

Package: `cat.vader` (via `r-universe`)

June 4, 2026

Title Social Media Content Classification with LLMs

Version 0.1.2

Description R interface to the Python `catvader` package. Classifies, extracts, and explores social media content using LLMs. A thin domain wrapper around `cat.stack` that adds social media sourcing parameters (`platform`, `handle`, `hashtags`, `credentials`).

License GPL (>= 3)

URL <https://christophersoria.com/cat-llm/cat.vader/>,
<https://github.com/chrissoria/cat-llm>

BugReports <https://github.com/chrissoria/cat-llm/issues>

Encoding UTF-8

Roxygen `list(markdown = TRUE)`

RoxygenNote 7.3.2

SystemRequirements Python (>= 3.9), pip

Imports `reticulate` (>= 1.28), `cat.stack` (>= 0.1.0)

Suggests `testthat` (>= 3.0.0), `knitr`, `rmarkdown`

VignetteBuilder `knitr`

Config/testthat/edition 3

Config/pak/sysreqs `libpng-dev` `python3`

Repository <https://chrissoria.r-universe.dev>

Date/Publication 2026-06-04 16:16:50 UTC

RemoteUrl <https://github.com/chrissoria/cat-llm>

RemoteRef `main`

RemoteSha `f2d83209be8d621fceb422d434fb5b3b98fe301b`

RemoteSubdir `r-package/cat.vader`

Contents

classify	2
explore	6
extract	8

Index	10
--------------	-----------

classify	<i>Classify social media content using LLMs</i>
----------	---

Description

Wraps the Python `catvader.classify()` function. Adds social media sourcing parameters to the base `cat.stack` classification engine.

Usage

```

classify(
    input_data = NULL,
    categories = NULL,
    api_key = NULL,
    sm_source = NULL,
    sm_limit = 50L,
    sm_months = NULL,
    sm_days = NULL,
    sm_credentials = NULL,
    sm_handle = NULL,
    sm_timezone = "UTC",
    sm_youtube_content = "video",
    sm_youtube_transcript = FALSE,
    sm_comments_per_video = 20L,
    sm_youtube_transcript_max_chars = 10000L,
    platform = NULL,
    handle = NULL,
    hashtags = NULL,
    post_metadata = NULL,
    description = "",
    feed_question = "",
    user_model = "gpt-4o",
    mode = "image",
    creativity = NULL,
    safety = FALSE,
    chain_of_verification = FALSE,
    chain_of_thought = FALSE,
    step_back_prompt = FALSE,
    context_prompt = FALSE,
    thinking_budget = 0L,

```

```

example1 = NULL,
example2 = NULL,
example3 = NULL,
example4 = NULL,
example5 = NULL,
example6 = NULL,
filename = NULL,
save_directory = NULL,
model_source = "auto",
max_categories = 12L,
categories_per_chunk = 10L,
divisions = 10L,
research_question = NULL,
models = NULL,
consensus_threshold = "unanimous",
use_json_schema = TRUE,
max_workers = NULL,
fail_strategy = "partial",
max_retries = 5L,
batch_retries = 2L,
retry_delay = 1,
row_delay = 0,
pdf_dpi = 150L,
auto_download = FALSE,
add_other = "prompt",
check_verbosity = TRUE,
prompt_tune = NULL,
tune_iterations = 1L,
tune_ui = "browser",
tune_optimize = "balanced"
)

```

Arguments

<code>input_data</code>	A character vector, list, or <code>data.frame</code> column, or <code>NULL</code> to fetch from social media. Default <code>NULL</code> .
<code>categories</code>	A character vector of category names, or <code>"auto"</code> . Default <code>NULL</code> .
<code>api_key</code>	Character or <code>NULL</code> . API key for the LLM provider.
<code>sm_source</code>	Character or <code>NULL</code> . Social media source (e.g., <code>"reddit"</code> , <code>"twitter"</code> , <code>"youtube"</code>).
<code>sm_limit</code>	Integer. Max posts to fetch. Default <code>50L</code> .
<code>sm_months</code>	Integer or <code>NULL</code> . Fetch posts from last N months.
<code>sm_days</code>	Integer or <code>NULL</code> . Fetch posts from last N days.
<code>sm_credentials</code>	Named list or <code>NULL</code> . API credentials for the social media platform.
<code>sm_handle</code>	Character or <code>NULL</code> . Social media handle to fetch from.
<code>sm_timezone</code>	Character. Timezone for date filtering. Default <code>"UTC"</code> .

sm_youtube_content Character. YouTube content type. Default "video".
 sm_youtube_transcript Logical. Include YouTube transcripts. Default FALSE.
 sm_comments_per_video Integer. Comments per YouTube video. Default 20L.
 sm_youtube_transcript_max_chars Integer. Max transcript chars. Default 10000L.
 platform Character or NULL. Alias for sm_source.
 handle Character or NULL. Alias for sm_handle.
 hashtags Character vector or NULL. Hashtags to filter by.
 post_metadata Named list or NULL. Additional post metadata.
 description Character. Context description. Default "".
 feed_question Character. Feed-specific question context. Default "".
 user_model Character. Model name. Default "gpt-4o".
 mode Character. Processing mode. Default "image".
 creativity Numeric or NULL. Temperature. Default NULL.
 safety Logical. Save progress after each item. Default FALSE.
 chain_of_verification Logical. Default FALSE.
 chain_of_thought Logical. Default FALSE.
 step_back_prompt Logical. Default FALSE.
 context_prompt Logical. Default FALSE.
 thinking_budget Integer. Default 0L.
 example1, example2, example3, example4, example5, example6
 Optional few-shot examples.
 filename Character or NULL. Output CSV filename.
 save_directory Character or NULL. Output directory.
 model_source Character. Provider hint. Default "auto".
 max_categories Integer. Default 12L.
 categories_per_chunk Integer. Default 10L.
 divisions Integer. Default 10L.
 research_question Character or NULL. Optional research context.
 models List of model specs for ensemble mode.
 consensus_threshold Character or numeric. Default "unanimous".

use_json_schema	Logical. Default TRUE.
max_workers	Integer or NULL. Default NULL.
fail_strategy	Character. Default "partial".
max_retries	Integer. Default 5L.
batch_retries	Integer. Default 2L.
retry_delay	Numeric. Default 1.0.
row_delay	Numeric. Default 0.0.
pdf_dpi	Integer. Default 150L.
auto_download	Logical. Default FALSE.
add_other	Logical or "prompt". Default "prompt".
check_verbosity	Logical. Default TRUE.
prompt_tune	Integer or NULL. Rows sampled per APO correction round. Default NULL.
tune_iterations	Integer. APO optimization passes. Default 1L.
tune_ui	Character. Correction UI: "browser" or "terminal". Default "browser".
tune_optimize	Character. Metric to optimize: "balanced", "sensitivity", or "precision". Default "balanced".

Value

A data.frame with classification results.

Examples

```
## Not run:  
# Classify your own text directly  
results <- classify(  
  input_data = c("Just had the best coffee ever!",  
                "Politicians are all the same",  
                "Looking forward to the game tonight"),  
  categories = c("Positive", "Negative", "Neutral"),  
  api_key    = Sys.getenv("OPENAI_API_KEY"),  
  user_model = "gpt-4o-mini"  
)  
  
## End(Not run)
```

`explore`*Explore raw categories in social media data*

Description

Wraps the Python `catvader.explore()` function. Returns every category string extracted from every chunk across every iteration – with duplicates intact.

Usage

```
explore(  
    input_data = NULL,  
    api_key = NULL,  
    description = "",  
    sm_source = NULL,  
    sm_limit = 50L,  
    sm_months = NULL,  
    sm_credentials = NULL,  
    platform = NULL,  
    handle = NULL,  
    hashtags = NULL,  
    post_metadata = NULL,  
    max_categories = 12L,  
    categories_per_chunk = 10L,  
    divisions = 12L,  
    user_model = "gpt-4o",  
    creativity = NULL,  
    specificity = "broad",  
    research_question = NULL,  
    filename = NULL,  
    model_source = "auto",  
    iterations = 8L,  
    random_state = NULL,  
    focus = NULL,  
    chunk_delay = 0  
)
```

Arguments

<code>input_data</code>	A character vector, list, or NULL to fetch from social media. Default NULL.
<code>api_key</code>	Character or NULL. API key for the LLM provider.
<code>description</code>	Character. Context description. Default "".
<code>sm_source</code>	Character or NULL. Social media source.
<code>sm_limit</code>	Integer. Max posts to fetch. Default 50L.
<code>sm_months</code>	Integer or NULL. Fetch posts from last N months.

sm_credentials Named list or NULL. API credentials.

platform Character or NULL. Alias for sm_source.

handle Character or NULL. Social media handle.

hashtags Character vector or NULL. Hashtags to filter by.

post_metadata Named list or NULL. Additional post metadata.

max_categories Integer. Default 12L.

categories_per_chunk Integer. Default 10L.

divisions Integer. Default 12L.

user_model Character. Default "gpt-4o".

creativity Numeric or NULL. Default NULL.

specificity Character. Default "broad".

research_question Character or NULL.

filename Character or NULL.

model_source Character. Default "auto".

iterations Integer. Default 8L.

random_state Integer or NULL.

focus Character or NULL.

chunk_delay Numeric. Default 0.0.

Value

A character vector of every category string extracted.

Examples

```
## Not run:
raw_cats <- explore(
  input_data = df$posts,
  api_key     = Sys.getenv("OPENAI_API_KEY"),
  user_model  = "gpt-4o-mini",
  iterations  = 4L
)
table(raw_cats)

## End(Not run)
```

`extract`*Extract categories from social media content using LLMs*

Description

Wraps the Python `catvader.extract()` function. Discovers and returns a normalised, deduplicated set of categories from social media data.

Usage

```
extract(  
    input_data = NULL,  
    api_key = NULL,  
    sm_source = NULL,  
    sm_limit = 50L,  
    sm_months = NULL,  
    sm_credentials = NULL,  
    platform = NULL,  
    handle = NULL,  
    hashtags = NULL,  
    post_metadata = NULL,  
    description = "",  
    max_categories = 12L,  
    categories_per_chunk = 10L,  
    divisions = 12L,  
    user_model = "gpt-4o",  
    creativity = NULL,  
    specificity = "broad",  
    research_question = NULL,  
    mode = "text",  
    filename = NULL,  
    model_source = "auto",  
    iterations = 8L,  
    random_state = NULL,  
    focus = NULL,  
    chunk_delay = 0  
)
```

Arguments

<code>input_data</code>	A character vector, list, or NULL to fetch from social media. Default NULL.
<code>api_key</code>	Character or NULL. API key for the LLM provider.
<code>sm_source</code>	Character or NULL. Social media source.
<code>sm_limit</code>	Integer. Max posts to fetch. Default 50L.
<code>sm_months</code>	Integer or NULL. Fetch posts from last N months.

sm_credentials Named list or NULL. API credentials.
 platform Character or NULL. Alias for sm_source.
 handle Character or NULL. Social media handle.
 hashtags Character vector or NULL. Hashtags to filter by.
 post_metadata Named list or NULL. Additional post metadata.
 description Character. Context description. Default "".
 max_categories Integer. Default 12L.
 categories_per_chunk Integer. Default 10L.
 divisions Integer. Default 12L.
 user_model Character. Default "gpt-4o".
 creativity Numeric or NULL. Default NULL.
 specificity Character. Default "broad".
 research_question Character or NULL.
 mode Character. Default "text".
 filename Character or NULL.
 model_source Character. Default "auto".
 iterations Integer. Default 8L.
 random_state Integer or NULL.
 focus Character or NULL.
 chunk_delay Numeric. Default 0.0.

Value

A named list with counts_df, top_categories, and raw_top_text.

Examples

```

## Not run:
result <- extract(
  input_data = df$posts,
  api_key    = Sys.getenv("OPENAI_API_KEY"),
  user_model = "gpt-4o-mini"
)
print(result$top_categories)

## End(Not run)

```

Index

classify, 2

explore, 6

extract, 8