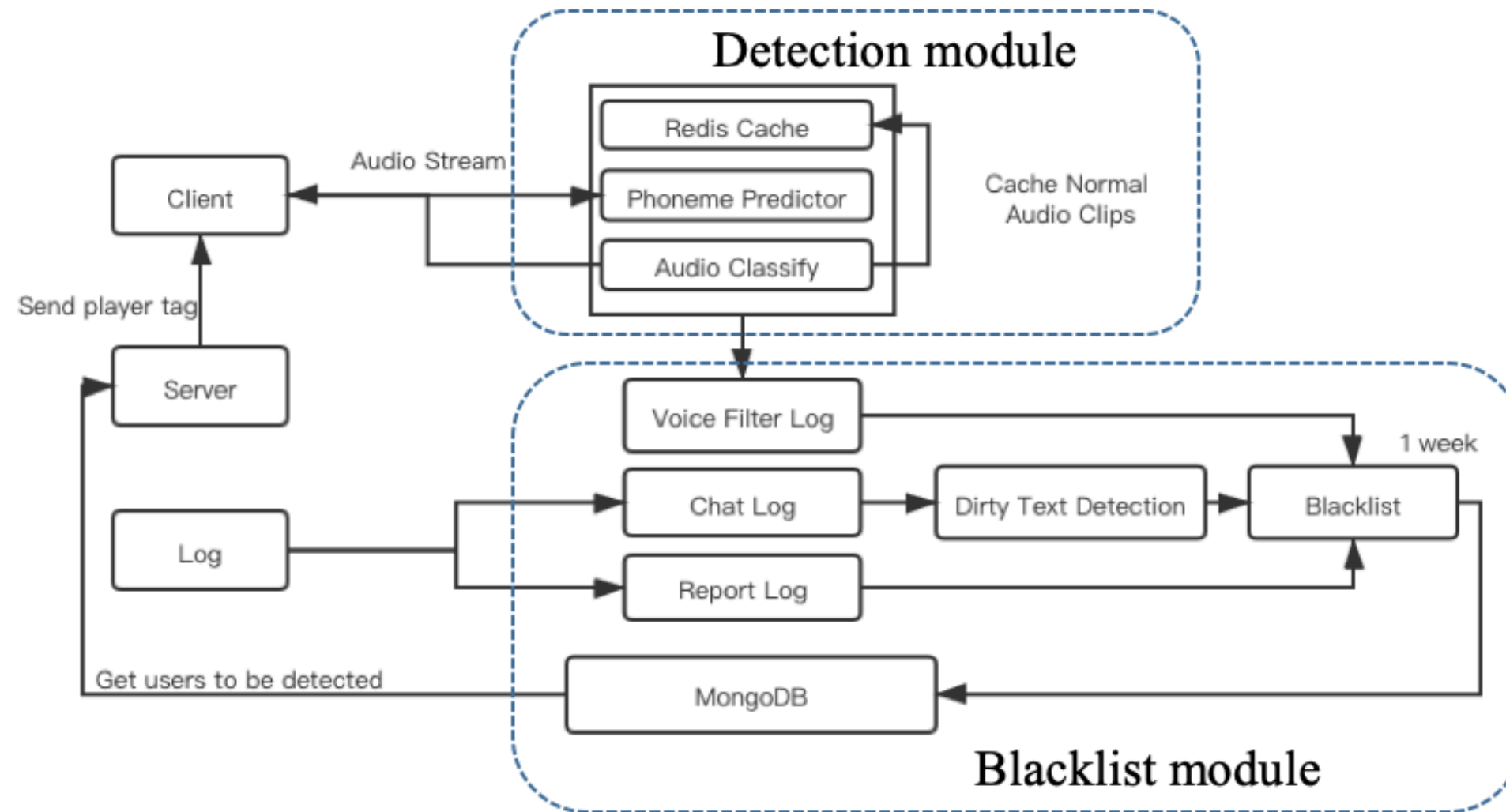

At time t-1, pass the detection, cache features

At time t , cannot pass the detection, filter

Voice Filtering System

- Blacklist to be detected
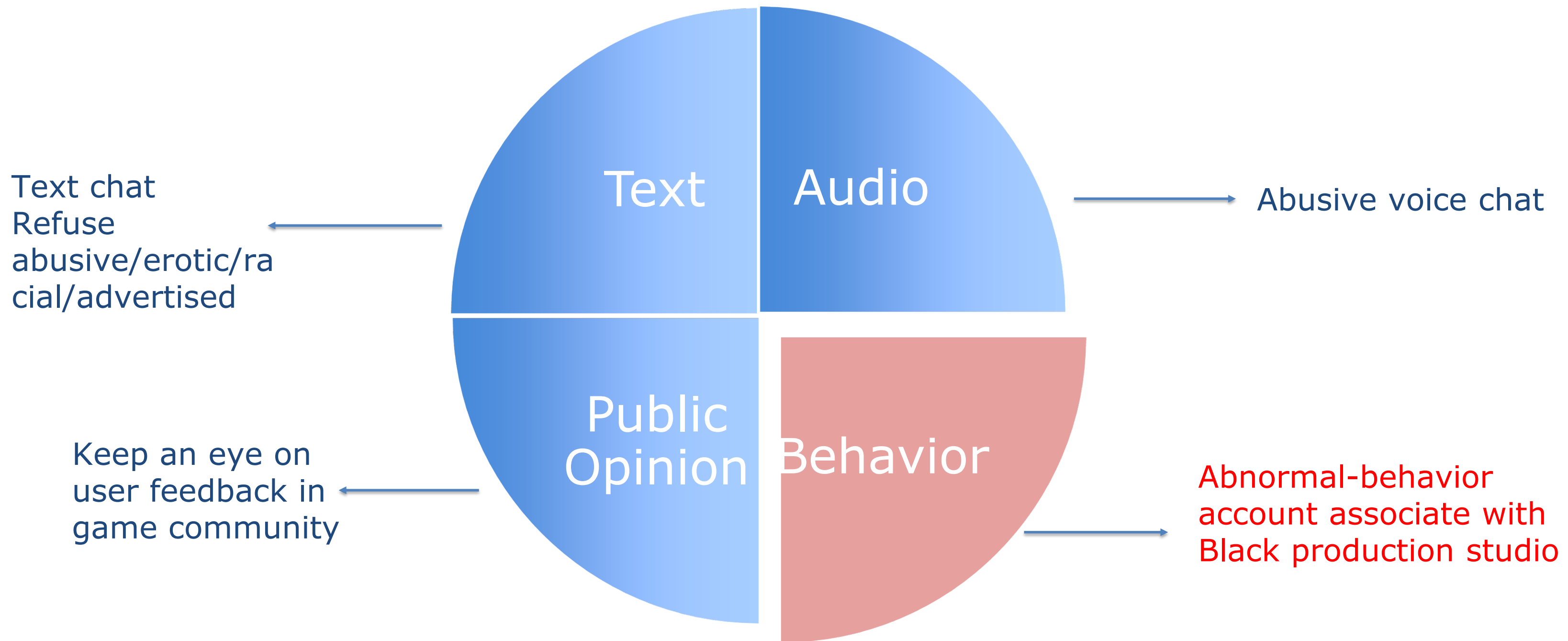- Reduce consumption of resources

Future Work

Not enough

1s latency (collect audio + process)
90% detection ratio (offline data)
76% detection ratio (online data)

- Surround noise
- Poor device
- Heavy accent
- Audio compression

More online data
More Efficient audio compression

Text chat
Refuse
abusive/erotic/ra
cial/advertised

Keep an eye on
user feedback in
game community

Text

Audio

Public
Opinion

Behavior

Abusive voice chat

Abnormal-behavior
account associate with
Black production studio

Game Environment Guardian System X AI

A case – spy

Even defraud

Loss user

Spy post his WeChat ID in the alliance notice

Hello? I know a wonderful game, shall we play together?

Inside Our game

Promote some underground games to High quality players

??

pirate or defraud

Behavior detect

create a role → play →
- create a alliance
- join a alliance
- keep in the wild

→ chat with players → make WeChat friends → promote other games

Chat Detect

Detect in advance of meeting the chat permission

"spy!!!"
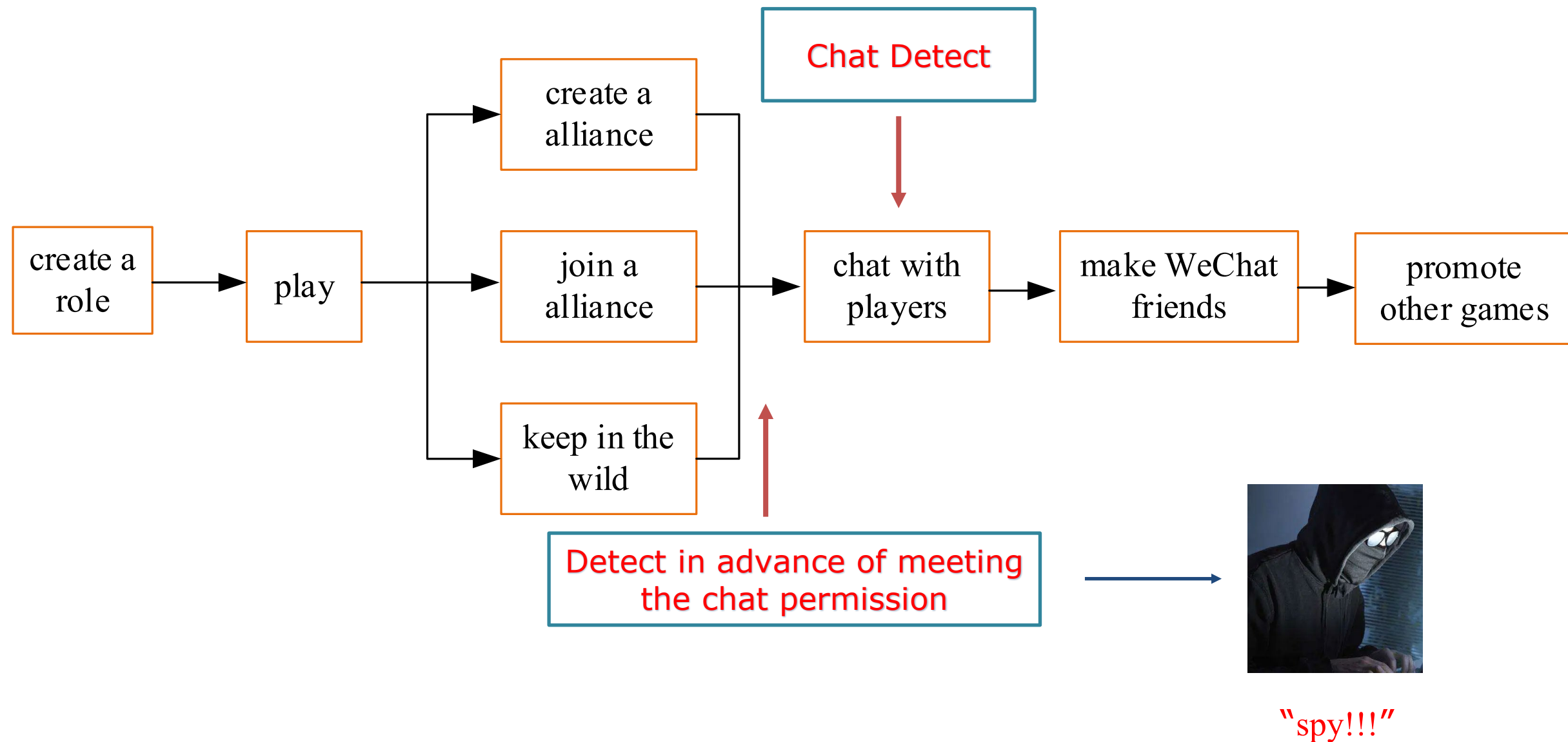
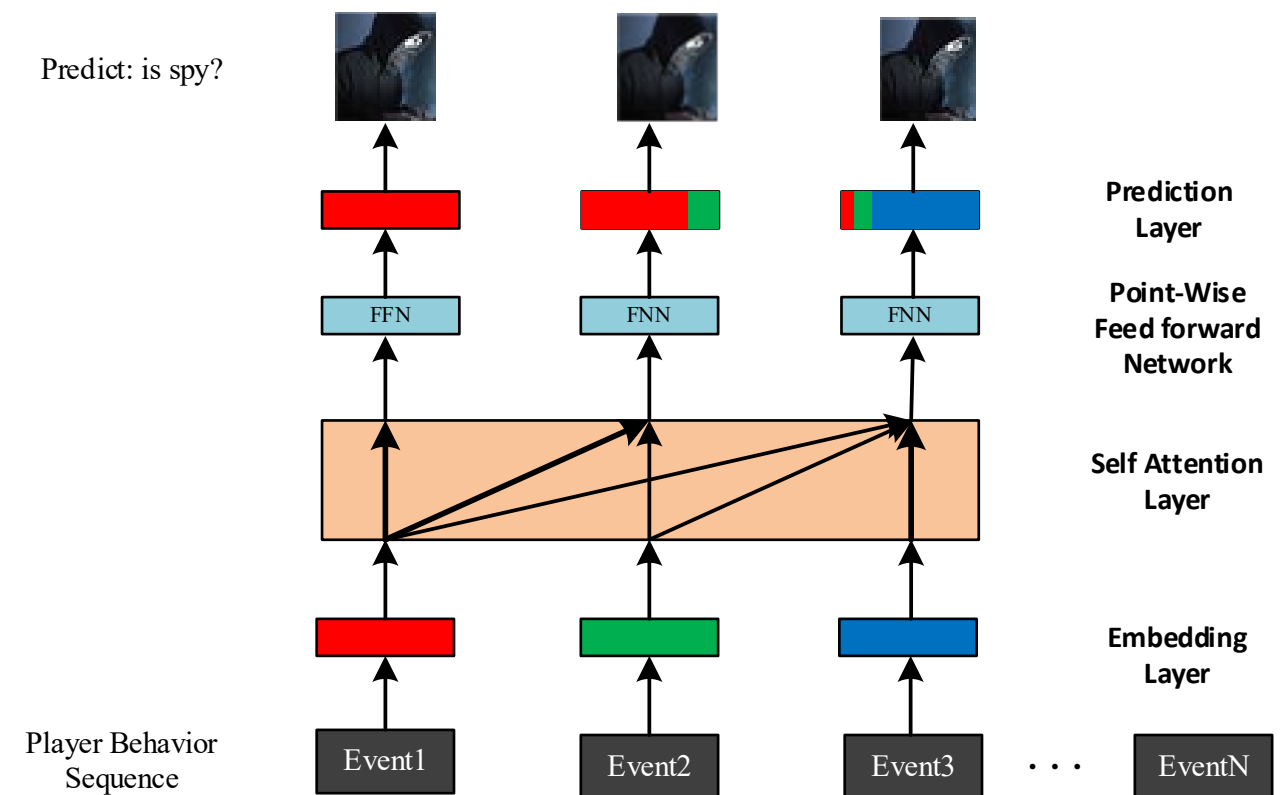# Spy Detect

How to detect spies in advance?

- Spies' behaviors come in special order, which differs from normal players

- Only use non-textual information——behavior sequence & status

- Inspired by the great success of time-series model in recommender system——detect spy based on user behavior sequence.

Time-series model

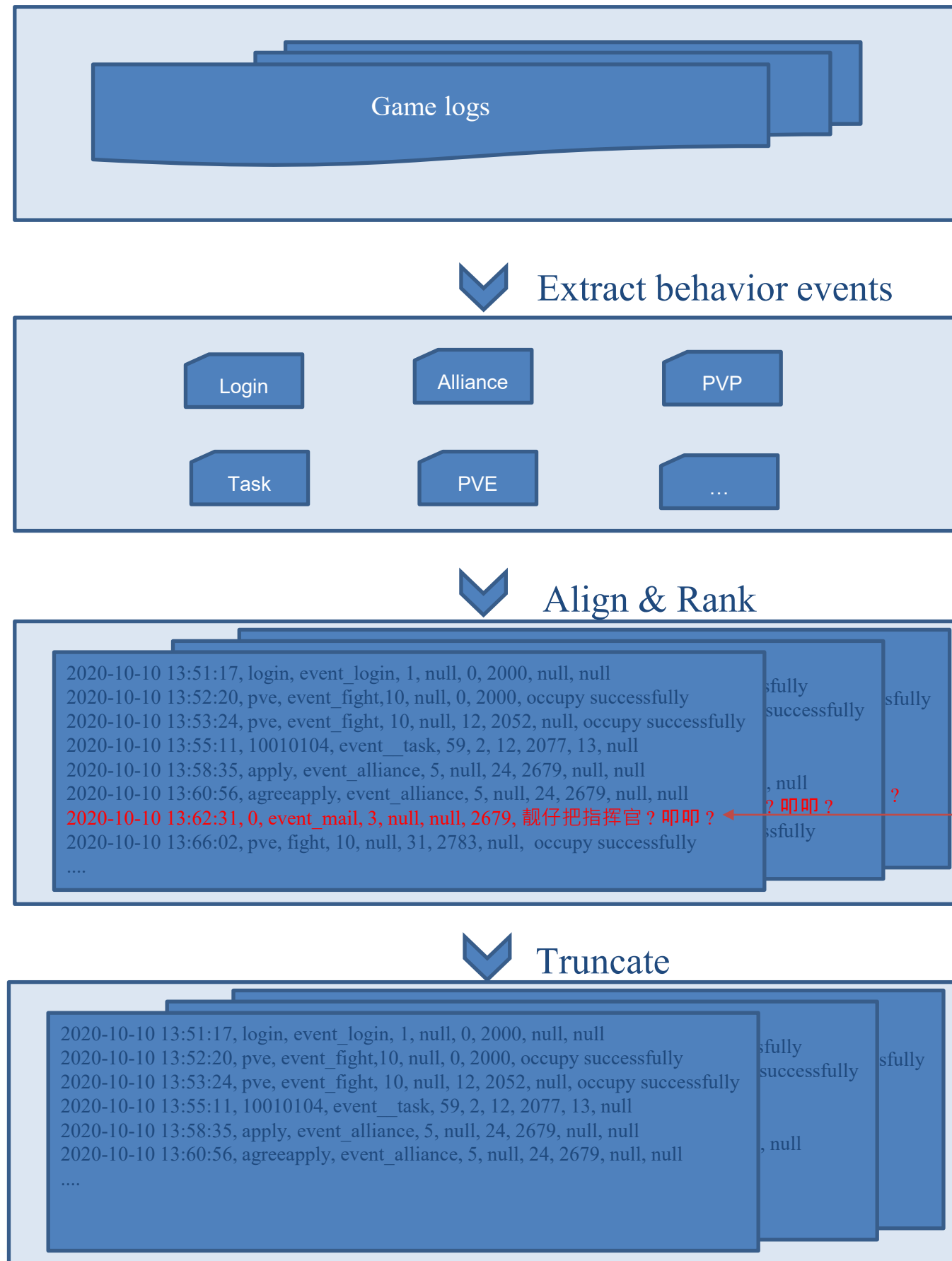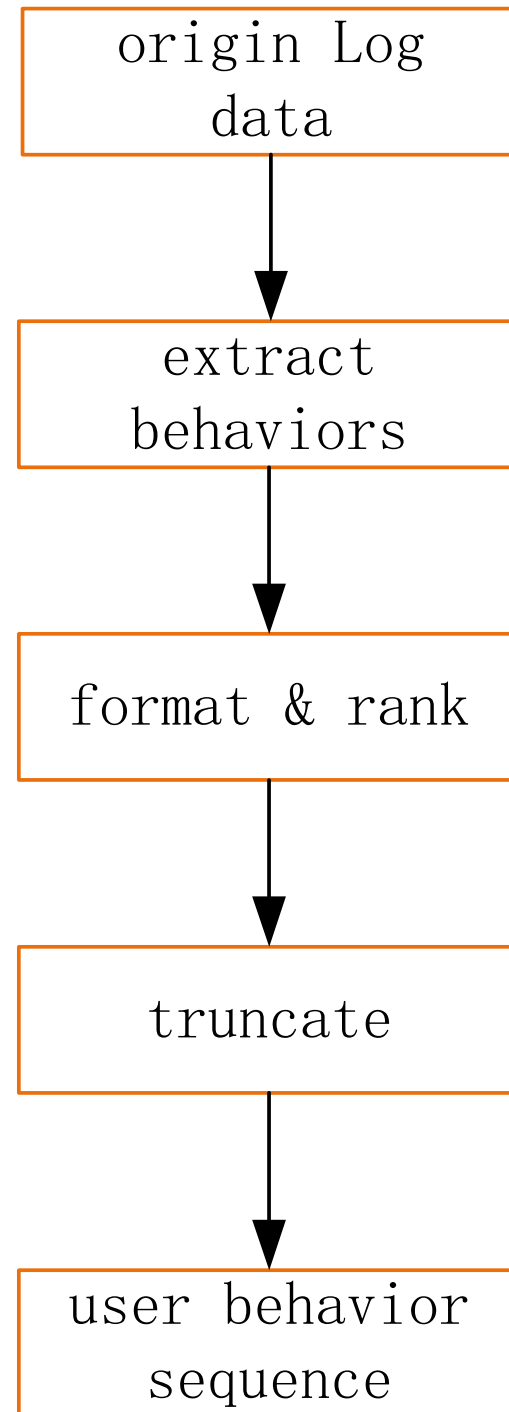time-series model in recommender systems
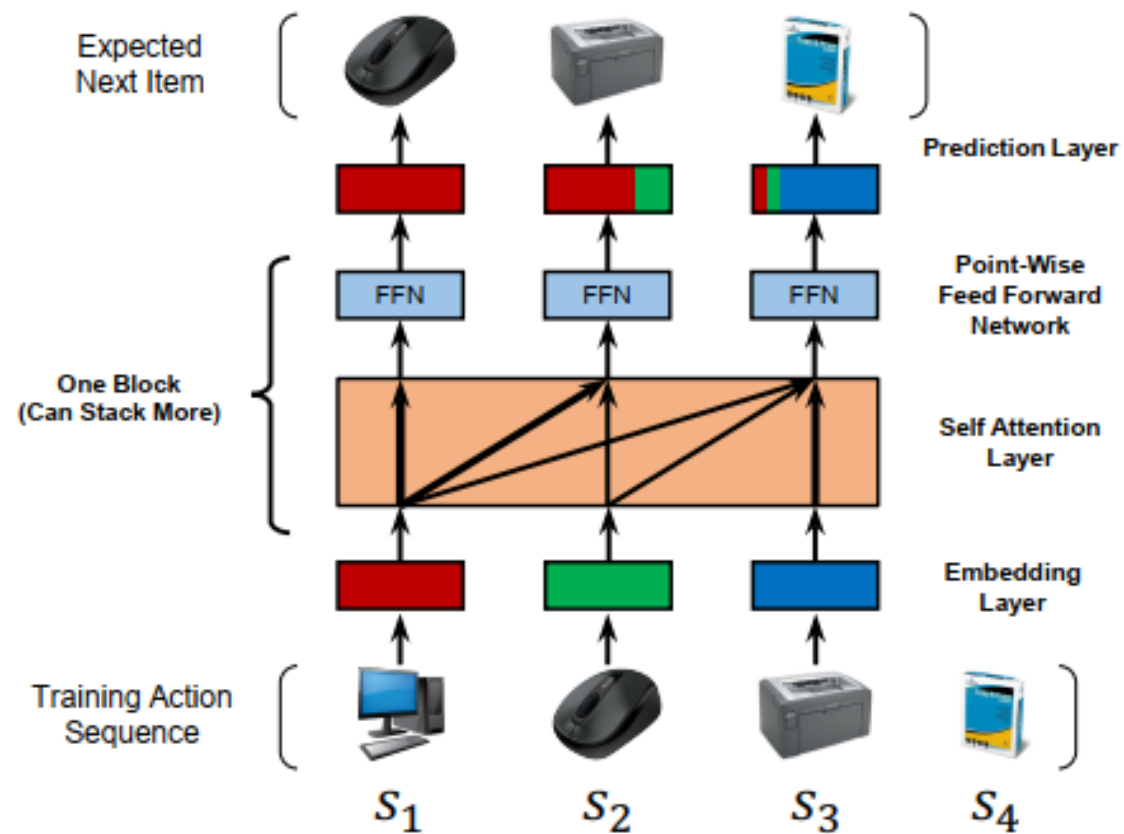
time-series model in spy detect

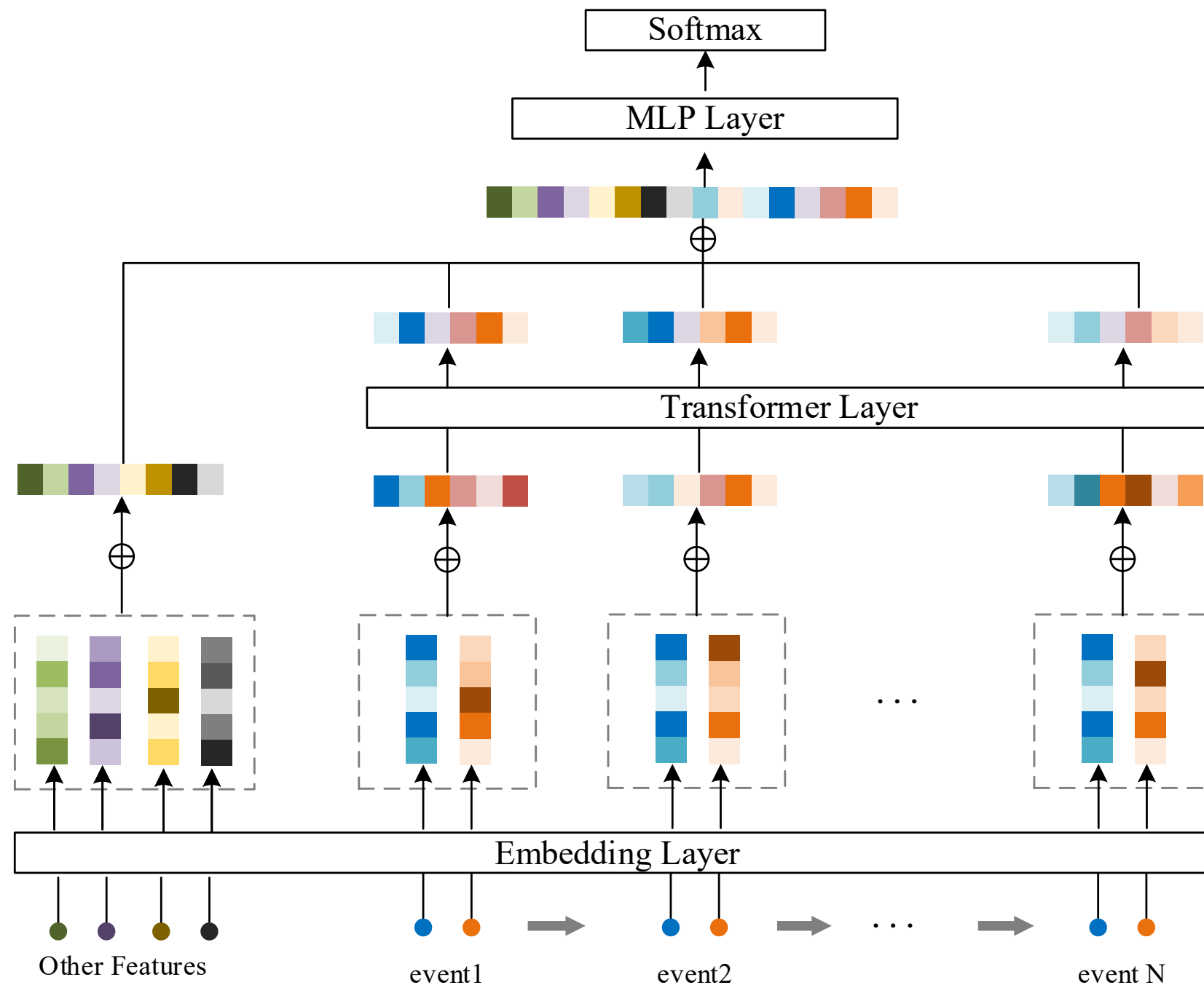Self-Attentive Sequential Recommendation (2018, ICDM)

Train data

origin Log data

↓

extract behaviors

↓

format & rank

↓

truncate

↓

user behavior sequence

Game logs

Extract behavior events

Login   Alliance   PVP

Task   PVE   ...

Align & Rank

```
2020-10-10 13:51:17, login, event_login, 1, null, 0, 2000, null, null
2020-10-10 13:52:20, pve, event_fight,10, null, 0, 2000, occupy successfully
2020-10-10 13:53:24, pve, event_fight, 10, null, 12, 2052, null, occupy successfully
2020-10-10 13:55:11, 10010104, event__task, 59, 2, 12, 2077, 13, null
2020-10-10 13:58:35, apply, event_alliance, 5, null, 24, 2679, null, null
2020-10-10 13:60:56, agreeapply, event_alliance, 5, null, 24, 2679, null, null
2020-10-10 13:62:31, 0, event_mail, 3, null, null, 2679, 靓仔把指挥官？叩叩？
2020-10-10 13:66:02, pve, fight, 10, null, 31, 2783, null,  occupy successfully
....
```

Chat indicates spy

Truncate

```
2020-10-10 13:51:17, login, event_login, 1, null, 0, 2000, null, null
2020-10-10 13:52:20, pve, event_fight,10, null, 0, 2000, occupy successfully
2020-10-10 13:53:24, pve, event_fight, 10, null, 12, 2052, null, occupy successfully
2020-10-10 13:55:11, 10010104, event__task, 59, 2, 12, 2077, 13, null
2020-10-10 13:58:35, apply, event_alliance, 5, null, 24, 2679, null, null
2020-10-10 13:60:56, agreeapply, event_alliance, 5, null, 24, 2679, null, null
....
```

Behaviors sequence belong to a positive label

Time-series model

Behavior sequence
$(event\_id, event\_type)_1$
$(event\_id, event\_type)_2$
$(event\_id, event\_type)_3$
…

Other numerical features
- Role characters (level/money/renown)
- Time passed after the role is created
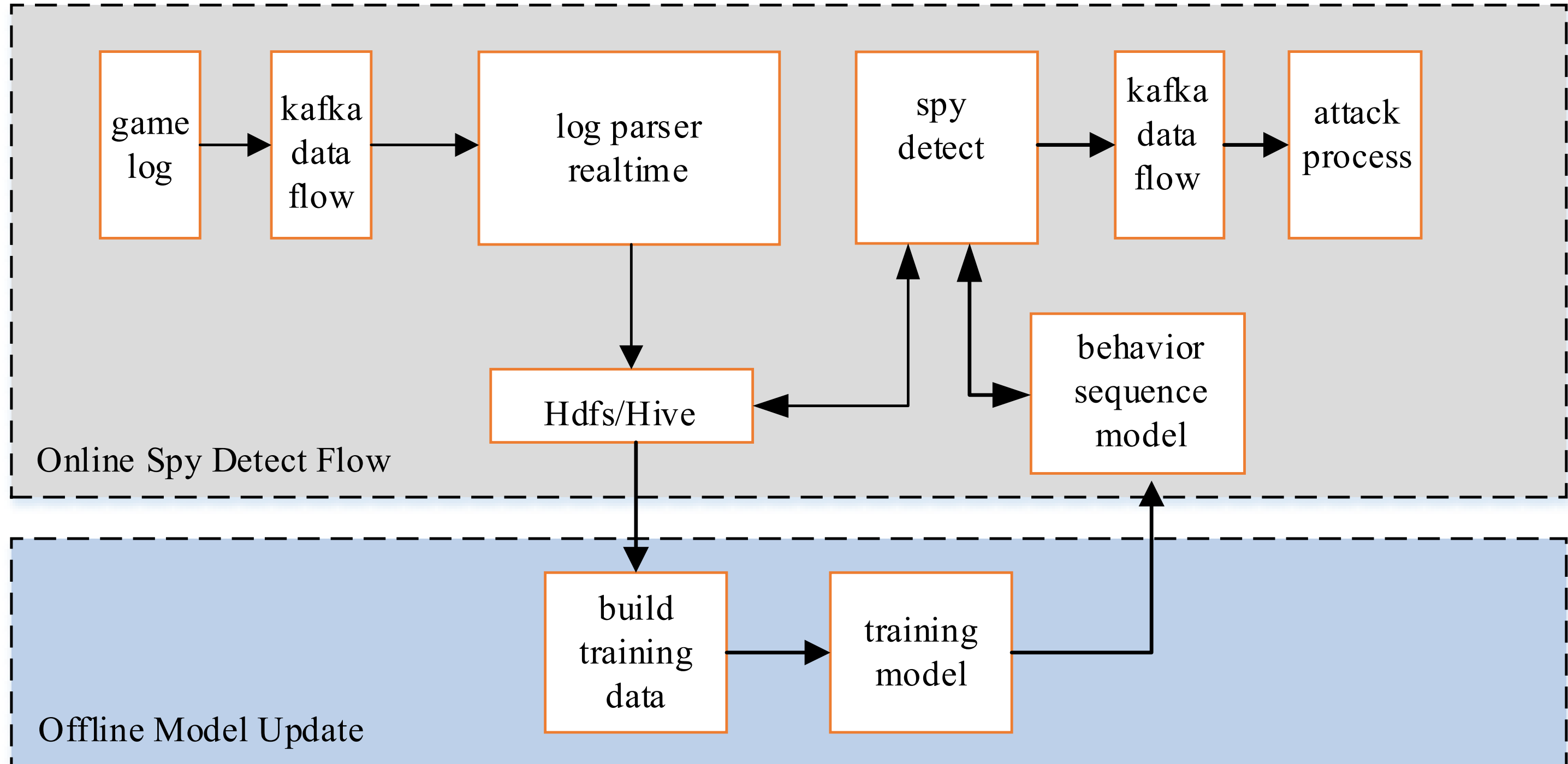- Login period

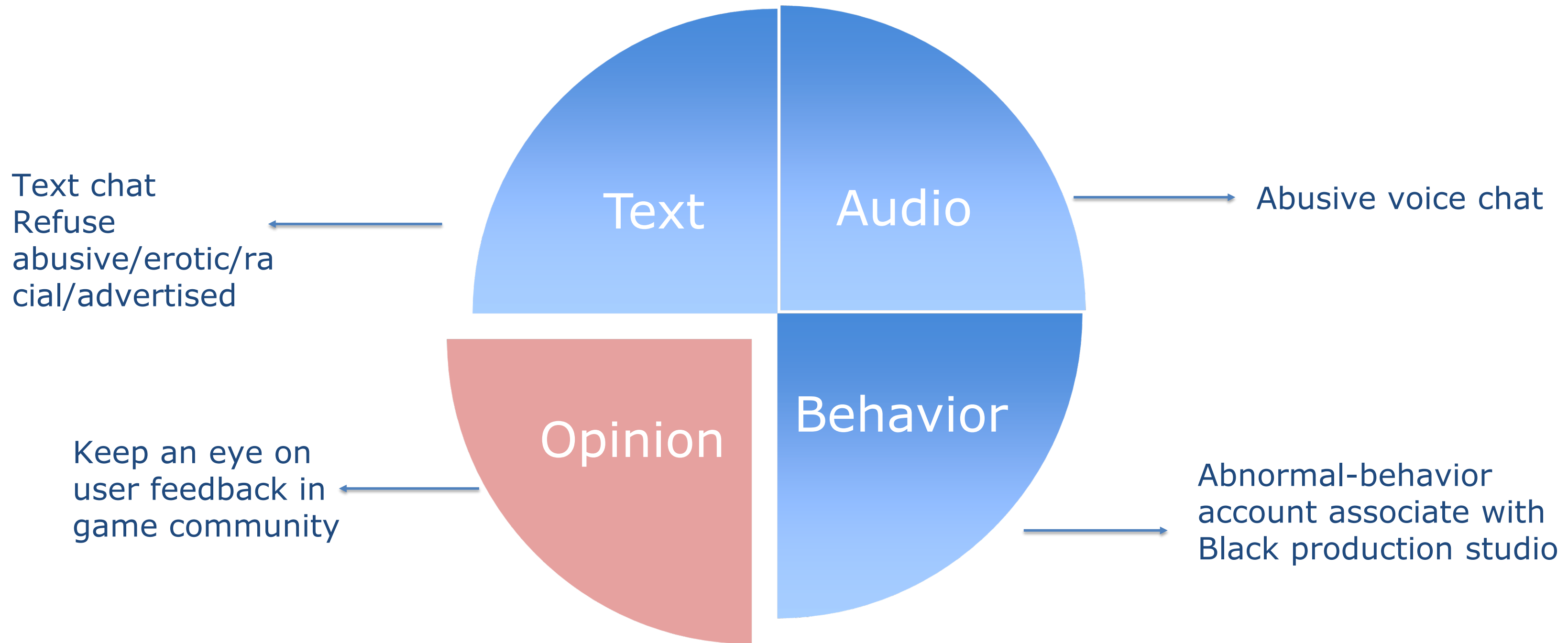Performance
AUC: 99.62%
ACC: 97.79%

Softmax

MLP Layer

Transformer Layer

Other Features

Embedding Layer

event1   event2   …   event N

behavior sequence detection system

Text chat
Refuse
abusive/erotic/ra
cial/advertised

Abusive voice chat

Keep an eye on
user feedback in
game community

Abnormal-behavior
account associate with
Black production studio

Text

Audio

Opinion

Behavior

Game Environment Guardian System X AI

Public Opinion Analysis Framework
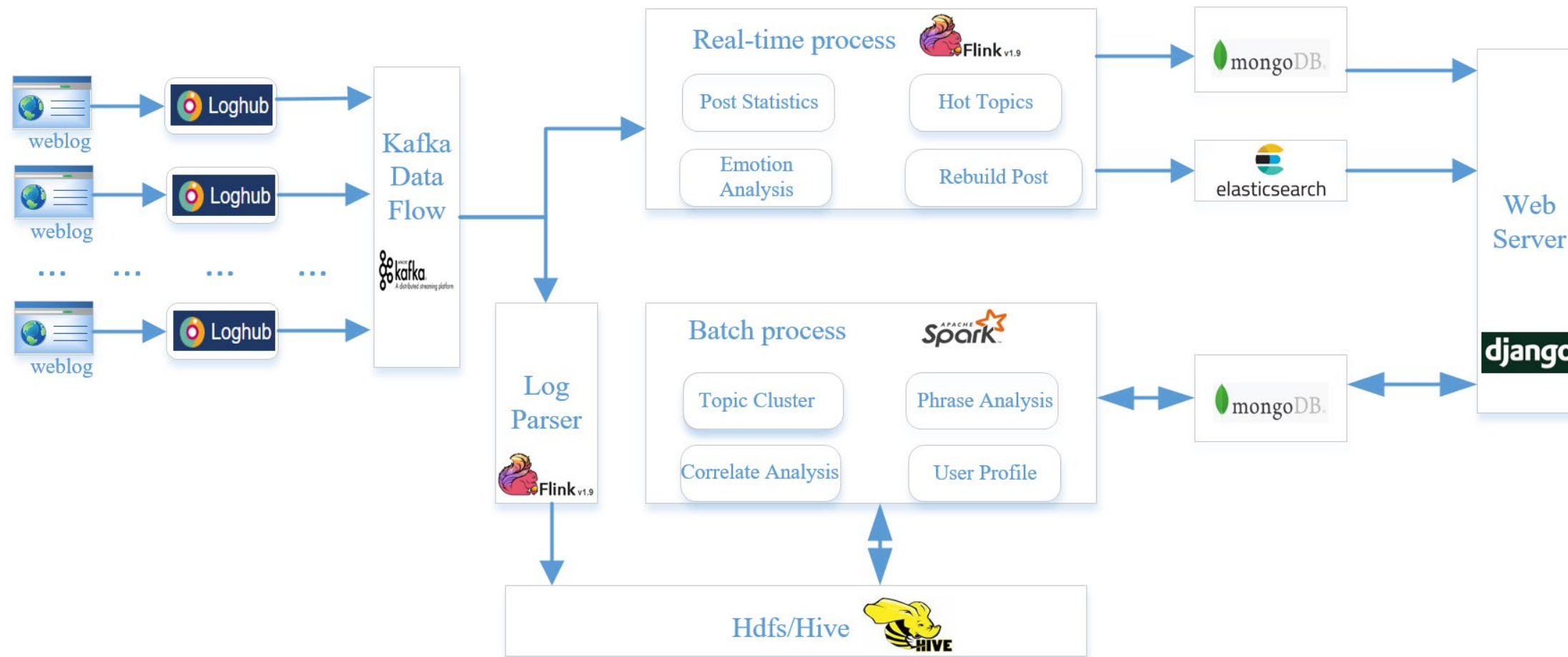
Overall framework of opinion analysis platform

Batch Process Analysis Jobs
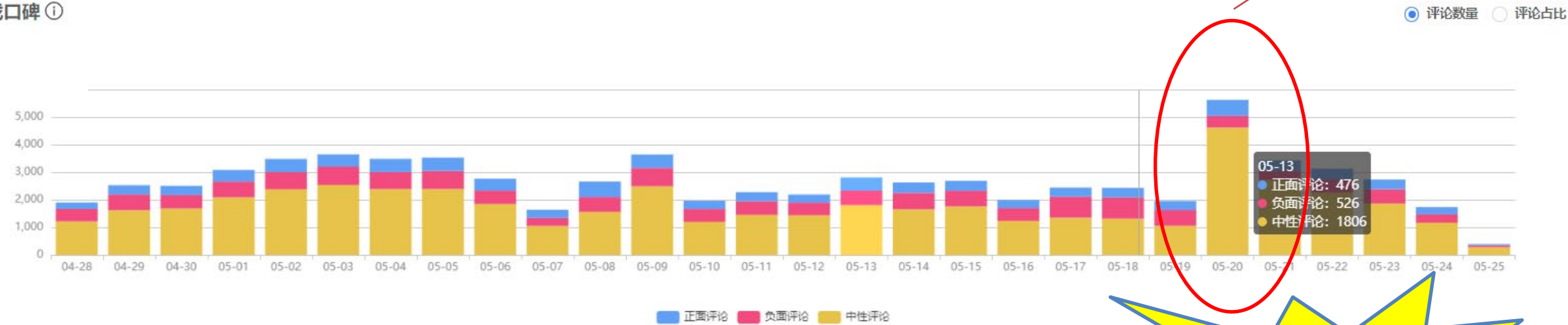- Topic Cluster
- Compare Analysis
- Phrase Analysis
- User Profile
- …

Real-time Analysis Jobs
- Emotion Analysis
- Post Statistics
- Hot Topics
- Rebuild Post
- …

Analysis Report
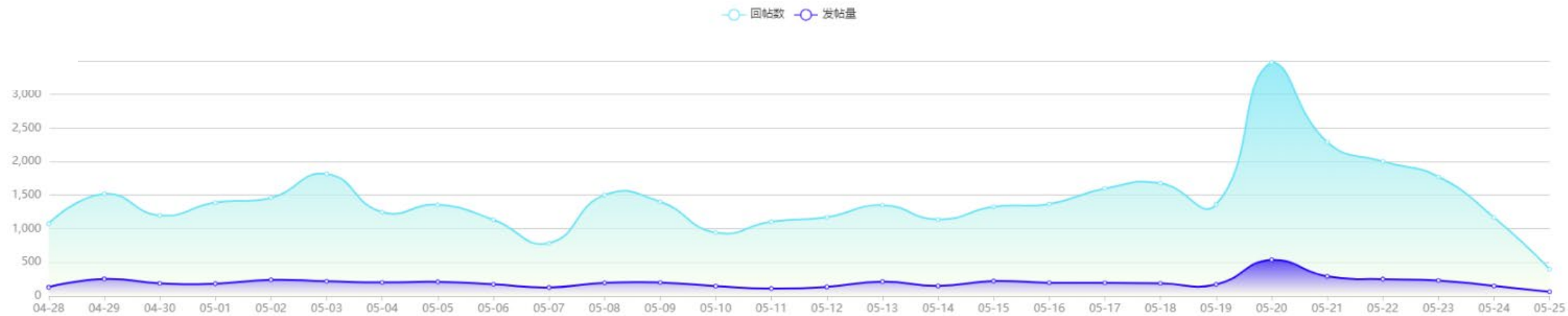
Variation Trend of Emotion
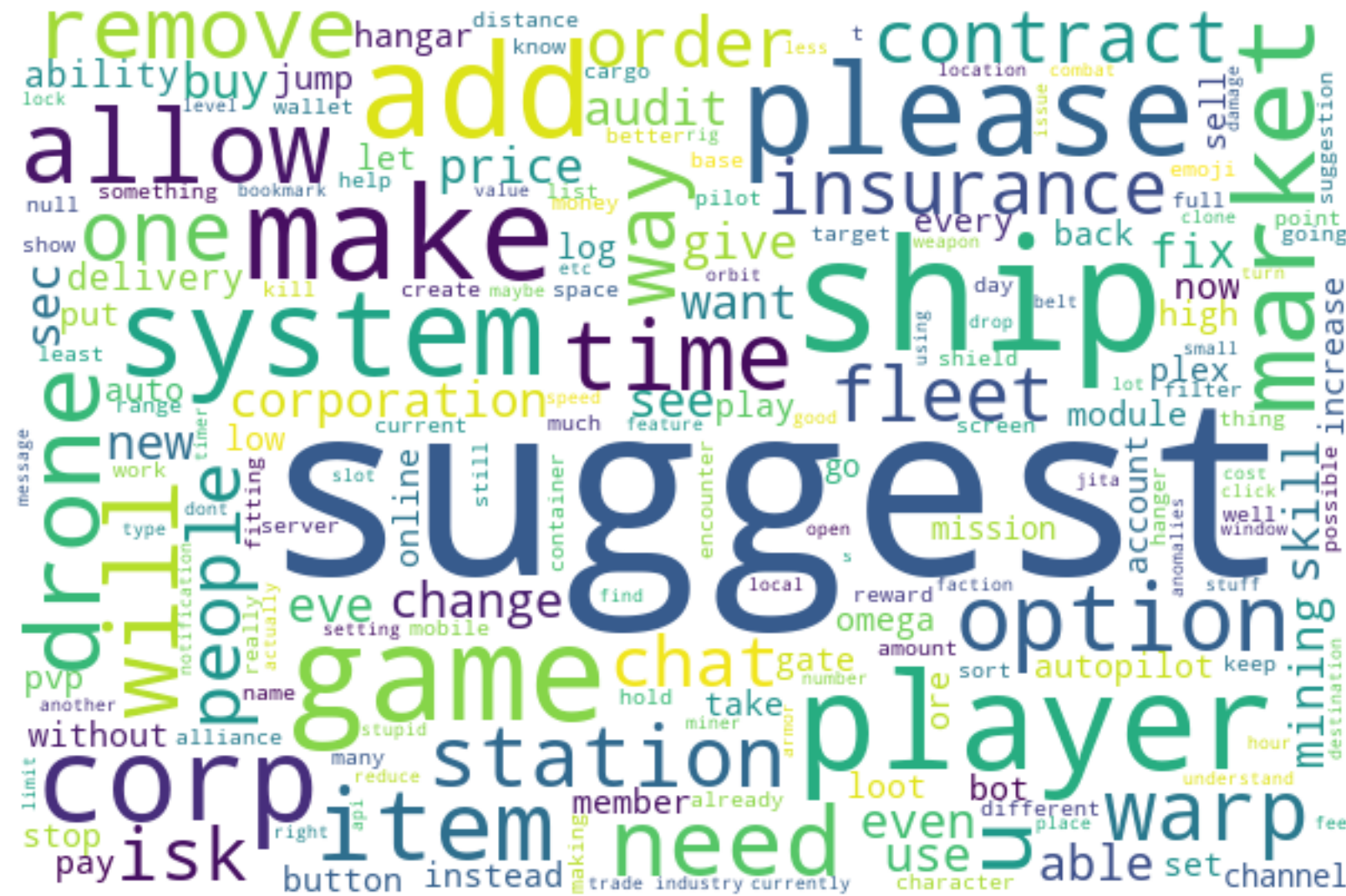

Update a gameplay

Variation Trend of Heat

Welcome!

Analysis Report

Hot Issue Analysis

Analysis Report

Compare Analysis on multiple dimensions
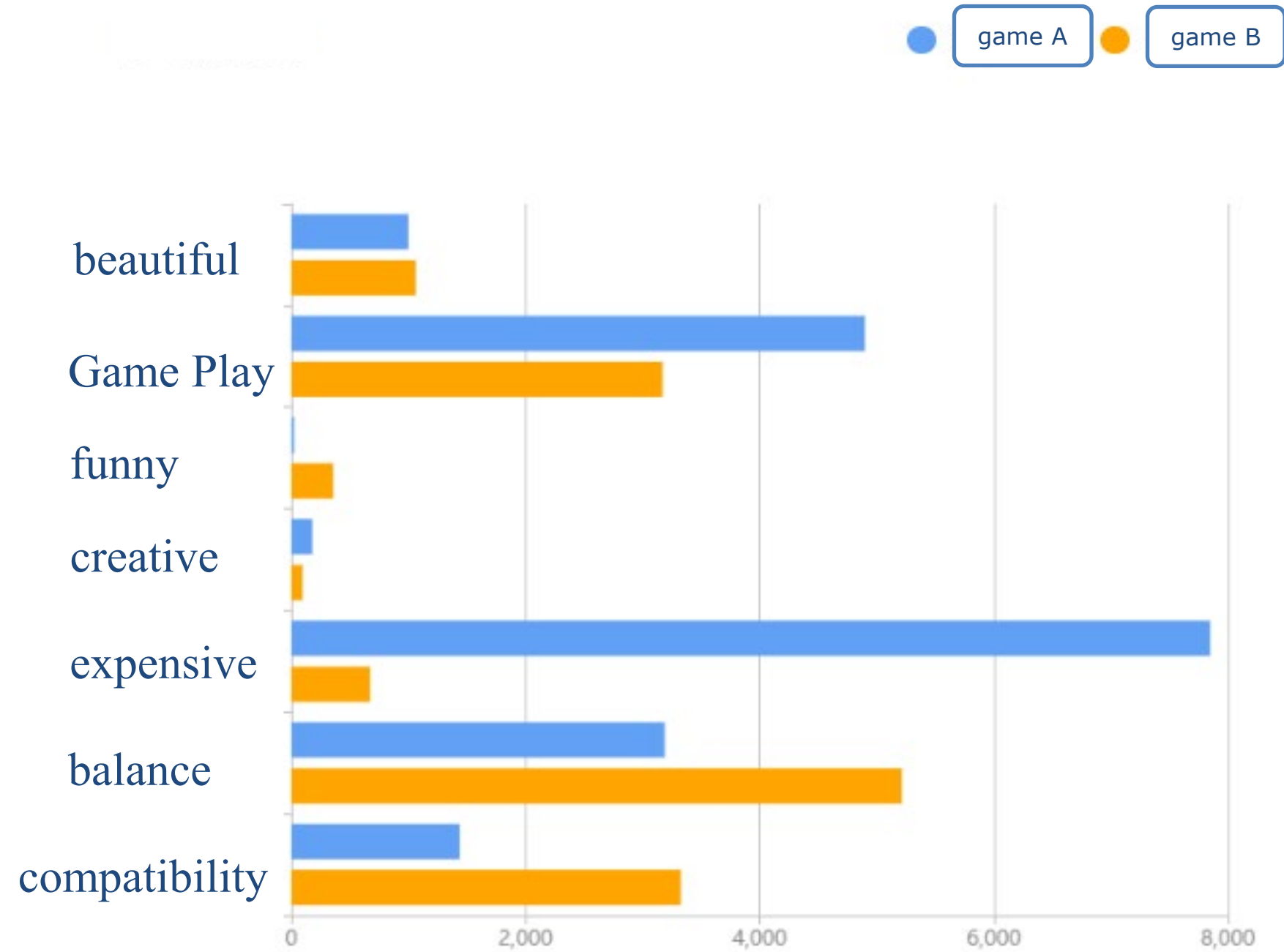
Analysis Report

Topic Cluster

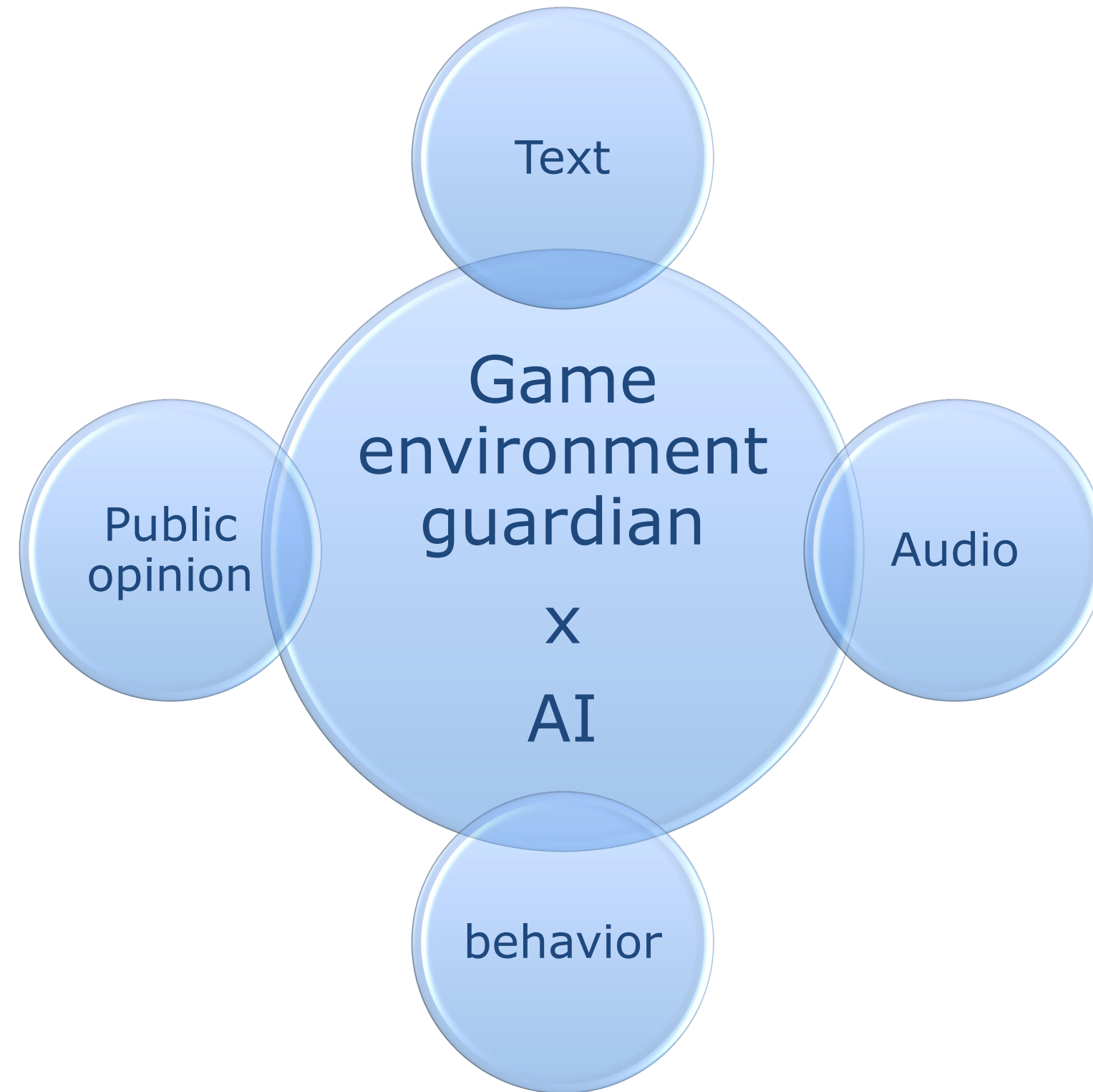| Suggest add feature to delete clone or character.		
suggest permit delete character/clone	2020-08-19	Zellgard#7944
suggest we should be able to delete characters	2020-08-17	Adrian Hue#5722
suggest make it possible to delete characters	2020-08-14	Romario#7984
suggest feature where you can delete your characters	2020-09-26	ylliqm#5859
suggest feature where you can delete your account or character	2020-09-28	ylliqm#5859
suggest option to delete clone/character	2020-08-21	Lancepro#6977
suggest let us delete character	2020-09-15	HavoHej#5783
suggest please add the ability to delete characters	2020-10-17	TaylorRayG#5524
suggest add feature to delete clone/character	2020-08-14	Riru#6036
add option for deleting character	2020-08-13	wintermute#0529
suggest feature for character delete	2020-10-04	ylliqm#5859
…	2020-09-18	ylliqm#5859

Suggest ban bots and players using bots.		
suggest ban coprs using bots n impairprs	2020-09-06	crazybull#0007
suggest maybe ban all those bots	2020-12-01	niksell#6768
suggest ban bots	2020-08-14	Murgen#8510
suggest ban bots before all players will leave the game	2020-11-16	Tamer#8990
suggest ban this bot!	2020-12-19	Ephyon#6868
suggest remove mining bots	2020-10-28	Mr-X-#2186
suggest ban bots and players using bots	2020-11-07	1111001
suggest plz remove bots from the game	2020-11-16	penticrack#9580
…	2020-10-09	Giovanni#6439

Abstract automatic generation   Cluster feedback of players

conclusion

Thank you!