AI Pro: Data Processing Framework for AI Models

Richie Frost, Debjyoti Paul, Feifei Li || {rsfrost, deb, lifeifei}@cs.utah.edu

INITIALDLAB, SCHOOL OF COMPUTING
UNIVERSITY OF UTAH

What is AI Pro ?

AI Pro is an open-source data processing framework for Artificial Intelligence (AI) models. It supports major deep learning frameworks and Open Neural Network Exchange (ONNX).

Why use AI Pro, what is unique about it?

AI Pro has all the quintessential features to perform end-to-end data processing pipelines *without a single line of code.* Yes, you read it right, AI Pro's working principle is *configuration as code* i.e. AI Pro empowers *non-data-scientist* to run prebuilt AI models without writing any code. Data pipelines are defined as directed acyclic grapha(DAGs).

How to use AI Pro?

AI Pro provides a Web User Interface (WebUI) for creating, managing and running data pipelines. Demos, tools, prebuilt models and configurations will be available from the website. Experts users can modify/update easy-entensible pluggable modules, custom models according to their needs and contribute in developement of AI Pro.

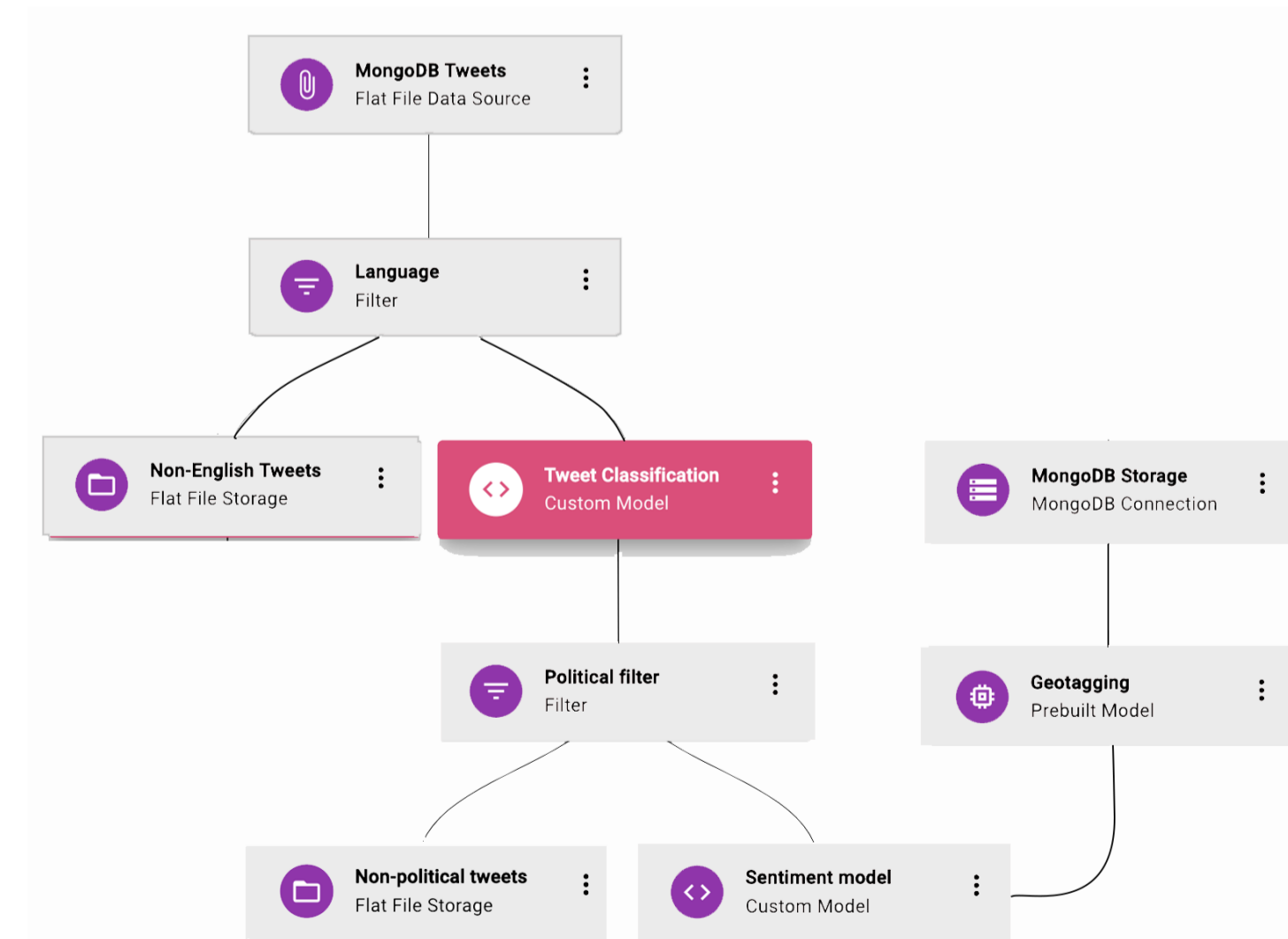## AI Pro Data Pipeline Example

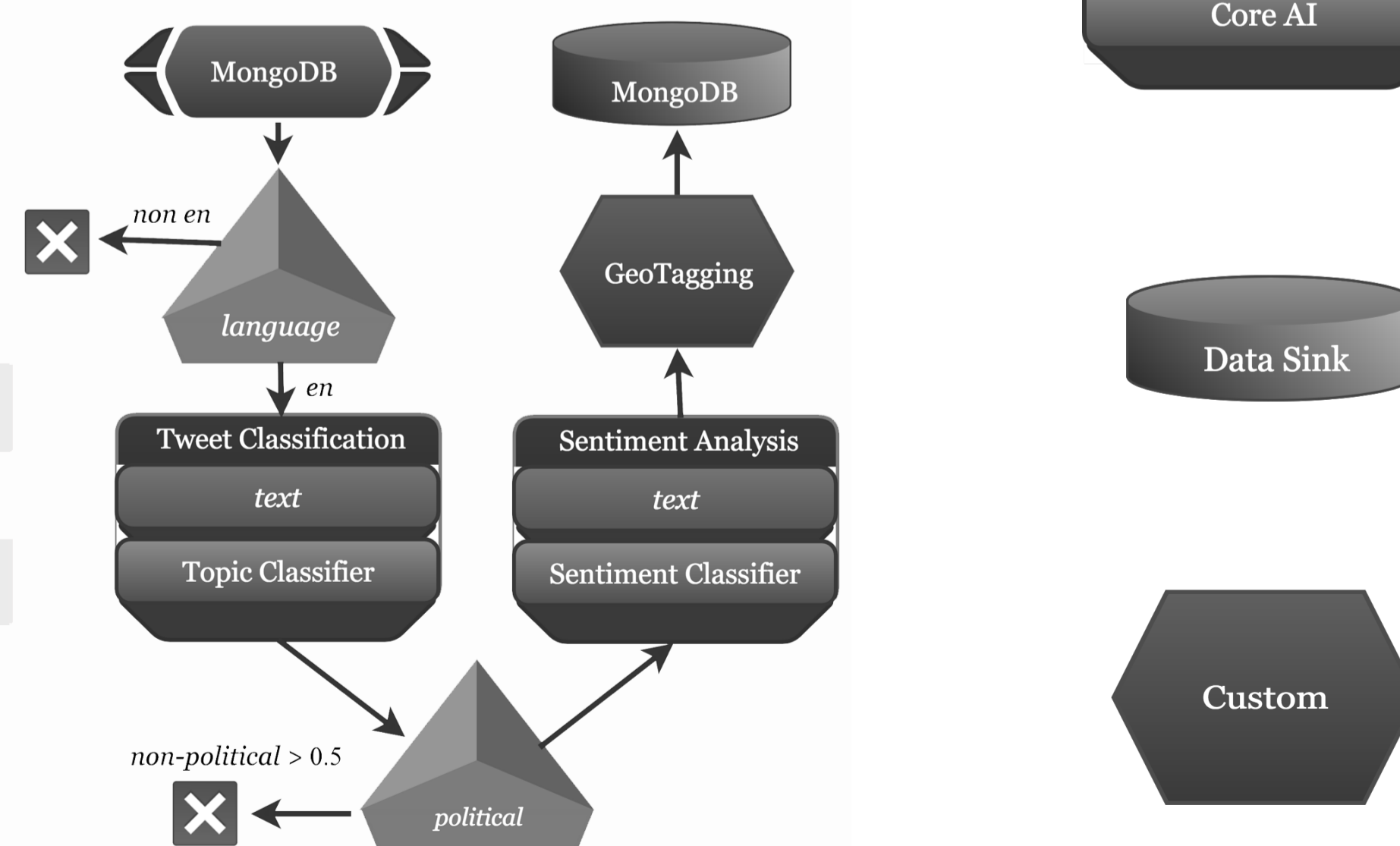Figure 1: Twitter Spatio-temporal Sentiment Analysis DAG from WebUI

Figure 2: Pictorial representation with module entities

AI Pro's Entities

We define **entity** as an abstract component responsible for either ingestion, transformation, addition, removal, or storage of data elements in the process of data flow. We present a brief introduction of each entity with example configuration.

Data Source Entity. Data pipeline always starts with a *data source entity*. AI Pro provides support for various types of data sources, such as (i) files, (ii) streaming APIs, and (iii) NoSQL Databases.

```
{
  "alias" : "Twitter Streaming Source",
  "api_key" : "XXXXXX",
  "url": "http://example.com/api/data/",
}
```
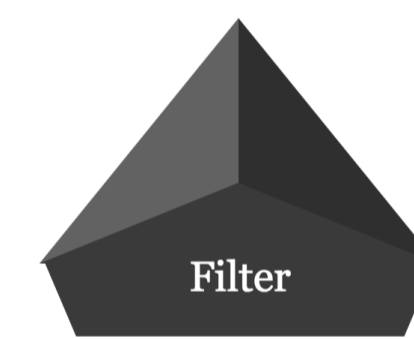
Filter Entity. A filter entity controls data flow in data pipeline with conditional mathematical statement on data attributes. The output of a filter entity flows through separate paths in the DAG, or even discard certain data elements based on configuration.

```
{
  "alias" : "Language filter",
  "attribute" : "lang",
  "value" : "en",
  "condition" : "=="
}
```
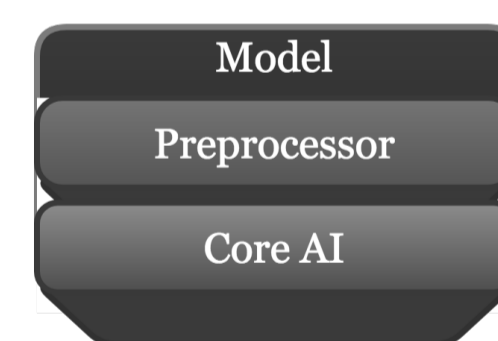
Model Entity. A model entity wraps an AI model with preprocessor and core model code. It has two sub-entitites: (i) Preprocessor (ii) Core AI. Prebuilt model configurations can be obtained from model zoo.

```
{
  "alias" : "Sentiment model, custom",
  "input_attribute" : "text",
  "module_file_path" : "uploads/sentiment.py",
  "method_name" : "predict",
  "module_classname" : "SentimenModel",
  "preprocessor_filename" : "tweet_preprocessor.py",
  "preprocessor_methodname": "preprocess"
  "output_attribute" : "sentiment",
}
```

Data Sink Entity. A data pipeline can have multiple storage entities that store processed data at different locations. AI Pro currently supports three types of storage entities: (a) Regular file, (b) Database (c) Standard I/O.

```
{
  "db" : "geotwitter",
  "collection" : "tweets",
  "alias" : "Tweets Mongo Connection",
  "host" : "localhost",
  "type" : "MongoDB",
  "port" : 27017
}
```

Custom Entity. Experts can create custom entities for customized transformation of data elements. One example of such a custom entity that is included in AI Pro is geo-location mapping.