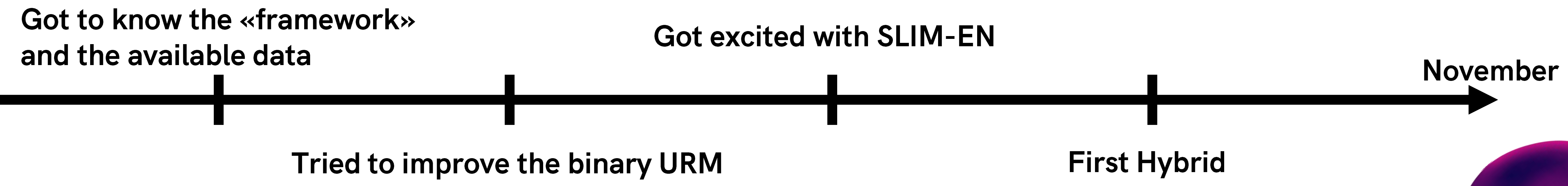
Recommender Systems

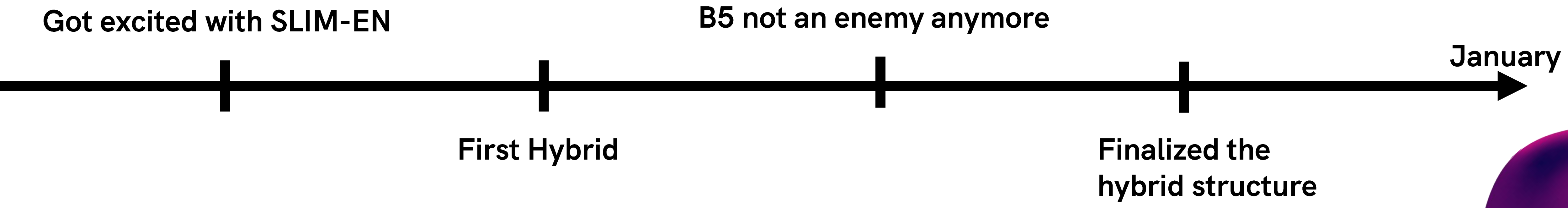
Challenge 2022/23

Nicola Cecere
Andrea Riboni

What we have done, briefly



What we have done, briefly



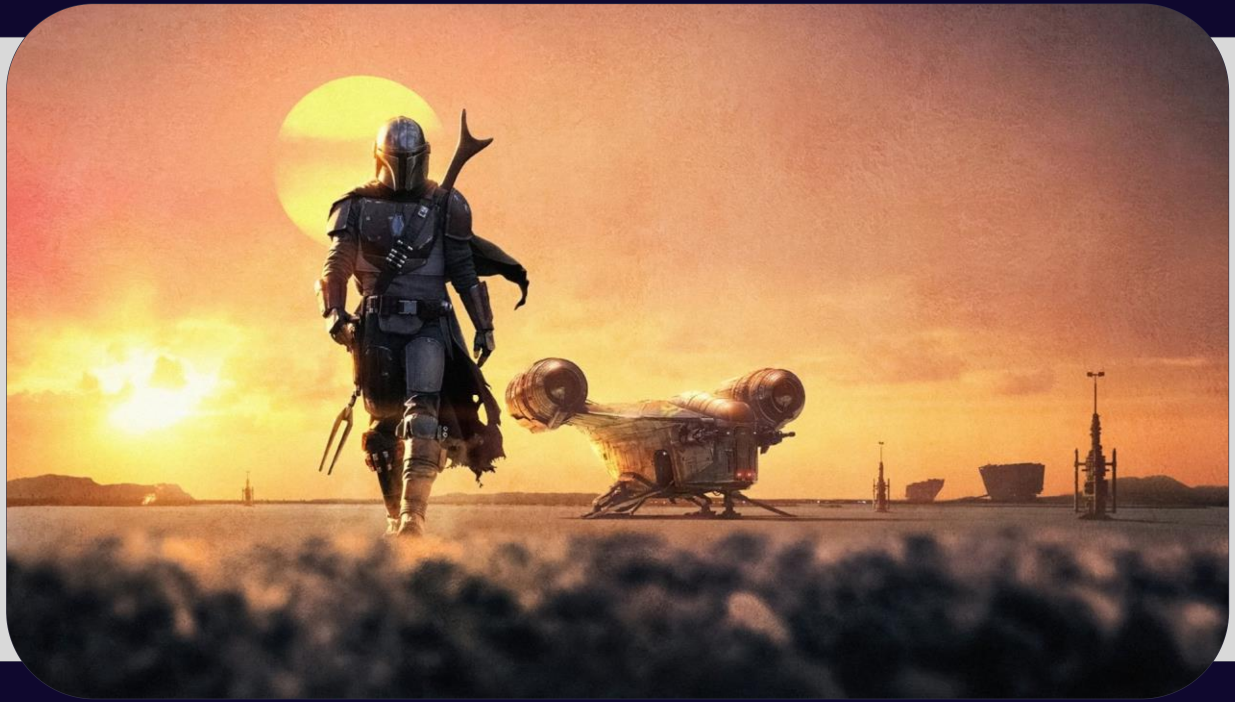
01

Dataset analysis: item length

There are movies or mono-episodes series

01

Dataset analysis: item length



There are series with more than one episode

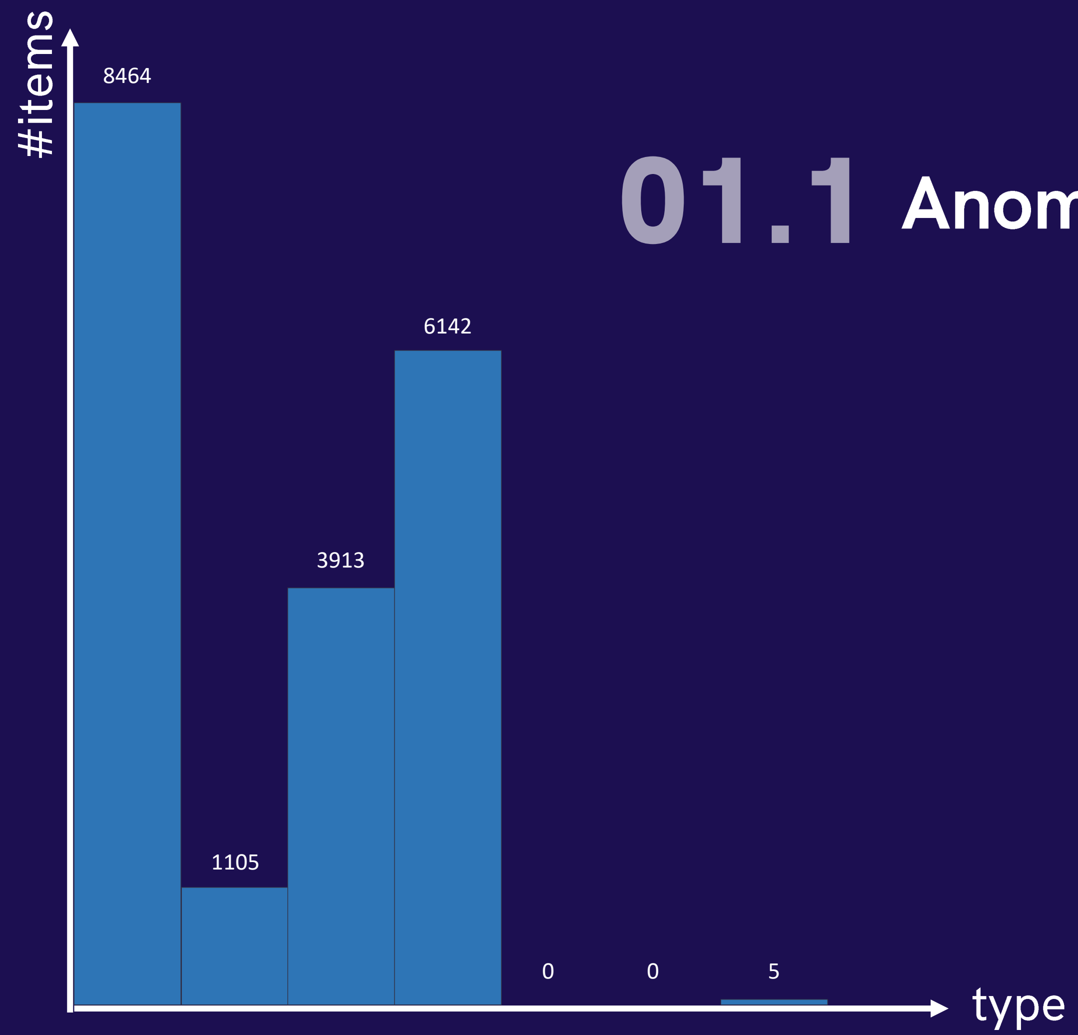
01

Dataset analysis: item length

There are series with probably “corrupted” values (10k)

01

Dataset analysis: item types, interactions



01.1 Anomalous item type distribution

01

Dataset analysis: item types, interactions

01.1 Anomalous item type distribution

01.2 No cold users

01

Dataset analysis: item types, interactions

01.1 Anomalous item type distribution

01.2 No cold users

01.3 3000 cold items

02

Our resources

02.1 Personal resources

Macbook Pro 2019

Lenovo Yoga Slim 7 Pro

02

Our resources

02.1 Personal resources

Macbook Pro 2019

Lenovo Yoga Slim 7 Pro

02.2 Cloud

Kaggle

Microsoft Azure Student

03 Approaches

What did work

Stacking URM + ICM
- Weighted stacking too

Binary URM and ICM

Non-binary URM
- Based on interaction count
- Normalized using tanh()

Hierarchical hybrids

What made us happy
at least once

List combination
- Round robin
- Condorcet-Schulze method

What did not work

Pipelined hybrids

Recommenders tailored on specific groups of users
- Using the profile length
- Using K-Means

Recommenders based on user-similarity

Feature weighting

03 Approaches

What did work

Stacking URM + ICM
Weighted stacking too

Binary URM and ICM

Non-binary URM
Based on interaction count
Normalized using tanh()

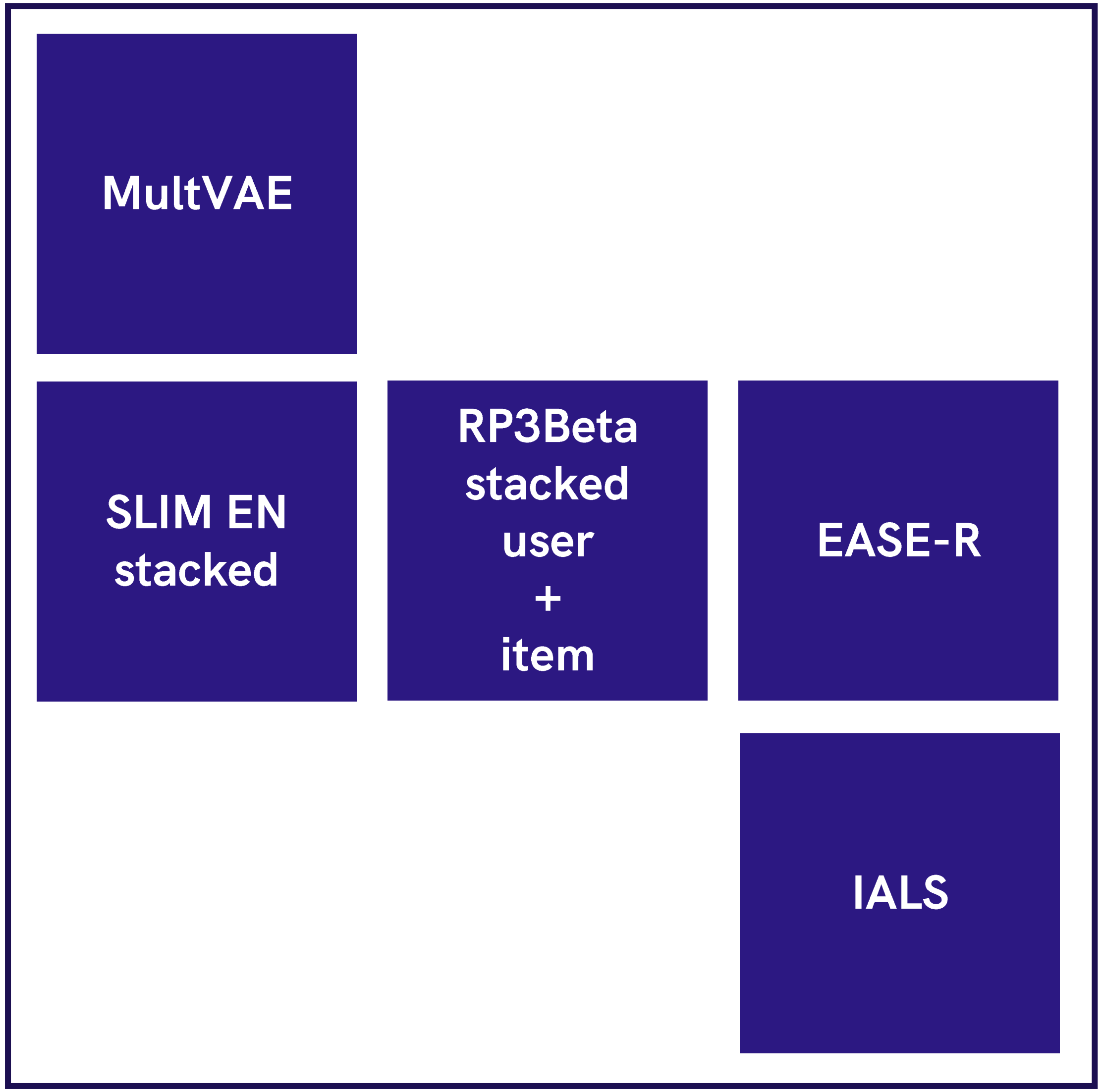
Hierarchical hybrids

But mostly

Trial and Error

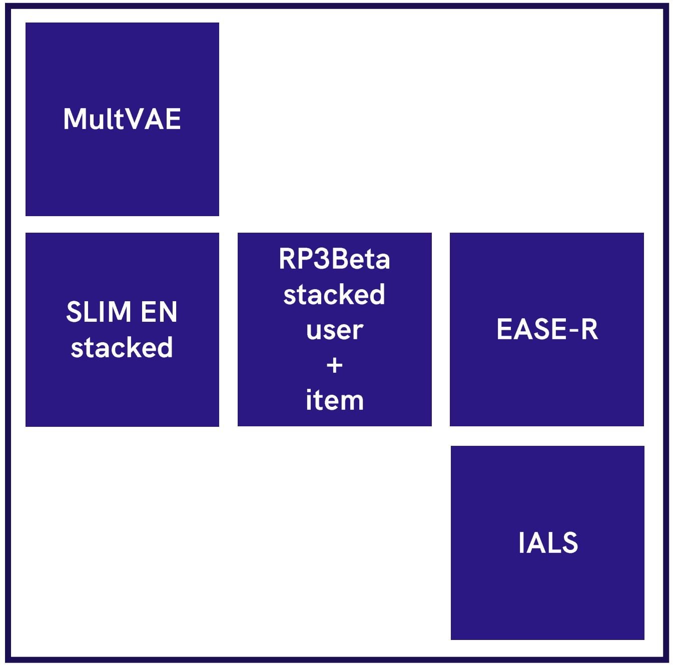
04 Our solution The “binary-trained” hybrid

04 Our solution The “binary-trained” hybrid

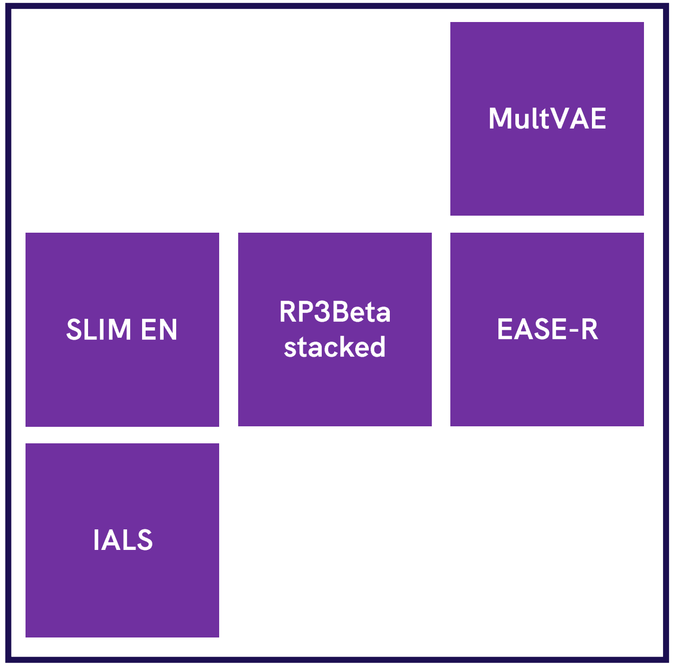


04 Our solution

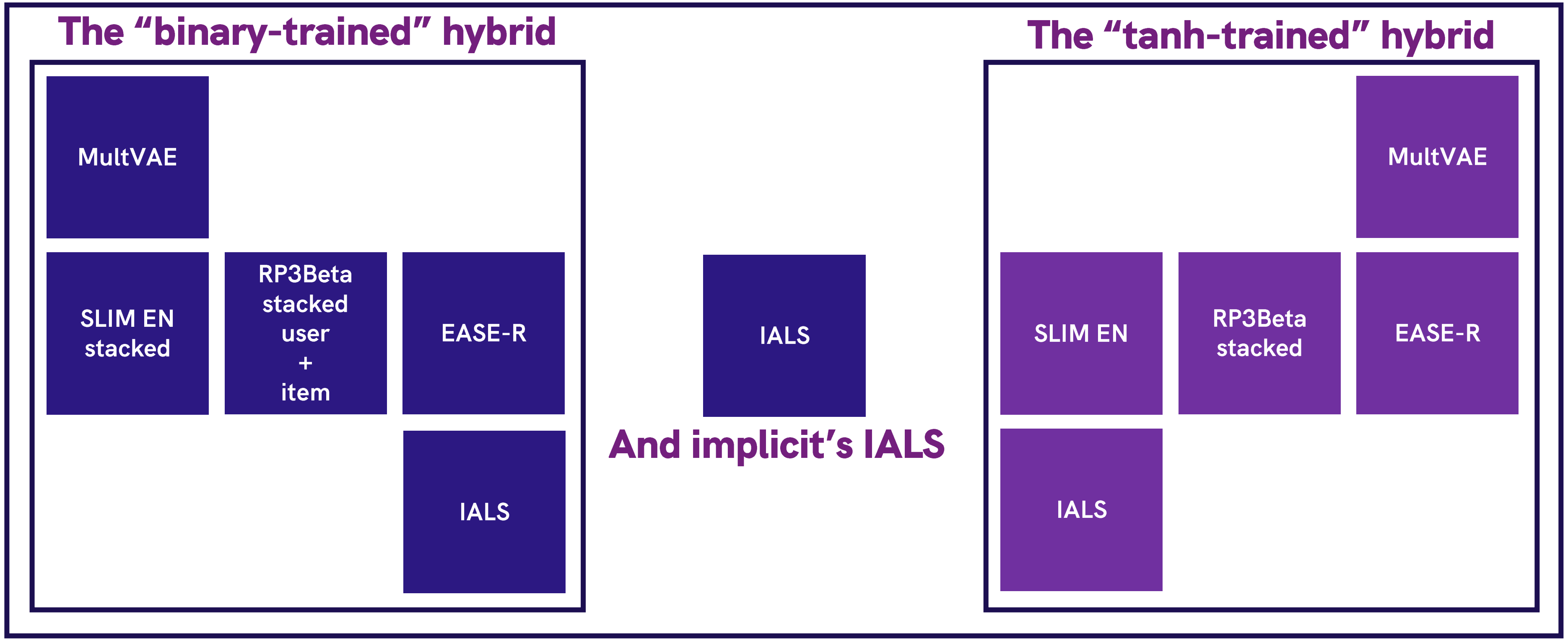
The “binary-trained” hybrid



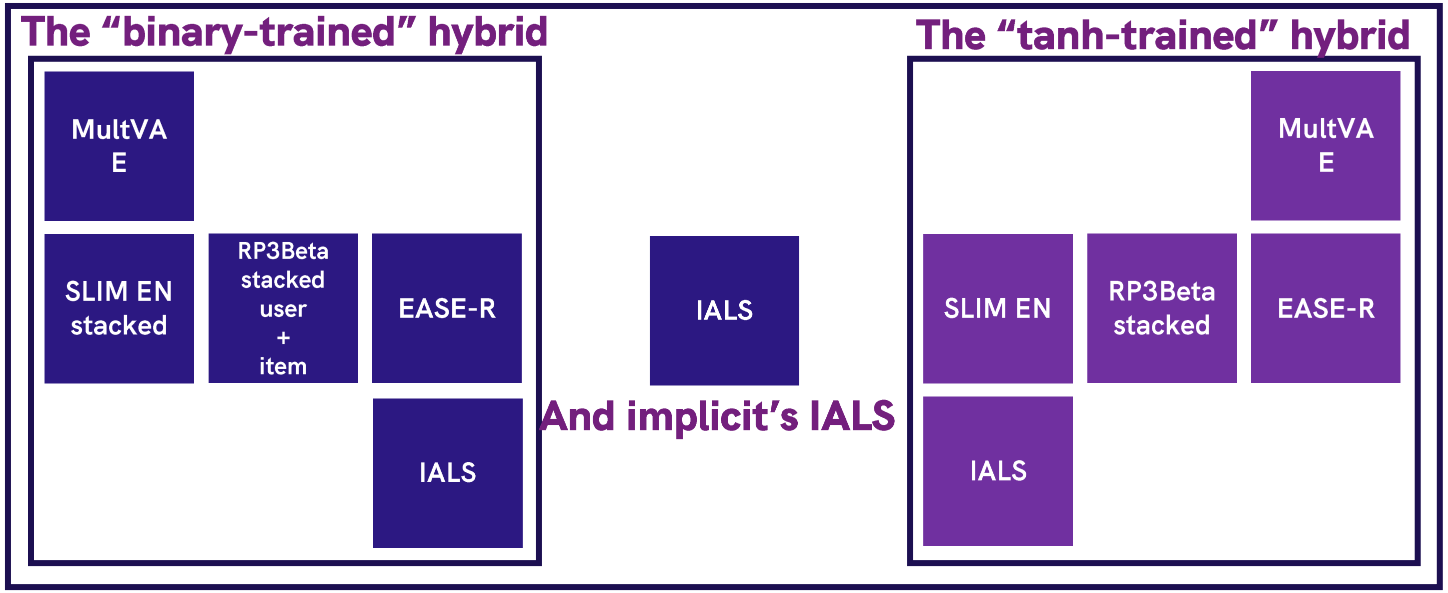
The “tanh-trained” hybrid



04 Our solution



04 Our solution



Final MAP@10:

0.06101 private
0.06200 public

Intuition:

Different URMs imply different results

Algorithms coefficients:

Binary:	6	
SLIM-EN:		400
IALS:		41
MultVAE:		4
RP3:		4
EASE:		27
TanH:	60	
SLIM-EN:		62
IALS:		22
MultVAE:		2
RP3:		200
EASE:		50
IALS:	169	

05 Further improvements

Reranking with XGBRanker

As seen in MaurizioFD's notebook

(even) More Tuning

Especially the «tanh» hybrid

Reranking with impressions

Such as impression discounting

Thanks

POLITECNICO
MILANO 1863

eit Digital