# Twitter's Recommendation Algorithm

Twitter's Recommendation Algorithm is a set of services and jobs that are responsible for serving feeds of Tweets and other content across all Twitter product surfaces (e.g. For You Timeline, Search, Explore, Notifications). For an introduction to how the algorithm works, please refer to our [engineering blog](https://blog.twitter.com/engineering/en\_us/topics/open-source/2023/twitter-recommendation-algorithm).

## Architecture

Product surfaces at Twitter are built on a shared set of data, models, and software frameworks. The shared components included in this repository are listed below:

| Type | Component | Description |

|------------|------------|------------|

| Data | [tweetypie](tweetypie/server/README.md) | Core Tweet service that handles the reading and writing of Tweet data. |

| | [unified-user-actions](unified\_user\_actions/README.md) | Real-time stream of user actions on Twitter. |

| | [user-signal-service](user-signal-service/README.md) | Centralized platform to retrieve explicit (e.g. likes, replies) and implicit (e.g. profile visits, tweet clicks) user signals. |

| Model | [SimClusters](src/scala/com/twitter/simclusters\_v2/README.md) | Community detection and sparse embeddings into those communities. |

| | [TwHIN](https://github.com/twitter/the-algorithm-ml/blob/main/projects/twhin/README.md) | Dense knowledge graph embeddings for Users and Tweets. |

| | [trust-and-safety-models](trust\_and\_safety\_models/README.md) | Models for detecting NSFW or abusive content. |

| | [real-graph](src/scala/com/twitter/interaction\_graph/README.md) | Model to predict the likelihood of a Twitter User interacting with another User. |

| | [tweepcred](src/scala/com/twitter/graph/batch/job/tweepcred/README) | Page-Rank algorithm for calculating Twitter User reputation. |

| | [recos-injector](recos-injector/README.md) | Streaming event processor for building input streams for [GraphJet](https://github.com/twitter/GraphJet) based services. |

| | [graph-feature-service](graph-feature-service/README.md) | Serves graph features for a directed pair of Users (e.g. how many of User A's following liked Tweets from User B). |

| | [topic-social-proof](topic-social-proof/README.md) | Identifies topics related to individual Tweets. |

| | [representation-scorer](representation-scorer/README.md) | Compute scores between pairs of entities (Users, Tweets, etc.) using embedding similarity. |

| Software framework | [navi](navi/README.md) | High performance, machine learning model serving written in Rust. |

| | [product-mixer](product-mixer/README.md) | Software framework for building feeds of content. |

| | [timelines-aggregation-framework](timelines/data\_processing/ml\_util/aggregation\_framework/README.md) | Framework for generating aggregate features in batch or real time. |

| | [representation-manager](representation-manager/README.md) | Service to retrieve embeddings (i.e. SimClusers and TwHIN). |

| | [twml](twml/README.md) | Legacy machine learning framework built on TensorFlow v1. |

The product surfaces currently included in this repository are the For You Timeline and Recommended Notifications.

### For You Timeline

The diagram below illustrates how major services and jobs interconnect to construct a For You Timeline.

![](docs/system-diagram.png)

The core components of the For You Timeline included in this repository are listed below:

| Type | Component | Description |

|------------|------------|------------|

| Candidate Source | [search-index](src/java/com/twitter/search/README.md) | Find and rank In-Network Tweets. ~50% of Tweets come from this candidate source. |

| | [cr-mixer](cr-mixer/README.md) | Coordination layer for fetching Out-of-Network tweet candidates from underlying compute services. |

| | [user-tweet-entity-graph](src/scala/com/twitter/recos/user\_tweet\_entity\_graph/README.md) (UTEG)| Maintains an in memory User to Tweet interaction graph, and finds candidates based on traversals of this graph. This is built on the [GraphJet](https://github.com/twitter/GraphJet) framework. Several other GraphJet based features and candidate sources are located [here](src/scala/com/twitter/recos). |

| | [follow-recommendation-service](follow-recommendations-service/README.md) (FRS)| Provides Users with recommendations for accounts to follow, and Tweets from those accounts. |

| Ranking | [light-ranker](src/python/twitter/deepbird/projects/timelines/scripts/models/earlybird/README.md) | Light Ranker model used by search index (Earlybird) to rank Tweets. |

| | [heavy-ranker](https://github.com/twitter/the-algorithm-ml/blob/main/projects/home/recap/README.md) | Neural network for ranking candidate tweets. One of the main signals used to select timeline Tweets post candidate sourcing. |

| Tweet mixing & filtering | [home-mixer](home-mixer/README.md) | Main service used to construct and serve the Home Timeline. Built on [product-mixer](product-mixer/README.md). |

| | [visibility-filters](visibilitylib/README.md) | Responsible for filtering Twitter content to support legal compliance, improve product quality, increase user trust, protect revenue through the use of hard-filtering, visible product treatments, and coarse-grained downranking. |

| | [timelineranker](timelineranker/README.md) | Legacy service which provides relevance-scored tweets from the Earlybird Search Index and UTEG service. |

### Recommended Notifications

The core components of Recommended Notifications included in this repository are listed below:

| Type | Component | Description |

|------------|------------|------------|

| Service | [pushservice](pushservice/README.md) | Main recommendation service at Twitter used to surface recommendations to our users via notifications.

| Ranking | [pushservice-light-ranker](pushservice/src/main/python/models/light\_ranking/README.md) | Light Ranker model used by pushservice to rank Tweets. Bridges candidate generation and heavy ranking by pre-selecting highly-relevant candidates from the initial huge candidate pool. |

| | [pushservice-heavy-ranker](pushservice/src/main/python/models/heavy\_ranking/README.md) | Multi-task learning model to predict the probabilities that the target users will open and engage with the sent notifications. |

## Build and test code

We include Bazel BUILD files for most components, but not a top-level BUILD or WORKSPACE file. We plan to add a more complete build and test system in the future.

## Contributing

We invite the community to submit GitHub issues and pull requests for suggestions on improving the recommendation algorithm. We are working on tools to manage these suggestions and sync changes to our internal repository. Any security concerns or issues should be routed to our official [bug bounty program](https://hackerone.com/twitter) through HackerOne. We hope to benefit from the collective intelligence and expertise of the global community in helping us identify issues and suggest improvements, ultimately leading to a better Twitter.

Read our blog on the open source initiative [here](https://blog.twitter.com/en\_us/topics/company/2023/a-new-era-of-transparency-for-twitter).