TECHFEST.PREFERRED.AI

23rd August 2019

Supported by:

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Office of
Postgraduate Research
Programmes

School of
Information Systems

About Us

PREFERRED.AI is a research group at SMU School of Information Systems (SIS). In this TechFest, we will share recent projects that the group is actively pursuing.

Mission

Our mission is to "push the envelope" on learning user preferences from data to improve the effectiveness and efficiency of recommendations using data mining, machine learning, and artificial intelligence. This encompasses designing algorithms for mining user-generated data of various modalities (e.g., ratings, text, images, social networks) for understanding the behaviours and preferences of users (individually and collectively), and applying the mined knowledge to develop user-centric intelligent applications.

Programme

SESSION I – TALKS (3.30PM to 5.00PM)

SMU School of Information Systems, Seminar Room B1-1

- **Preferred.AI:** Preferences and Recommendations from Data & AI
  – an overview of our activities and how you can get involved
- **PCRL:** Jointly Modeling User Preferences and Learning Deep Item Features from Auxiliary Data
  – a publication at The Conference on Uncertainty in Artificial Intelligence (**UAI-18**)
- **MRG:** Multimodal Review Generation for Recommender Systems
  – a publication at The Web Conference (**WWW-19**)
- **CompareLDA:** A Topic Model for Document Comparison
  – a publication at The AAAI Conference on Artificial Intelligence (**AAAI-19**)
- **MP-SimRank:** Multiperspective Graph-theoretic Similarity Measure
  – a publication at The ACM Conference on Information and Knowledge Management (**CIKM-18**)

SESSION II – POSTERS AND DEMOS (5.00PM to 6.30PM)

SMU School of Information Systems, Concourse opposite OCBC

- **Cerebro**: Closed-loop recommendation retrieval engine
- **Cornac**: Multimodal recommender system library
- **JioApp**: Recommendations for group meetups
- **Propedia**: Web-mined product encyclopedia
- **SentiVec**: Sentiment-infused word embeddings
- **ThriftCity**: Web-sourced price comparisons
- **Venom**: Focused crawler for the deep Web
- **Butler**: Conversational recommender system with natural language explanations
- **FaceInMotion**: Face-based intelligent emotion detection
- **MindReader**: News recommendation app based on reading history
- **Neural Network Lab**: Machine learning in your browser

PCRL: Jointly Modeling User Preferences and Learning Deep Item Features from Auxiliary Data

https://cornac.preferred.ai

Personalized recommendation

	5	3	?	?
	?	?	2	?
	?	?	5	4

Preference data: user-item interactions
e.g., **clicks, ratings, purchases**

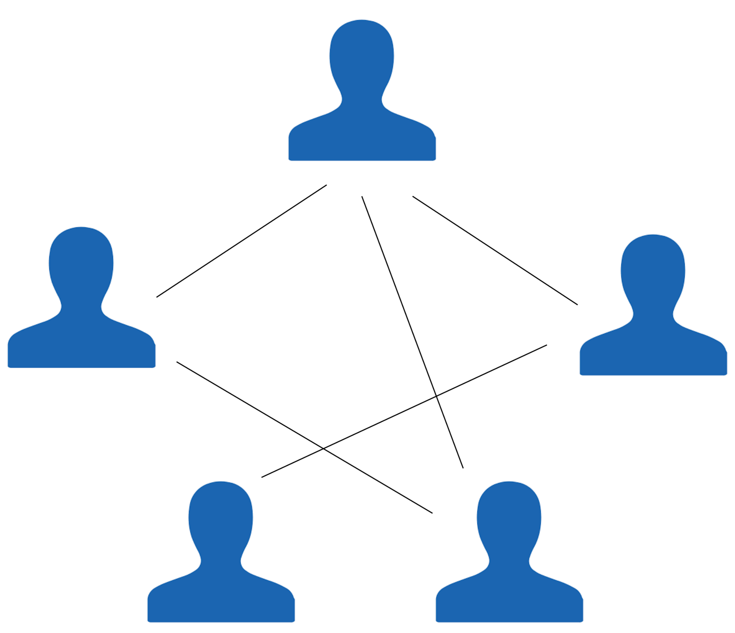
Objective

- ▶Predict unknown user-item interactions.

Challenge

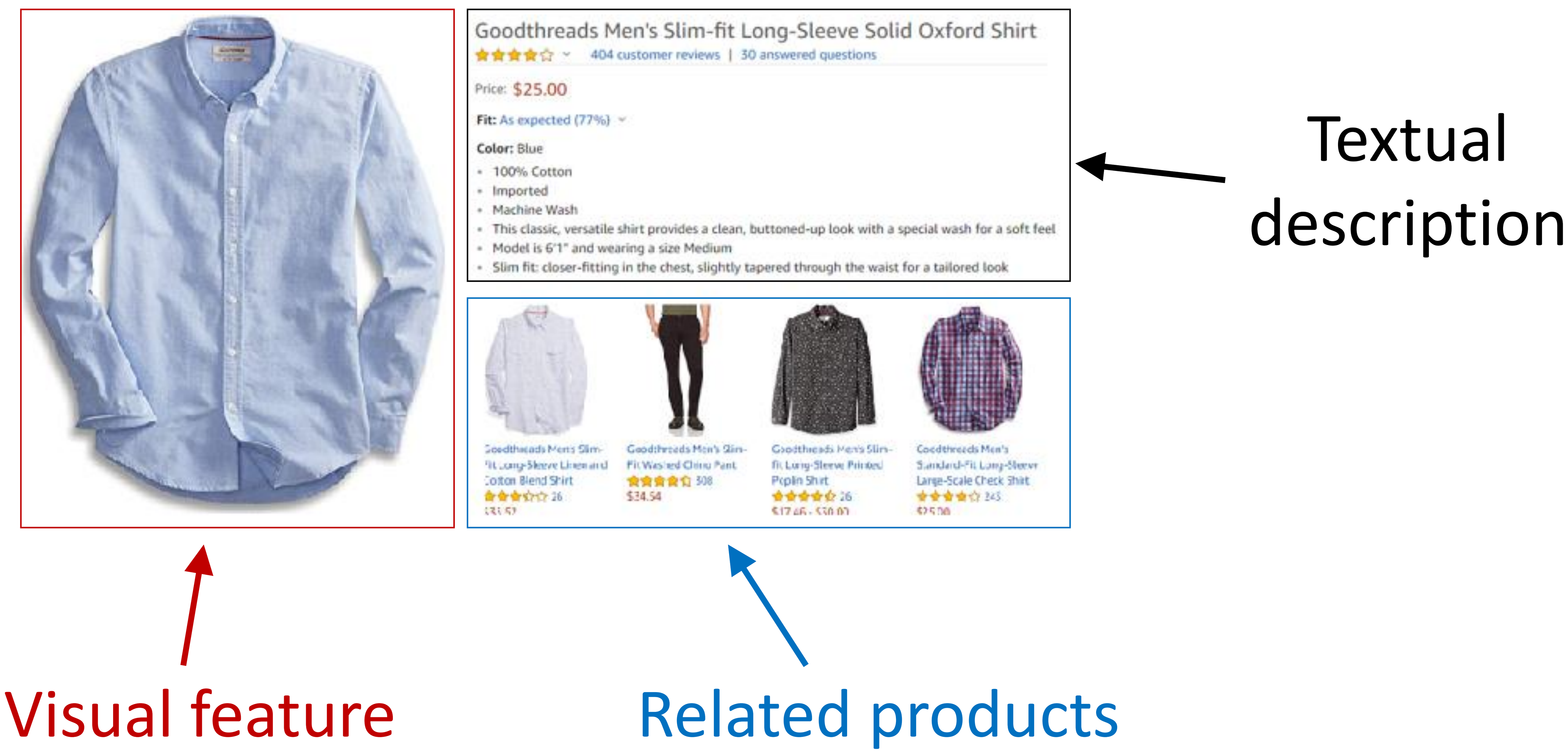
- ▶Data is extremely sparse
  - ▶Do not cover all aspects of user behavior
  - ▶Difficult to generalize a user's preference

Personalized recommendation



Social network

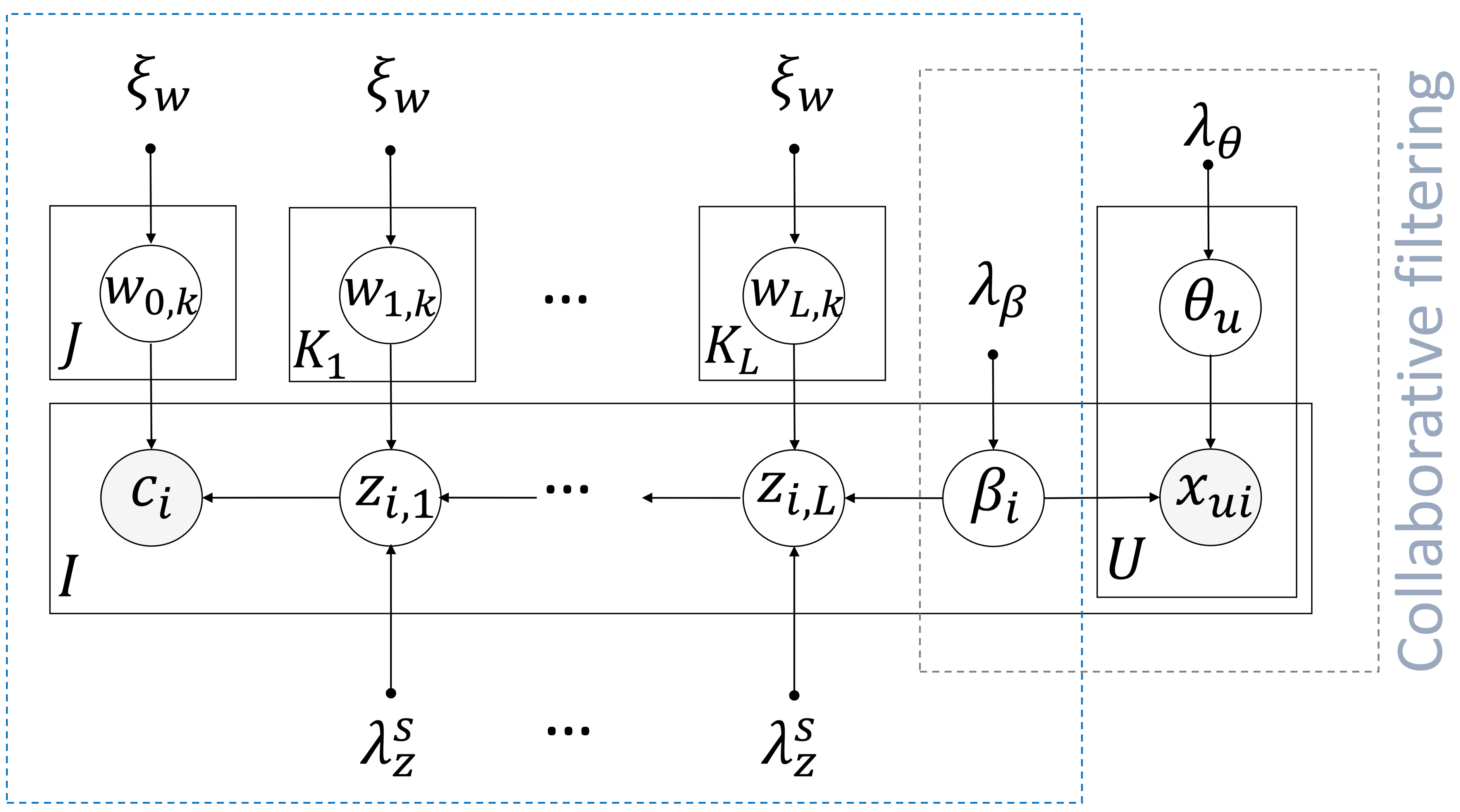
User auxiliary data



Item auxiliary information

PCRL: putting it all together

Deep item representation



PCRL's graphical model

Intuition

- ▶ Preferences guide representation learning
- ▶ Content helps in predicting preferences

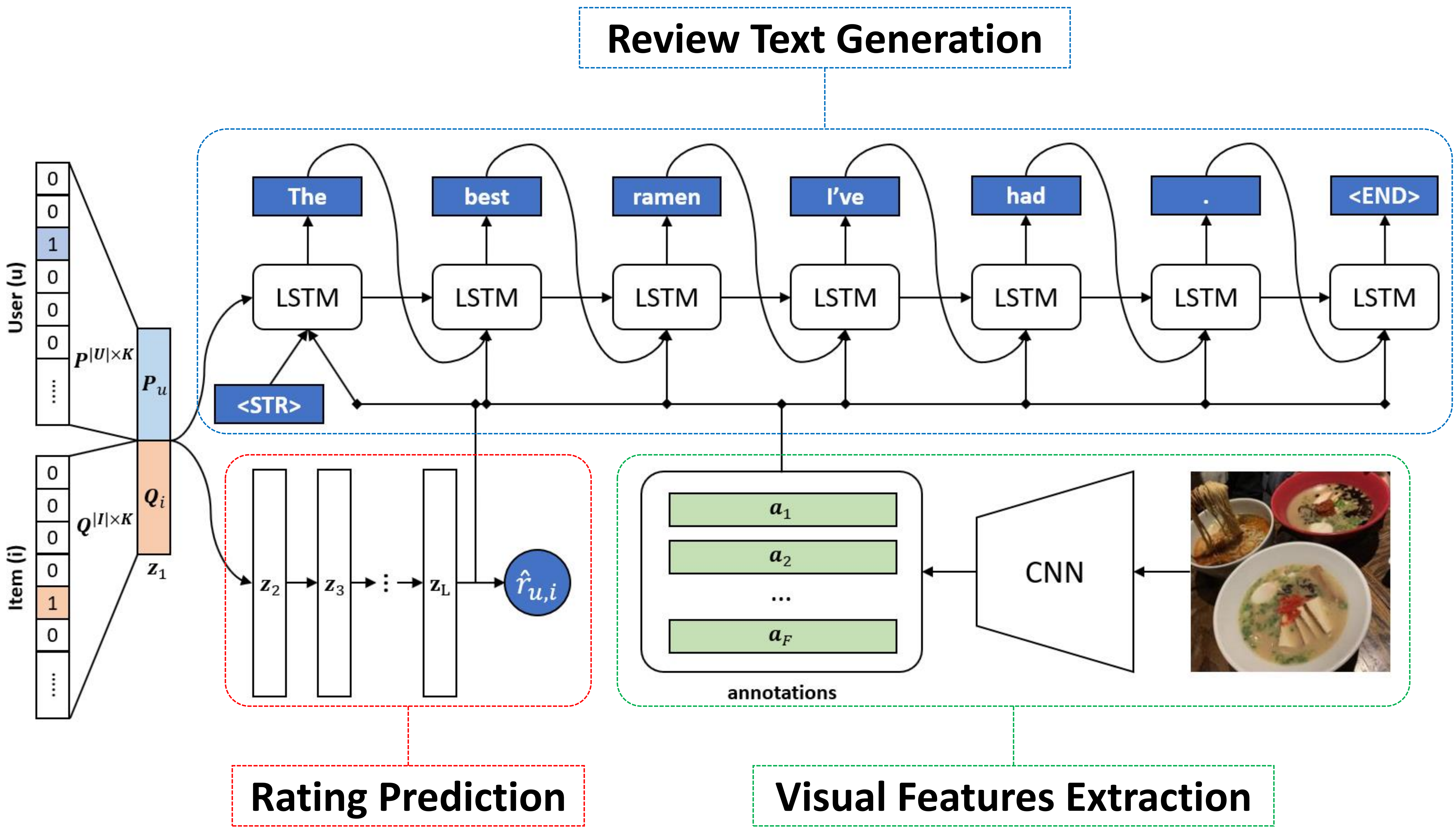
MGR: Multimodal Review Generation for Recommender Systems

https://code.preferred.ai/mrg

Problem

- Given:
  - a user
  - an item
  - an image (*optional*)
- Output:
  - rating (for *recommendation*)
  - review text (potentially for *explanation*)

Multimodal Review Generation



CompareLDA: A Topic Model for Document Classification

https://code.preferred.ai/Compare-LDA

Product Comparison

>

80% **Topic 1:** low-light, matrix, sensor, pictures, … 10%

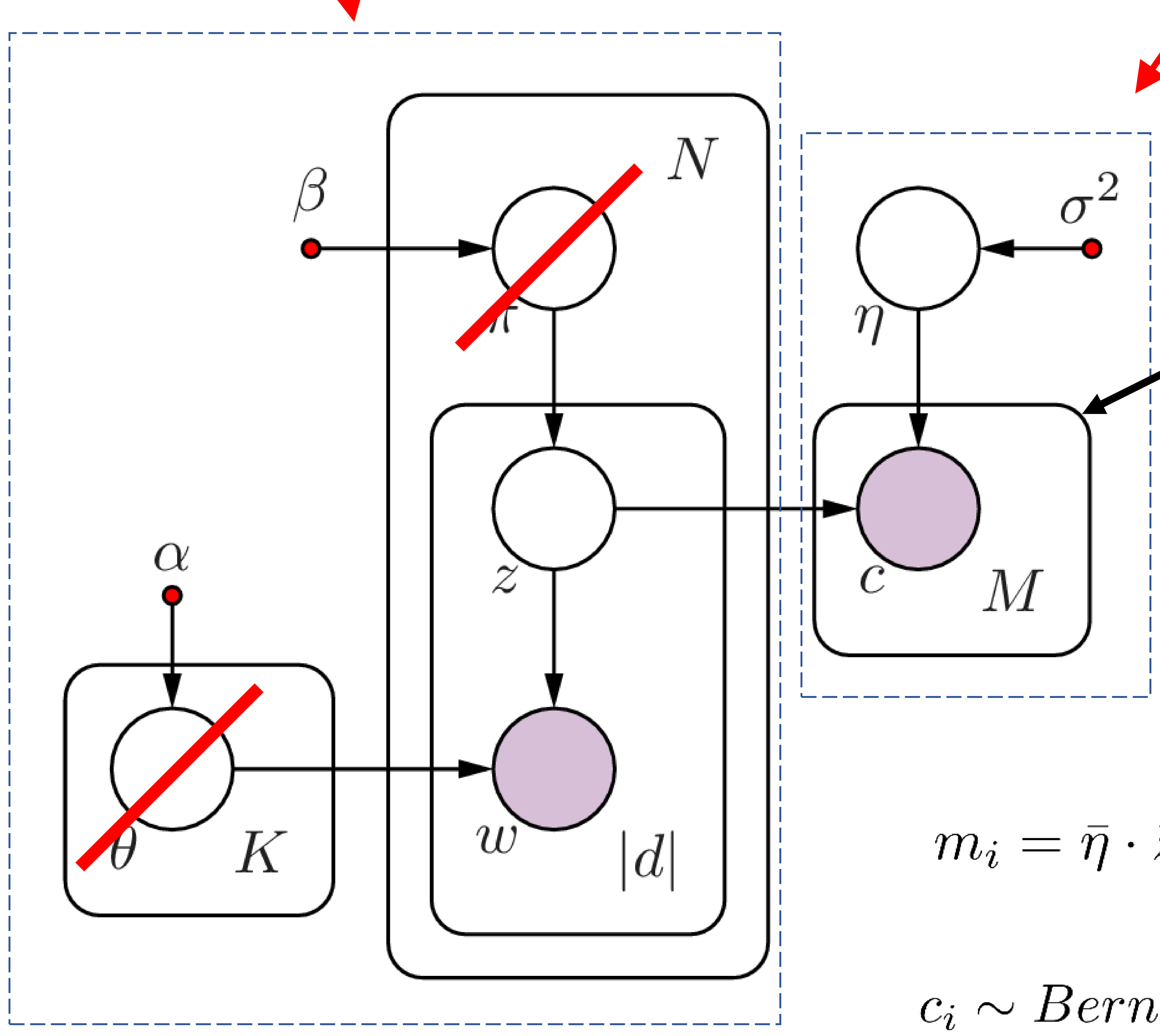
9% **Topic 2:** compact, light, handy, carry, … 12%

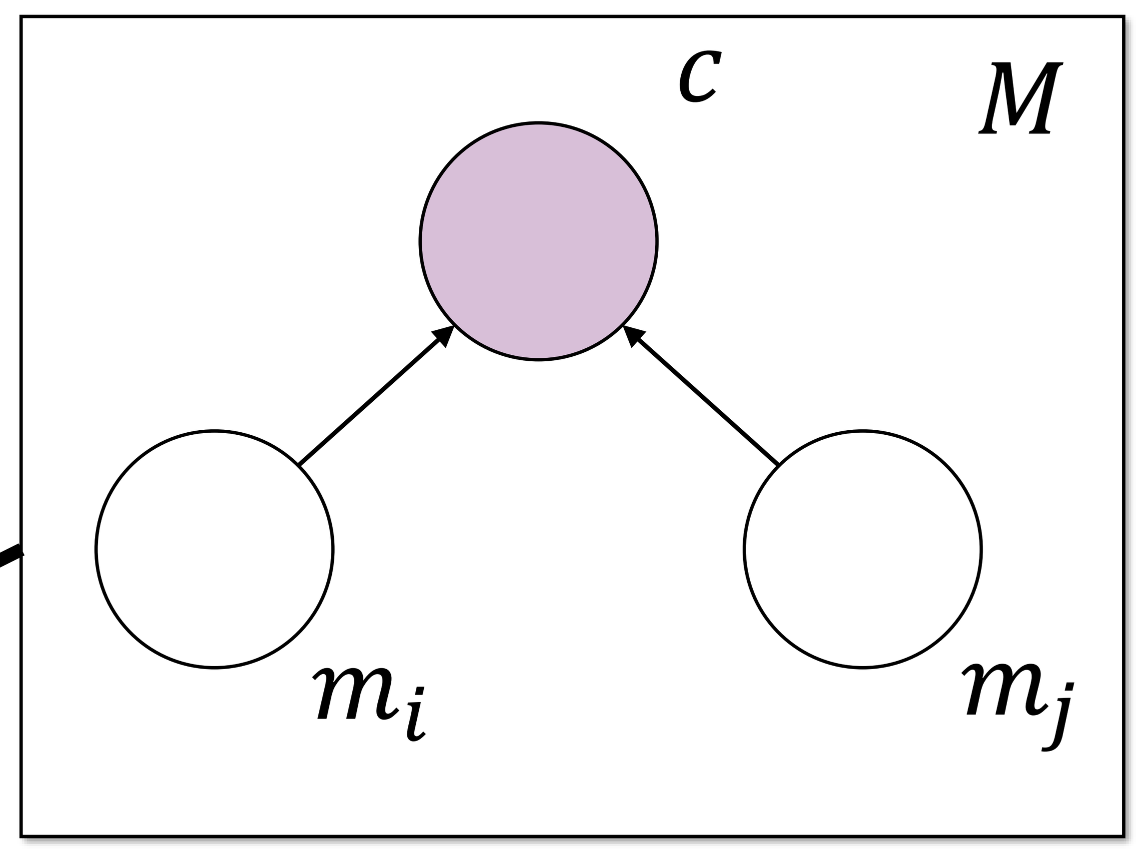
Topic 3: …

Multimodal Review Generation

Latent Dirichlet Allocation (LDA)

BTL-R Component





$$m_i = \bar{\eta} \cdot \bar{z}_i = \bar{\eta} \cdot \left(\frac{1}{|d_i|} \sum_{i=1}^{|d_i|} z_{ij} \right)$$

$$c_i \sim Bernoulli(\sigma(m_{r_i[1]} - m_{r_i[2]}))$$

MP-SimRank: Multiperspective Graph-Theoretic Similarity Measure

https://code.preferred.ai/mp-simrank

Questions

How to tell when two objects are similar?

According to what perspective?

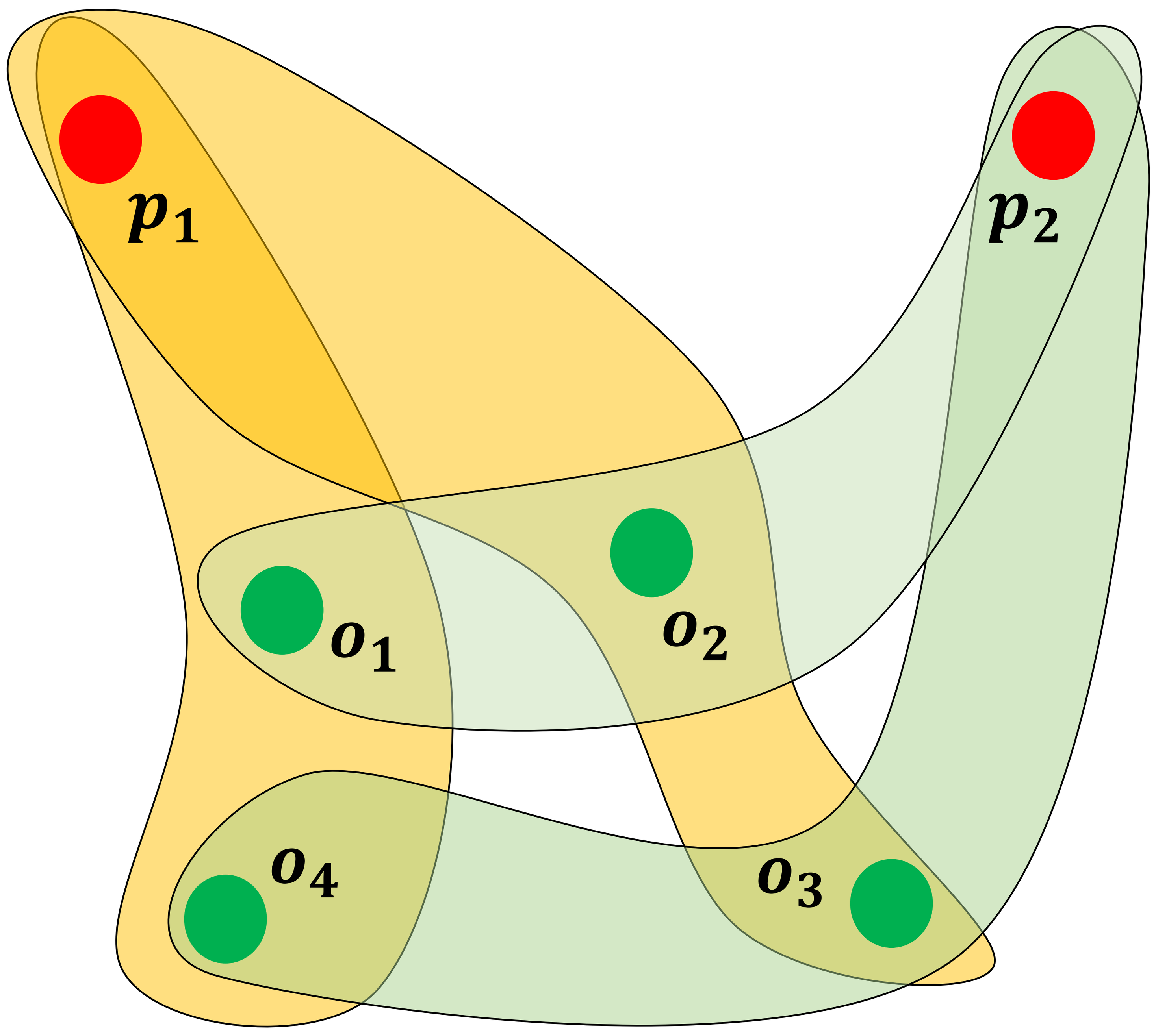
Challenges

- ❑ Similarity observations for each perspective might be under-sampled, do not cover all objects.
- ❑ Two objects are similar according to one perspective, but not to the other.

Which picture is more similar to picture **(a)**?
(b) or **(c)**?

Multiperspective Similarity Measure

$$\mathbf{S}_{\mathbf{p}}^{(t+1)}(o_i, o_j) = \frac{C}{|P|} \sum_{p^* \in P} \mathbf{sim}^{(t)}(p, p^*) \times \sum_{o_k \in \mathrm{N}_{p^*}(o_i)} \sum_{o_l \in \mathrm{N}_{p^*}(o_j)} \frac{\boldsymbol{S}_{\boldsymbol{p}^*}^{(t)}(o_k, o_l)}{|\mathrm{N}_{p^*}(o_i)||\mathrm{N}_{p^*}(o_j)|}$$



Base cases: $\begin{cases} \mathbf{sim}^{(0)}(\mathrm{p}, \mathrm{p}^*) = 1 \text{ if } \mathrm{p} = \mathrm{p}^* \text{ and } 0 \text{ otherwise} \\ \mathbf{S}_{\mathbf{p}}^{(0)}(\mathrm{o_i}, \mathrm{o_j}) = 1 \text{ if } \mathrm{i} = \mathrm{j}, \forall p \text{ and } 0 \text{ otherwise} \end{cases}$

Need to measure: $\mathbf{sim}(p, p^*)$ and $\boldsymbol{S}_{\boldsymbol{p}}(o_i, o_j)$

Cerebro

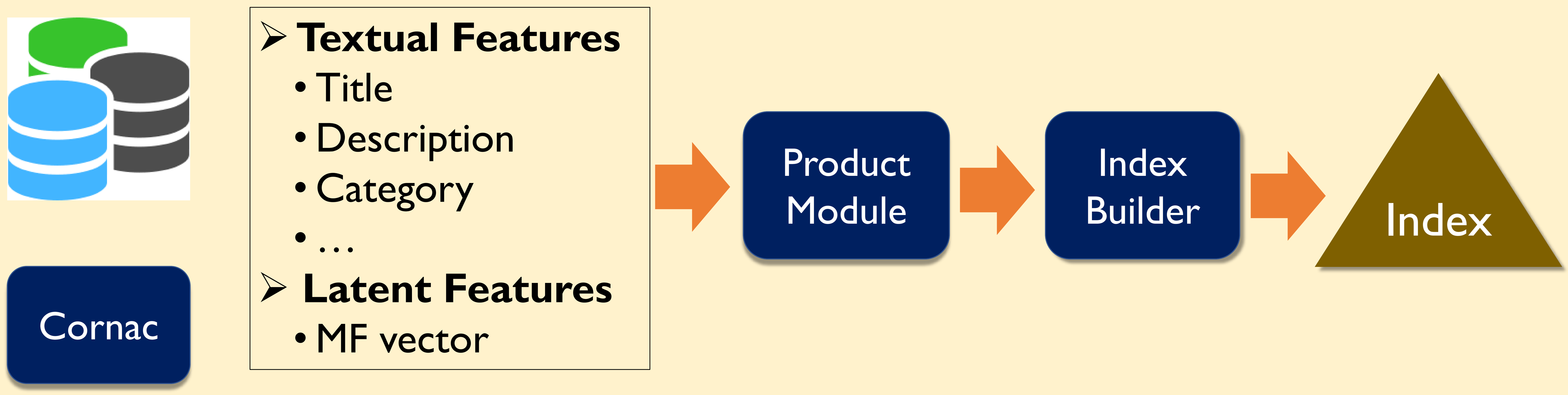
https://cerebro.preferred.ai/

PREFERRED.AI

Closed-Loop Recommendation Retrieval Engine

Fast Retrieval

Implementing indexing structures to support various types of information retrieval, e.g., personalized recommendation, personalized keyword search, and similar products search.

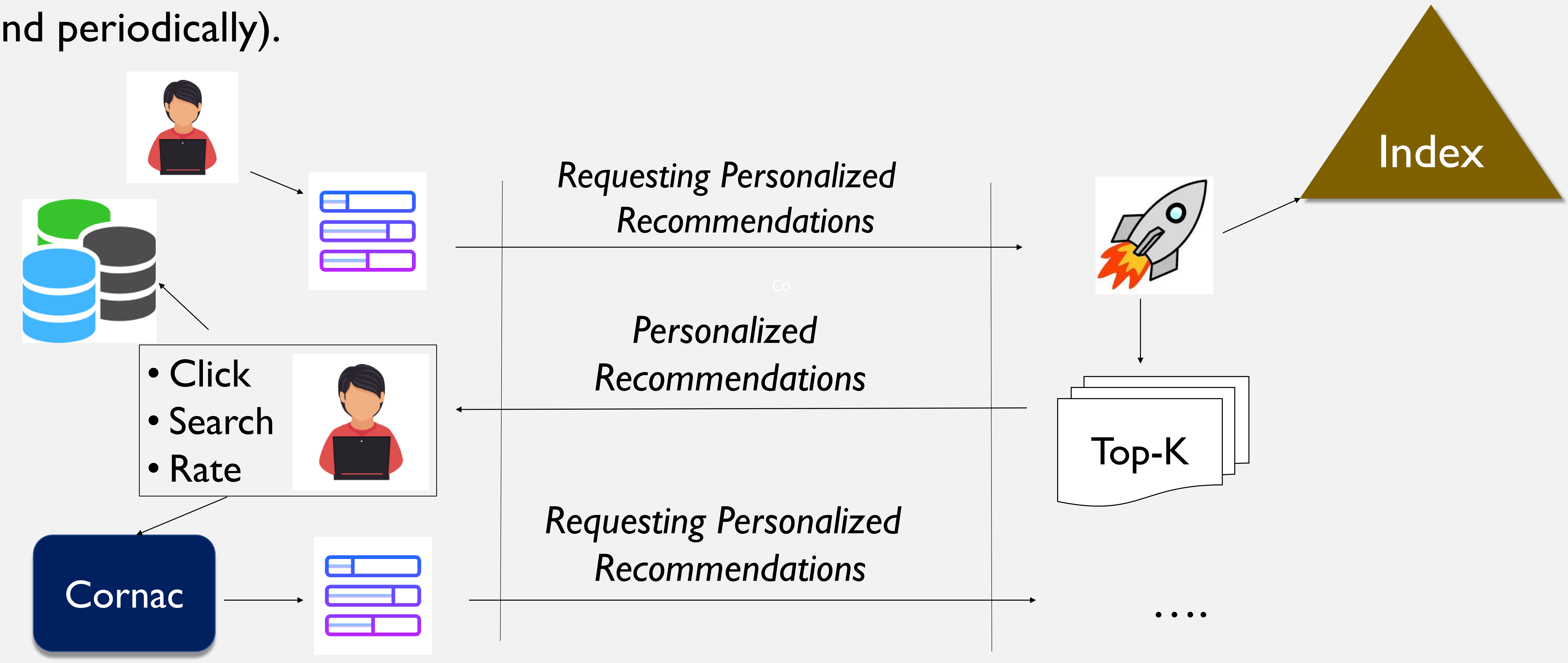


https://cornac.preferred.ai/

Sub-linear Time Retrieval via Indexing

Recommendation Framework

Initiating and managing recommendation session, tracking users' actions (e.g., clicking, rating, searching, etc.), updating users' personalized recommendations (within session and periodically).



https://cornac.preferred.ai/

Closed-loop and End-to-end

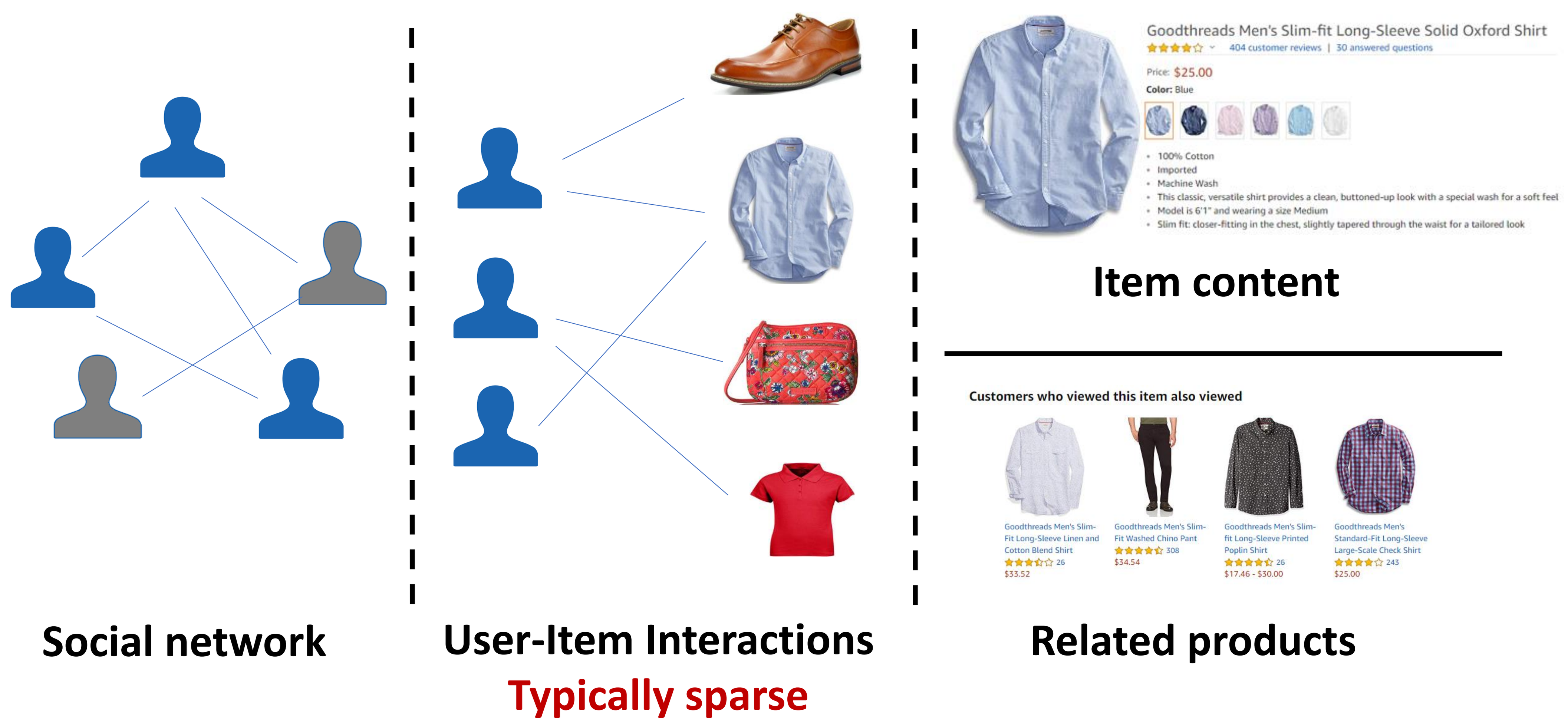
Cornac

http://cornac.preferred.ai

PREFERRED.AI

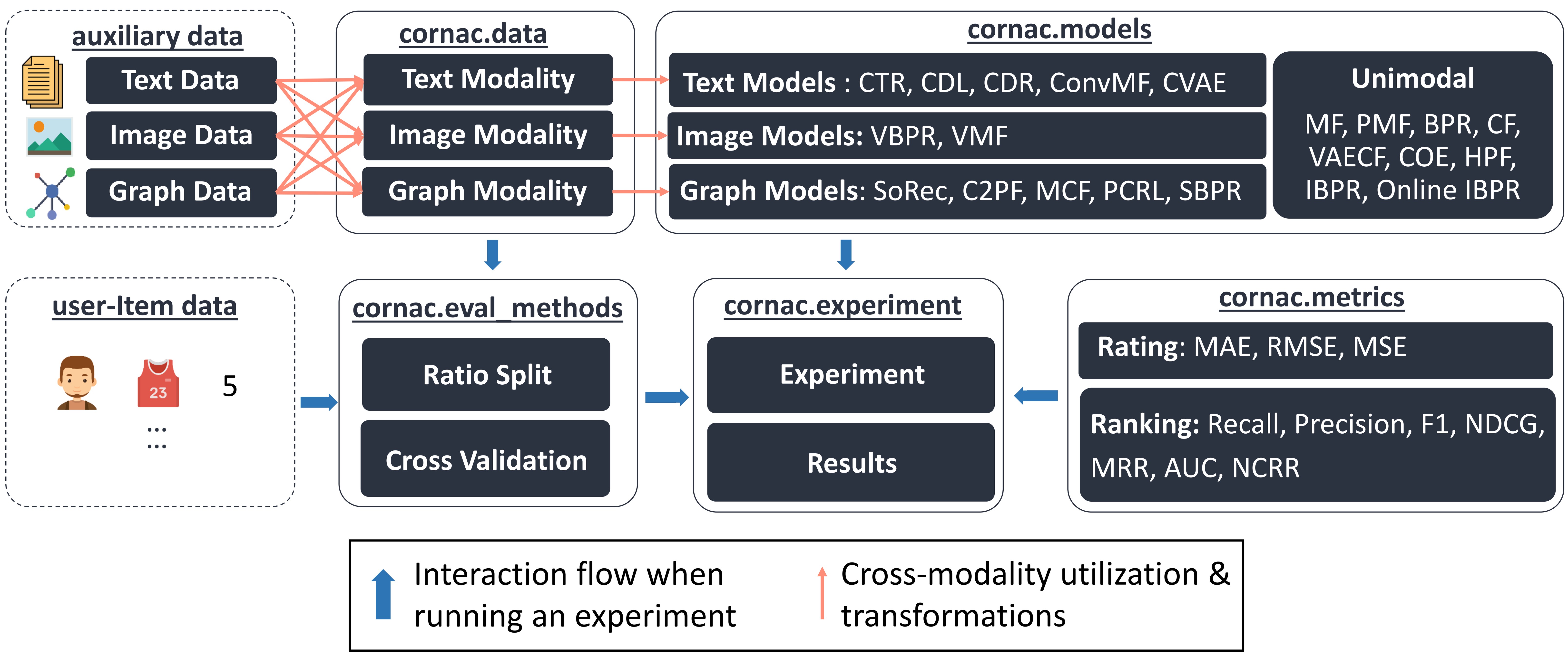
A Library for Multimodal Recommender System

▶ *Multimodal Recommender Systems* leverage **auxiliary information** to alleviate data sparsity



- Fast experimentation, exploration, and comparisons
- Convenient development of new multimodal recommenders
- Open-access to a rich collection of recommendation models
- Straightforward usage of real-world benchmark datasets

Structures



Key Features

Multimodality. Data infrastructures make it convenient to work with auxiliary information and enable seamless cross-modality comparisons.

Scalability. A rich collection of iterators for easy stochastic optimization. Model implementations make use of Cython to achieve C/C++ performance.

Reproducibility. Full control over random number generators, open-access to existing algorithms and built-in datasets for reproducible research.

SMU SINGAPORE MANAGEMENT UNIVERSITY
School of Information Systems

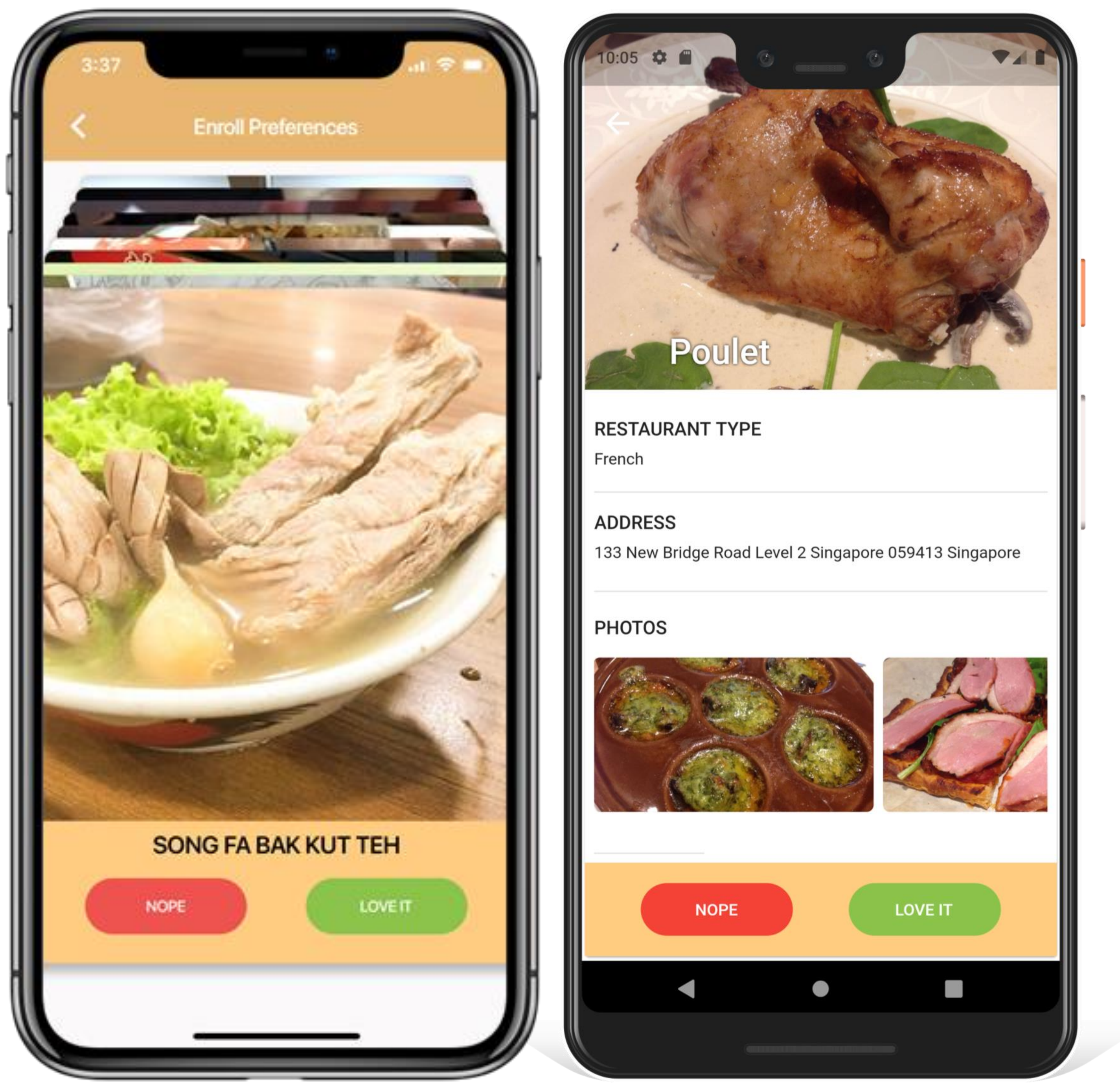
JioApp

https://jioapp.preferred.ai

PREFERRED.AI

Food Recommendations for Group Meetups

- Personalised group-based food recommendations application
- Aggregates food preferences of users in a group
- Provides list of food establishments based on the overall preference of the group



Download on the App Store

GET IT ON Google Play

Why JioApp?

Ease of planning group meetups

Know everyone's food preferences

Find a food place that most group members likes

Key Features

EASY ACCESS TO FOOD OPTIONS

More than 7000 food establishments over 26 Singapore regions

SIMPLE EVENT CREATION

Create events by selecting attendees from your contact list

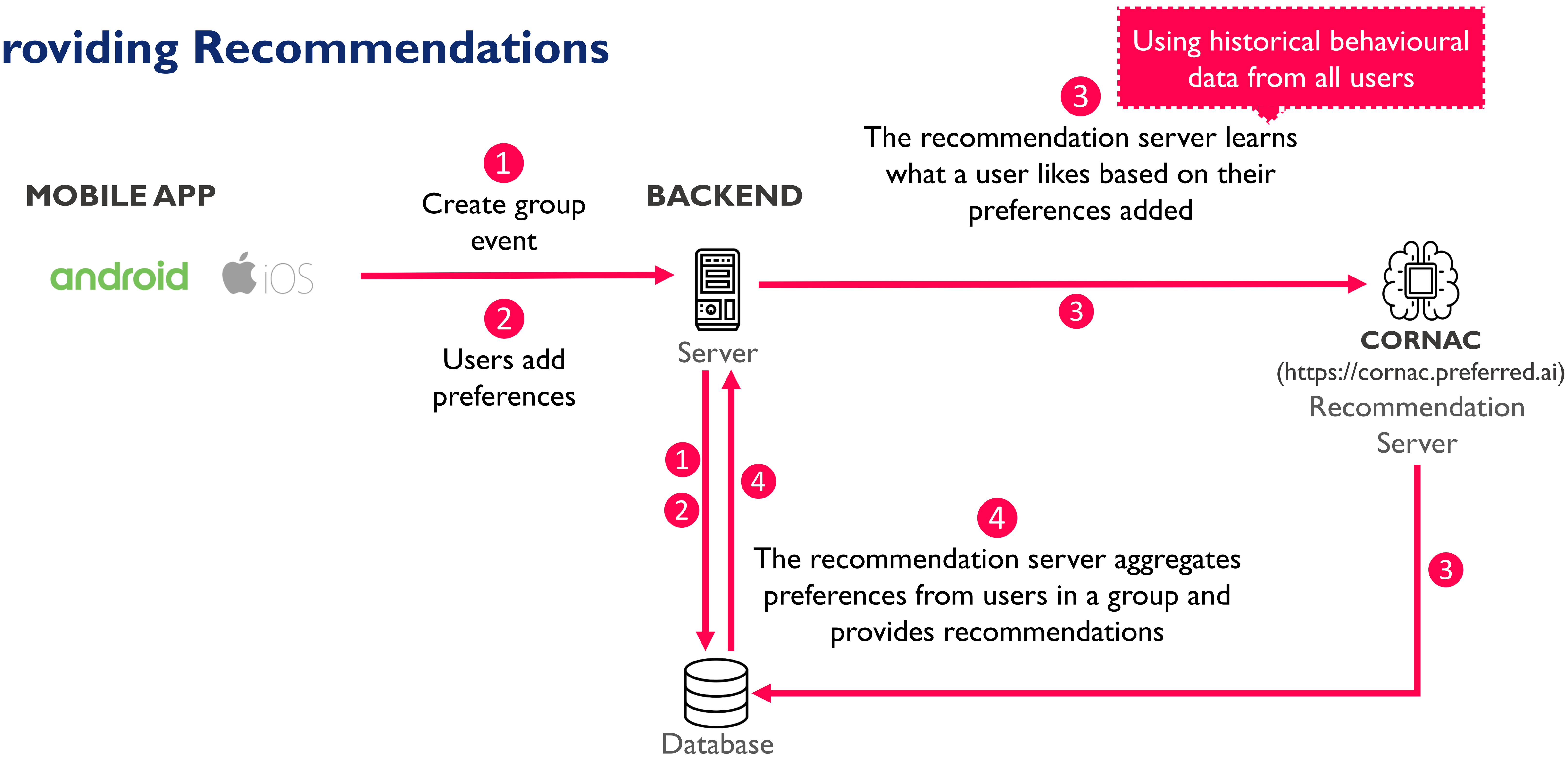
SWIPE TO YOUR DESIRES

Swipe your screen to have your preferences recorded

FEWER DECISIONS, MORE INTERACTION

Our machine learning algorithm recommends the best locations for your group

Providing Recommendations



Propedia

https://propedia.preferred.ai

Web-Mined Product Encyclopedia

Multi-Source

Comprehensive catalogue integrated from different e-commerce platforms

amazon.com

淘宝网 Taobao.com

Lazada

亚马逊 amazon.cn

Multi-Lingual

Ease of accessibility to product information mined from different languages

中文 | English

Feed1: BERT Sentiment Analysis	P 96. / R 97. F1 96.
Feed 2: LingPipe Sentiment Analysis	P 96. / R 90. F1 93.

PROPEDIA

Multi-Feed

Parallel versions of product information reflecting diverse viewpoints and methodologies

Up-to-Date

Continual data collection using Venom crawler (https://venom.preferred.ai)

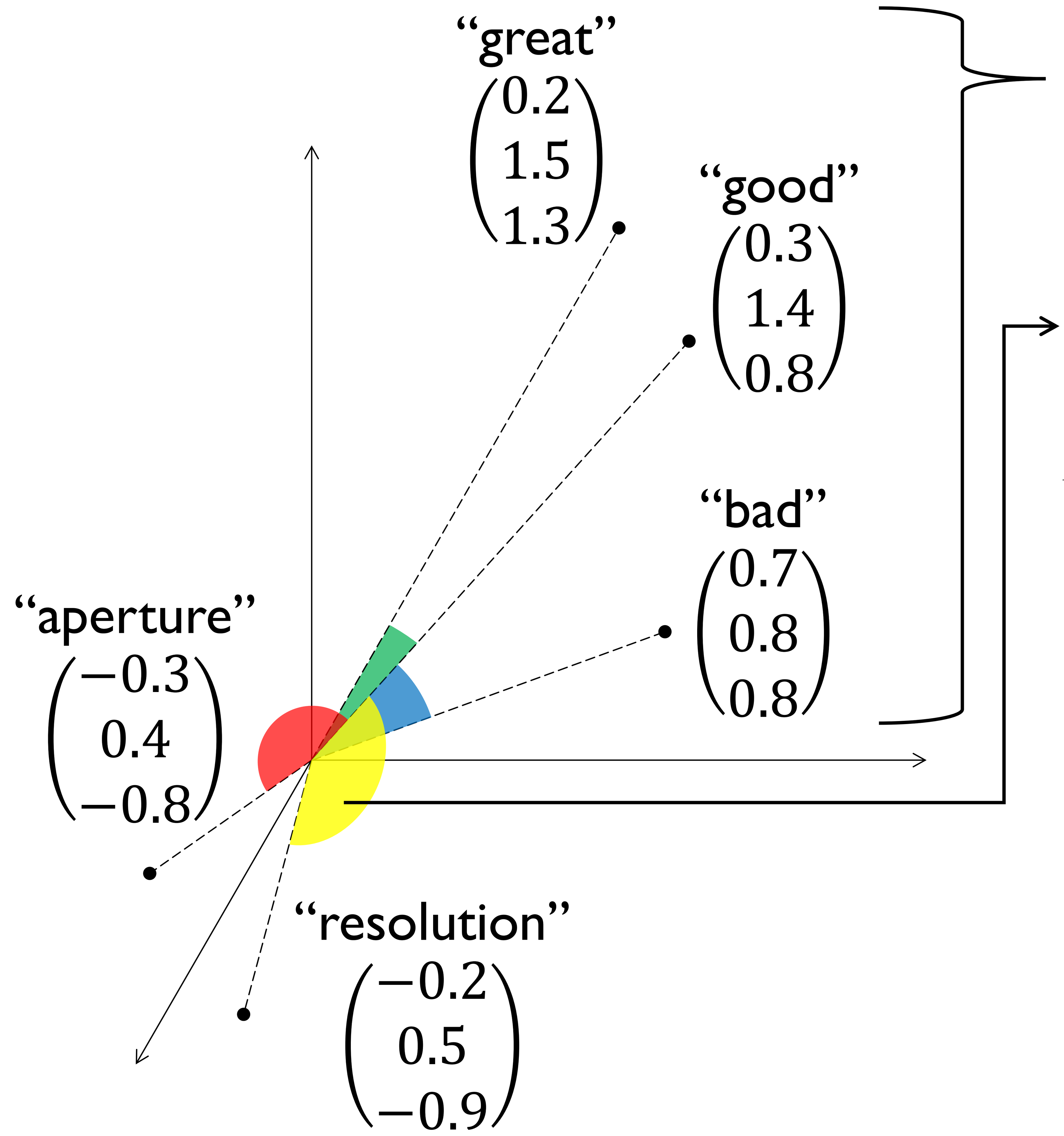
SMU SINGAPORE MANAGEMENT UNIVERSITY
School of Information Systems

SentiVec

http://sentivec.preferred.ai

PREFERRED.AI

Word Embeddings



Words with similar neighbours should have similar embeddings

Similarity is given by the angle between two embeddings

Applications

- Used in text classification tasks, such as sentiment analysis
  - Customer Profiling
  - Market Segmentation
- Typically obtained from distributional analysis methods e.g., Word2Vec

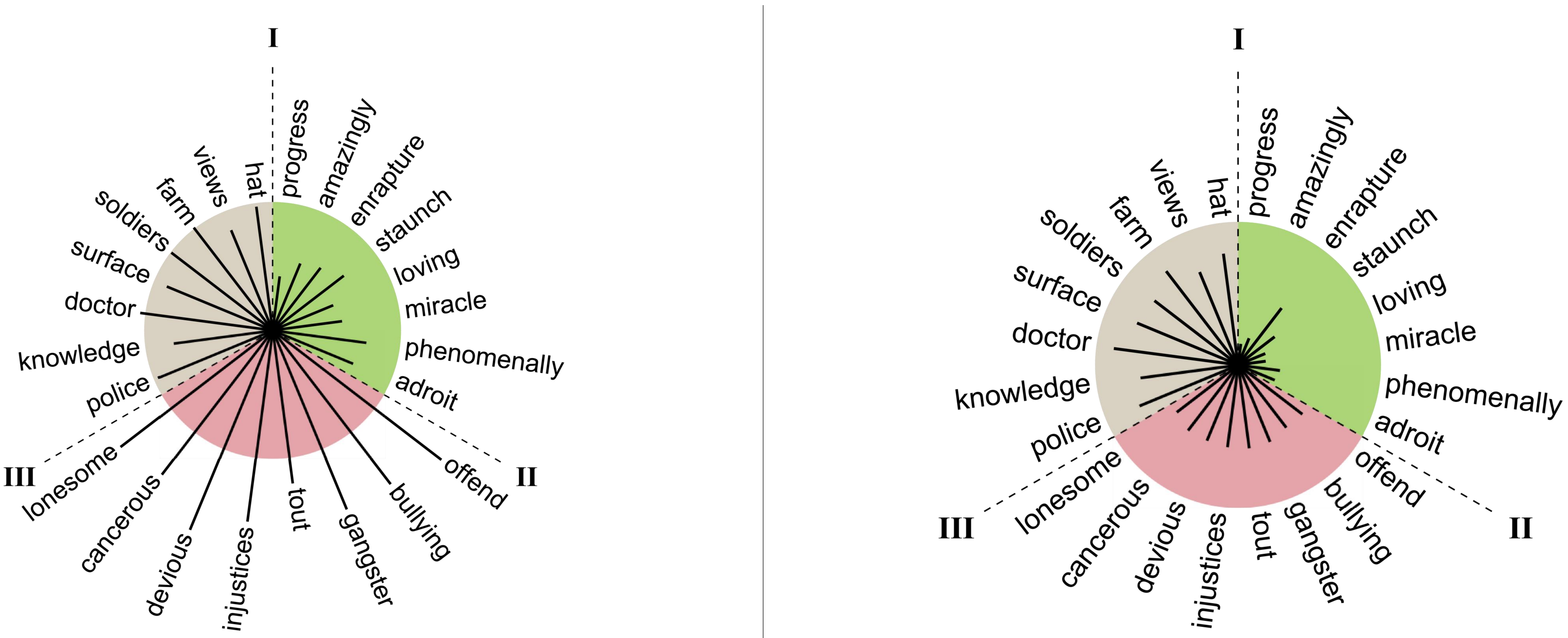
SentiVec: Sentiment-Infused Word Embeddings

Sentiment words (i.e. good, bad) have similar neighbours

Incorporate alignment of words to sentiment, from external lexicon L

$$\log \mathcal{L}_{sentivec} = \log \mathcal{L}_{word2vec}(W; C) + \lambda \log \mathcal{L}_{sentiment}(W, L)$$

Up to 85.7% accuracy for sentiment classification, higher than Word2Vec



Target word: “good”

Target word: “evil”

Relative changes in vector similarity, contrasted with Word2Vec

SMU
SINGAPORE MANAGEMENT UNIVERSITY
School of Information Systems

ThriftCity

https://thriftcity.preferred.ai

PREFERRED.AI

Getting the best prices from retailers worldwide is just a click away

Lazada amazon 天猫 TMALL.COM

ThriftCity T

Wide Variety

Search a unified catalogue of products, integrated from multiple retailers using Propedia (https://propedia.preferred.ai). Bringing to you all your favourite products on one platform.

Fast and Convenient

Check the reviews as well as the latest prices with shipping and taxes included. Bringing to you comprehensive information about products.

Enjoy Savings $

Compare prices all-in between multiple retailers for the same product in your preferred currency. Linking you to retailers with the best offers.

Tailored for You, to You

ThriftCity remembers your perferences and fetches the most relevant products for you using Cerebro (https://cerebro.preferred.ai), a closed-loop recommendation retrieval engine.

Personalised Homepage

Personalised Search

Personalised Related Products

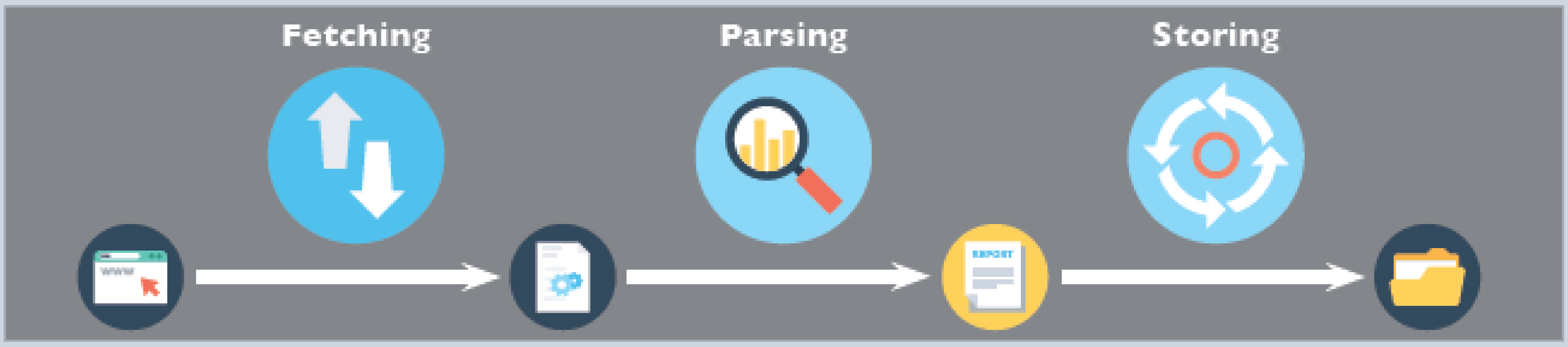
Venom

https://venom.preferred.ai

PREFERRED.AI

SMU SINGAPORE MANAGEMENT UNIVERSITY
School of Information Systems

Your preferred open source focused crawler for the deep Web



Venom is a feature-packed crawling framework built with essential Web scraping features that are seamlessly integrated with content parsing and storage. Easy to use in both prototyping and production, it is available on all operating systems that support JAVA.

Blazing Fast Performance

Leverages an asynchronous, event driven I/O model, that can send and process massive number of requests while parsing and indexing massive amount of data with built-in multithreading.

Fully Customisable

Combines high-level API with low-level fine tuning, providing different users with the right amount of control they need over their crawlers.

Highly Robust

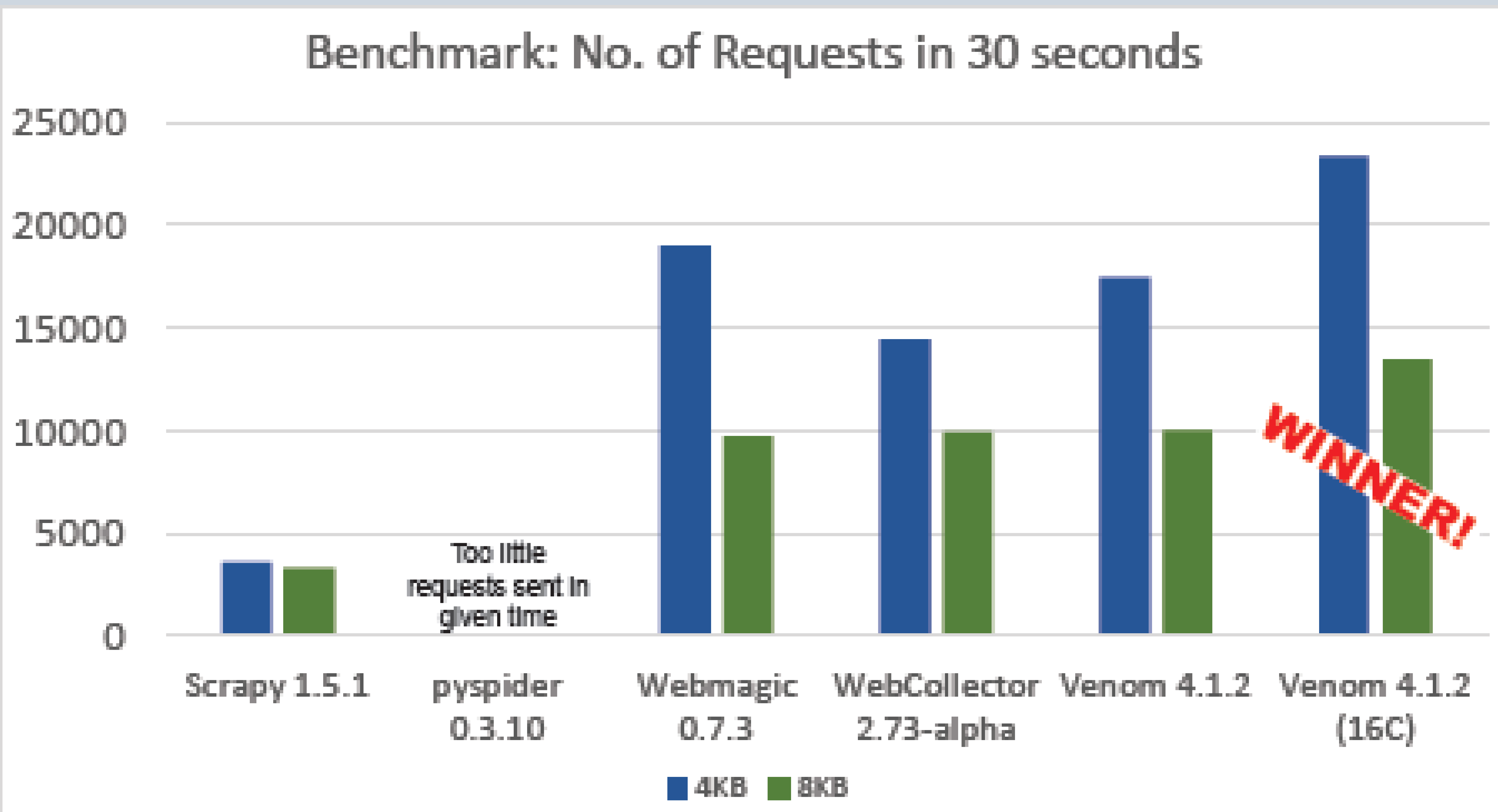
Handles issues gracefully with built-in header and content validation, which ensures data correctness through auto re-fetching. Fully integrated into the request-handler scheme.

Simple and Handy to Use

Provides all the essential features required to scrape the web, allowing you to write a full-fledged crawler in just a few lines of code. We do all the work so you don't have to.

Notable features

- Structured crawling with jsoup integration
- Page validation and retry handling
- Built-in raw file storage system and reparsing
- Proxy support
- Open source and free!



Butler

http://t.me/ButlerPreferredAI_Bot

Conversational Recommender System

Explore by chatting with a user-friendly smart-system

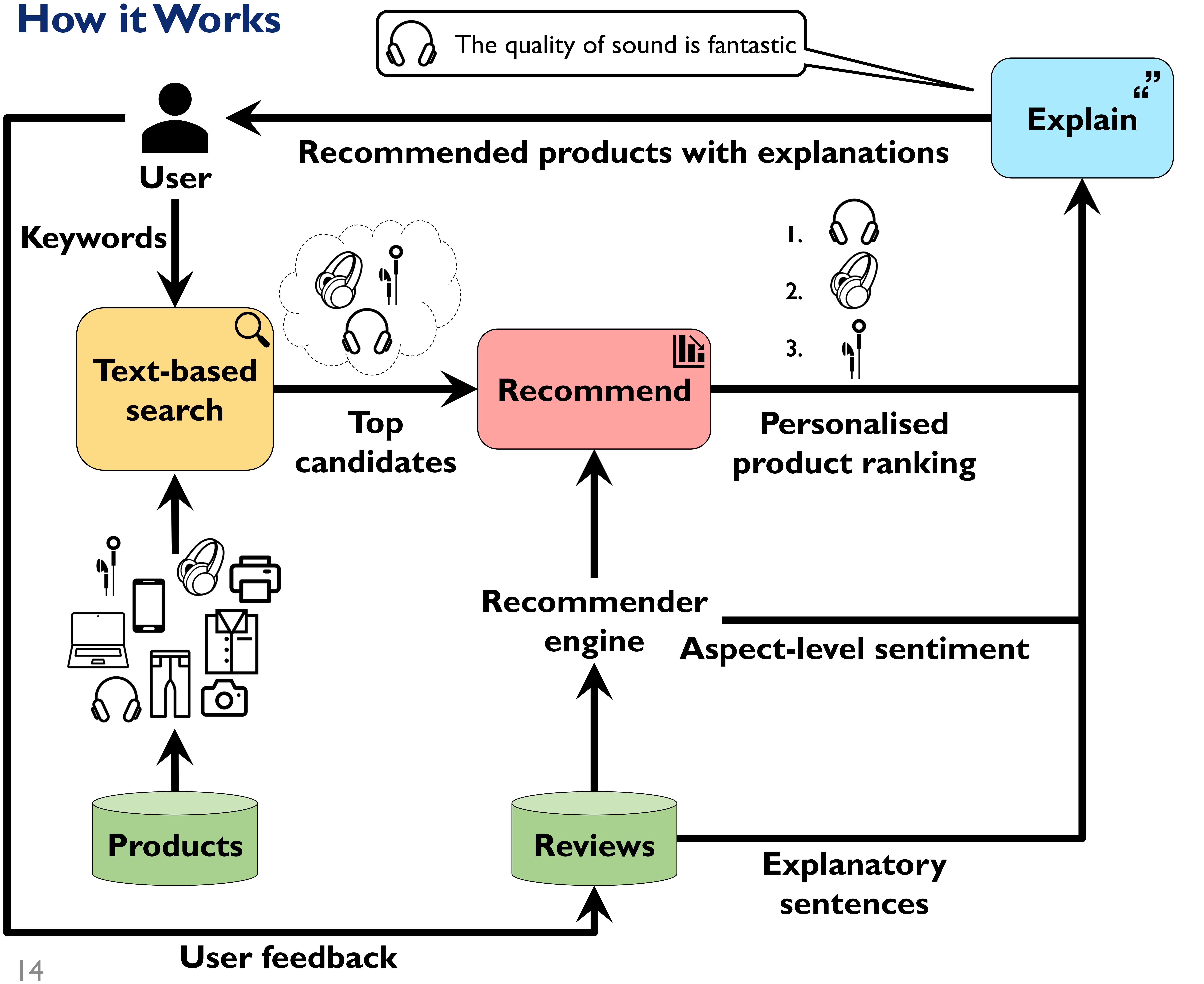
- Clutter free – focus only on relevant products
- Convenient – view and edit past reviews at any time
- Customized – analyse user preferences over time

Natural Language Explanations

Meaningful explanations for smarter and swifter decisions

- Efficient – understand your recommendations at a glance
- Effective – compare only the features that matter most to you
- Evaluable – verify that our recommendations are tailored to your needs

How it Works



FaceInMotion

https://faceinmotion.preferred.ai

PREFERRED.AI

SMU Singapore Management University
School of Information Systems

Lite Emotion Detection

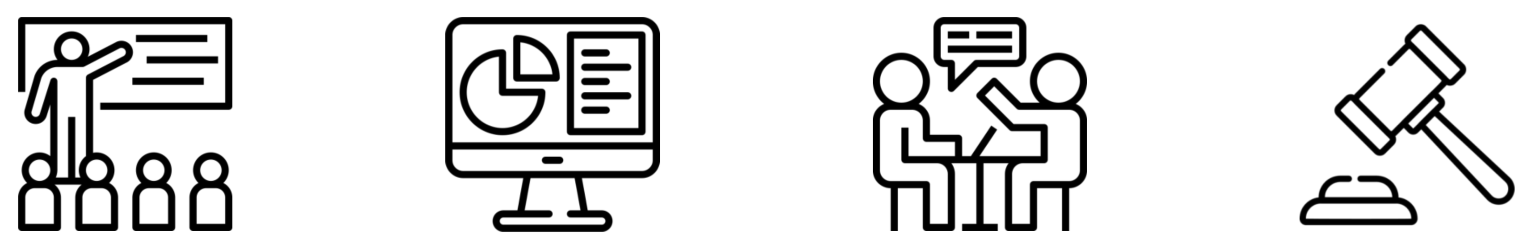
- FaceInMotion is built to detect human emotions from facial expressions
- The compact system is deployable on multiple platforms (e.g., mobile, IoT) with low latency yet high accuracy

Angry *Disgust* *Fear* *Happy* *Sad* *Surprise* *Neutral*

Supported Platform

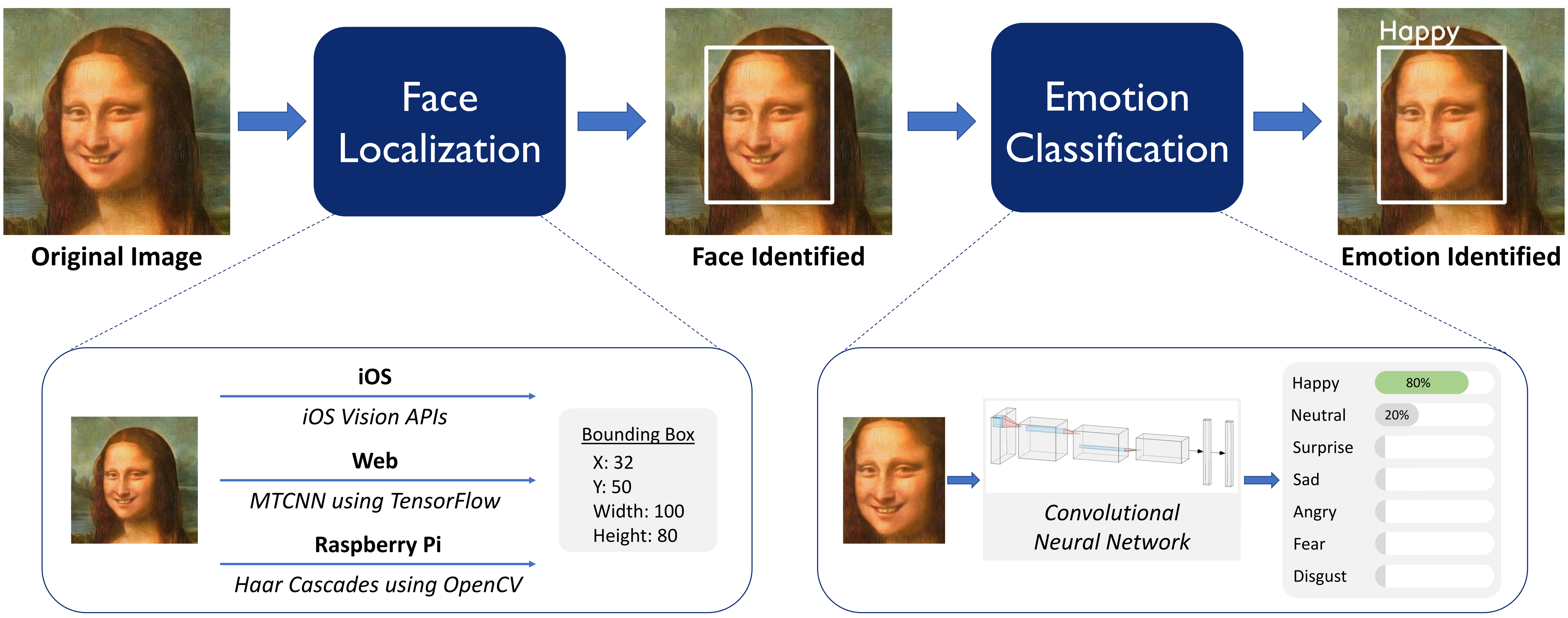
iOS **Web** **Raspberry Pi**

Potential Applications



- Education
  Monitoring students' learning (e.g., identifying learning difficulties)
- Market Research
  Analysing customers' sentiment (e.g., customers' response to products)
- Interviews
  Profiling interviewees (e.g., confidence level)
- Law Enforcement
  Detecting malicious intent (e.g., hostility)

Emotion Detection Process



Performance

- FER2013 dataset *(Goodfellow et al. 2013)*

	Our Model (Ensembled)	Kaggle 1st rank	Our Model (Single Model)	Kaggle 2nd rank
Public Leaderboard	70.94%	69.77%	69.66%	69.07%
Private Leaderboard	73.25%	71.16%	70.24%	69.27%

- Deployed Model

		Original	Optimized
Model Size		18.4MB	4.6MB (4x smaller)
Accuracy		70.24%	70.02% (-0.22%)
Speed	iOS	186 FPS	231 FPS (1.24x faster)
	Rasp. Pi	14 FPS	19 FPS (1.35x faster)

MindReader

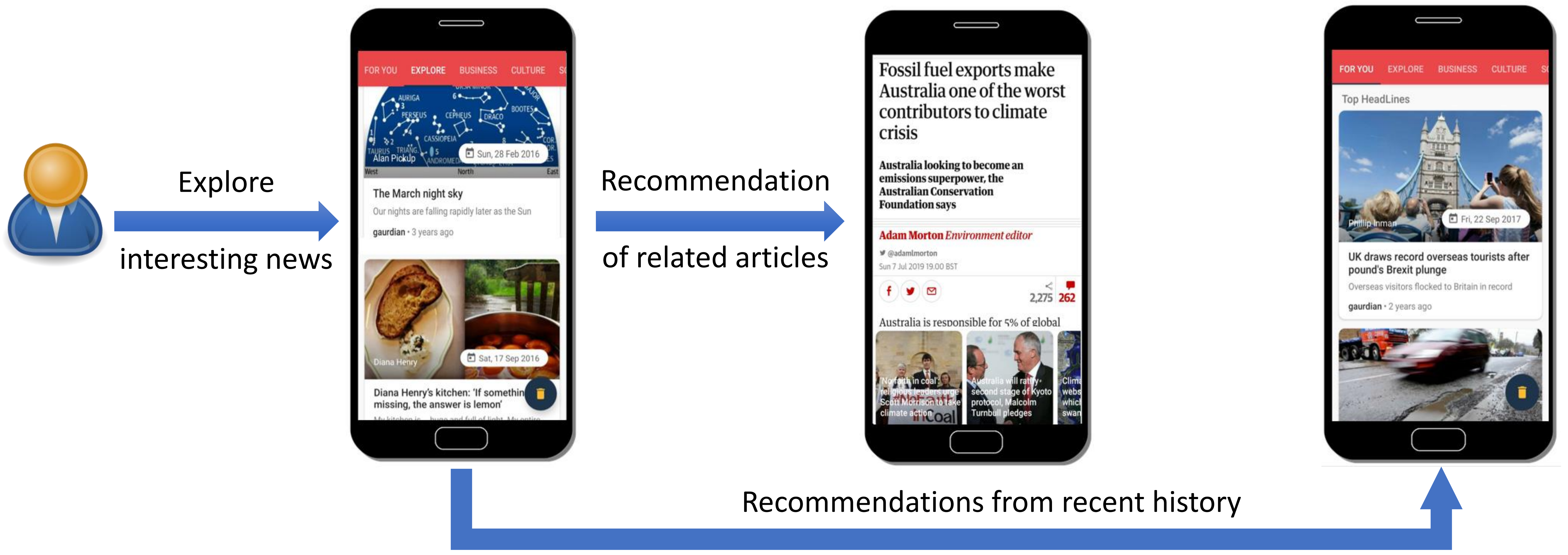
https://mindreader.preferred.ai

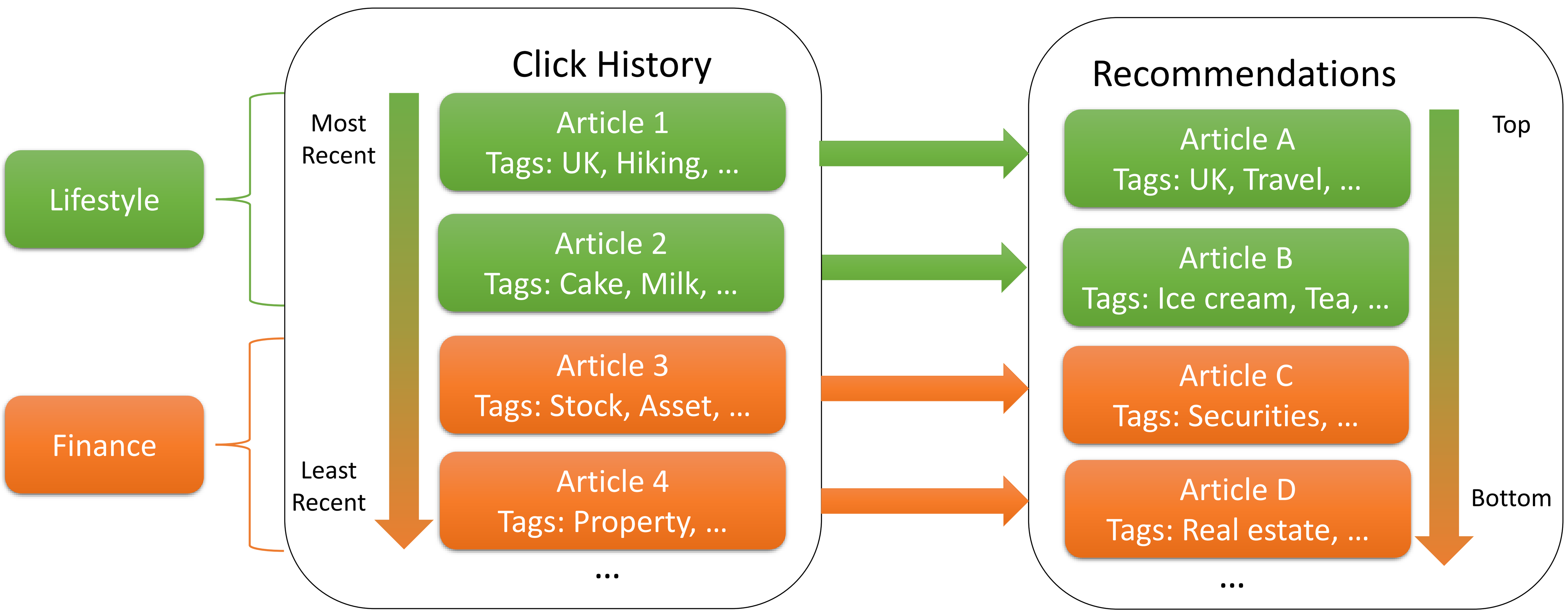
SMU SINGAPORE MANAGEMENT UNIVERSITY
School of Information Systems

PREFERRED.AI

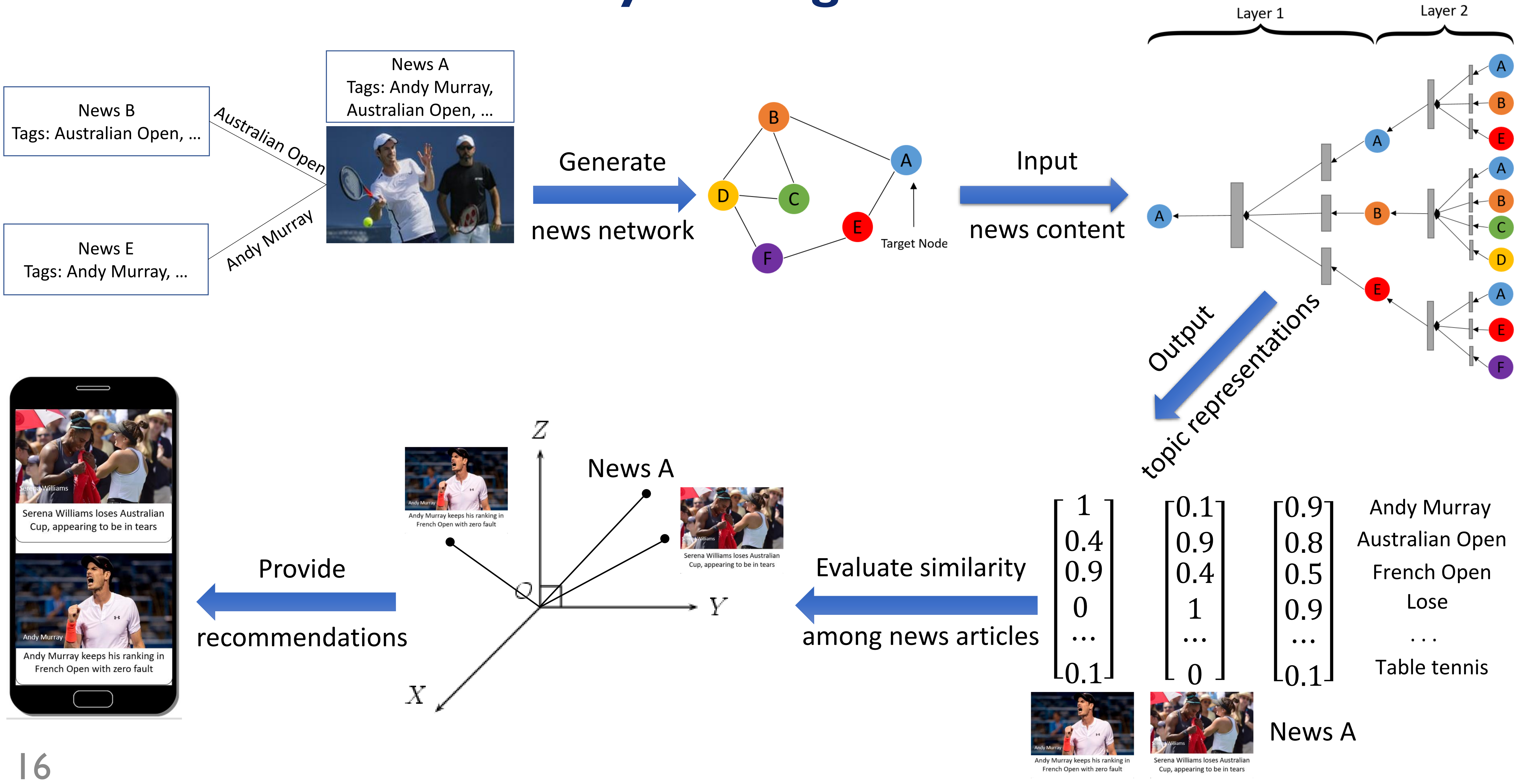
MindReader is a news recommendation app that provides **personalized recommendations** based on an individual's **reading history**.

Generating Recommendations





How to Assess Similarity Among News Articles



SMU
SINGAPORE MANAGEMENT UNIVERSITY
School of Information Systems

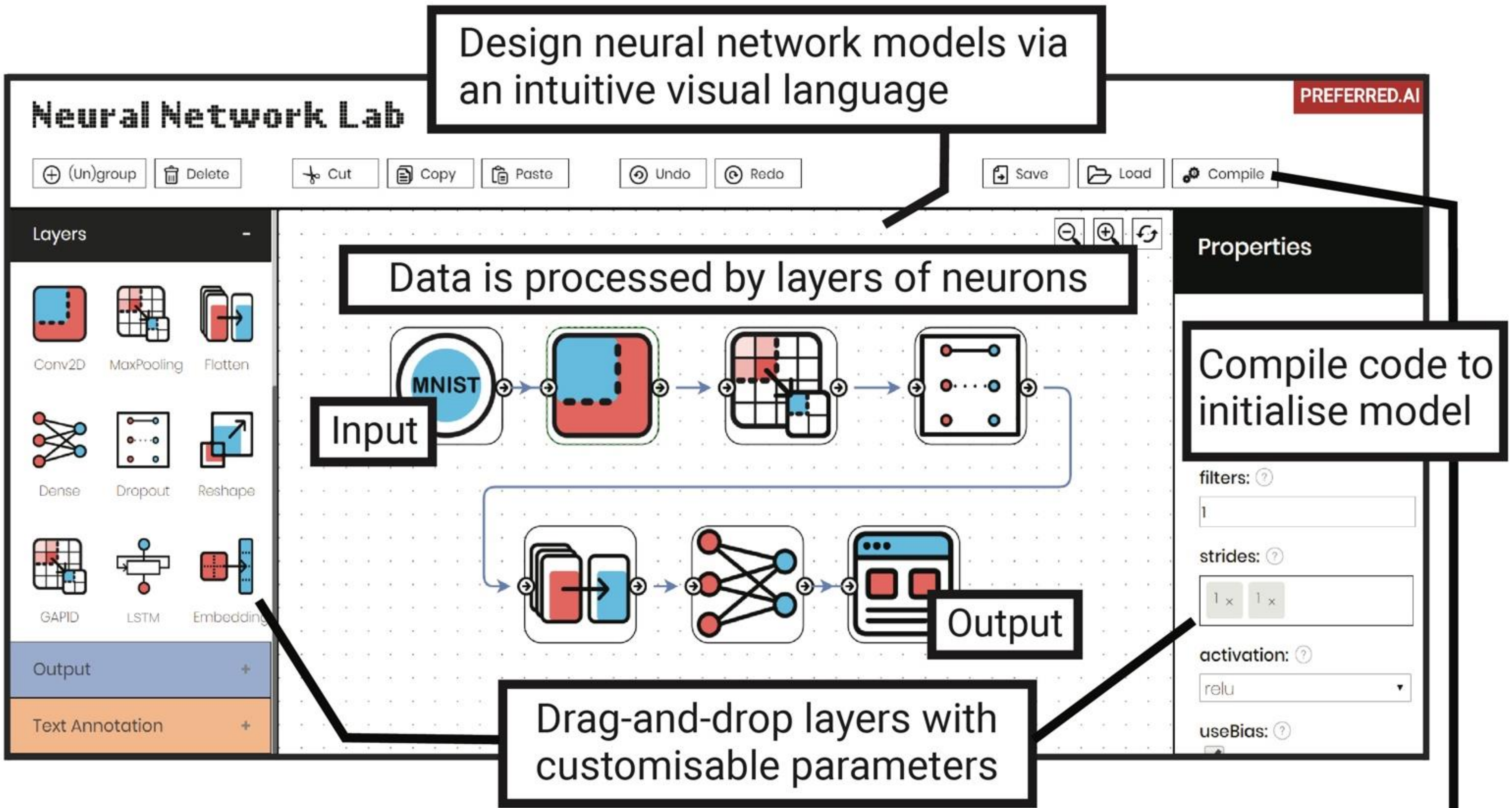
Neural Network Lab

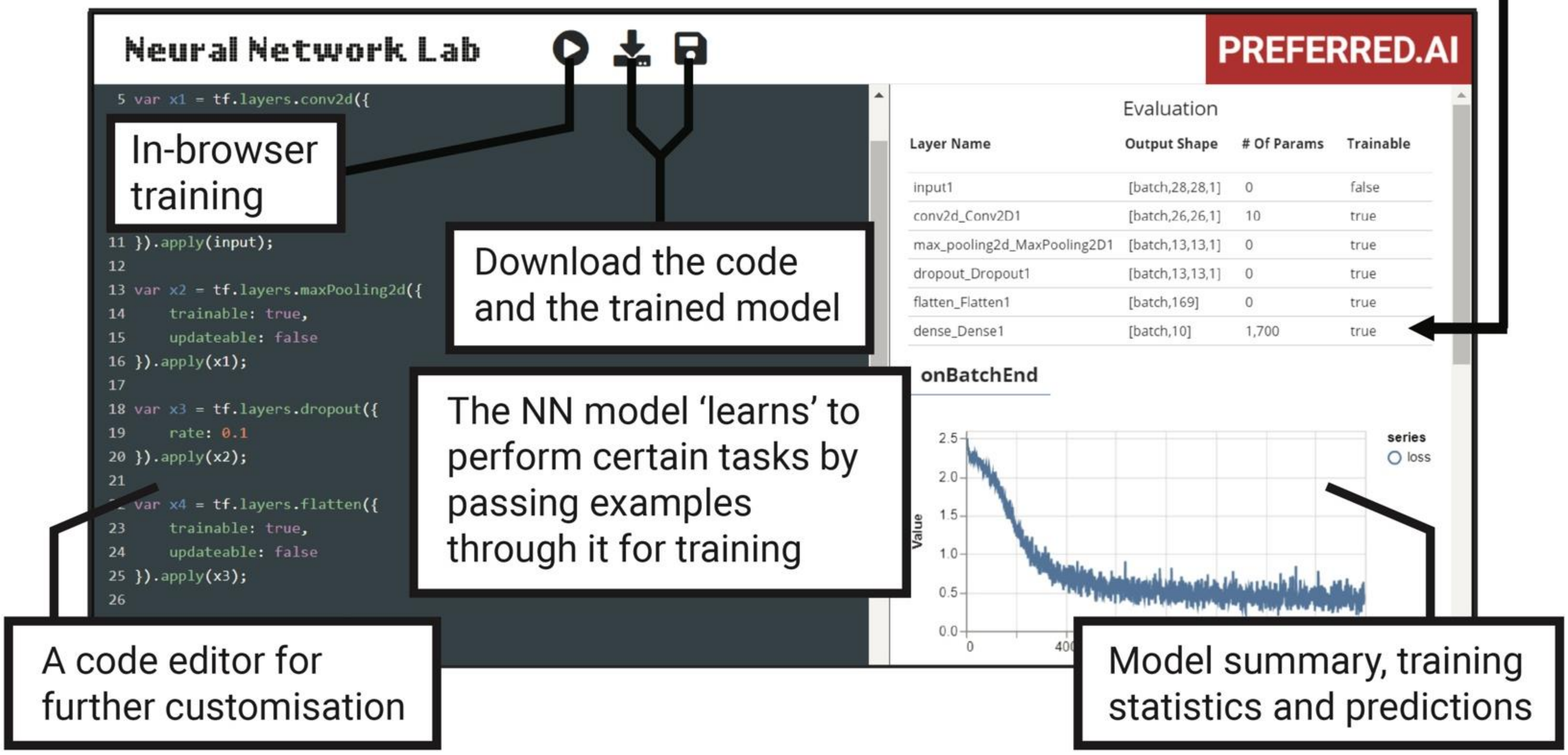
https://nnlab.preferred.ai

PREFERRED.AI

- Provides a browser-friendly, time-saving code generator for neural network (NN) models, customizable to suit your needs
- Facilitates NN education and builds a better understanding of the models
- Generates ready-to-use-and-deploy systems for businesses

Machine Learning in Your Browser





Links

Contact us to:

get involved in our projects as interns or research assistants

use our libraries or license our technologies

or just join us already

We are also on:

https://preferred.ai/publications/

https://preferred.ai/

https://code.preferred.ai

https://preferred.ai/videos

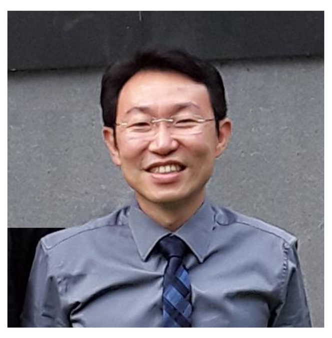
https://preferred.ai/join/

Learn more about SMU's Postgraduate Programmes:

Office of Postgraduate Research Programmes https://graduatestudies.smu.edu.sg

School of Information Systems https://sis.smu.edu.sg/programmes/postgraduate

TECHFEST.PREFERRED.AI 2019 Organizing Team

 Hady Lauw

 Aghiles Salah

 Maksim Tkachenko

 Andrew Le Duy Dung

 Trong Quoc Tuan

 Le Trung Hoang

 Chia Chong Cher

 Zhang Ce

 Lee Ween Jiann

 Darryl Ong

 Guo Jingyao

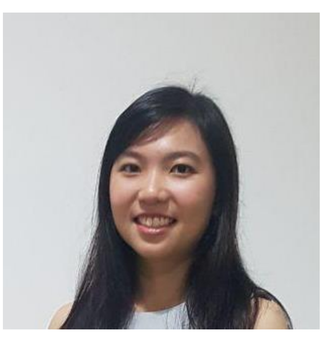
 Huynh Phu Minh

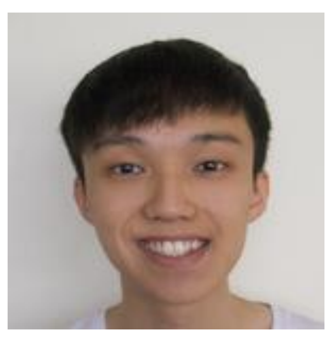
 Tran Thanh Binh

 Abhyuday Sammadder

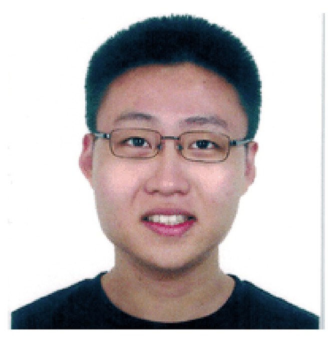
 Aw Jiayu

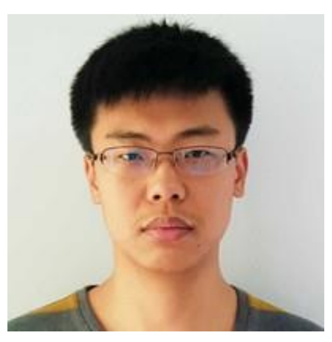
 Cao Wanyue

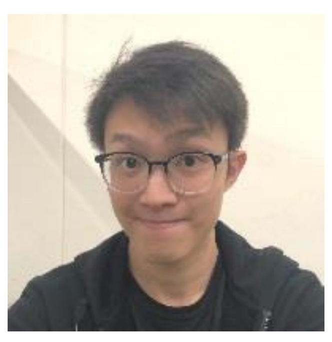
 Cheryl Lim Wei Lin

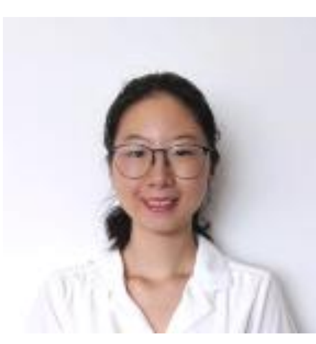
 Choy Kar Sen

 Hee Ming Shan

 Liu Ziyuan

 Lyu Cheng

 Ngoh Yi Long

 Steffi Tan Xin Rong

 Wei Ming

We would also like to acknowledge the other contributors to the various projects exhibited in this TechFest, who though unnamed are much appreciated.

PREFERRED.AI

We would love to hear your comments:

https://techfest.preferred.ai/feedback