

# Module Desing - Cross-module contract



⚙ Estado	Pendiente
----------	-----------

- **Preprocessing** returns:
    - `normalized` (lowercase/single-spaced) and `tokens` that include URLs and hashtags.
  - **Classification:**
    - uses `tokens` and handles punctuation when matching keywords (via `categorize`).
    - provides symbol sequence + simple counts.
  - **Transformation:**
    - reuses `classify` so results stay consistent (no duplicated logic).
  - All modules are **pure** (no I/O), except optional CLI.
- 

## Minimal CLI (optional but handy)

- Preprocessing:

```
python src/moderation/post_preprocessing.py "text here"
```

- Classification:

```
python src/moderation/content_classification_dfa.py "text here"
```

- Transformation:

```
python -m src.moderation.content_transformation_fst "text here"
```

## Quick interface diagram (text)

```
raw post (str)
  |
  ▼
[Preprocessing] extract_all
  ↳ { normalized, tokens, mentions, hashtags, urls, emojis }
    |
    ▼
[Classification DFAs] classify
  ↳ ClassificationReport { hate, offensive, spam, details: tokens, symbols, counts }
    |
    ▼
[Transformation FST] transform
  ↳ TransformResult { transformed_text, masked_tokens, suggestions, categories, original_tokens }

"
```