



WUNDERMAN

RecSys-DK
March 9th 2023

WiFi
WPP_Guest
“guest@wpp”

18:00	Opening by Wunderman
18:10	Jobmatch Project v/ Mesut Kaya
19:00	Break
19:10	Wunderman v/ Morten Arngren and the gang
20:00	Intro Round v/ Kim Falk
20:15	Networking & Drinks
	:



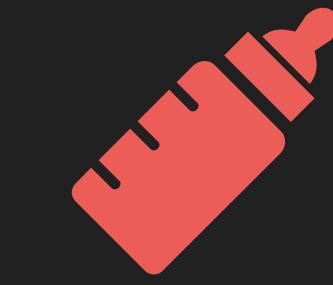
Guest cards
leave at desk
or
give to me

.....



Meat
Vegetarian
Vegan
Lactose-free

.....



Coffee
Tee
Soda
Beer
Wine
Snacks



WUNDERMAN

A
Media Agency



1960s

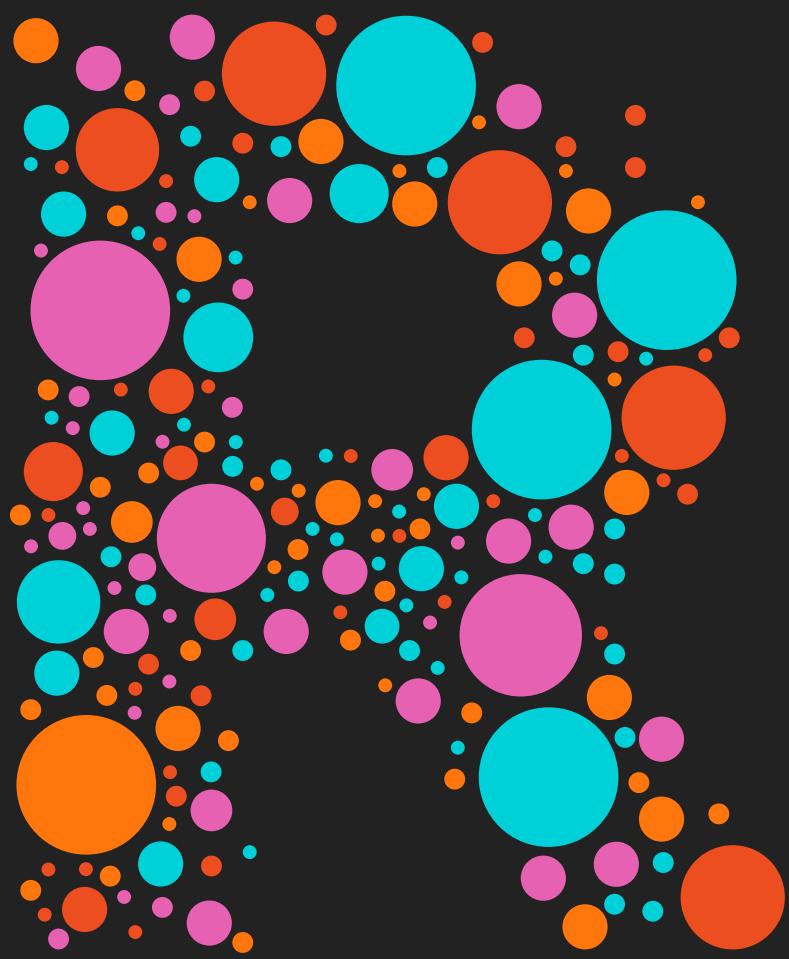


2020s



WUNDERMAN

A
Media Agency



RESOLVE

A privacy-first approach
in Digital advertisement

Data Science

Applied
Research
Team

Federated
Learning Team

Graph Team

Cloud

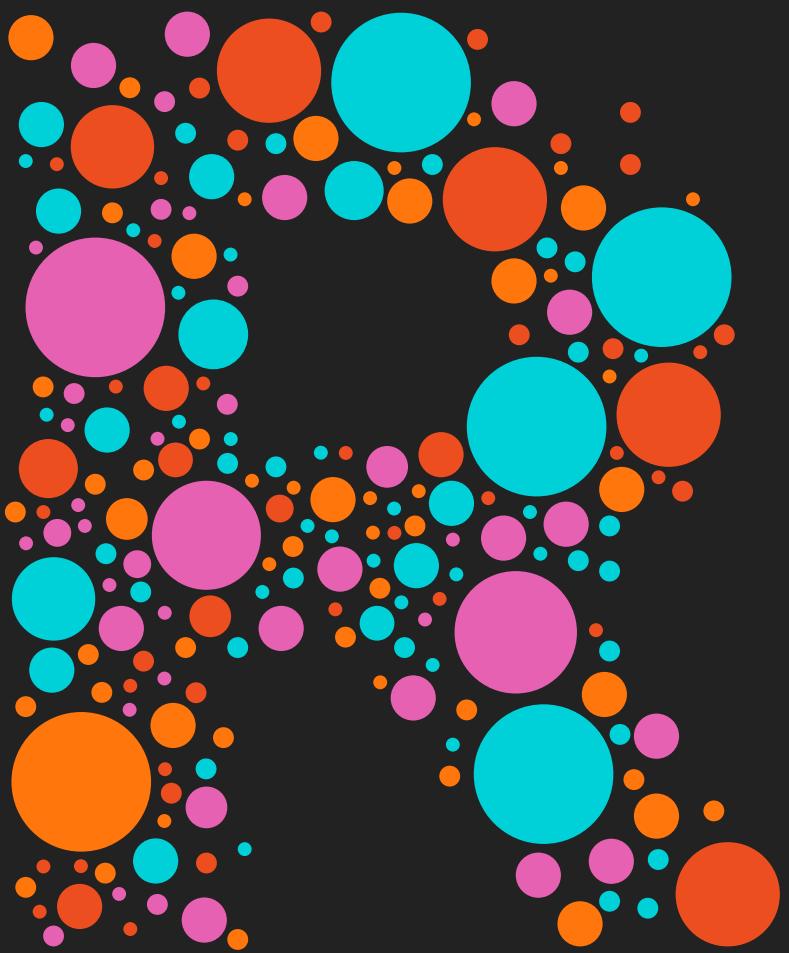
Platform

NLP



WUNDERMAN

A
Media Agency



RESOLVE

A privacy-first approach
in Digital advertisement

Recommender
systems

NLP

Topic modeling

Textual
Embeddings

Node2Bits

GraphSage

Federated
Learning

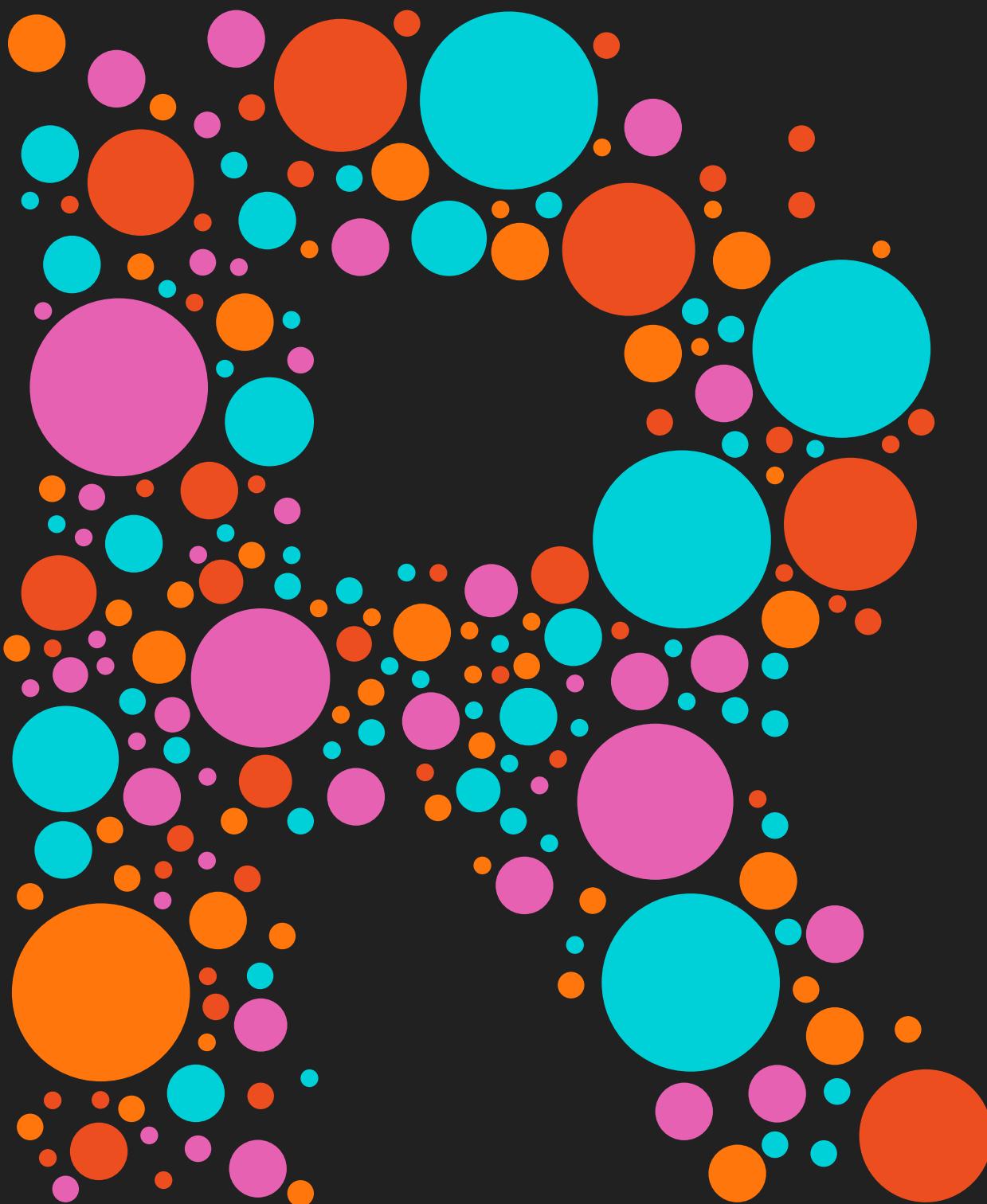
Look-Alike modeling

User
Profiling

ID Stitching

Jobmatch Project

Mesut Kaya



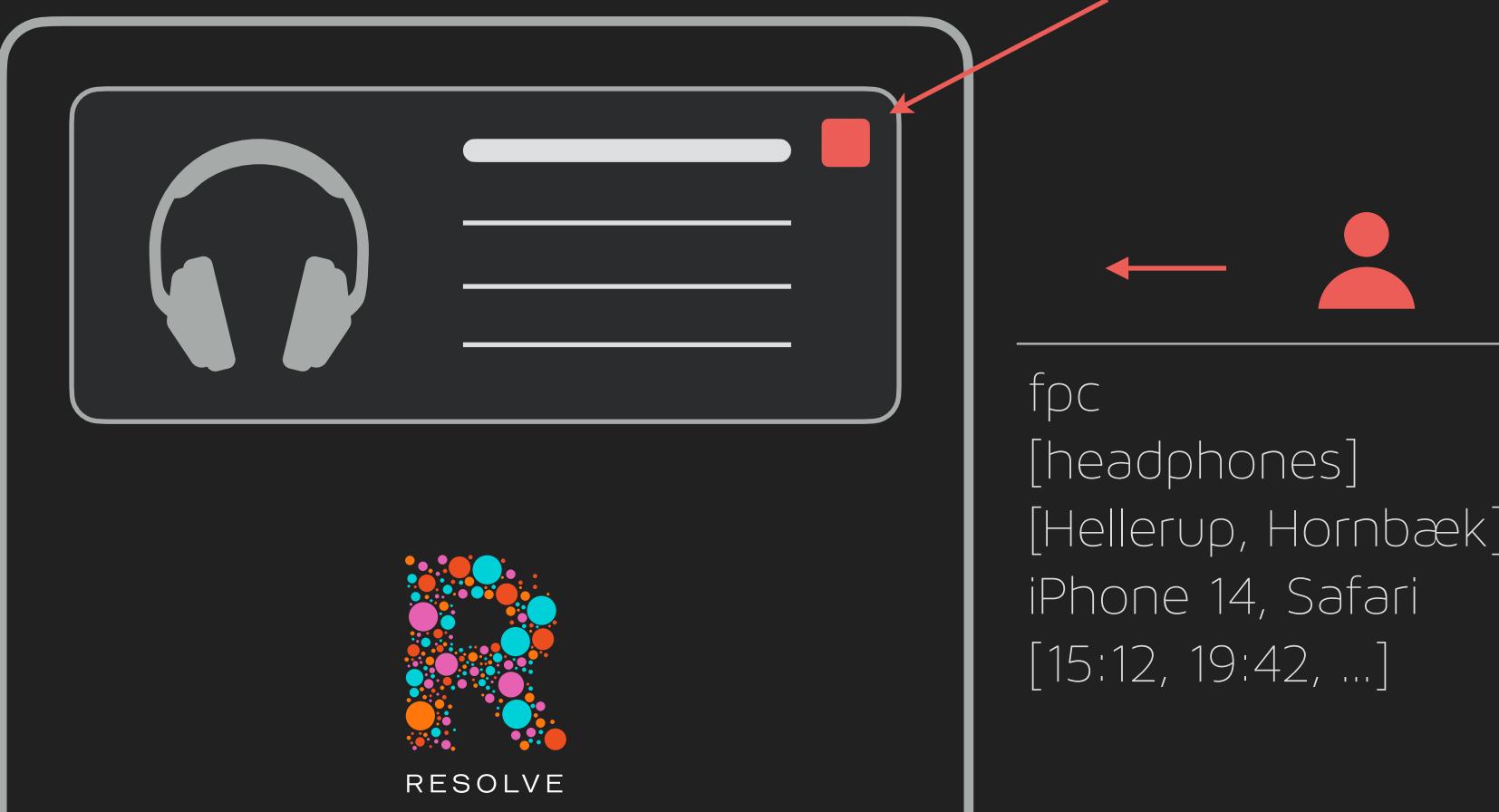
RESOLVE



Resolve AdTech Ecosystem

...

Advertiser



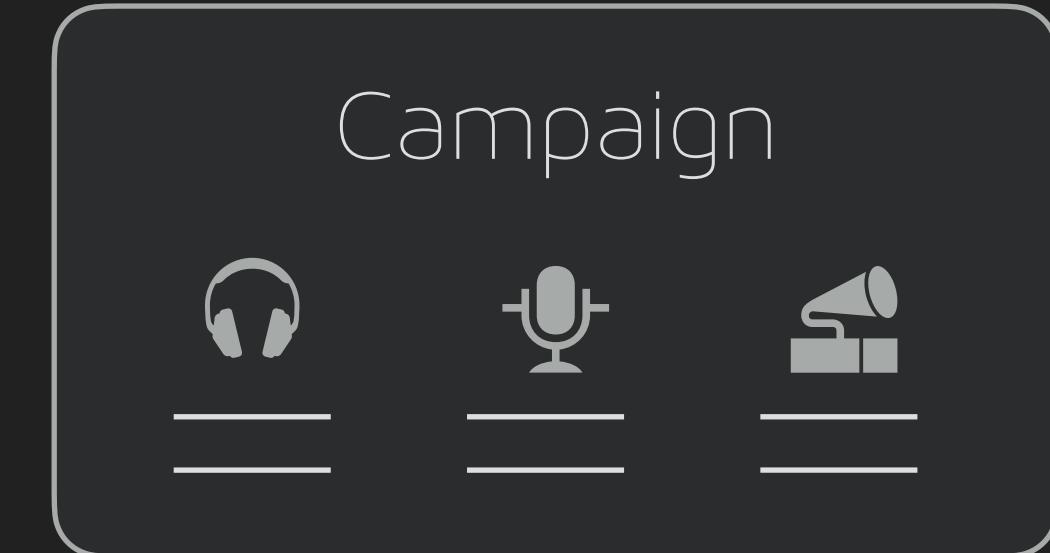
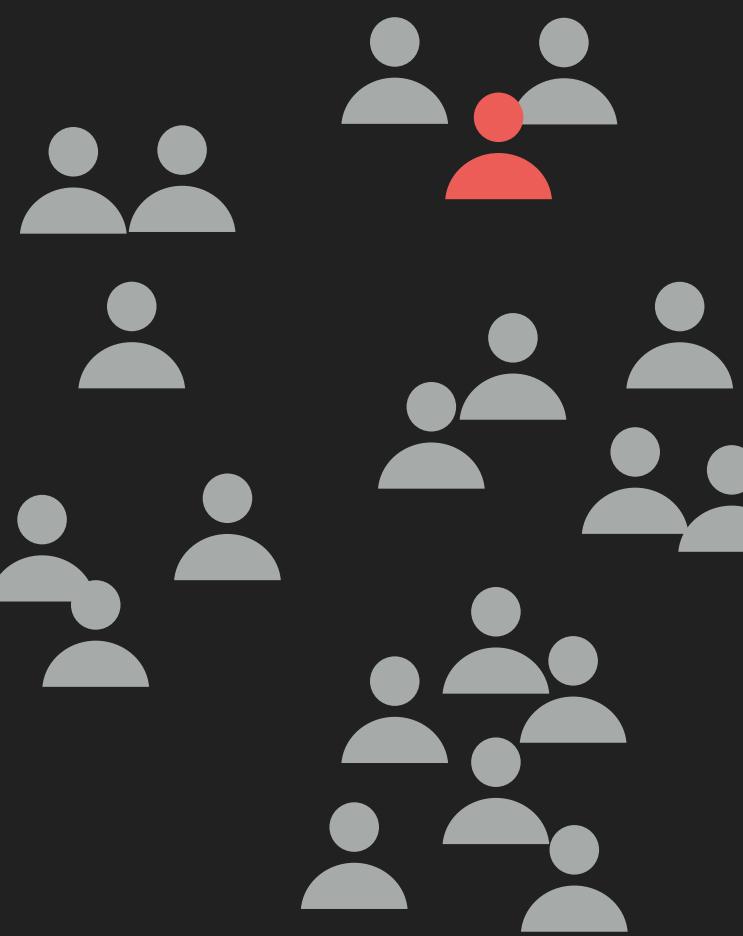
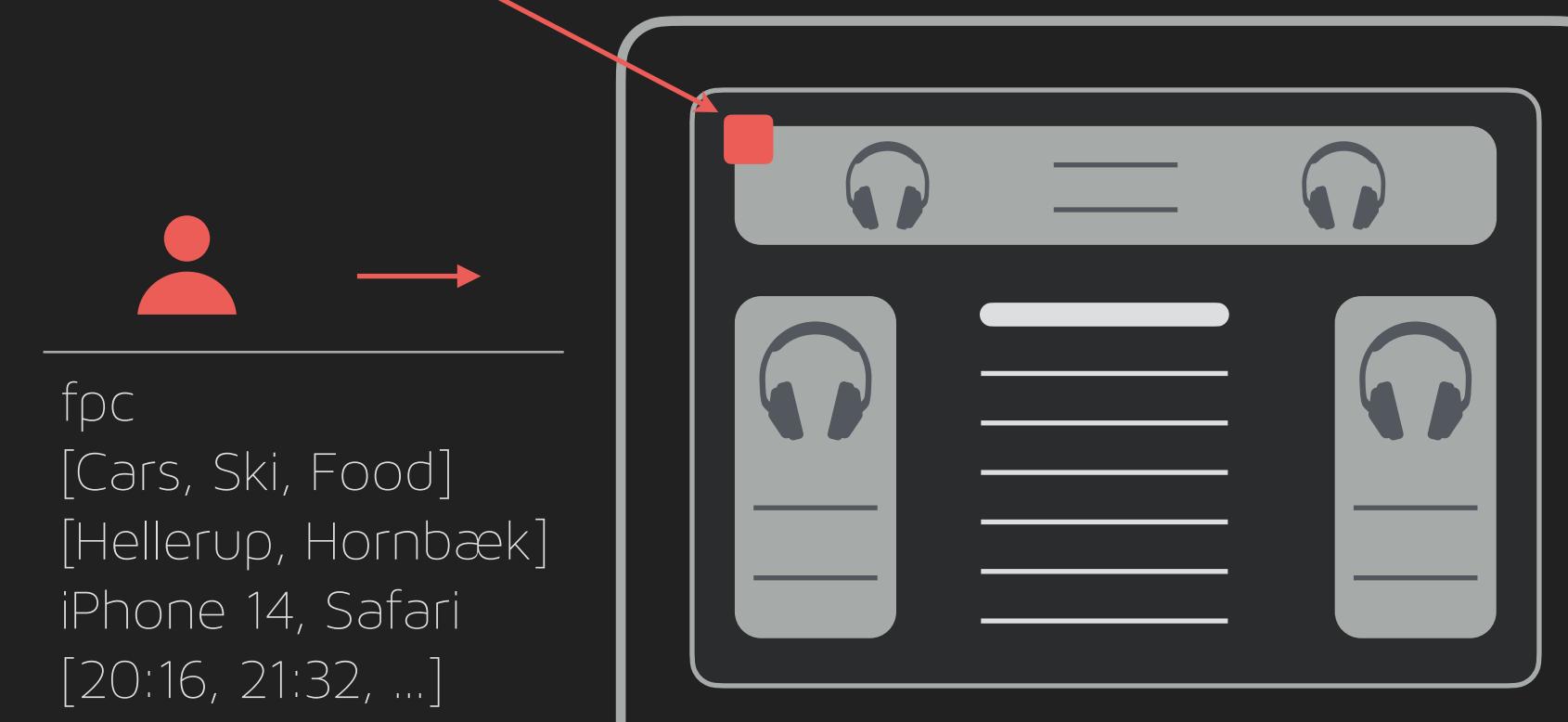
Tracking pixel



3rd party cookie

Tracking pixel

Publisher



DSP



SSP

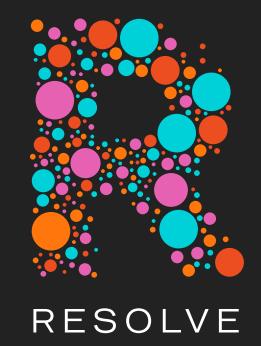
⋮

⋮

⋮

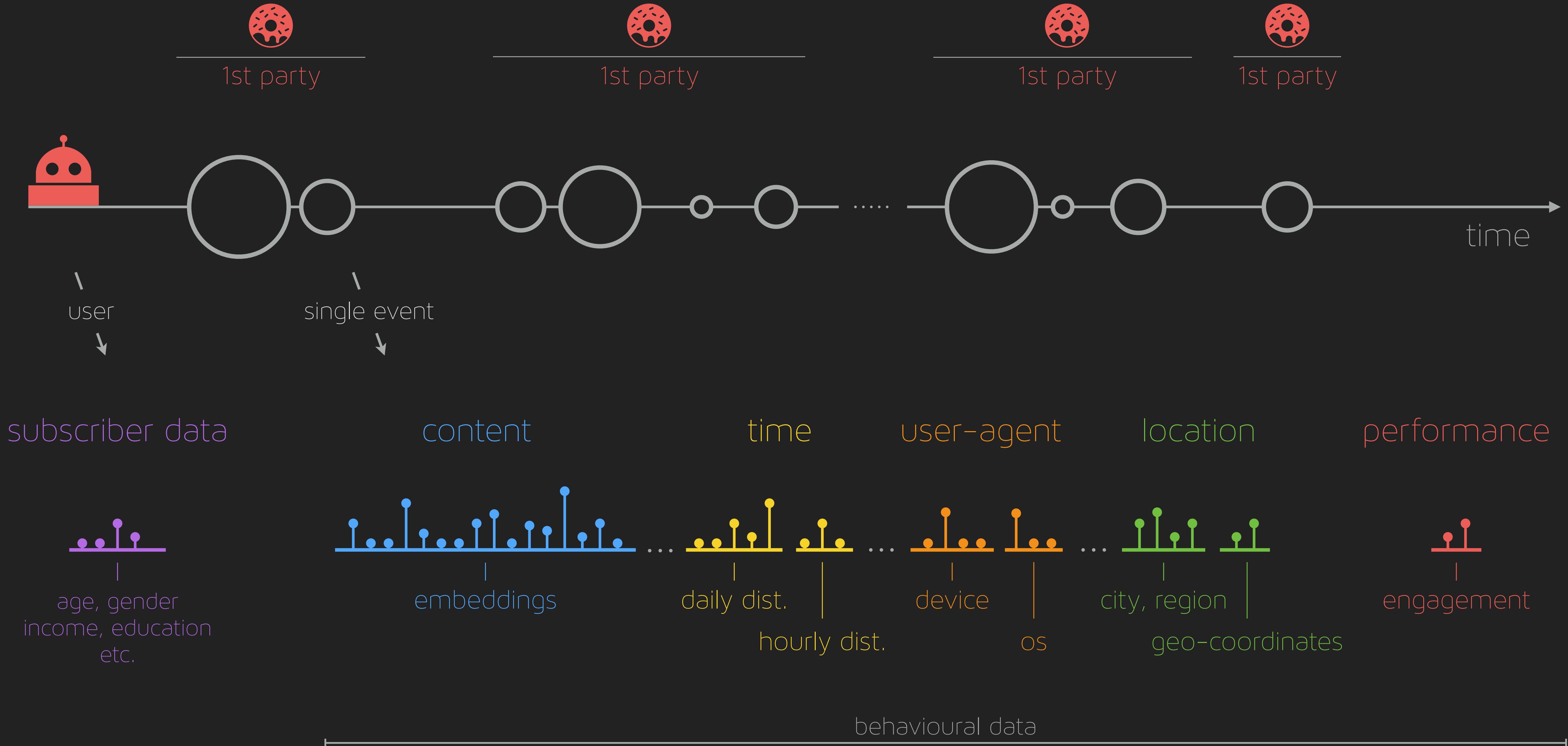
⋮

⋮



User History

1st party data





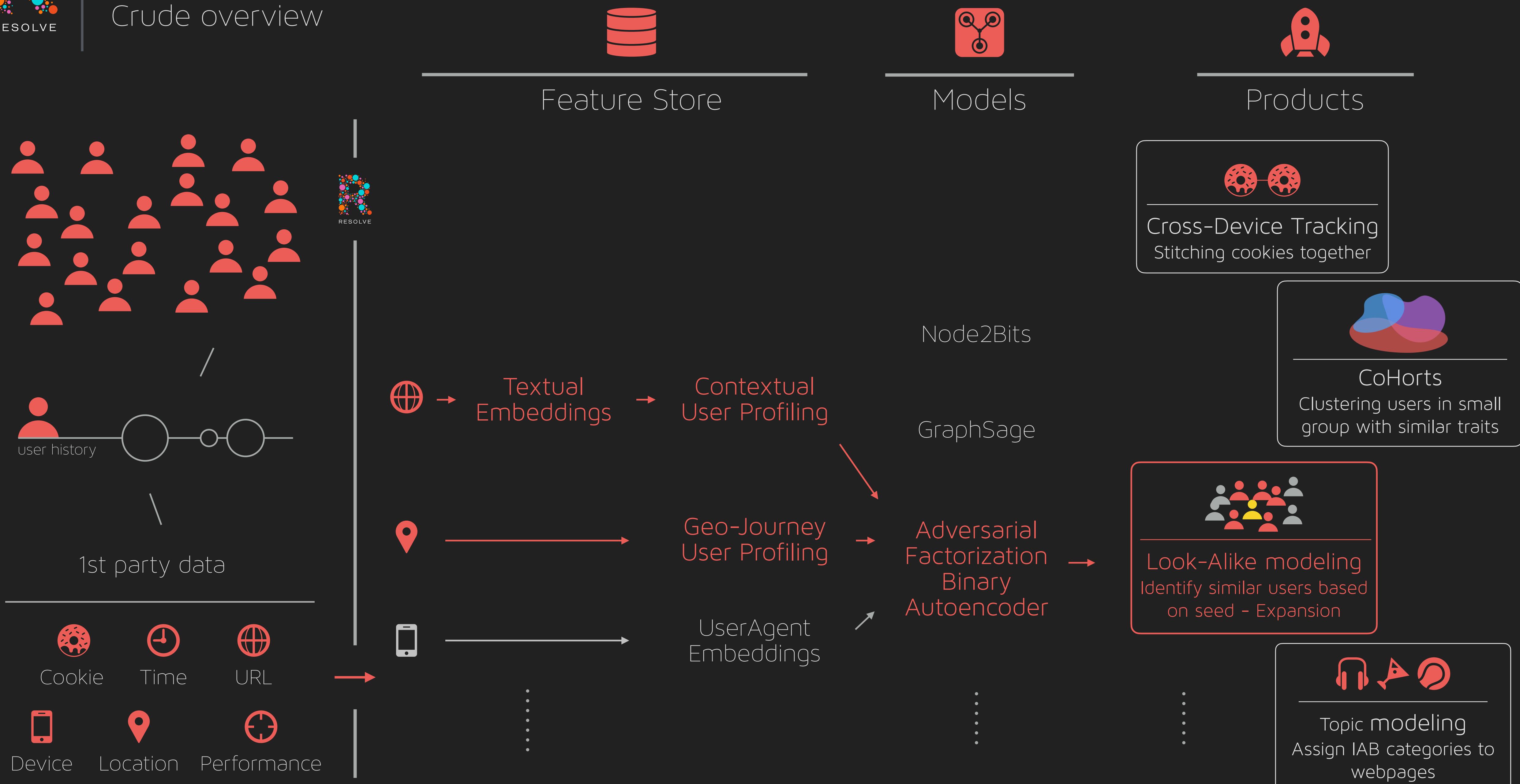
Hexagon Geo-Locations

Random user's journey



Resolve Data Science

Crude overview





Textual
embeddings



Feature Extraction



Textual Embeddings

- URL-level embeddings
- Embedding textual webpage information



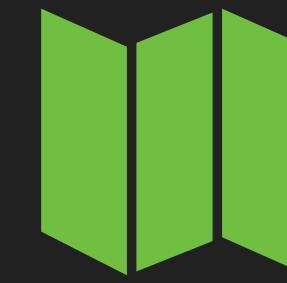
Contextual

Focuses on the article reading pattern of users



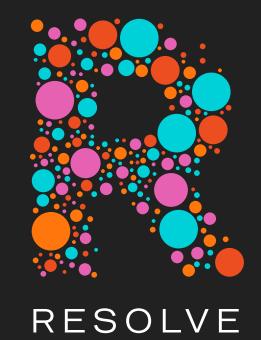
User Profiles

- User-level embeddings
- Embeddings of user behaviour
- Sequential models

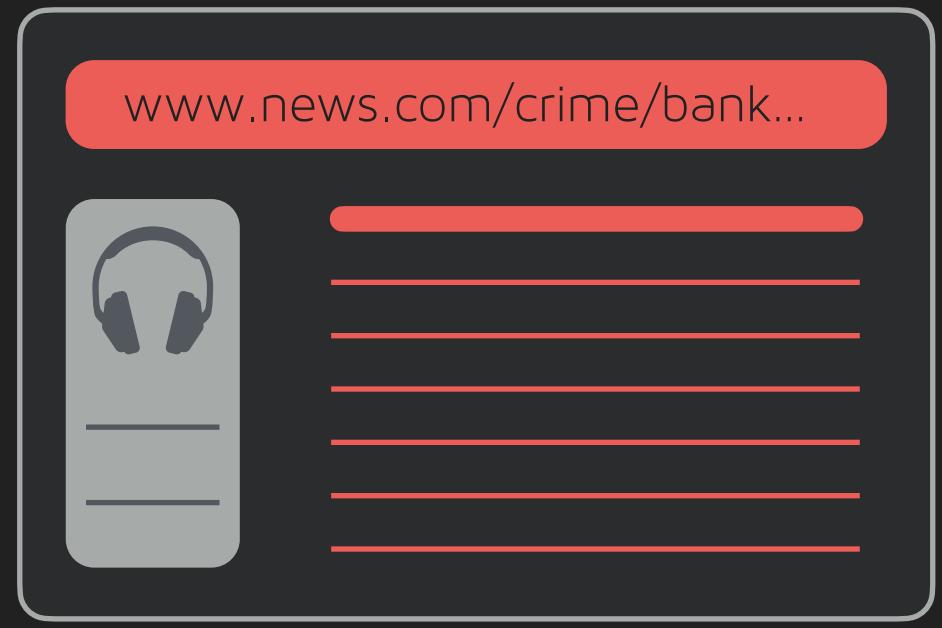


Geo-Journey

Focuses on the movement patterns of users

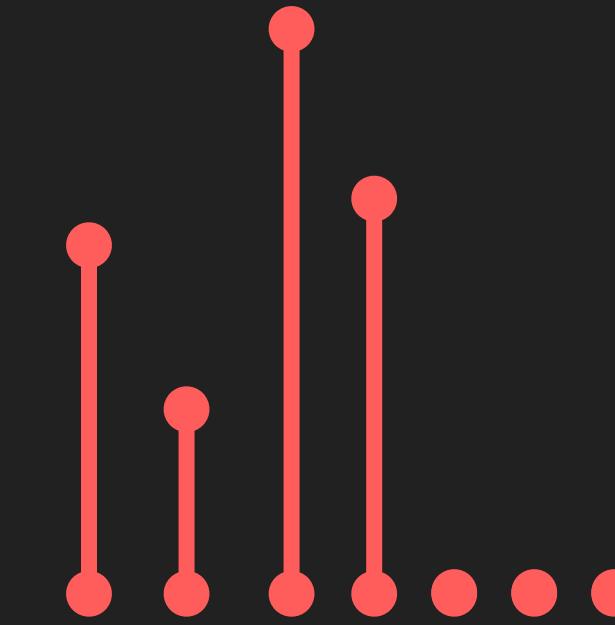


What is a textual embedding?

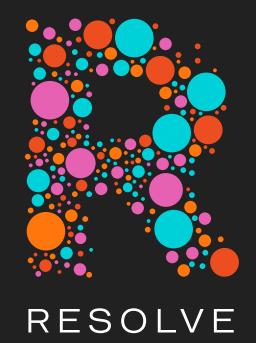


Textual webpage information

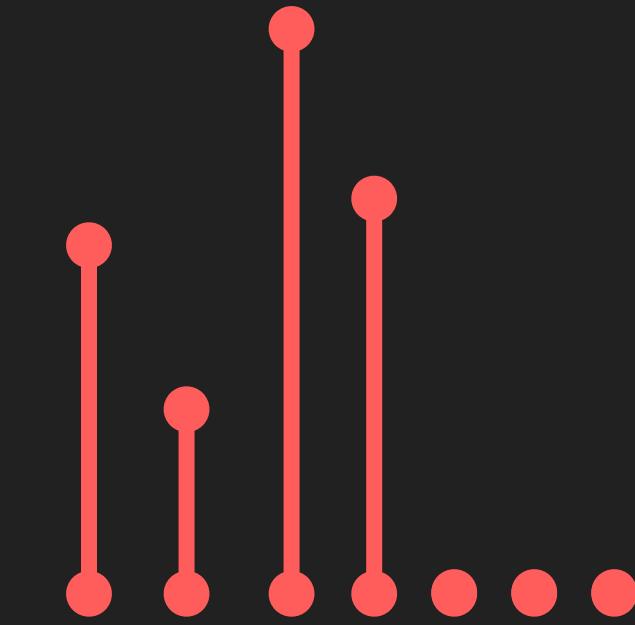
- URL
- Page title
- Page text
- Image text



[0.39, 0.31, 0.25, ...]



URL embedding

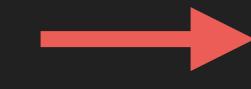


[https://domain.dk/fritid/musikinstrumenter/guitar/...](https://domain.dk/fritid/musikinstrumenter/guitar/)



[0.39, 0.31, 0.25, ...]

[https://domain.dk/fritid/musikinstrumenter/bas/...](https://domain.dk/fritid/musikinstrumenter/bas/)

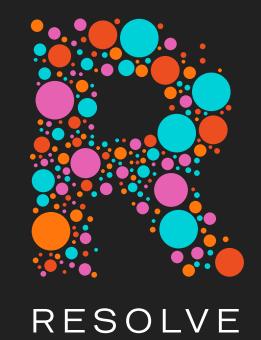


[0.39, 0.30, 0.20, ...]

[https://domain.dk/camping/campingvogne/...](https://domain.dk/camping/campingvogne/)



[0.89, 0.55, 0.01, ...]



Preprocessing

Beam events:

User ID	URL	Page title	...
123	... /musikinstrumenter/guitar/...	Mega fed guitar manner	...



Split URL:

URL	Taxonomy 0	Taxonomy 1	Taxonomy 2	Page title
... /musikinstrumenter/guitar/...	fritid	musikinstrumenter	guitar	Mega fed guitar manner

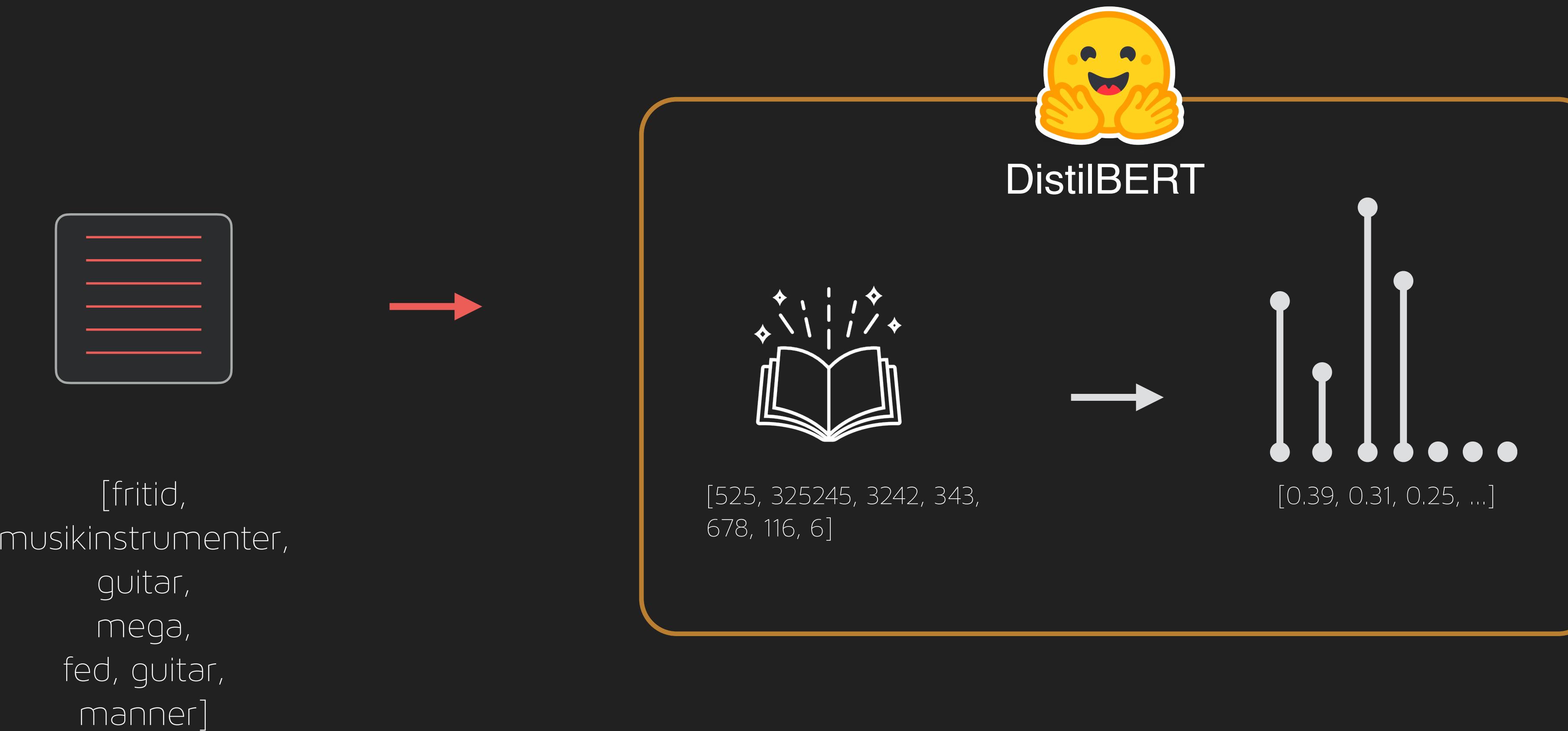


Combined text:

URL	Combined text
... /musikinstrumenter/guitar/...	fritid, musikinstrumenter, guitar, mega, fed, guitar, manner



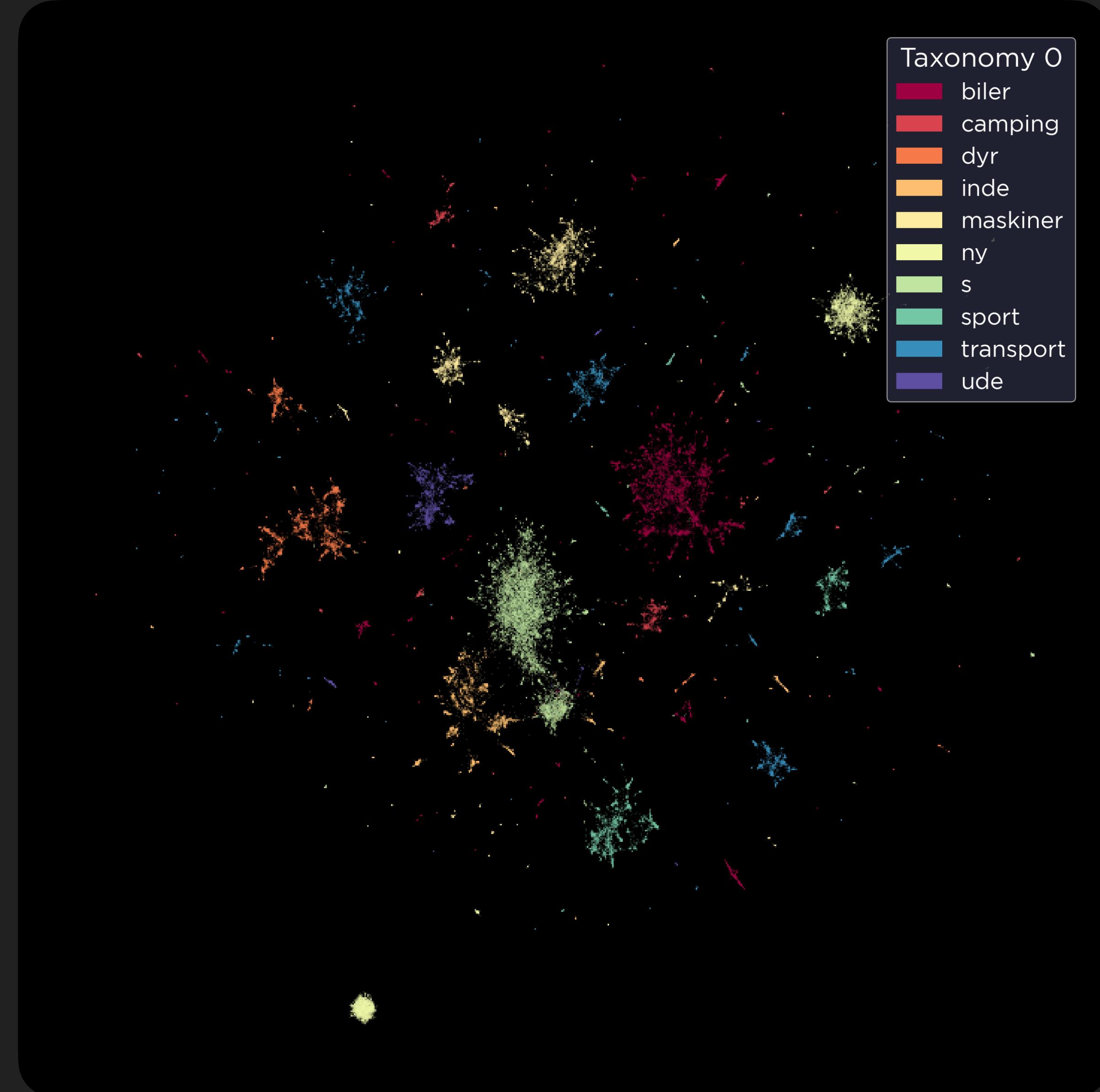
Tokenization and model

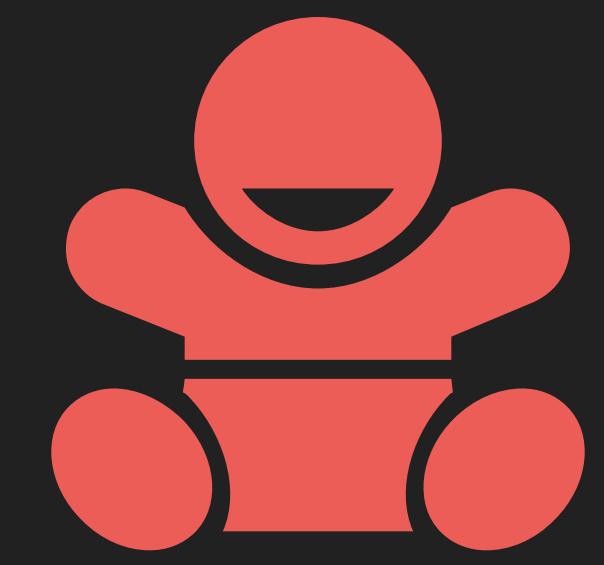




Use case: Gul & Gratis

**GUL OG
GRATIS.DK**

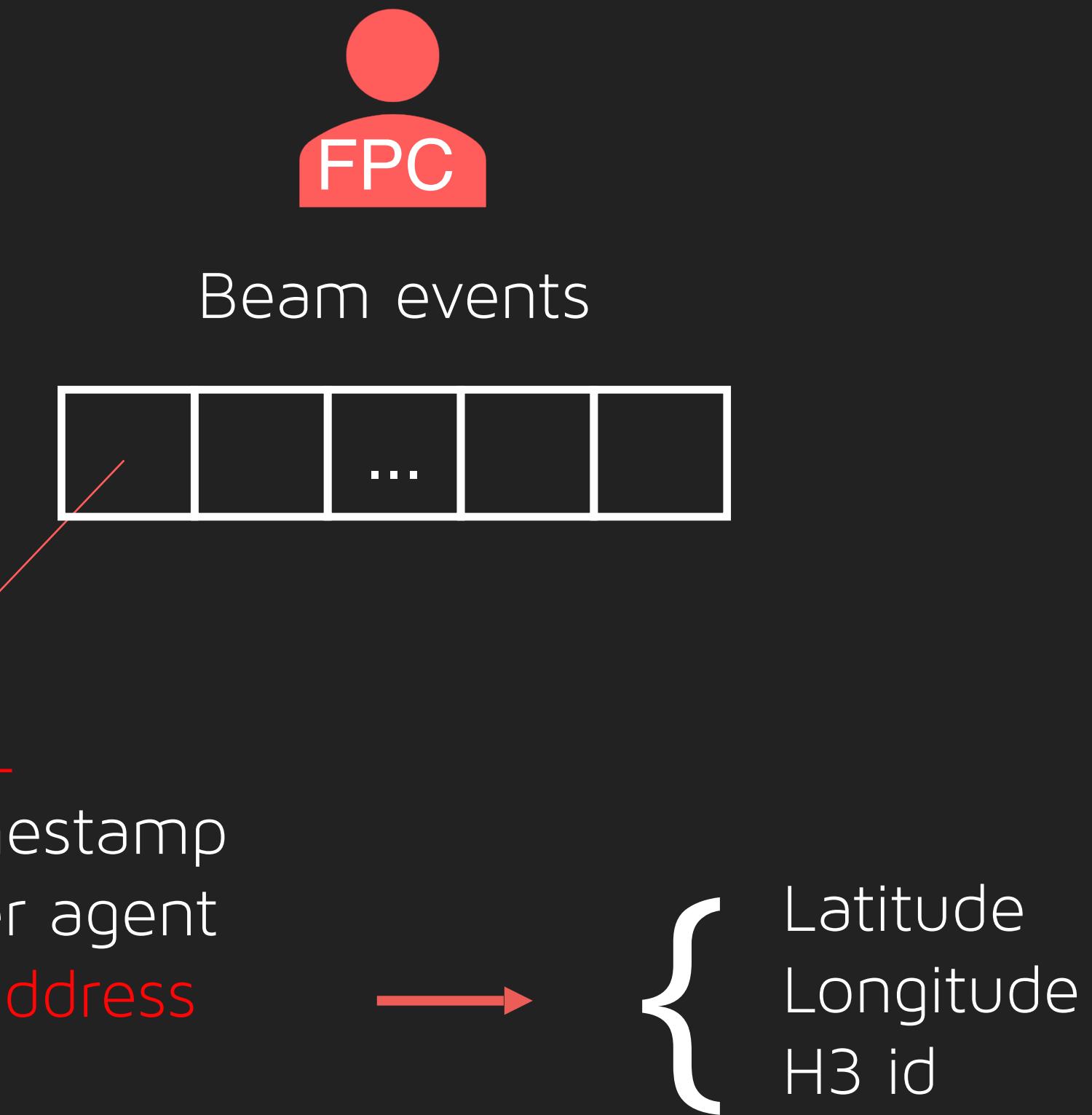




User Profiling



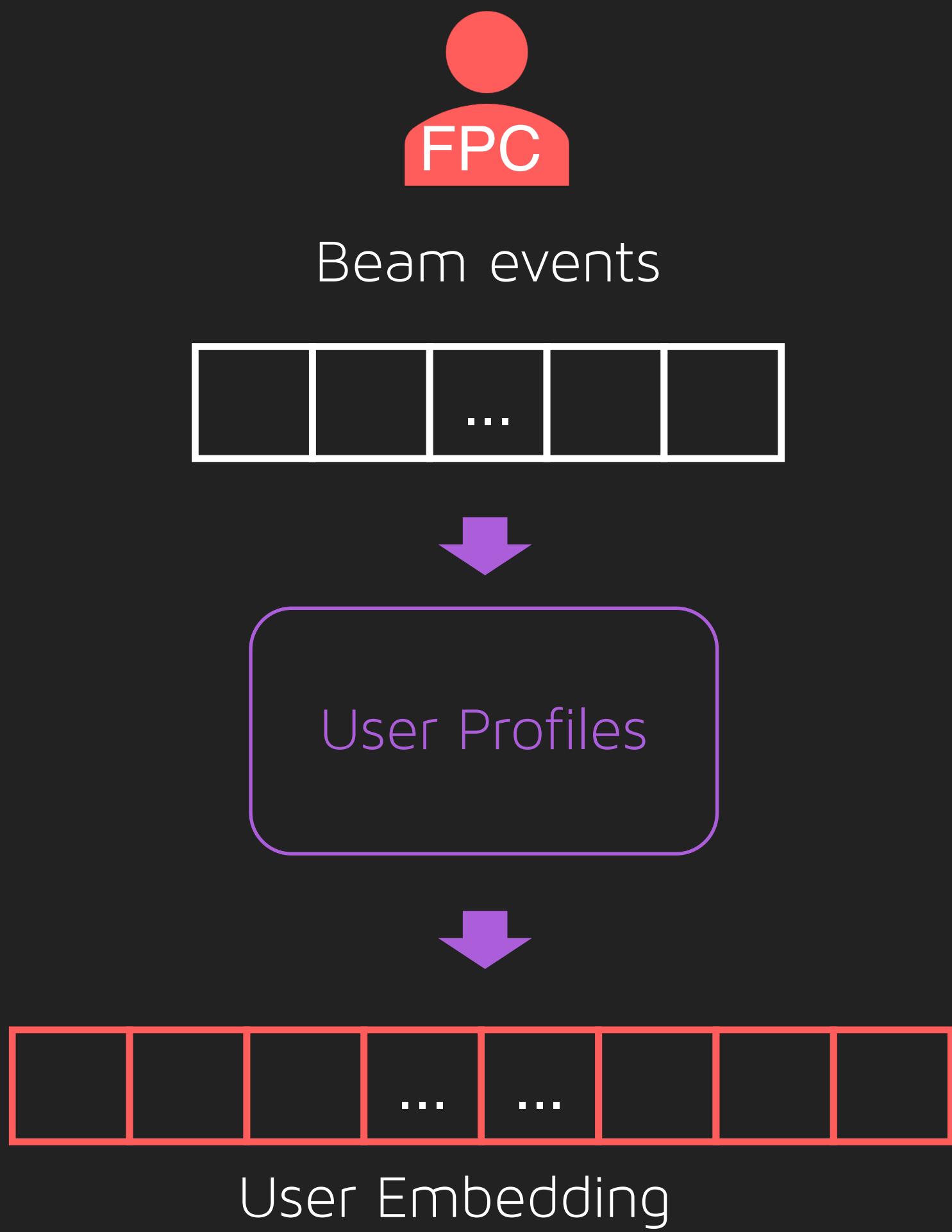
Beam-Events





Feature extraction

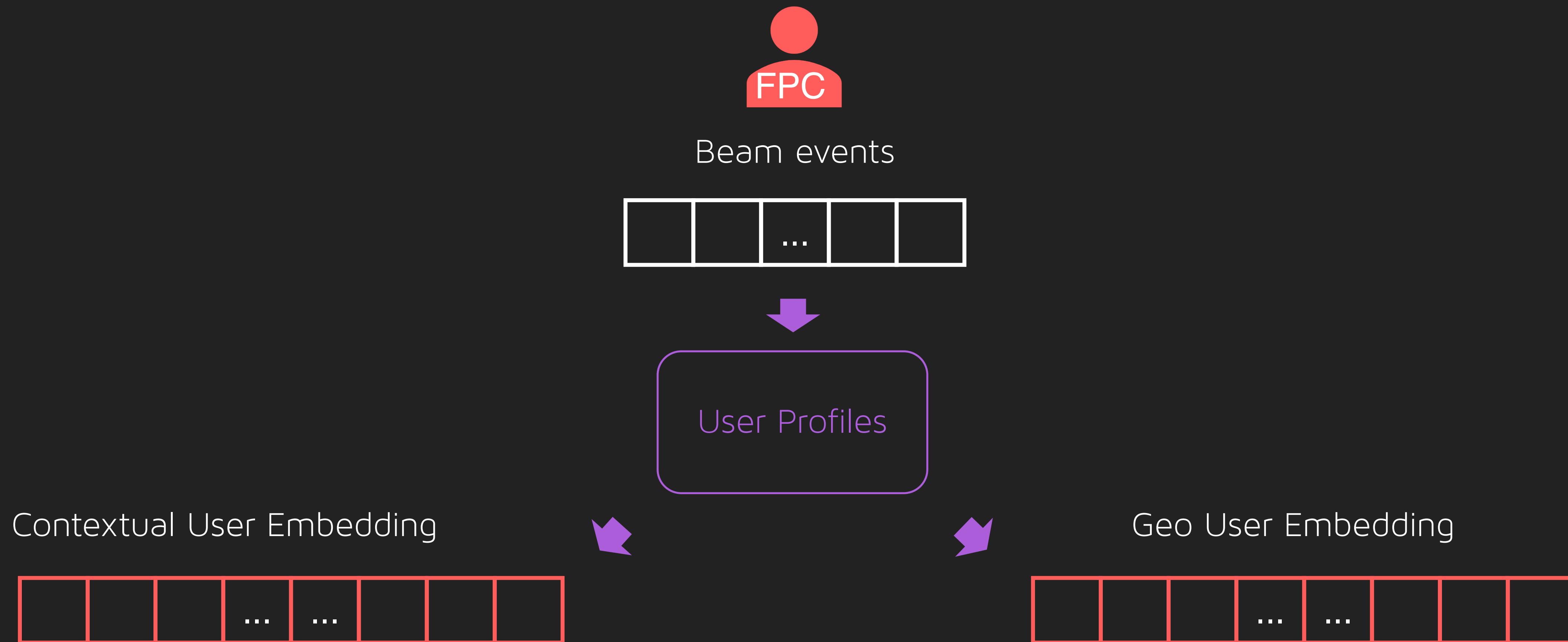
User Profiles

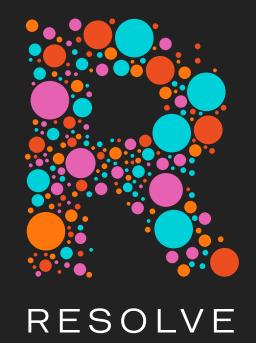




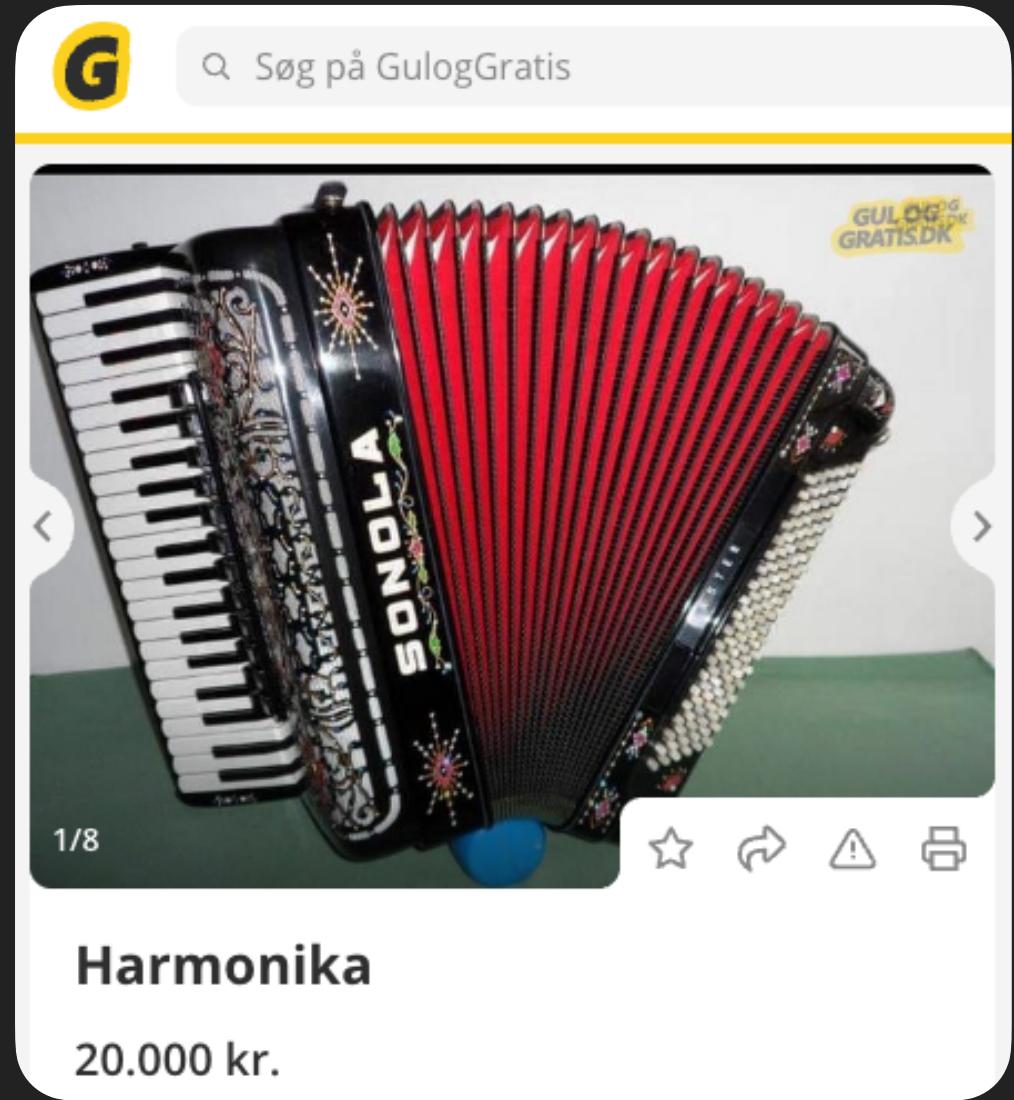
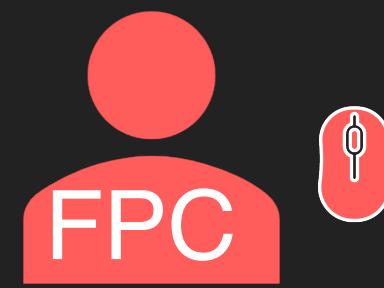
Feature extraction

User Profiles





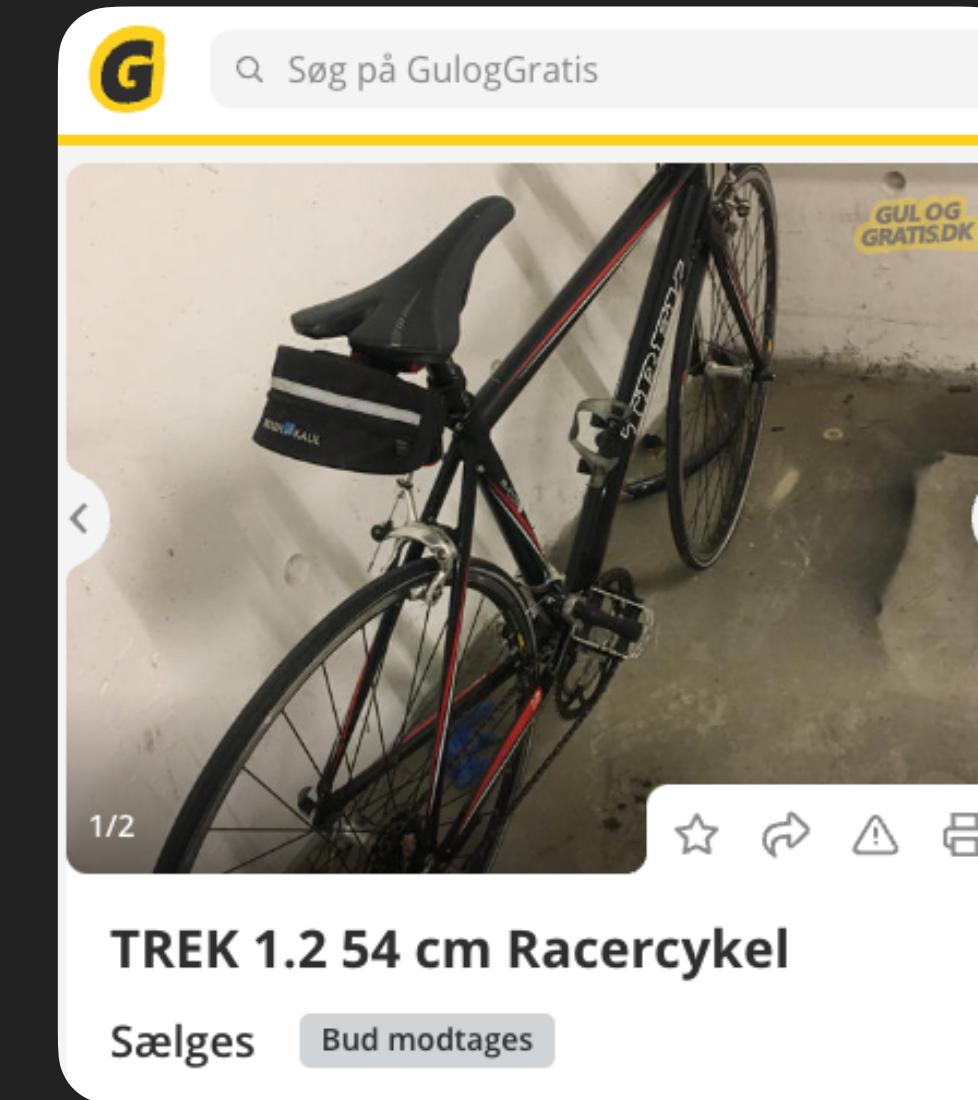
Contextual User Profile



<https://www.guloggratis.dk/annonce/3c55c290...>

0.3	0.15	-2.2	...	-0.4
-----	------	------	-----	------

URL embedding



TREK 1.2 54 cm Racercykel
Sælges Bud modtages

<https://www.guloggratis.dk/annonce/9e3385ca...>

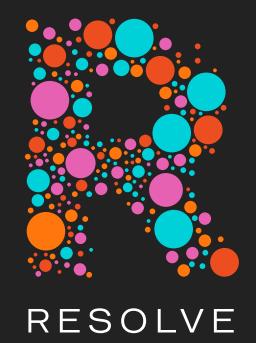
0.45	-1.2	-0.2	...	1.9
------	------	------	-----	-----



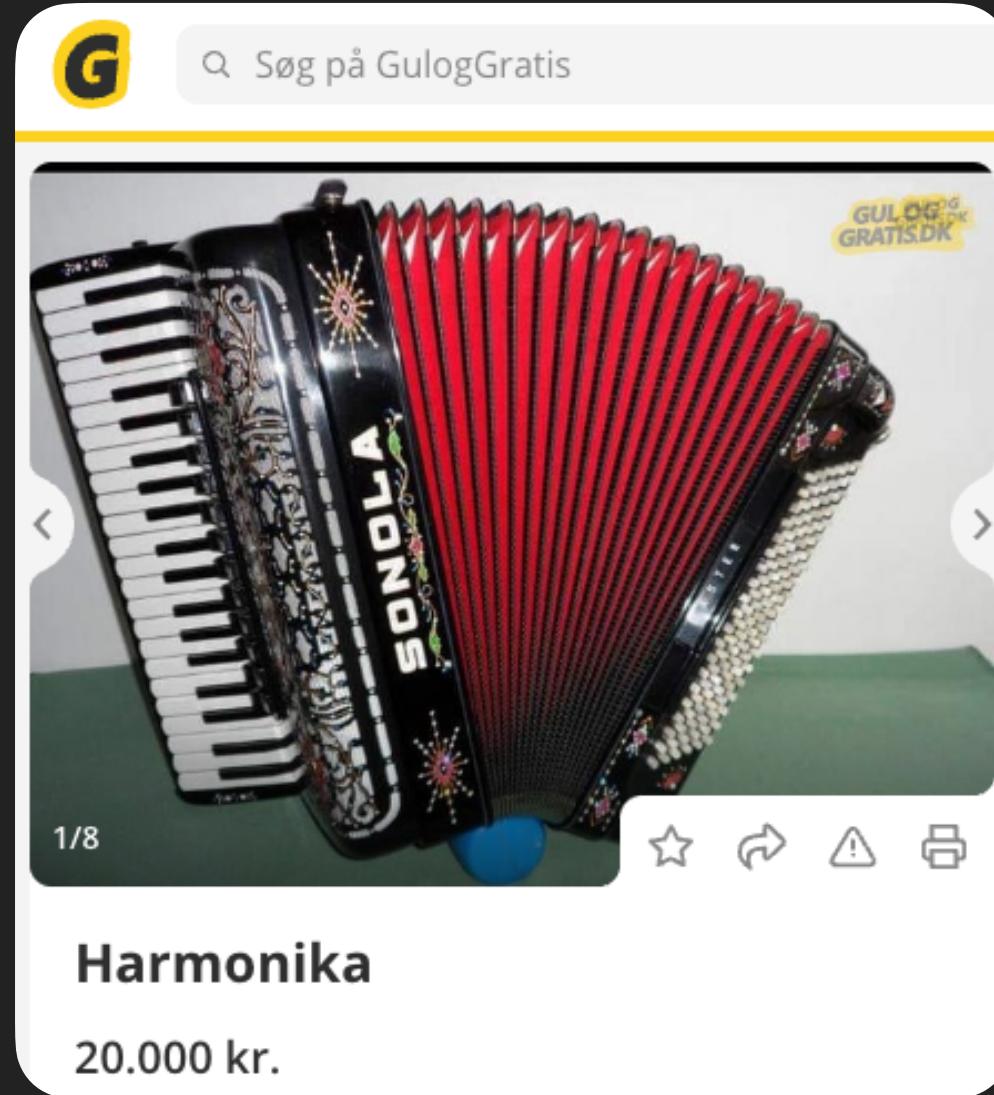
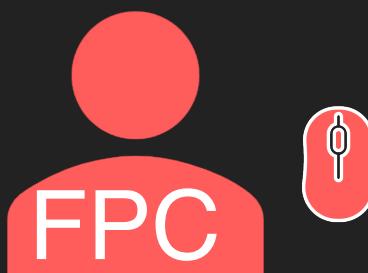
Knaus 450 sudwind
67.800 kr.

<https://www.guloggratis.dk/annonce/d0c64a94...>

-2.7	1.0	-3.8	...	-0.9
------	-----	------	-----	------

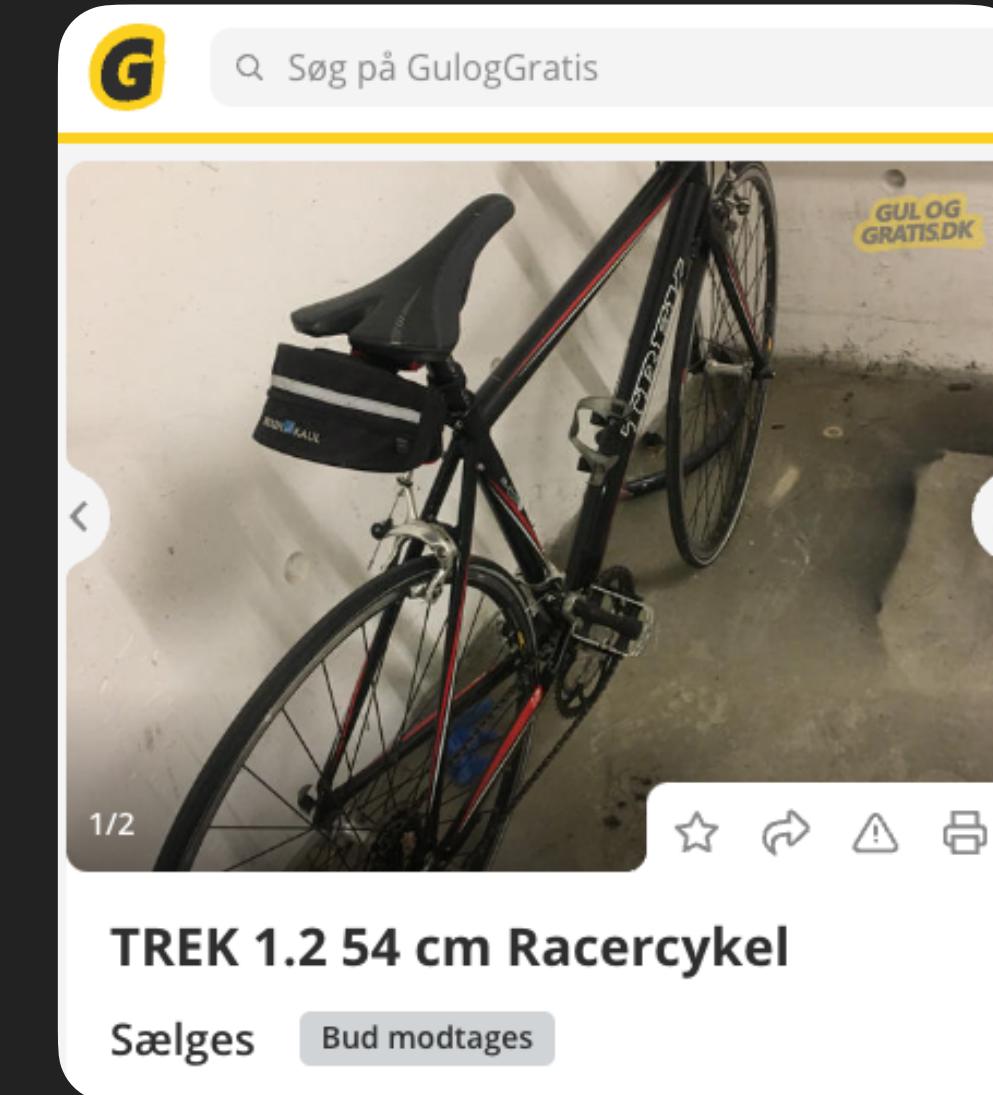


Contextual User Profile



<https://www.guloggratis.dk/annonce/3c55c290...>

0.3	0.15	-2.2	...	-0.4
-----	------	------	-----	------



<https://www.guloggratis.dk/annonce/9e3385ca...>

0.45	-1.2	-0.2	...	1.9
------	------	------	-----	-----



<https://www.guloggratis.dk/annonce/d0c64a94...>

-2.7	1.0	-3.8	...	-0.9
------	-----	------	-----	------

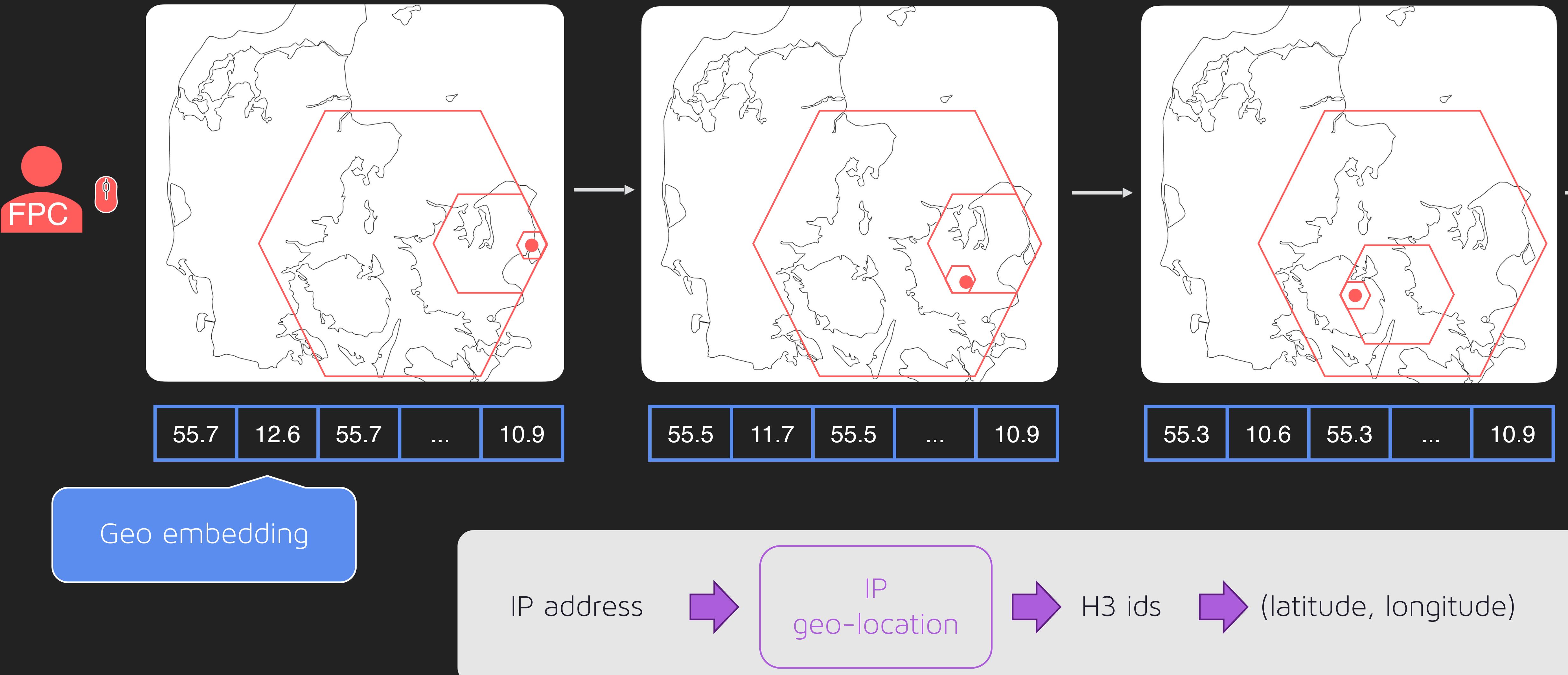
0.12	-0.56	1.10	...	-0.71
------	-------	------	-----	-------

(Contextual) User Embedding



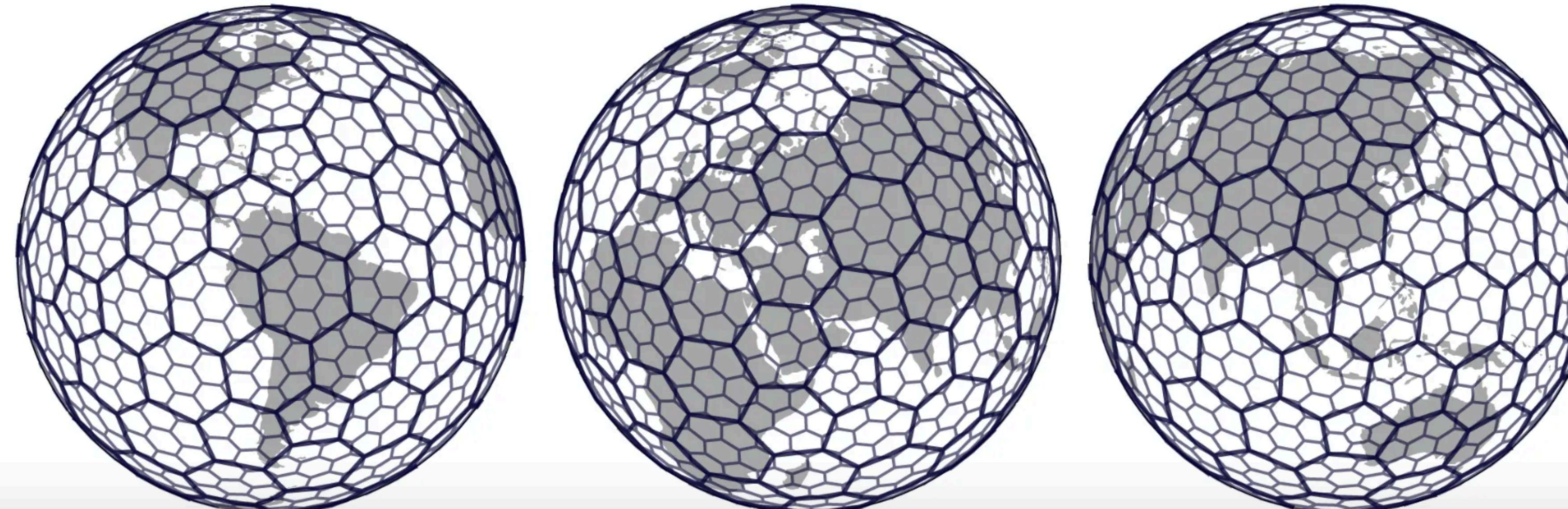
Geo Journey

User Profiles



H3: Uber's Hexagonal Hierarchical Spatial Index

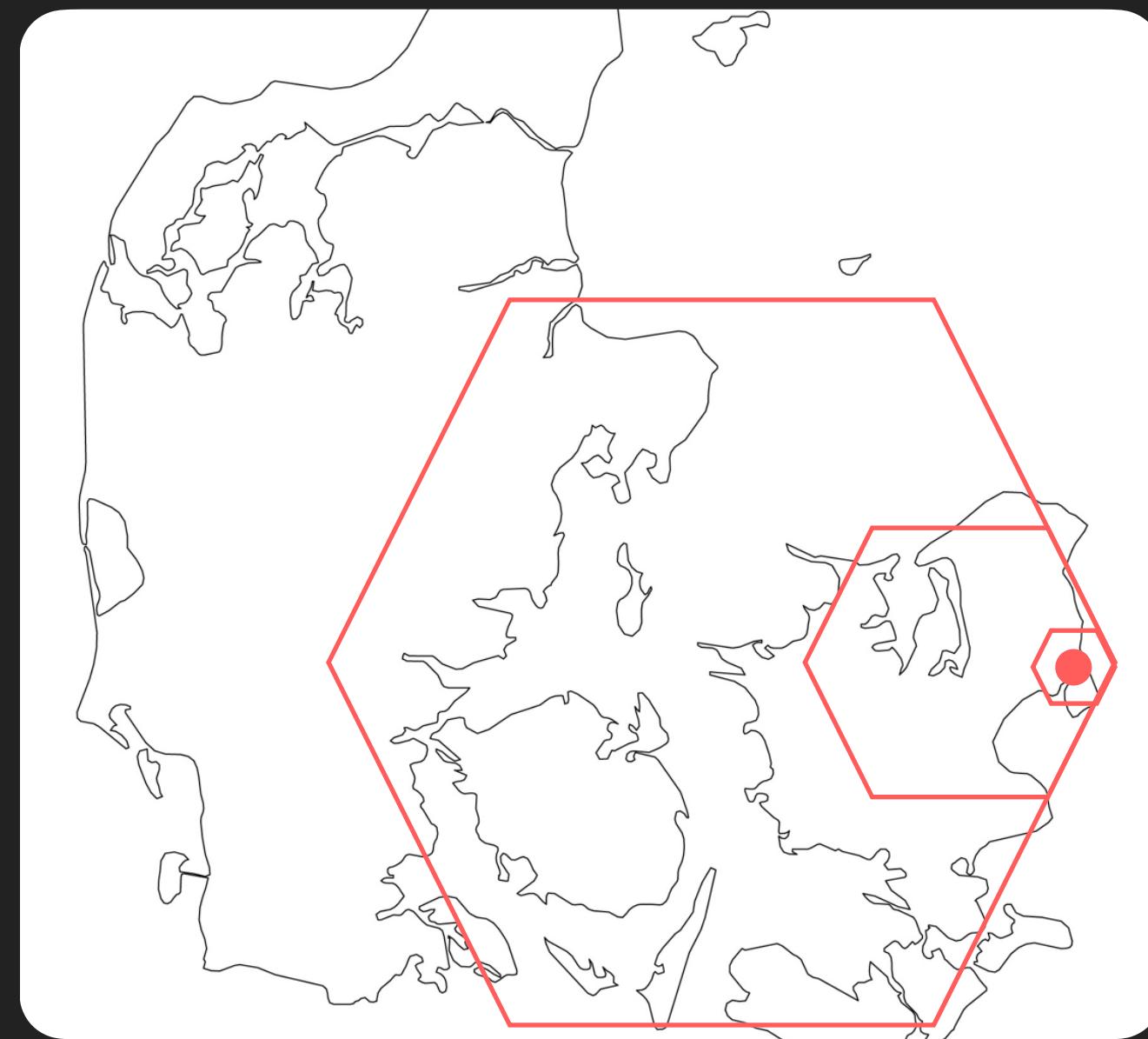
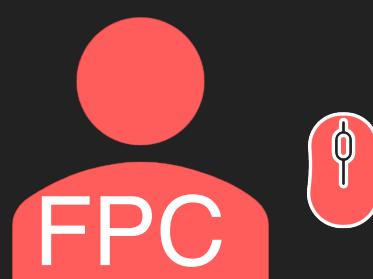
June 27, 2018 / Global



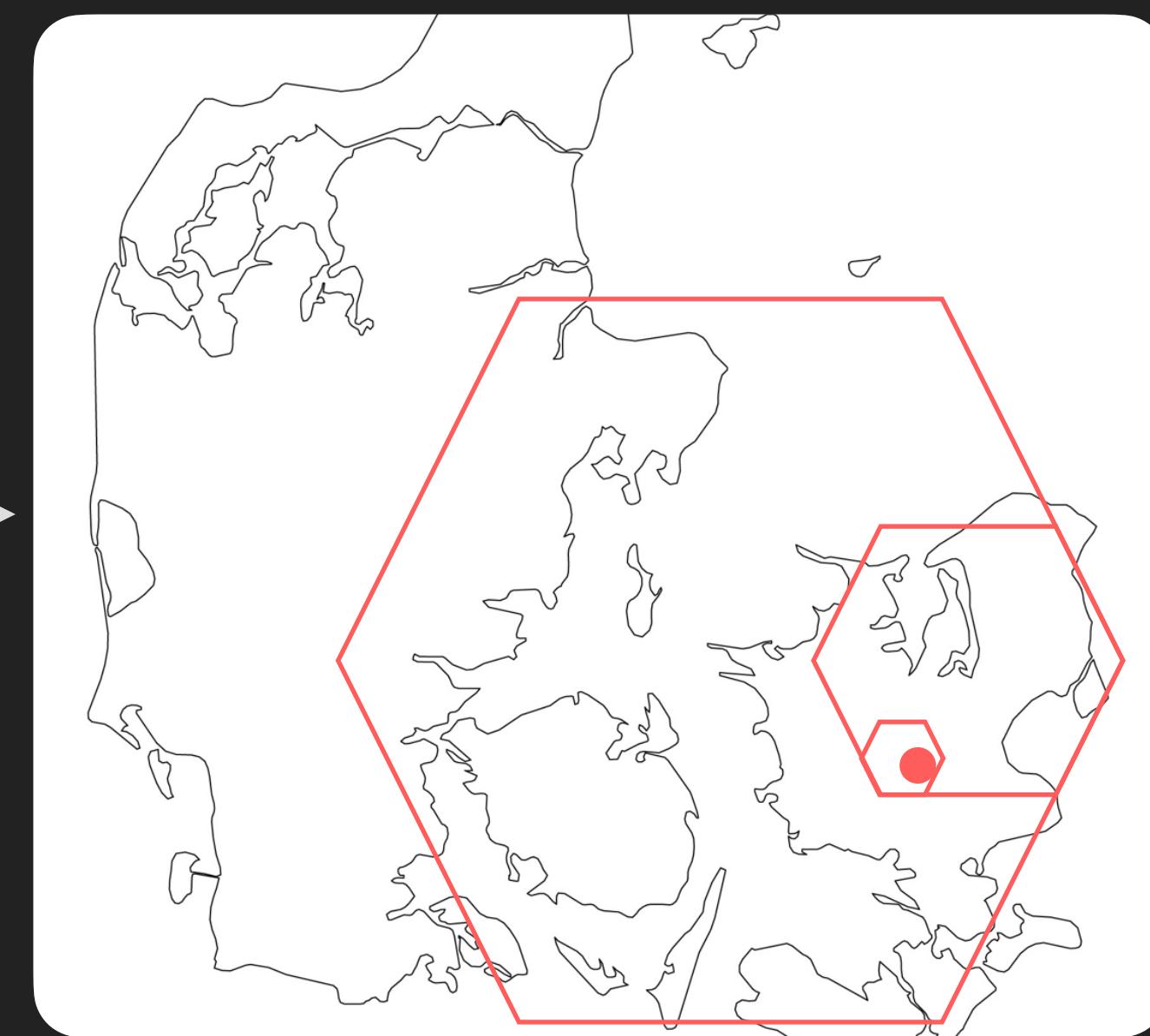


Geo Journey

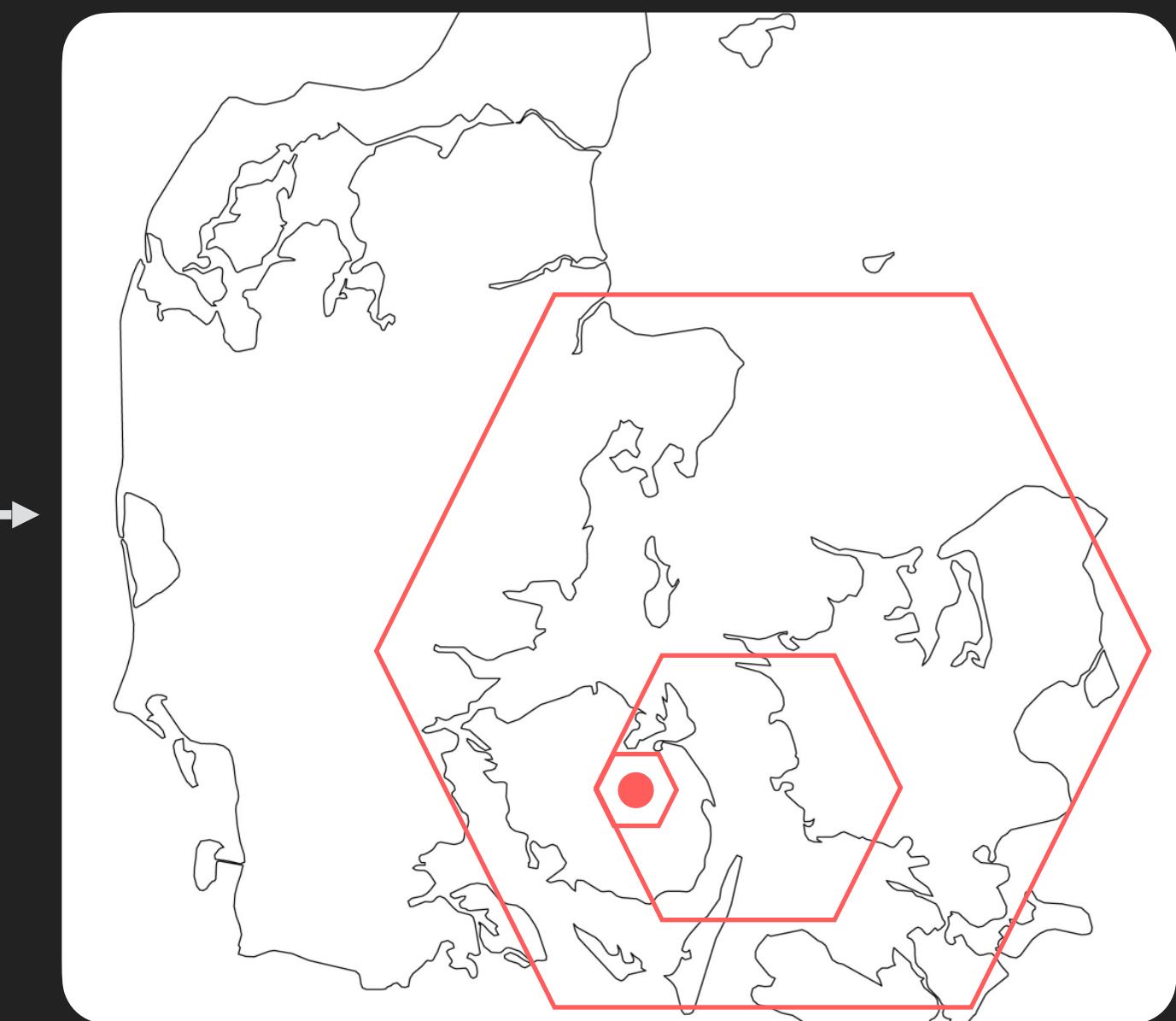
User Profiles



55.7	12.6	55.7	...	10.9
------	------	------	-----	------



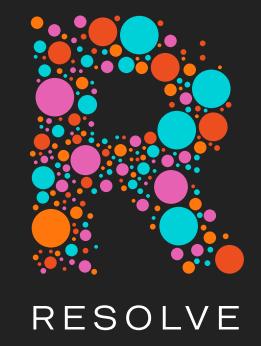
55.5	11.7	55.5	...	10.9
------	------	------	-----	------



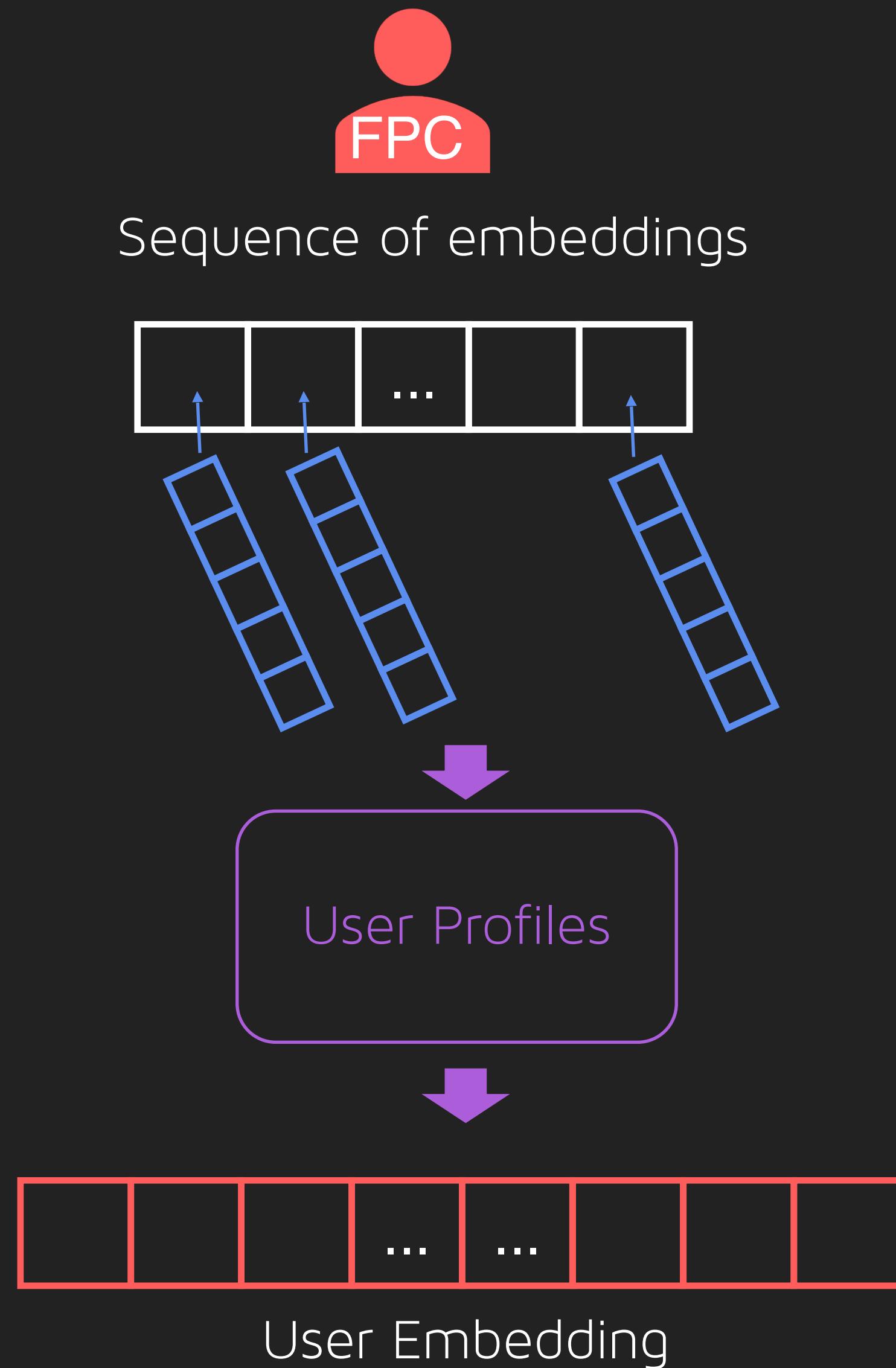
55.3	10.6	55.3	...	10.9
------	------	------	-----	------

0.12	-0.56	1.10	...	-0.71
------	-------	------	-----	-------

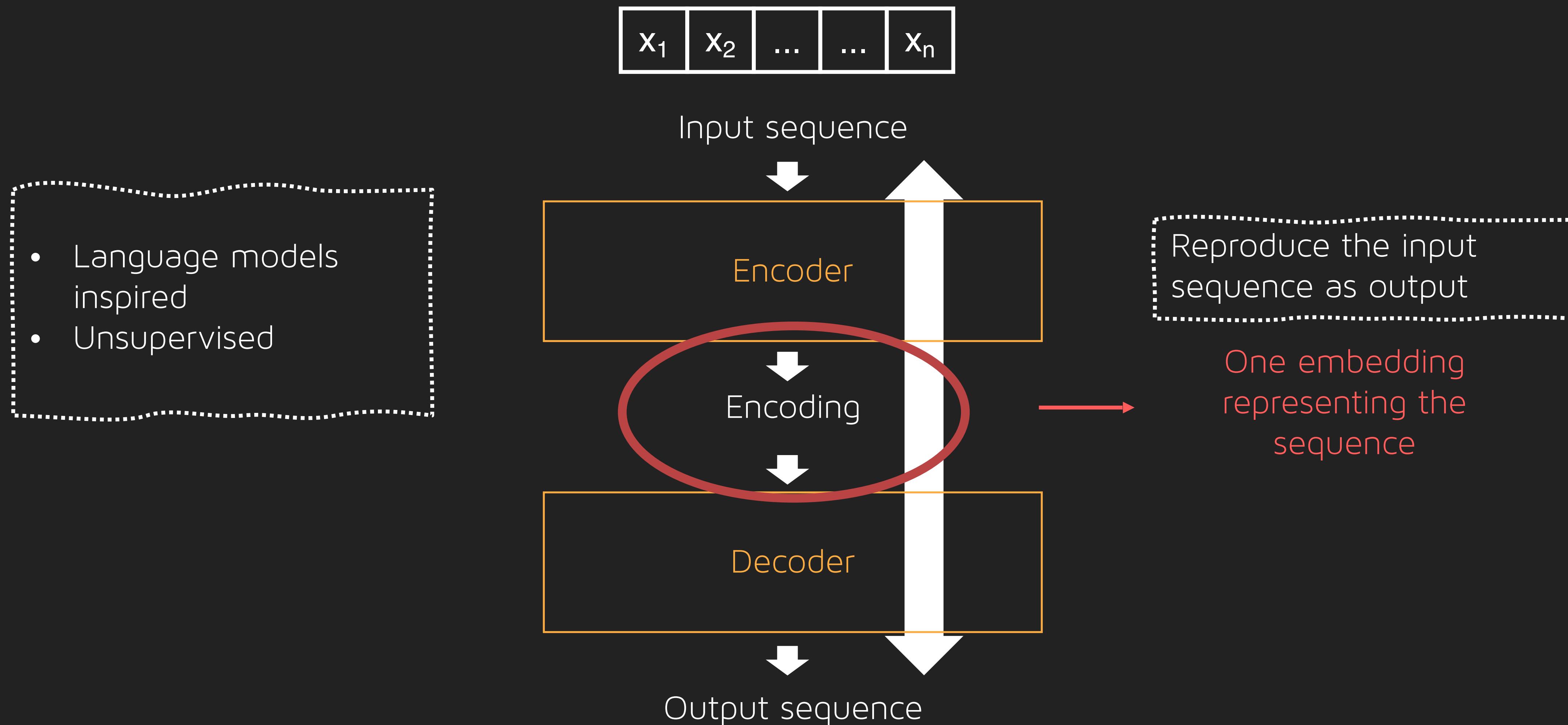
(Geo) User Embedding



User Profiles Model

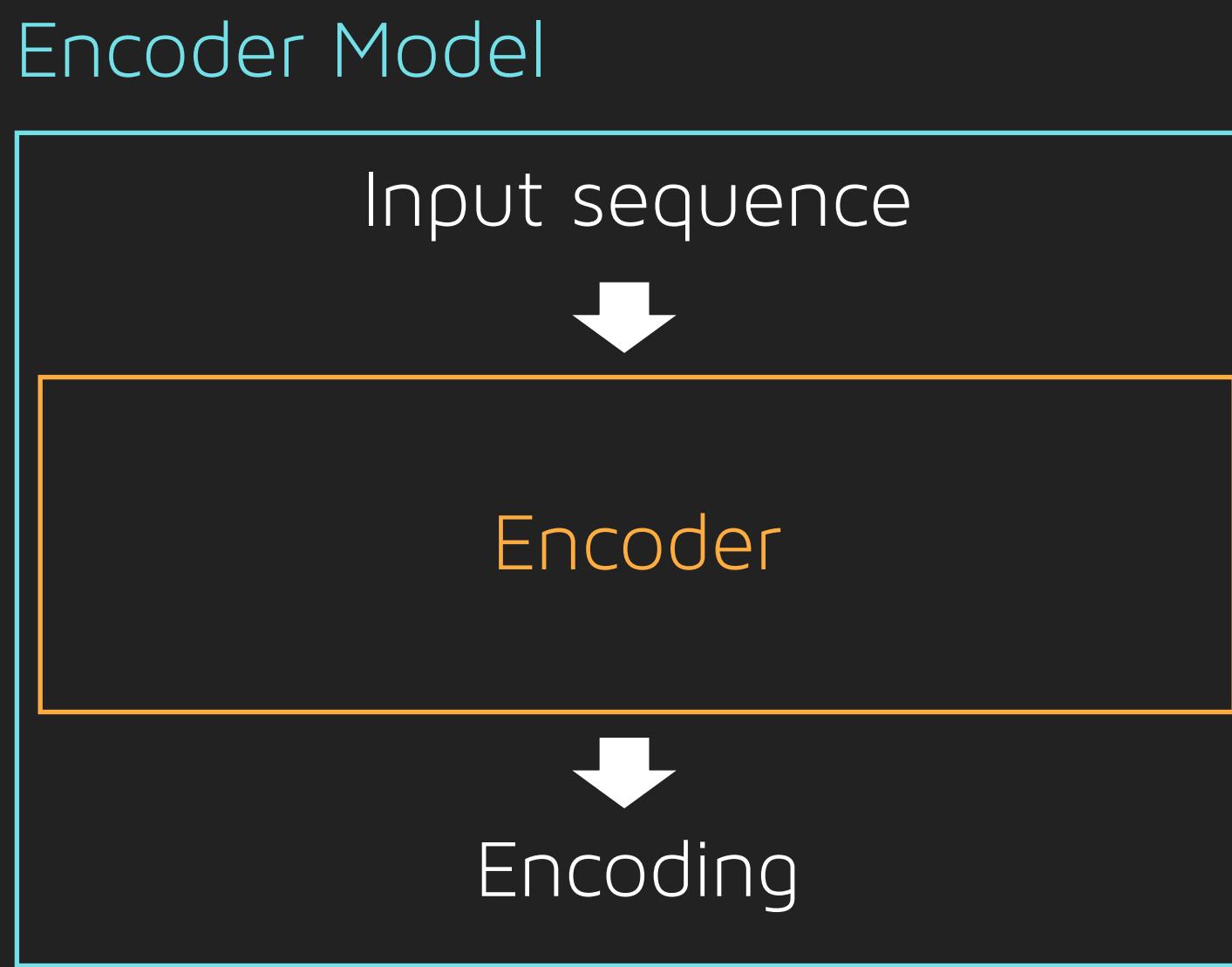


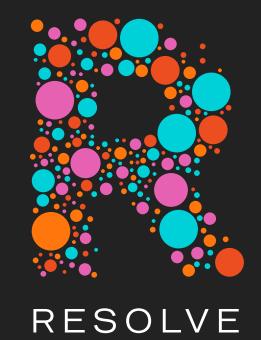
Autoencoder architecture





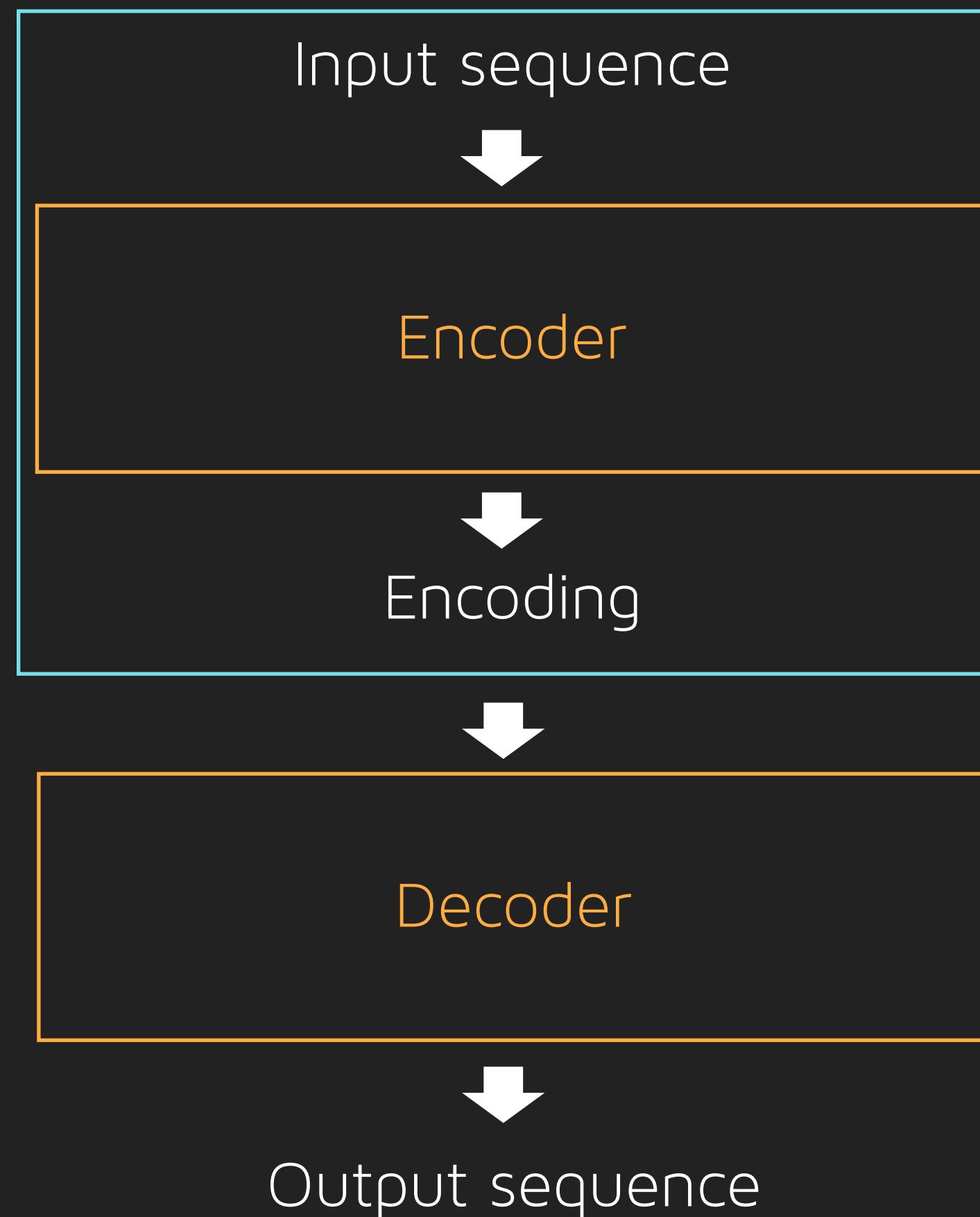
Encoder model for inference





Model Layers

Encoder Model



Encoder Layer(s)

Multi Head Attention
Add & Norm
Feed Forward

Attention Average Layer

Multi Head Attention
Dense

NN Decoder

(De-) Encoder Layer(s)

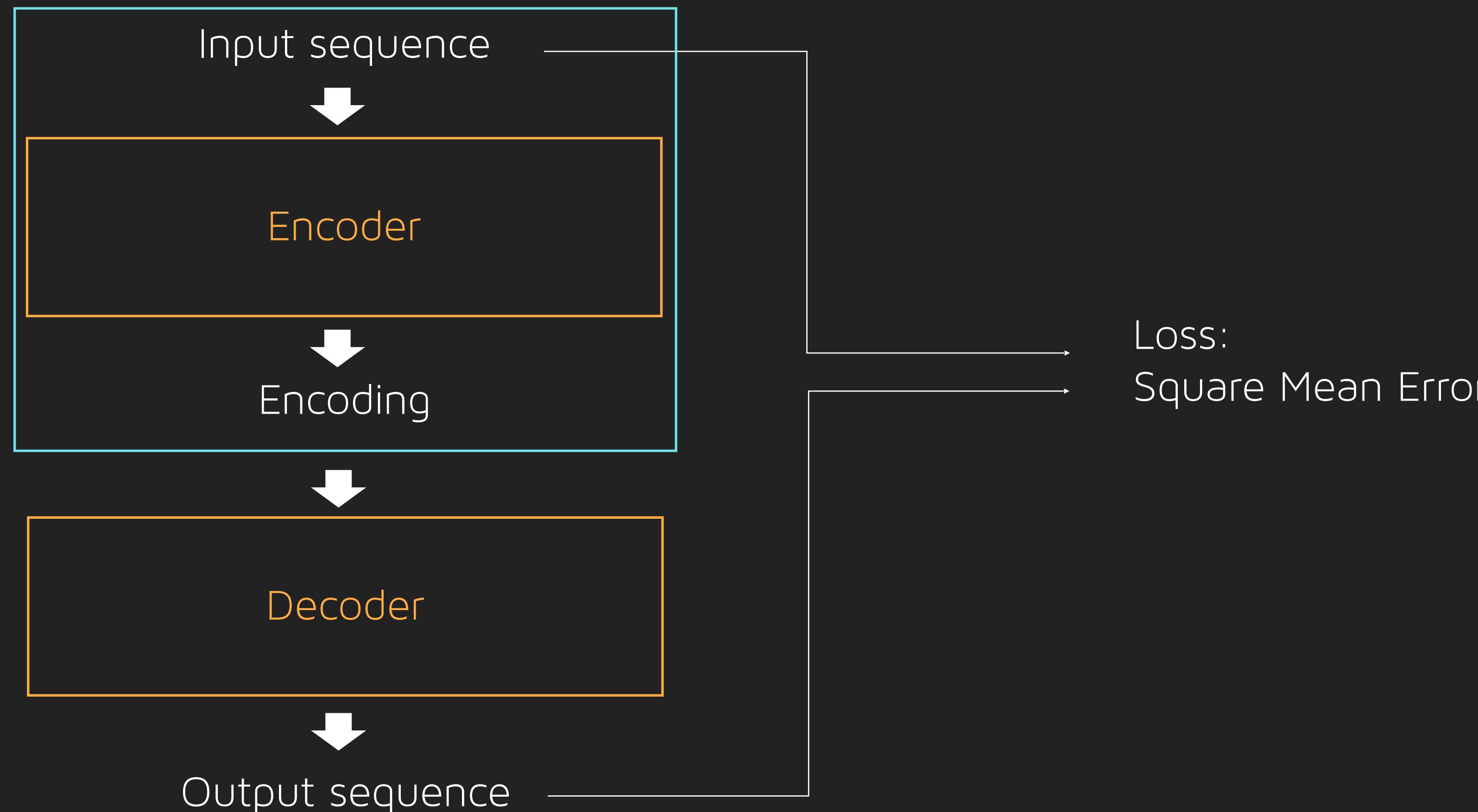
Linear

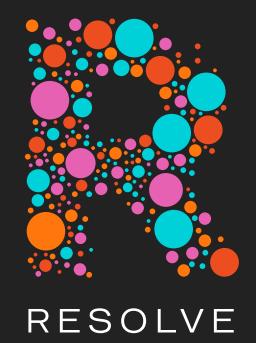


Training objective

Autoencoder

Encoder Model





Learning Strategies

Masked

Inspired by RoBERTa's
dynamic masking
learning strategy

x_1	0	\dots	\dots	x_n
0	1	\dots	\dots	0

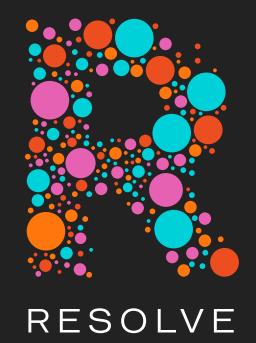
NCP
(next cookie prediction)

Inspired by BERT
next sentence prediction
learning strategy



Segment 1

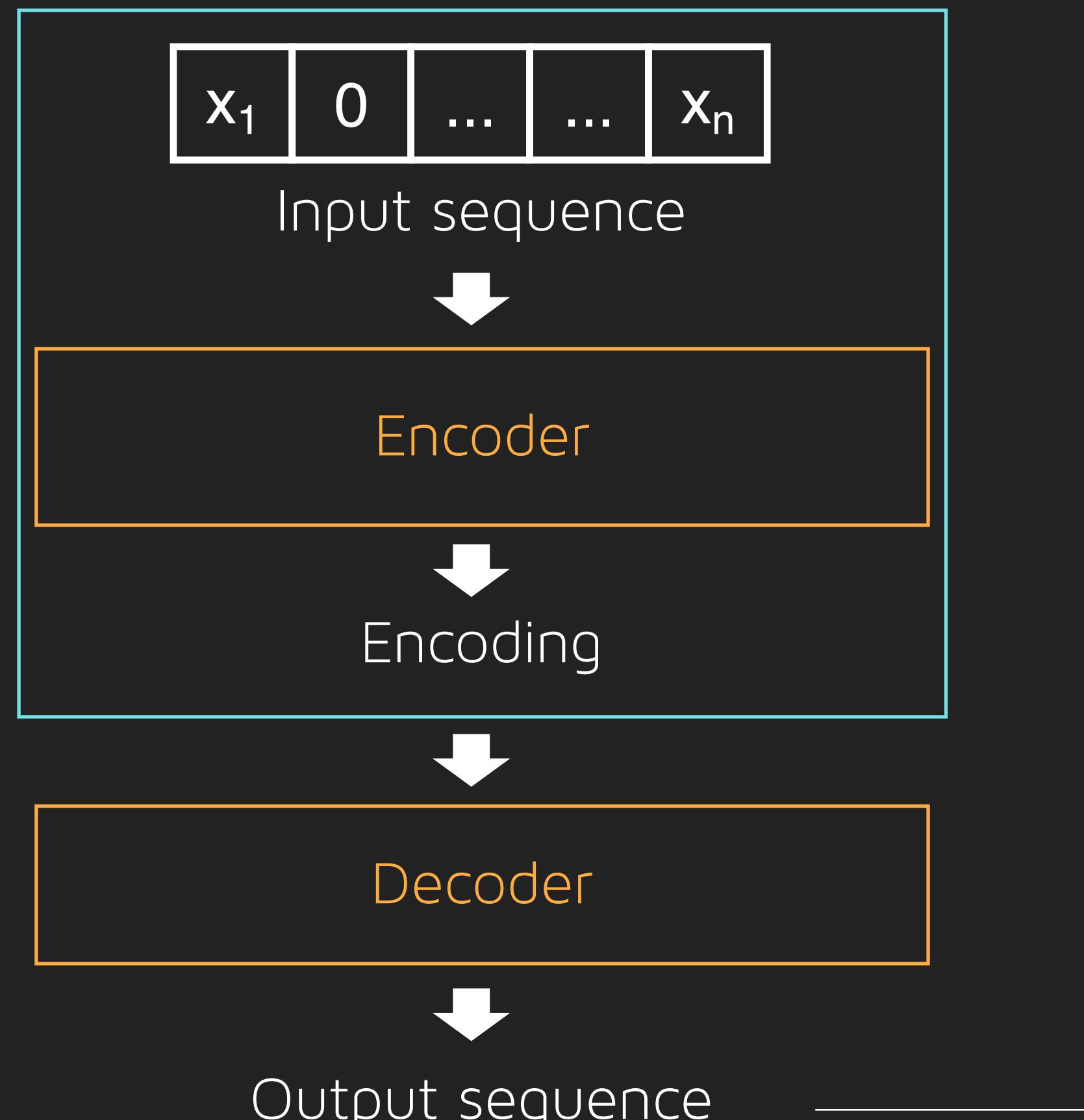
Segment 2



Learning Strategies

Masked

Encoder Model



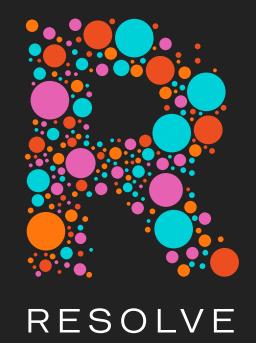
$x_1 \ x_2 \ \dots \ \dots \ x_n$

Original sequence

0 1 \dots \dots 0

Mask

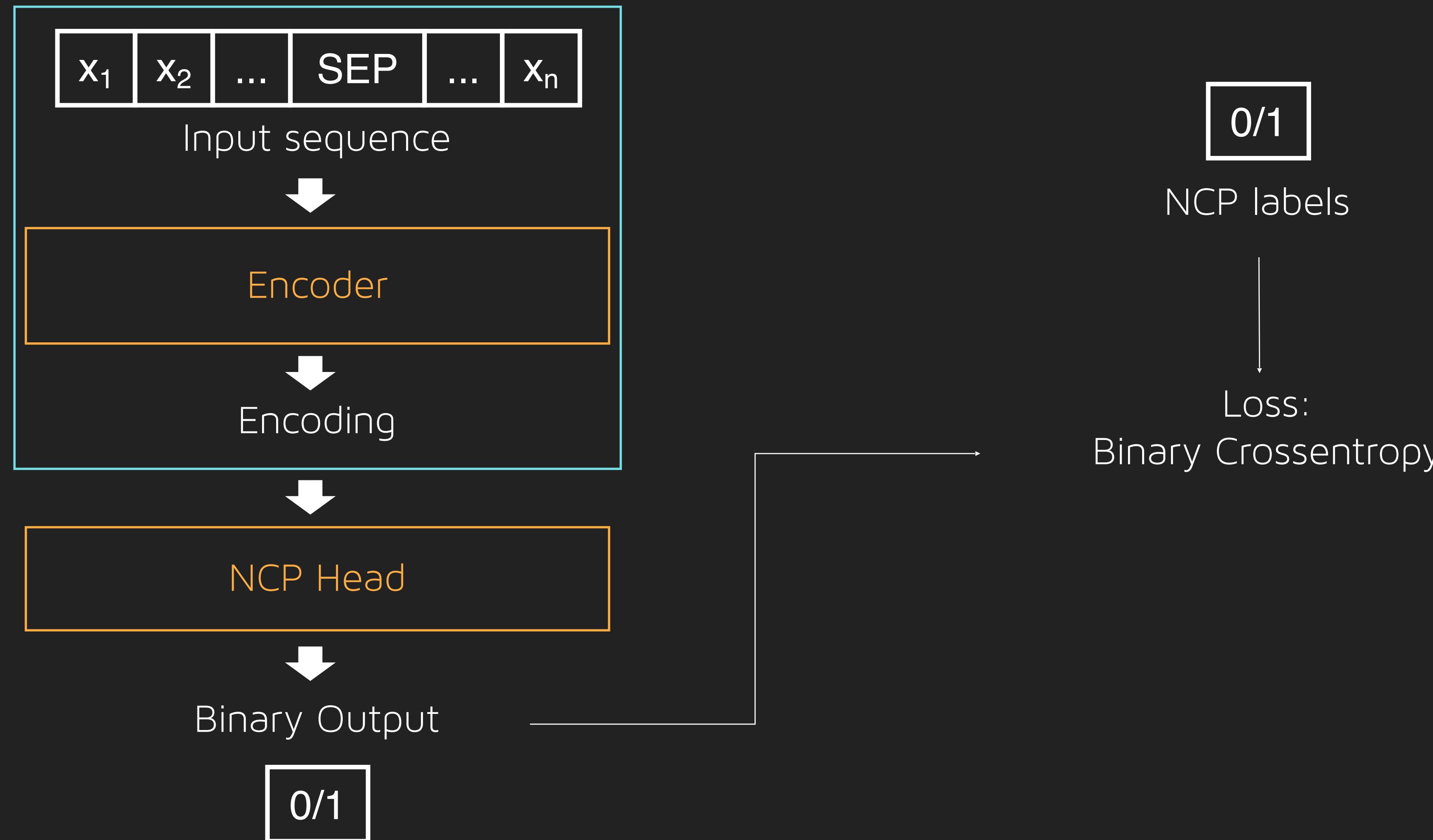
Loss:
Square Mean Error



Learning Strategies

NCP

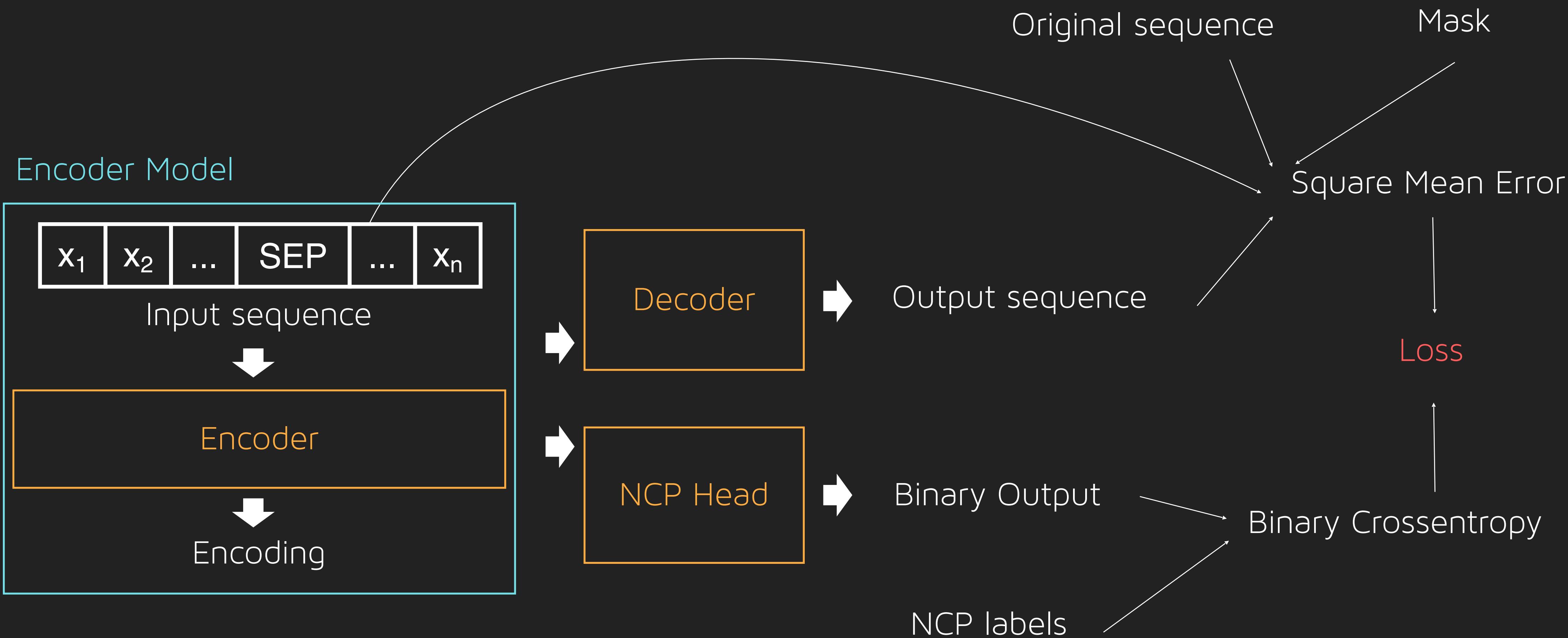
Encoder Model





Learning Strategies

Masked + NCP





Learning Strategy selection

Contextual User Profiles

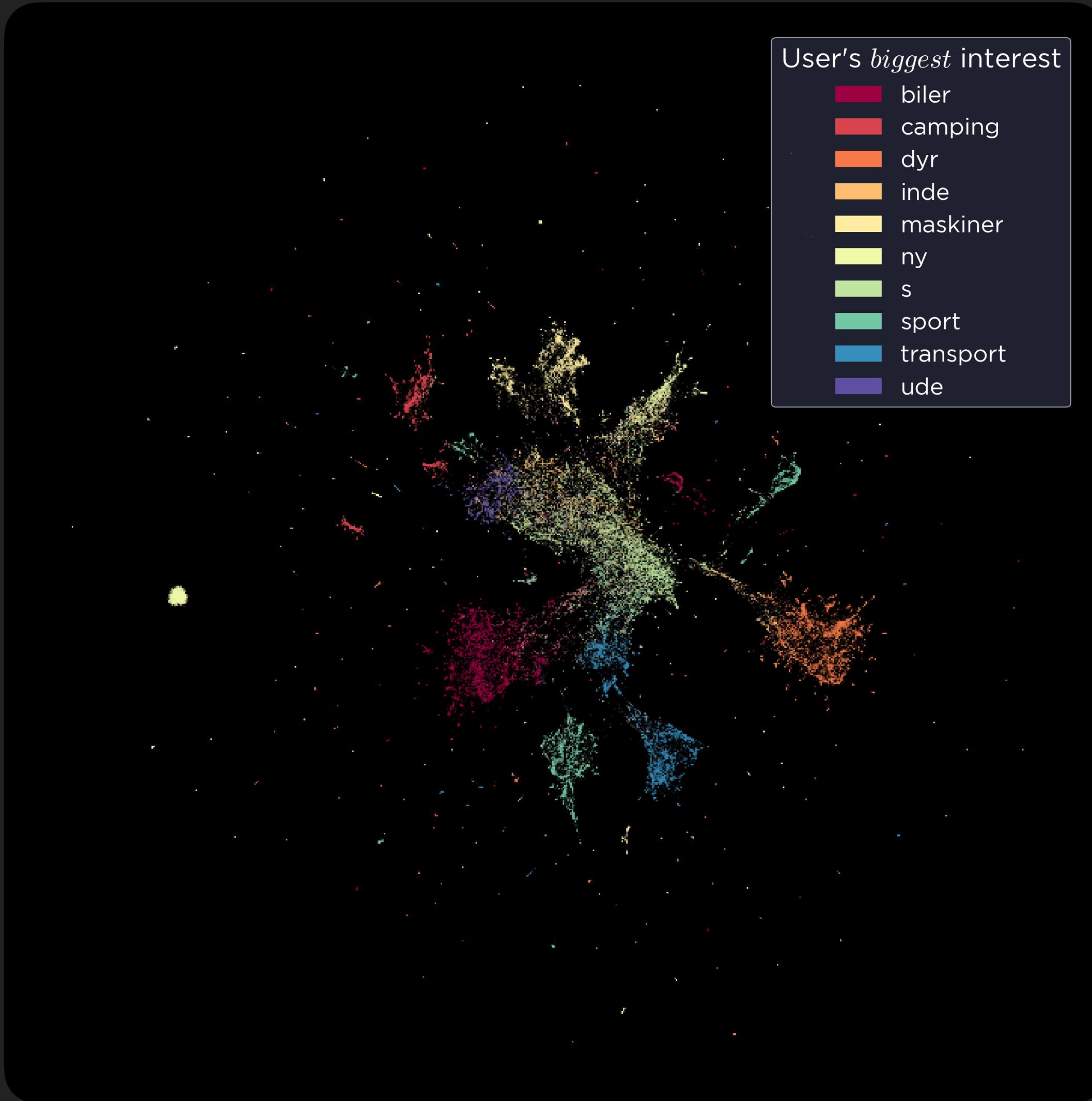
- Masked

Geo journey User Profiles

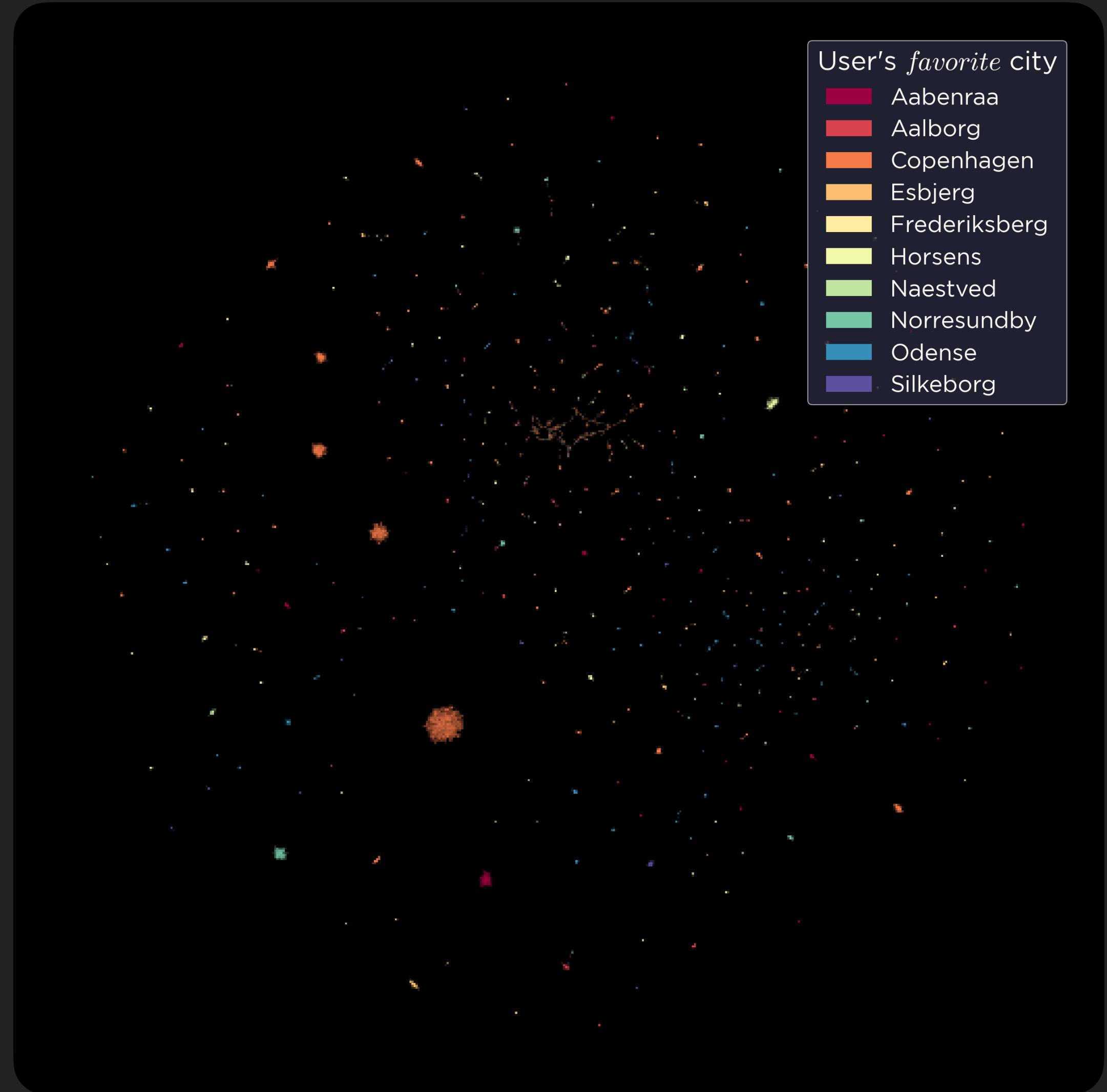
- Masked + NCP

Visualising embeddings

Contextual



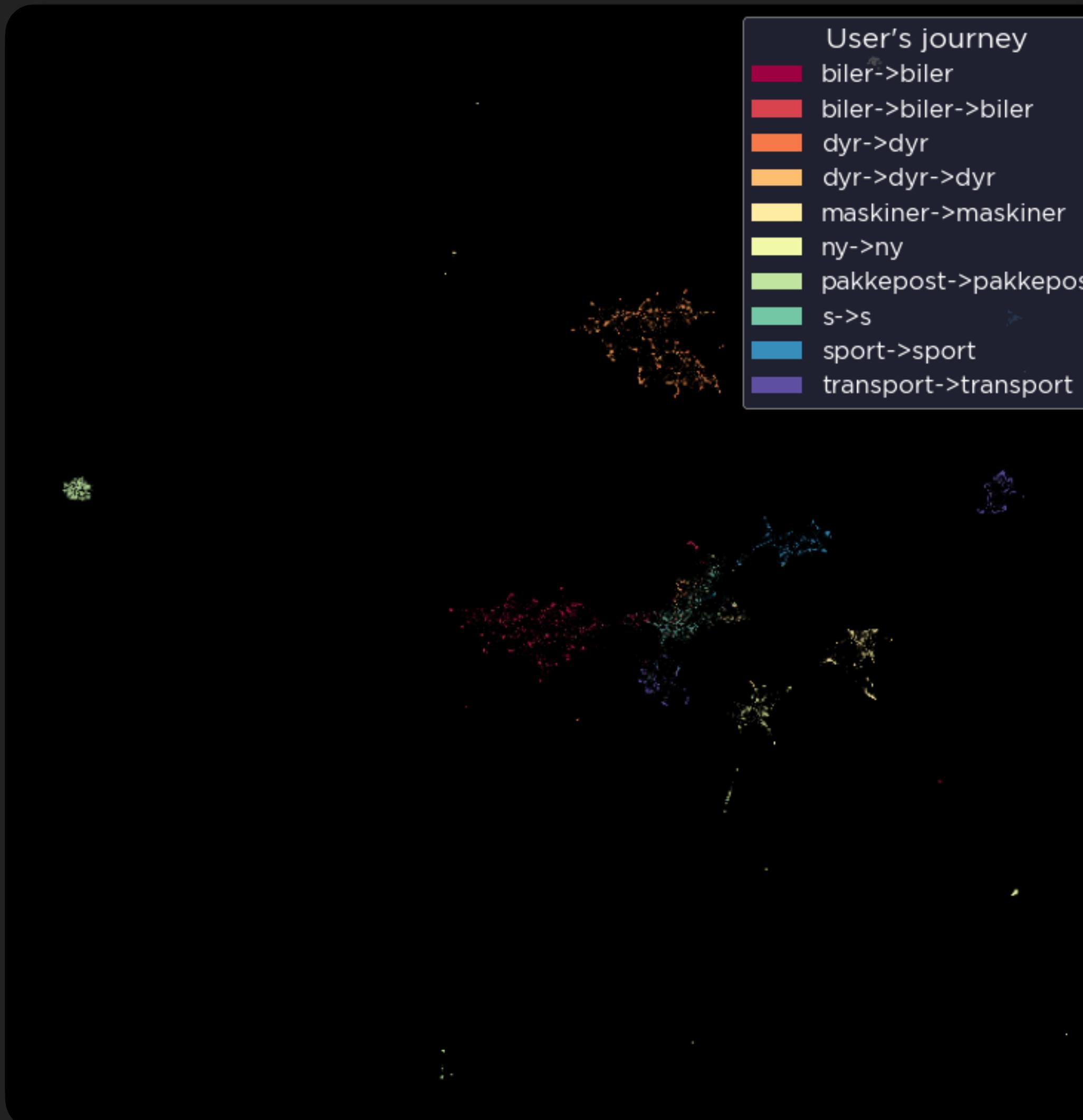
Geo-Journey



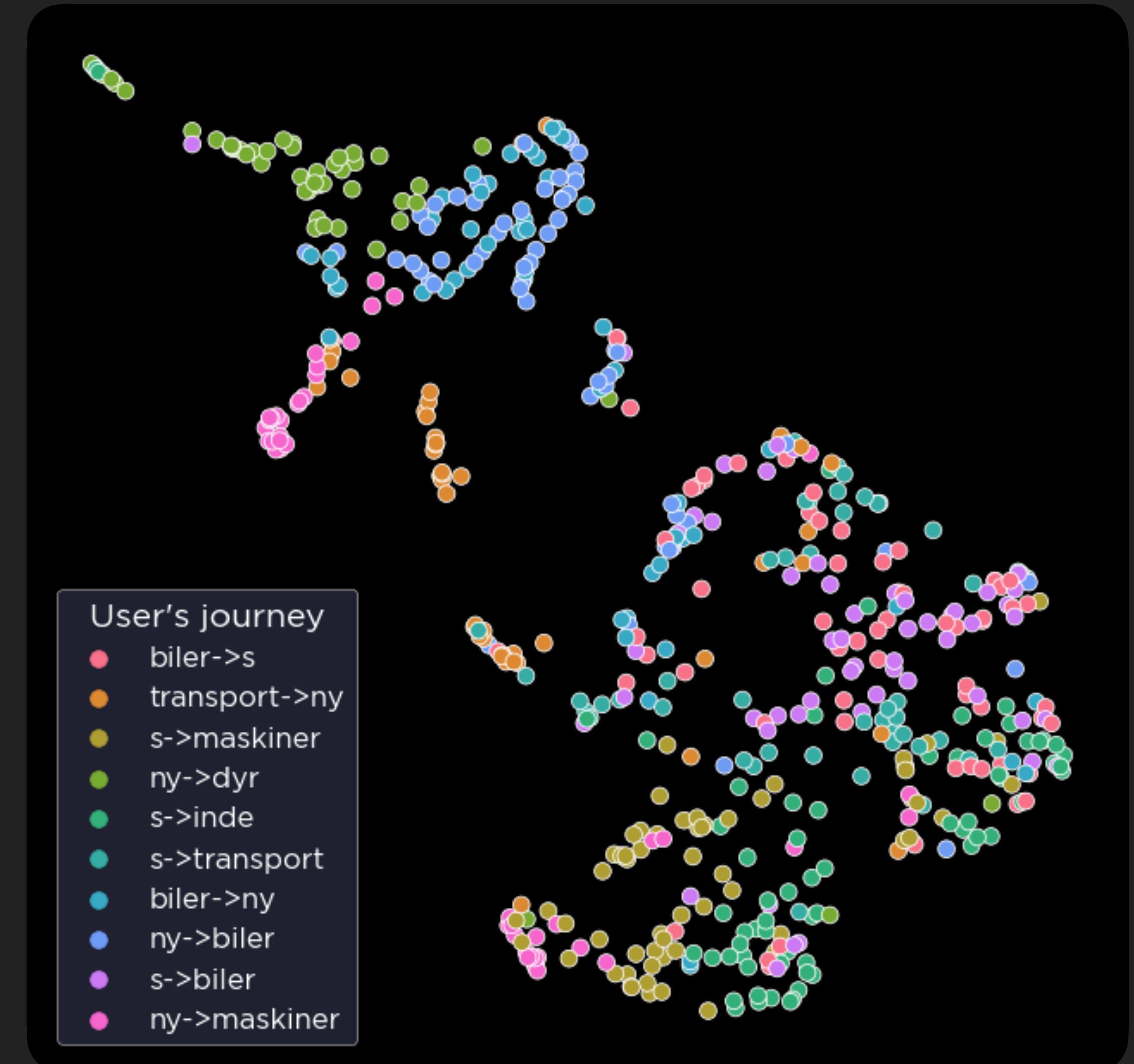
Visualising embeddings

Contextual

Most frequent topic sequences



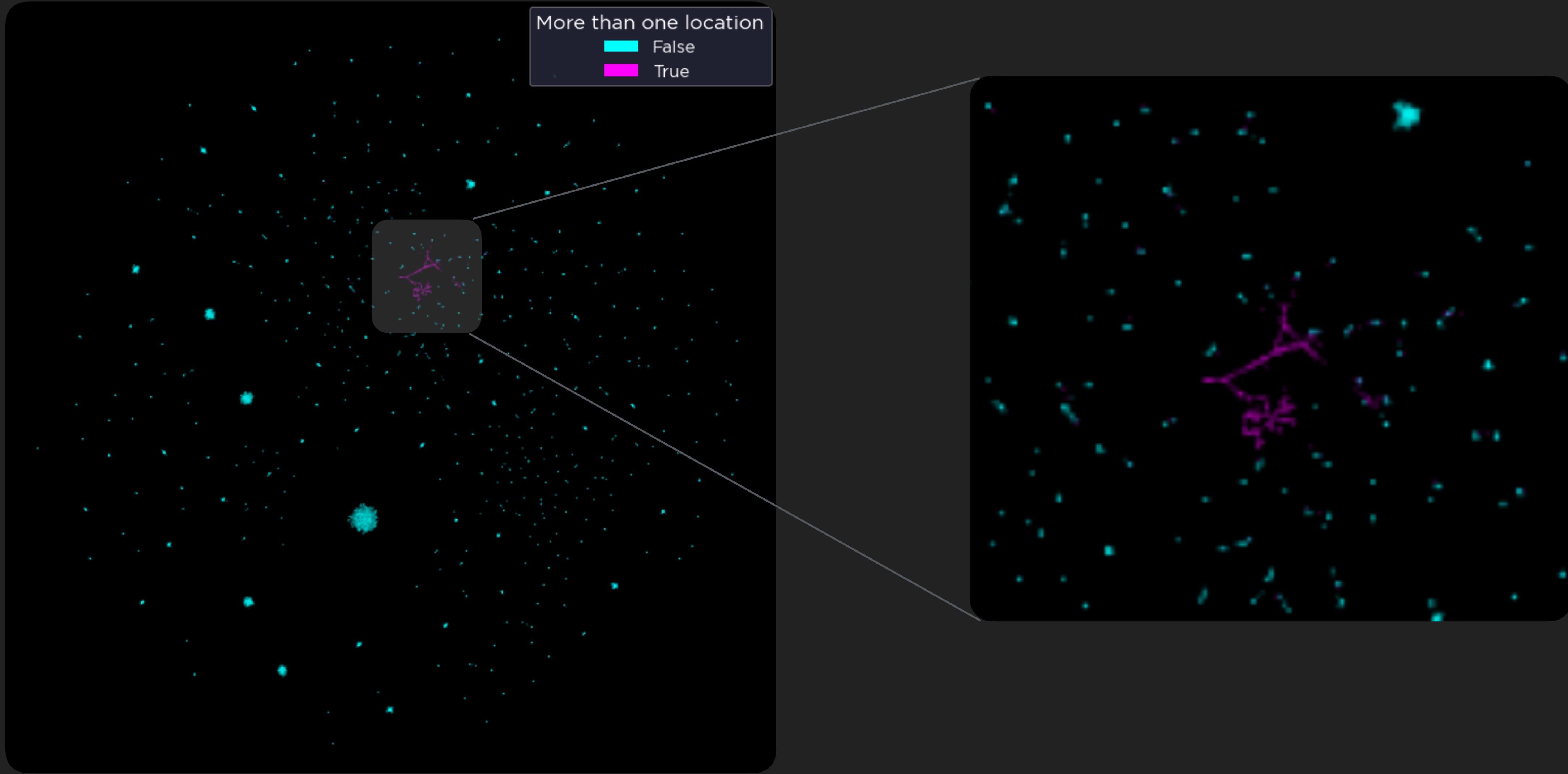
Most diverse topic sequences



Visualising embeddings

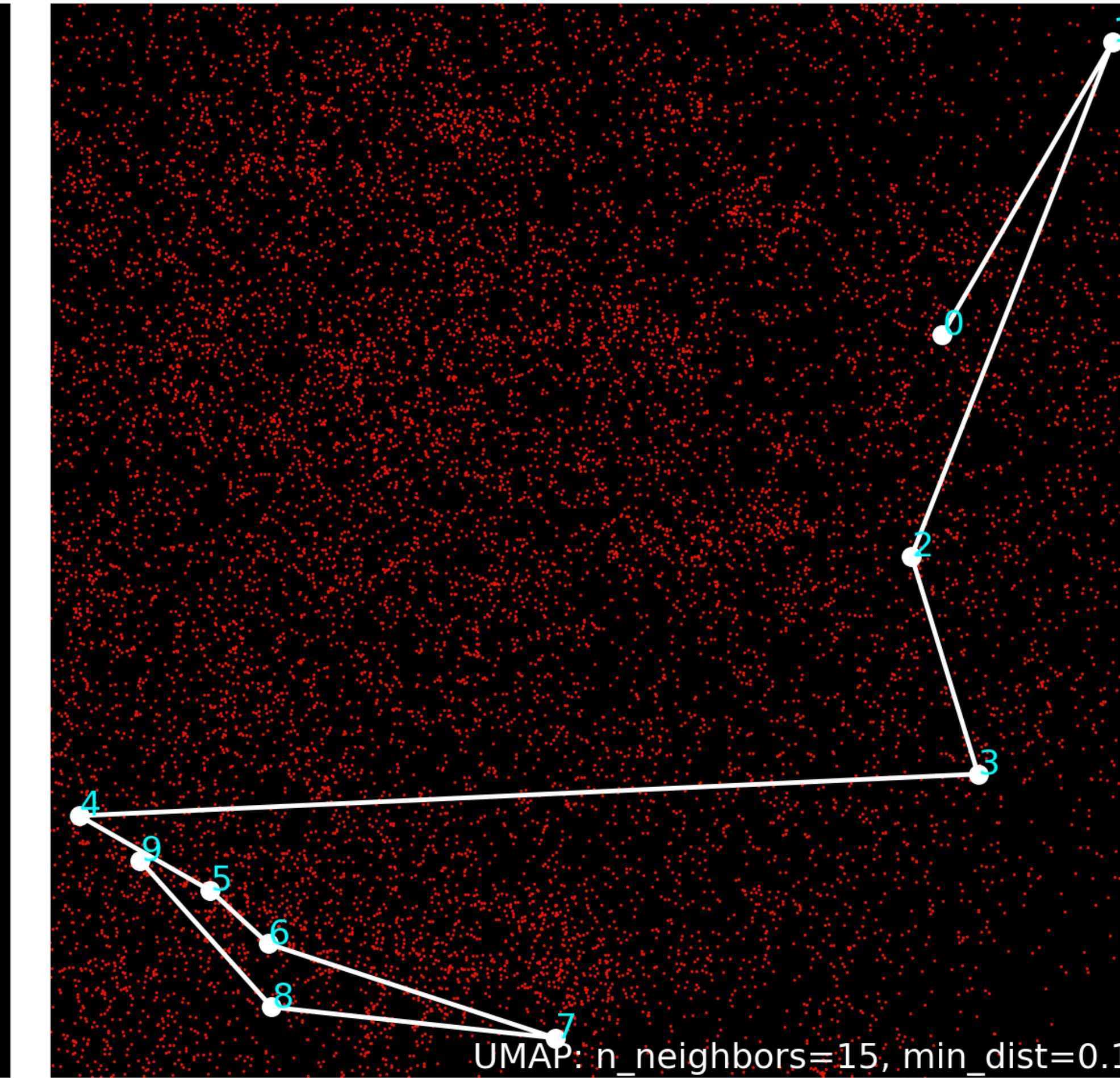
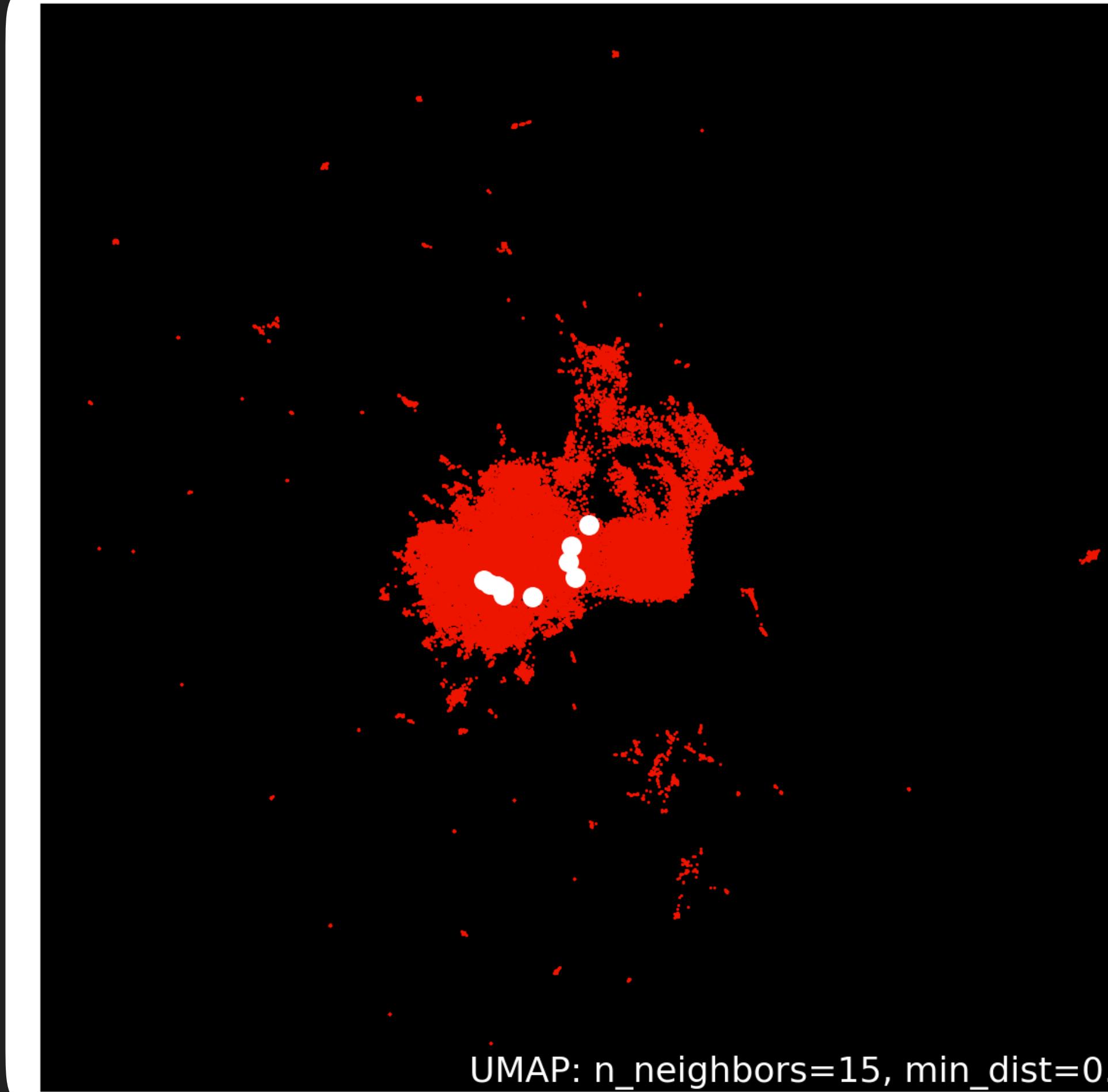
Geo-Journey

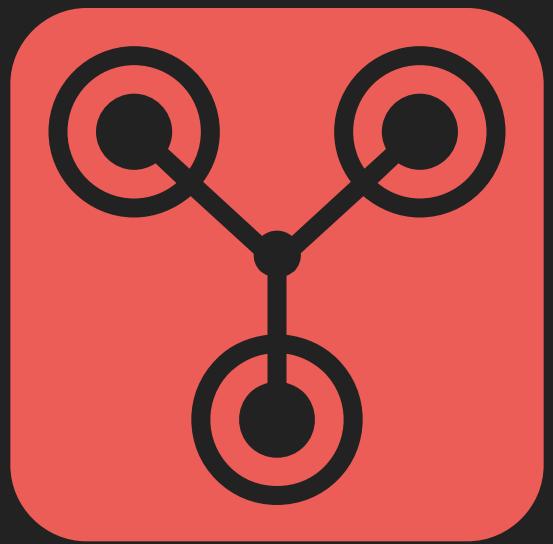
most people don't move...



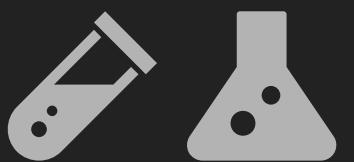
Use case: Gul & Gratis

Visualising the sequence...





Look-Alike Modeling

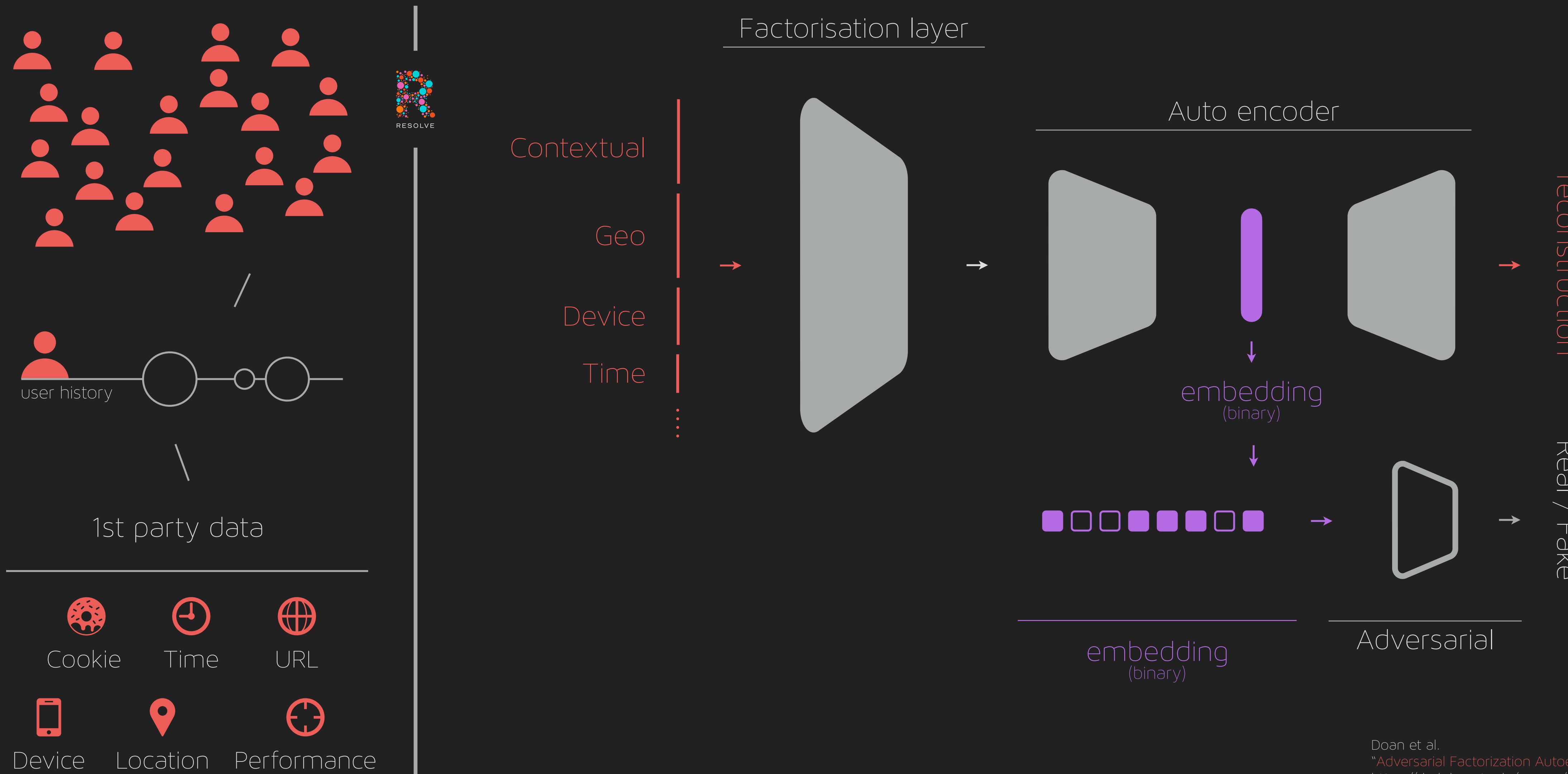


Still under development



Look-Alike Modeling

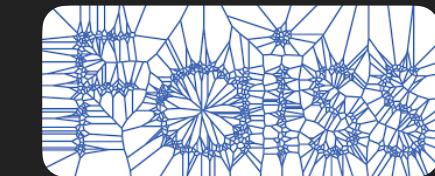
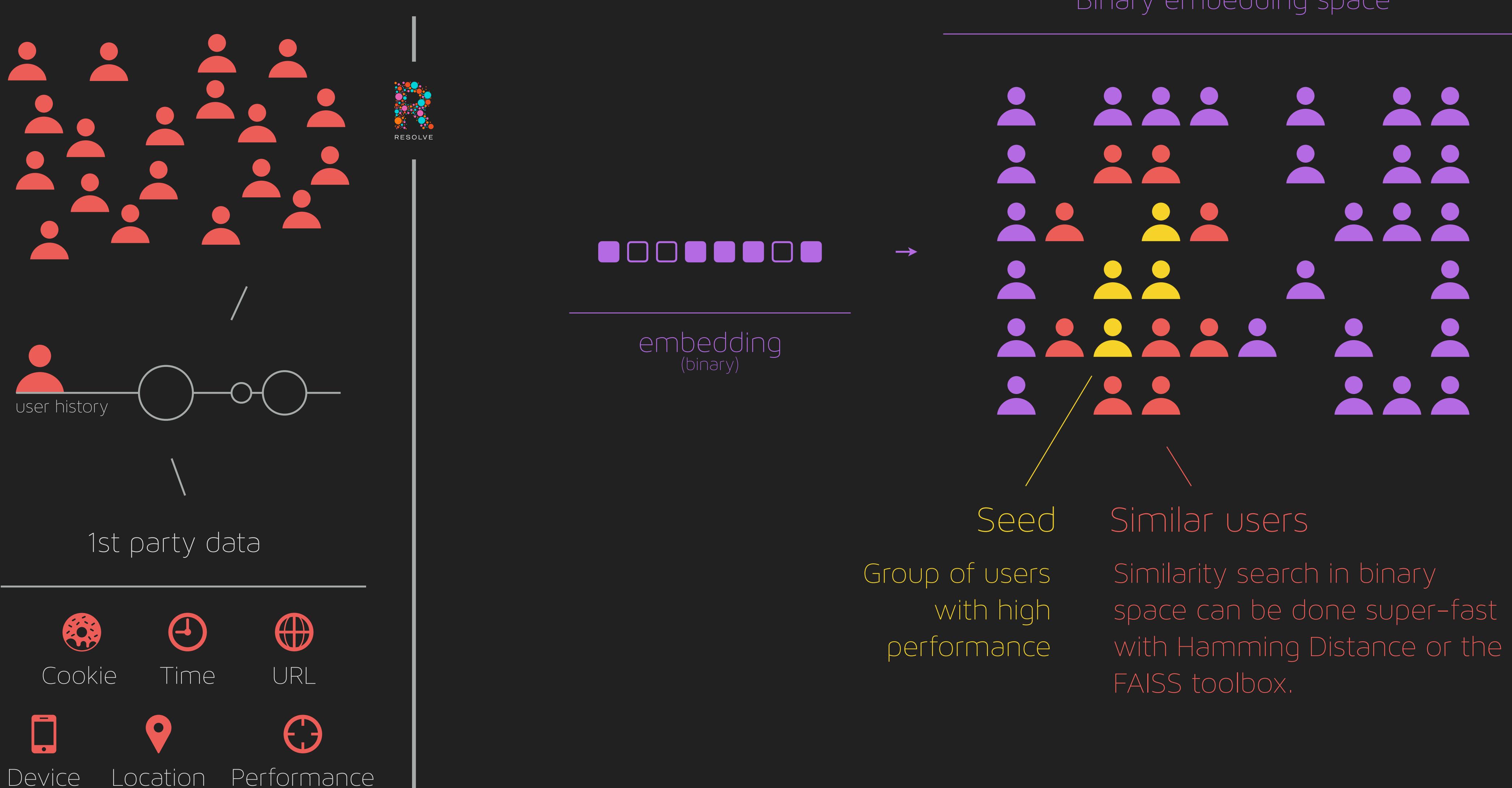
Adversarial Factorization Binary Autoencoder

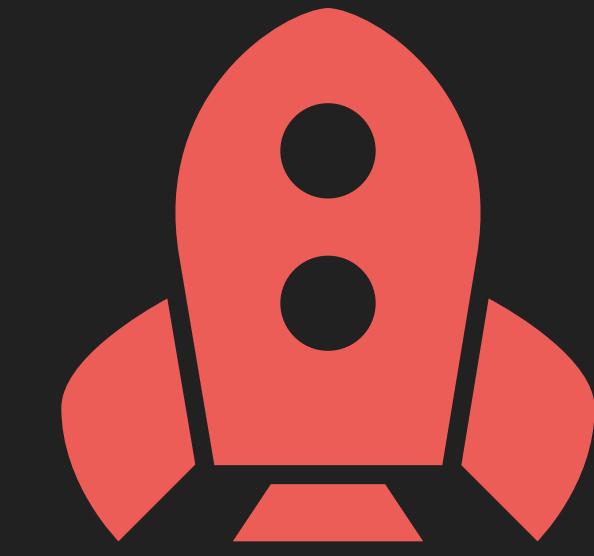




Look-Alike Modeling

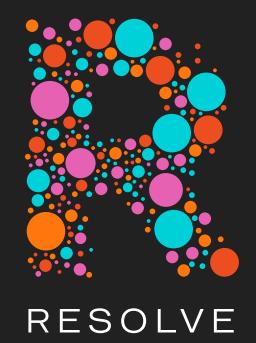
Adversarial Factorization Binary Autoencoder





Production

Moving Research into the real world



Production

What needs to be thought of...

ML projects need planning for Production.



Data

- Where is it stored
- How much data
- How do we retrieve it.
(i.e., batch vs real-time)



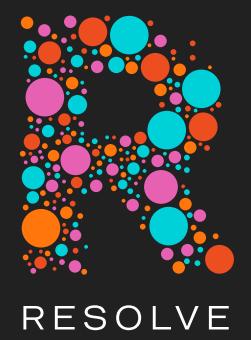
Frameworks

- Tools – which is best for the task at hand



Monitoring

- feedback from production models
- test new models
- iterate model without interrupting production



Production

Data and Frameworks



Data

Sources

LLD/Tracking data, CRM, etc.

Size

Scale
Batch vs. Real time

Store

Data lake

Our own SDK
Wundermagic

Orchestration



Tools

Argo Workflow



Production Monitoring...



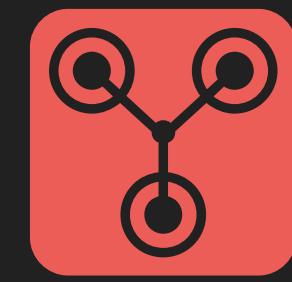
Feedback for
production models

... Heavily investigating
WhyLabs/WhyLogs



Monitoring
MLFlow

Tracking the training
of new models



Bump Models in production
MLFlow

We have different filters to promote
the models

Newest or different matrices
depending on what is needed.



Questions