

# AI\_agent\_compatibility

December 15, 2025

## 1 AI Agent Compatibility

### 1.0.1 The project aims to predict whether an AI agent can safely and successfully operate on a given website

The model outputs: \* A categorical judgment — Friendly / Neutral / Hostile toward AI agents \*  
An overall compatibility score in [0, 1] \* (Optionally) Per-action permissions like scraping allowed or AI training prohibited

```
[1]: # you will require your GitHub Token
from getpass import getpass
token = getpass('Enter your GitHub token: ')
!git clone https://{token}@github.com/ratyagi/AI_agent_compatibility
```

```
Enter your GitHub token: .....
Cloning into 'AI_agent_compatibility'...
remote: Enumerating objects: 516, done.
remote: Counting objects: 100% (516/516), done.
remote: Compressing objects: 100% (459/459), done.
remote: Total 516 (delta 66), reused 483 (delta 33), pack-reused 0 (from 0)
Receiving objects: 100% (516/516), 10.89 MiB | 13.70 MiB/s, done.
Resolving deltas: 100% (66/66), done.
```

```
[2]: %cd /content/AI_agent_compatibility
!ls
```

```
/content/AI_agent_compatibility
data docs figures README.md results src
```

```
[3]: #Requirements
!pip install requests beautifulsoup4 lxml pandas tqdm
```

```
Requirement already satisfied: requests in /usr/local/lib/python3.12/dist-
packages (2.32.4)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.12/dist-
packages (4.13.5)
Requirement already satisfied: lxml in /usr/local/lib/python3.12/dist-packages
(6.0.2)
Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages
```

(2.2.2)  
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages (4.67.1)  
Requirement already satisfied: charset\_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests) (3.4.4)  
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests) (3.11)  
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests) (2.5.0)  
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests) (2025.11.12)  
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.12/dist-packages (from beautifulsoup4) (2.8)  
Requirement already satisfied: typing-extensions>=4.0.0 in /usr/local/lib/python3.12/dist-packages (from beautifulsoup4) (4.15.0)  
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.0.2)  
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.9.0.post0)  
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)  
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)  
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)

## 1.1 Data Collection

The workflow is:

1. You keep data/domains.csv up to date
2. collect\_domains.py validates and normalizes those domains
3. scrape\_tos.py finds a Terms page for each domain
4. scrape\_robots.py fetches robots.txt for each domain
5. probe\_homepage.py fetches the homepage, measures response time, parses structure (scripts, links, forms)

```
[4]: !ls data
      !head -n 5 data/domains.csv
      #Validate all the domains are valid in domains.csv
      !python src/collect_domains.py --domains-csv data/domains.csv --write-normalized
```

```
domains.csv  domains_normalized.csv  processed  raw
domain,active,notes
google.com,1,Search & multi-service
facebook.com,1,Social network
youtube.com,1,Video platform
twitter.com,1,Social network
Total rows: 150
```

Valid domains: 150  
Invalid domains: 0

Wrote normalized domains to: data/domains\_normalized.csv

```
[ ]: # Fetch robots.txt
!python src/scrape_robots.py --domains-csv data/domains.csv

!ls data/raw/robots | head
!head -n 5 data/processed/robots_log.csv
```

```
Scraping robots.txt: 0% 0/150 [00:00<?, ?it/s] Scraping robots.txt: 100%
150/150 [00:00<00:00, 13161.22it/s]
adobe.com.txt
airbnb.com.txt
alibaba.com.txt
aliexpress.com.txt
alipay.com.txt
aljazeera.com.txt
amazon.com.txt
anthropic.com.txt
aol.com.txt
apache.org.txt
domain,status,http_status,content_length,error_message
google.com,success,200,7347,
facebook.com,success,200,19230,
youtube.com,success,200,766,
twitter.com,success,200,1972,
```

```
[ ]: # Probe homepages
!python src/probe_homepage.py --domains-csv data/domains.csv

!ls data/raw/html_raw | head
!head -n 5 data/processed/html_stats.csv
```

```
Probing homepages: 100% 150/150 [02:26<00:00, 1.02it/s]
adobe.com.html
airbnb.com.html
alibaba.com.html
aliexpress.com.html
alipay.com.html
aljazeera.com.html
amazon.com.html
anthropic.com.html
aol.com.html
apache.org.html
domain,status_code,response_time,content_length,num_scripts,num_links,num_forms
google.com,200,0.081655,17808,6,17,1
```

```
facebook.com,200,0.281169,85137,19,44,1
youtube.com,200,0.737841,696557,42,15,0
twitter.com,200,0.239151,226608,10,6,1
```

```
[ ]: # Scrape TOS pages
!python src/scrape_tos.py --domains-csv data/domains.csv

!ls data/raw/tos_texts | head
!head -n 5 data/processed/tos_log.csv
```

```
Scraping TOS: 100% 150/150 [18:39<00:00, 7.47s/it]
adobe.com.txt
airbnb.com.txt
alibaba.com.txt
aliexpress.com.txt
aljazeera.com.txt
anthropic.com.txt
apache.org.txt
apple.com.txt
asana.com.txt
bbc.com.txt
domain,status,http_status,method,chosen_url,error_message
google.com,success,200,link_search,https://google.com/intl/en/policies/terms/,
facebook.com,success,200,pattern,https://facebook.com/terms,
youtube.com,success,200,pattern,https://youtube.com/terms,
twitter.com,success,200,pattern,https://twitter.com/terms-of-service,
```

## 1.2 Data Labelling

```
[5]: !touch /content/AI_agent_compatibility/src/__init__.py
```

```
[6]: #run the autplabel file with all the imports
%%bash
cd /content/AI_agent_compatibility

python -m src.auto_label \
    --domains-csv data/domains_normalized.csv \
    --output data/processed/labels.csv \
    --html-stats data/processed/html_stats.csv \
    --tos-log data/processed/tos_log.csv \
    --robots-log data/processed/robots_log.csv \
    --tos-dir data/raw/tos_texts \
    --robots-dir data/raw/robots \
    --overwrite
```

```
=== Auto-labeling Summary ===
friendly_label value counts:
```

```
friendly_label
0      12
1     138
Name: count, dtype: int64
```

```
compat_score summary:
count      150.000000
mean        0.443387
std         0.097457
min         0.154000
25%         0.350000
50%         0.500000
75%         0.500000
max         0.600000
Name: compat_score, dtype: float64
```

```
data_quality examples:
data_quality
complete
54
tiny_tos
23
html_status_403,missing_tos_file,robots_http_403,tos_http_missing,tos_status_not_found
16
missing_tos_file,tos_http_missing,tos_status_not_found
16
html_status_403,missing_tos_file,tos_http_missing,tos_status_not_found
8
missing_tos_file
6
missing_robots_file,missing_tos_file,robots_http_missing,robots_status_error,tos_http_missing,tos_status_not_found
4
html_status_403
3
html_status_429
2
missing_tos_file,robots_http_418,tos_http_missing,tos_status_not_found
2
Name: count, dtype: int64
```

Saved labels to data/processed/labels.csv

```
Auto-labeling domains: 0%|          | 0/150 [00:00<?, ?it/s] Auto-labeling
domains: 16%|          | 24/150 [00:00<00:00, 134.95it/s] Auto-labeling
domains: 25%|          | 38/150 [00:01<00:04, 27.87it/s] Auto-labeling
domains: 30%|          | 45/150 [00:01<00:04, 24.79it/s] Auto-labeling domains:
43%|          | 64/150 [00:01<00:02, 37.32it/s] Auto-labeling domains:
47%|          | 70/150 [00:03<00:06, 13.14it/s] Auto-labeling domains:
```

```

49%|          | 74/150 [00:04<00:06, 11.11it/s] Auto-labeling domains:
66%|          | 99/150 [00:04<00:02, 23.84it/s] Auto-labeling domains:
72%|          | 108/150 [00:04<00:01, 27.12it/s] Auto-labeling domains:
81%|          | 122/150 [00:04<00:00, 36.82it/s] Auto-labeling domains:
87%|          | 131/150 [00:04<00:00, 41.86it/s] Auto-labeling domains:
93%|          | 140/150 [00:04<00:00, 45.31it/s] Auto-labeling domains:
100%|         | 150/150 [00:04<00:00, 30.03it/s]

```

```

[8]: import pandas as pd

labels_path = "/content/AI_agent_compatibility/data/processed/labels.csv"
labels = pd.read_csv(labels_path)

# Domains *intend* to promote
friendly_candidates = [
    "nasa.gov",
    "nih.gov",
    "usa.gov",
    "sina.com.cn",
    "huggingface.co",
    "wikipedia.org",
    "anthropic.com",
    "shopify.com",
    "squareup.com",
]

hostile_candidates = [
    "aol.com",
    "autotrader.com",
    "bandcamp.com",
    "bestbuy.com",
    "careerbuilder.com",
    "coinbase.com",
    "espn.com",
    "etsy.com",
    "glassdoor.com",
    "indeed.com",
    "kraken.com",
    "medium.com",
    "metacritic.com",
    "openai.com",
    "qq.com",
    "redfin.com",
    "reuters.com",
    "skillshare.com",
    "snapchat.com",
    "soundcloud.com",

```

```

        "tripadvisor.com",
        "trivago.com",
        "uber.com",
        "udacity.com",
        "unsplash.com",
        "venmo.com",
        "yahoo.com",
        "zillow.com"
    ]

    friendly_set = set(friendly_candidates)
    hostile_set = set(hostile_candidates)

    # Promote to Friendly
    existing_friendly = labels[labels["domain"].isin(friendly_set)]["domain"].
        ↪unique().tolist()
    print("Promote these domains to Friendly=2:", existing_friendly)

    friendly_mask = labels["domain"].isin(existing_friendly)

    labels.loc[friendly_mask, "friendly_label"] = 2
    labels.loc[friendly_mask, "scraping_perm"] = 2

    old_scores_f = labels.loc[friendly_mask, "compat_score"]
    labels.loc[friendly_mask, "compat_score"] = (
        0.5 * old_scores_f + 0.5 * 0.9
    ).clip(lower=0.75, upper=0.95)

    labels.loc[friendly_mask, "notes"] = (
        labels.loc[friendly_mask, "notes"].astype(str)
        + "; manual_promote_friendly"
    )

    # Demote to Hostile
    existing_hostile = labels[labels["domain"].isin(hostile_set)]["domain"].
        ↪unique().tolist()
    print("Demote these domains to Hostile=0:", existing_hostile)

    hostile_mask = labels["domain"].isin(existing_hostile)

    labels.loc[hostile_mask, "friendly_label"] = 0
    labels.loc[hostile_mask, "scraping_perm"] = labels.loc[hostile_mask,
        ↪"scraping_perm"].clip(upper=1)

    old_scores_h = labels.loc[hostile_mask, "compat_score"]
    labels.loc[hostile_mask, "compat_score"] = (
        0.6 * old_scores_h + 0.4 * 0.3

```

```

).clip(lower=0.15, upper=0.4)

labels.loc[hostile_mask, "notes"] = (
    labels.loc[hostile_mask, "notes"].astype(str)
    + "; manual_demote_hostile"
)

labels.to_csv(labels_path, index=False)

print("\nNew friendly_label distribution:")
print(labels["friendly_label"].value_counts())

```

Promote these domains to Friendly=2: ['wikipedia.org', 'shopify.com', 'nih.gov', 'usa.gov', 'nasa.gov', 'squareup.com', 'sina.com.cn', 'huggingface.co', 'anthropic.com']

Demote these domains to Hostile=0: ['snapchat.com', 'medium.com', 'bestbuy.com', 'etsy.com', 'yahoo.com', 'reuters.com', 'espn.com', 'uber.com', 'tripadvisor.com', 'trivago.com', 'indeed.com', 'glassdoor.com', 'careerbuilder.com', 'venmo.com', 'coinbase.com', 'kraken.com', 'metacritic.com', 'soundcloud.com', 'bandcamp.com', 'qq.com', 'zillow.com', 'redfin.com', 'autotrader.com', 'skillshare.com', 'udacity.com', 'aol.com', 'openai.com', 'unsplash.com']

New friendly\_label distribution:

friendly\_label

1     101

0     40

2     9

Name: count, dtype: int64

[9]: labels.sample(5)

```

[9]:
      domain  friendly_label  compat_score  scraping_perm  \
110  yandex.ru              1          0.500              1
44   npr.org                1          0.500              1
130  aol.com                0          0.318              1
18   stackoverflow.com      0          0.220              0
96   soundcloud.com         0          0.318              1

      ai_training_perm  derivative_perm  \
110                  1                1
44                   1                1
130                  1                1
18                   1                1
96                   1                1

```

notes \

```

110 robots_has_allow
44 auto-labeled
130 robots_has_allow; manual_demote_hostile; manua...
18 robots_disallow_root
96 auto-labeled; manual_demote_hostile; manual_de...

data_quality
110 complete
44 complete
130 html_status_429,missing_tos_file,tos_http_miss...
18 missing_tos_file,robots_http_418,tos_http_miss...
96 html_status_403,missing_tos_file,robots_http_4...

```

```

[10]: sus = labels[
      (labels["friendly_label"].isin([0, 2])) &
      (labels["data_quality"] != "complete")
    ]
sus.head(10)

```

```

[10]:
      domain  friendly_label  compat_score  scraping_perm  \
1    facebook.com           0         0.220             0
3    twitter.com           0         0.220             0
4    instagram.com          0         0.220             0
6    reddit.com            0         0.154             0
9    snapchat.com          0         0.318             1
10   wikipedia.org         2         0.825             2
18  stackoverflow.com        0         0.220             0
19  stackexchange.com        0         0.220             0
20    quora.com             0         0.154             0
21   medium.com             0         0.318             1

```

```

      ai_training_perm  derivative_perm  \
1                   1                 1
3                   1                 1
4                   1                 1
6                   1                 1
9                   1                 1
10                  1                 1
18                  1                 1
19                  1                 1
20                  1                 1
21                  1                 1

```

```

      notes  \
1 robots_disallow_root
3 robots_disallow_root
4 robots_disallow_root

```

```

6             robots_disallow_root
9  auto-labeled; manual_demote_hostile; manual_de...
10 auto-labeled; manual_promote_friendly; manual_...
18             robots_disallow_root
19             robots_disallow_root
20             robots_disallow_root; robots_has_allow
21 auto-labeled; manual_demote_hostile; manual_de...

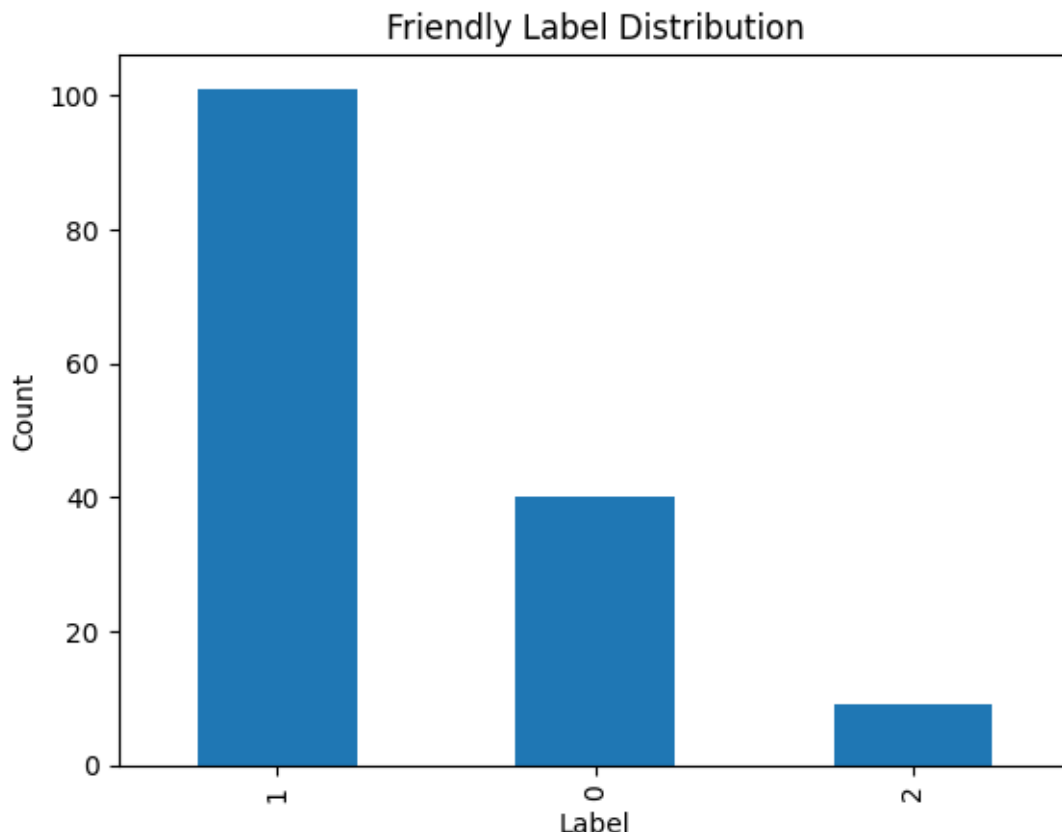
                                data_quality
1                                tiny_tos
3                                tiny_tos
4                                tiny_tos
6  html_status_403,missing_tos_file,tos_http_miss...
9  html_status_403,missing_tos_file,robots_http_4...
10 html_status_403,missing_tos_file,robots_http_4...
18 missing_tos_file,robots_http_418,tos_http_miss...
19 missing_tos_file,robots_http_418,tos_http_miss...
20 html_status_403,missing_tos_file,tos_http_miss...
21 html_status_403,missing_tos_file,robots_http_4...

```

```

[11]: import matplotlib.pyplot as plt
labels["friendly_label"].value_counts().plot(kind="bar", title="Friendly Label_
↪Distribution")
plt.xlabel("Label")
plt.ylabel("Count")
plt.show()

```



### 1.3 Text/ Feature Representation for ML tasks

1. TF-IDF: givs sparse high dimensional vector per domain.
2. SBERT Embeddings: use MiniLM to get dense embeddings. we will compare later in the project.

```
[12]: !pip install sentence-transformers scikit-learn scipy tqdm
```

```
Requirement already satisfied: sentence-transformers in  
/usr/local/lib/python3.12/dist-packages (5.1.2)  
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.12/dist-  
packages (1.6.1)  
Requirement already satisfied: scipy in /usr/local/lib/python3.12/dist-packages  
(1.16.3)  
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages  
(4.67.1)  
Requirement already satisfied: transformers<5.0.0,>=4.41.0 in  
/usr/local/lib/python3.12/dist-packages (from sentence-transformers) (4.57.3)  
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.12/dist-  
packages (from sentence-transformers) (2.9.0+cu126)  
Requirement already satisfied: huggingface-hub>=0.20.0 in
```

/usr/local/lib/python3.12/dist-packages (from sentence-transformers) (0.36.0)  
 Requirement already satisfied: Pillow in /usr/local/lib/python3.12/dist-packages  
 (from sentence-transformers) (11.3.0)  
 Requirement already satisfied: typing\_extensions>=4.5.0 in  
 /usr/local/lib/python3.12/dist-packages (from sentence-transformers) (4.15.0)  
 Requirement already satisfied: numpy>=1.19.5 in /usr/local/lib/python3.12/dist-  
 packages (from scikit-learn) (2.0.2)  
 Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.12/dist-  
 packages (from scikit-learn) (1.5.2)  
 Requirement already satisfied: threadpoolctl>=3.1.0 in  
 /usr/local/lib/python3.12/dist-packages (from scikit-learn) (3.6.0)  
 Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-  
 packages (from huggingface-hub>=0.20.0->sentence-transformers) (3.20.0)  
 Requirement already satisfied: fsspec>=2023.5.0 in  
 /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sentence-  
 transformers) (2025.3.0)  
 Requirement already satisfied: packaging>=20.9 in  
 /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sentence-  
 transformers) (25.0)  
 Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.12/dist-  
 packages (from huggingface-hub>=0.20.0->sentence-transformers) (6.0.3)  
 Requirement already satisfied: requests in /usr/local/lib/python3.12/dist-  
 packages (from huggingface-hub>=0.20.0->sentence-transformers) (2.32.4)  
 Requirement already satisfied: hf-xet<2.0.0,>=1.1.3 in  
 /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sentence-  
 transformers) (1.2.0)  
 Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-  
 packages (from torch>=1.11.0->sentence-transformers) (75.2.0)  
 Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-  
 packages (from torch>=1.11.0->sentence-transformers) (1.14.0)  
 Requirement already satisfied: networkx>=2.5.1 in  
 /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-  
 transformers) (3.6.1)  
 Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages  
 (from torch>=1.11.0->sentence-transformers) (3.1.6)  
 Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in  
 /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-  
 transformers) (12.6.77)  
 Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in  
 /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-  
 transformers) (12.6.77)  
 Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in  
 /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-  
 transformers) (12.6.80)  
 Requirement already satisfied: nvidia-cudnn-cu12==9.10.2.21 in  
 /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-  
 transformers) (9.10.2.21)  
 Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in

/usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (12.6.4.1)  
 Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (11.3.0.4)  
 Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (10.3.7.77)  
 Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (11.7.1.2)  
 Requirement already satisfied: nvidia-cuspars-cu12==12.5.4.2 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (12.5.4.2)  
 Requirement already satisfied: nvidia-cusparse-cu12==12.5.4.2 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (12.5.4.2)  
 Requirement already satisfied: nvidia-cusparselt-cu12==0.7.1 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (0.7.1)  
 Requirement already satisfied: nvidia-nccl-cu12==2.27.5 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (2.27.5)  
 Requirement already satisfied: nvidia-nvshmem-cu12==3.3.20 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (3.3.20)  
 Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (12.6.77)  
 Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (12.6.85)  
 Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (1.11.1.6)  
 Requirement already satisfied: triton==3.5.0 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers) (3.5.0)  
 Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.12/dist-packages (from transformers<5.0.0,>=4.41.0->sentence-transformers) (2025.11.3)  
 Requirement already satisfied: tokenizers<0.23.0,>=0.22.0 in /usr/local/lib/python3.12/dist-packages (from transformers<5.0.0,>=4.41.0->sentence-transformers) (0.22.1)  
 Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.12/dist-packages (from transformers<5.0.0,>=4.41.0->sentence-transformers) (0.7.0)  
 Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.12/dist-packages (from sympy>=1.13.3->torch>=1.11.0->sentence-transformers) (1.3.0)  
 Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2->torch>=1.11.0->sentence-

```
transformers) (3.0.3)
Requirement already satisfied: charset_normalizer<4,>=2 in
/usr/local/lib/python3.12/dist-packages (from requests->huggingface-
hub>=0.20.0->sentence-transformers) (3.4.4)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-
packages (from requests->huggingface-hub>=0.20.0->sentence-transformers) (3.11)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.12/dist-packages (from requests->huggingface-
hub>=0.20.0->sentence-transformers) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.12/dist-packages (from requests->huggingface-
hub>=0.20.0->sentence-transformers) (2025.11.12)
```

[13]: `python -m src.build_features`

```
2025-12-15 01:57:43.129909: E
external/local_xla/xla/stream_executor/cuda/cuda_fft.cc:467] Unable to register
cuFFT factory: Attempting to register factory for plugin cuFFT when one has
already been registered
WARNING: All log messages before absl::InitializeLog() is called are written to
STDERR
E0000 00:00:1765763863.149384      8187 cuda_dnn.cc:8579] Unable to register cuDNN
factory: Attempting to register factory for plugin cuDNN when one has already
been registered
E0000 00:00:1765763863.155333      8187 cuda_blas.cc:1407] Unable to register
cuBLAS factory: Attempting to register factory for plugin cuBLAS when one has
already been registered
W0000 00:00:1765763863.170342      8187 computation_placer.cc:177] computation
placer already registered. Please check linkage and avoid linking the same
target more than once.
W0000 00:00:1765763863.170366      8187 computation_placer.cc:177] computation
placer already registered. Please check linkage and avoid linking the same
target more than once.
W0000 00:00:1765763863.170370      8187 computation_placer.cc:177] computation
placer already registered. Please check linkage and avoid linking the same
target more than once.
W0000 00:00:1765763863.170373      8187 computation_placer.cc:177] computation
placer already registered. Please check linkage and avoid linking the same
target more than once.
2025-12-15 01:57:43.174784: I tensorflow/core/platform/cpu_feature_guard.cc:210]
This TensorFlow binary is optimized to use available CPU instructions in
performance-critical operations.
To enable the following instructions: AVX2 AVX512F FMA, in other operations,
rebuild TensorFlow with the appropriate compiler flags.
Number of domains: 150
robots.txt features.
robots.txt + HTML - numeric features.
X_numeric_raw shape: (150, 12)
```

```

TF-IDF on TOS texts.
TF-IDF shape: (150, 105989)
Loading SBERT model: all-MiniLM-L6-v2
modules.json: 100% 349/349 [00:00<00:00, 2.22MB/s]
config_sentence_transformers.json: 100% 116/116 [00:00<00:00, 916kB/s]
README.md: 10.5kB [00:00, 35.4MB/s]
sentence_bert_config.json: 100% 53.0/53.0 [00:00<00:00, 541kB/s]
config.json: 100% 612/612 [00:00<00:00, 6.04MB/s]
model.safetensors: 100% 90.9M/90.9M [00:00<00:00, 115MB/s]
tokenizer_config.json: 100% 350/350 [00:00<00:00, 3.39MB/s]
vocab.txt: 232kB [00:00, 25.2MB/s]
tokenizer.json: 466kB [00:00, 49.8MB/s]
special_tokens_map.json: 100% 112/112 [00:00<00:00, 1.17MB/s]
config.json: 100% 190/190 [00:00<00:00, 1.89MB/s]
Encoding TOS texts with SBERT...
Batches: 100% 5/5 [00:02<00:00, 2.27it/s]
SBERT embedding shape: (150, 384)
aligning labels.
y_class shape: (150,), y_score shape: (150,)
standardizing numeric features.
X_numeric (scaled) shape: (150, 12)
Combining TF-IDF text with numeric features.
Final X_tfidf shape: (150, 106001)
Combining SBERT embeddings with numeric features.
Final X_embed shape: (150, 396)
Saving feature to data/processed/...

```

```

[14]: !ls -lh data/processed/
# shape
import numpy as np, scipy.sparse as sp

X_tfidf = sp.load_npz("data/processed/X_tfidf.npz")
X_embed = np.load("data/processed/X_embed.npy")
y_class = np.load("data/processed/y_class.npy")

print("X_tfidf:", X_tfidf.shape, "sparse")
print("X_embed:", X_embed.shape, "dense")
print("y_class:", y_class.shape)

```

```

total 1.1M
-rw-r--r-- 1 root root 921 Dec 15 01:28 clf_results.csv
-rw-r--r-- 1 root root 5.7K Dec 15 01:28 html_stats.csv
-rw-r--r-- 1 root root 14K Dec 15 01:55 labels.csv
-rw-r--r-- 1 root root 586 Dec 15 01:28 reg_results.csv
-rw-r--r-- 1 root root 5.4K Dec 15 01:28 robots_log.csv
-rw-r--r-- 1 root root 2.6K Dec 15 01:28 splits.csv
-rw-r--r-- 1 root root 2.0K Dec 15 01:28 splits.npz
-rw-r--r-- 1 root root 7.5K Dec 15 01:28 tos_log.csv

```

```
-rw-r--r-- 1 root root 233K Dec 15 01:57 X_embed.npy
-rw-r--r-- 1 root root 7.2K Dec 15 01:57 X_numeric.npy
-rw-r--r-- 1 root root 729K Dec 15 01:57 X_tfidf.npz
-rw-r--r-- 1 root root 1.3K Dec 15 01:57 y_class.npy
-rw-r--r-- 1 root root 728 Dec 15 01:57 y_score.npy
X_tfidf: (150, 106001) sparse
X_embed: (150, 396) dense
y_class: (150,)
```

## 1.4 Splitting the data

```
[15]: !python -m src.split_data
      !ls data/processed/
```

```
[split_data] Loading data...
[split_data] Loaded 150 domains.
[split_data] Creating Train/Val/Test splits...
Train: 105 | Val: 22 | Test: 23
[split_data] Saving results...
[split_data] Done.
clf_results.csv  reg_results.csv  splits.npz  X_numeric.npy  y_score.npy
html_stats.csv  robots_log.csv   tos_log.csv X_tfidf.npz
labels.csv      splits.csv       X_embed.npy y_class.npy
```

```
[16]: import numpy as np, pandas as pd

splits = np.load("data/processed/splits.npz")
print("Train:", len(splits["train_idx"]))
print("Val:", len(splits["val_idx"]))
print("Test:", len(splits["test_idx"]))

# See the csv mapping
pd.read_csv("data/processed/splits.csv").head()
```

```
Train: 105
Val: 22
Test: 23
```

```
[16]:      domain  split
0    google.com  train
1  facebook.com  train
2  youtube.com   test
3  twitter.com   train
4  instagram.com  train
```

## 1.5 Supervised Learning (Model training)

```
[17]: %cd /content/AI_agent_compatibility
      !python -m src.train_models
```

```
/content/AI_agent_compatibility
Running classification experiments...
```

```

===== NB_TFIDF =====
Best params: {'alpha': 0.1}
Train acc=0.819, F1=0.523 | Val acc=0.818, F1=0.534 | Test acc=0.826, F1=0.561
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
    warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
    warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
    warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
    warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
    warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
    warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
    warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
    warnings.warn(

```

in 1.7. From then on, it will always use 'multinomial'. Leave it to its default value to avoid this warning.

```
warnings.warn(  
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:  
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.
```

```
warnings.warn(  
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:  
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.
```

```
warnings.warn(  
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:  
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.
```

```
warnings.warn(  
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:  
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.
```

```
warnings.warn(  
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:  
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.
```

```
warnings.warn(  
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:  
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.
```

```
warnings.warn(  
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:  
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.
```

```
warnings.warn(  
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:  
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.
```

```
warnings.warn(  
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:  
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.
```

```
warnings.warn(  

```

```
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
```

```
warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
```

```
warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
```

```
warnings.warn(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247:
FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default
value to avoid this warning.
```

```
warnings.warn(
```

```
==== LR_TFIDF ====
```

```
Best params: {'C': 0.1}
```

```
Train acc=0.819, F1=0.540 | Val acc=0.864, F1=0.585 | Test acc=0.870, F1=0.606
```

```
==== LinearSVC_TFIDF ====
```

```
Best params: {'C': 0.01}
```

```
Train acc=0.819, F1=0.540 | Val acc=0.864, F1=0.585 | Test acc=0.870, F1=0.606
```

```
==== RF_DENSE ====
```

```
Best params: {'max_depth': None, 'n_estimators': 100}
```

```
Train acc=1.000, F1=1.000 | Val acc=0.864, F1=0.585 | Test acc=0.783, F1=0.525
```

```
==== GB_DENSE ====
```

```
Best params: {'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100}
```

```
Train acc=1.000, F1=1.000 | Val acc=0.864, F1=0.585 | Test acc=0.783, F1=0.534
```

```
==== MLP_DENSE ====
```

```
Best params: {'alpha': 0.0001, 'hidden_layer_sizes': (128, 64)}
```

```
Train acc=0.990, F1=0.964 | Val acc=0.909, F1=0.619 | Test acc=0.870, F1=0.606
```

```
Saved classification results to data/processed/clf_results.csv
```

```
Running regression experiments...
```

```
==== Ridge_TFIDF (regression) ====
```

```
Best params: {'alpha': 10.0}
```

```
Train MAE=0.070, RMSE=0.103 | Val MAE=0.051, RMSE=0.081 | Test MAE=0.071,
```

RMSE=0.108

==== RFReg\_DENSE (regression) ====

Best params: {'max\_depth': 5, 'n\_estimators': 300}

Train MAE=0.033, RMSE=0.053 | Val MAE=0.039, RMSE=0.076 | Test MAE=0.078, RMSE=0.146

==== MLPReg\_DENSE (regression) ====

Best params: {'alpha': 0.001, 'hidden\_layer\_sizes': (64,)}

Train MAE=0.031, RMSE=0.057 | Val MAE=0.087, RMSE=0.108 | Test MAE=0.099, RMSE=0.141

Saved regression results to data/processed/reg\_results.csv

complete.

## 1.6 Discovering latent patterns

```
[18]: !python -m src.unsupervised_analysis
```

```
/content/AI_agent_compatibility/data/processed
/content/AI_agent_compatibility/data/domains.csv
```

```
X_embed shape: (150, 396)
```

```
y_class shape: (150,)
```

```
Number of domains: 150
```

```
PC1 variance: 0.2256
```

```
PC2 variance: 0.1690
```

```
Total variance (PC1+PC2): 0.3947
```

```
PCA scatter plot saved to:
```

```
/content/AI_agent_compatibility/figures/pca_scatter.png
```

```
k-means with k = 3
```

```
Cluster vs true-label:
```

```
true_label  0  1  2
```

```
cluster
```

```
0           7  54  4
```

```
1          32  47  5
```

```
2           1   0  0
```

```
Saved contingency table to:
```

```
/content/AI_agent_compatibility/results/cluster_summary_k3.csv
```

```
summary of clusters by main label:
```

- Cluster 0: mostly Neutral, label counts = {1: 54, 0: 7, 2: 4}
- Cluster 1: mostly Neutral, label counts = {1: 47, 0: 32, 2: 5}
- Cluster 2: mostly Hostile, label counts = {0: 1}

```
Cluster scatter plot saved to:
```

```
/content/AI_agent_compatibility/figures/kmeans_k3_scatter.png
```

```
k-means with k = 4
```

Cluster vs true-label:

true_label	0	1	2
cluster			
0	1	14	3
1	23	11	1
2	1	0	0
3	15	76	5

Saved contingency table to:

/content/AI\_agent\_compatibility/results/cluster\_summary\_k4.csv

summary of clusters by main label:

- Cluster 0: mostly Neutral, label counts = {1: 14, 2: 3, 0: 1}
- Cluster 1: mostly Hostile, label counts = {0: 23, 1: 11, 2: 1}
- Cluster 2: mostly Hostile, label counts = {0: 1}
- Cluster 3: mostly Neutral, label counts = {1: 76, 0: 15, 2: 5}

Cluster scatter plot saved to:

/content/AI\_agent\_compatibility/figures/kmeans\_k4\_scatter.png

k-means with k = 5

Cluster vs true-label:

true_label	0	1	2
cluster			
0	7	69	6
1	8	11	0
2	1	0	0
3	23	11	1
4	1	10	2

Saved contingency table to:

/content/AI\_agent\_compatibility/results/cluster\_summary\_k5.csv

summary of clusters by main label:

- Cluster 0: mostly Neutral, label counts = {1: 69, 0: 7, 2: 6}
- Cluster 1: mostly Neutral, label counts = {1: 11, 0: 8}
- Cluster 2: mostly Hostile, label counts = {0: 1}
- Cluster 3: mostly Hostile, label counts = {0: 23, 1: 11, 2: 1}
- Cluster 4: mostly Neutral, label counts = {1: 10, 2: 2, 0: 1}

Cluster scatter plot saved to:

/content/AI\_agent\_compatibility/figures/kmeans\_k5\_scatter.png

Anomaly scores saved to:

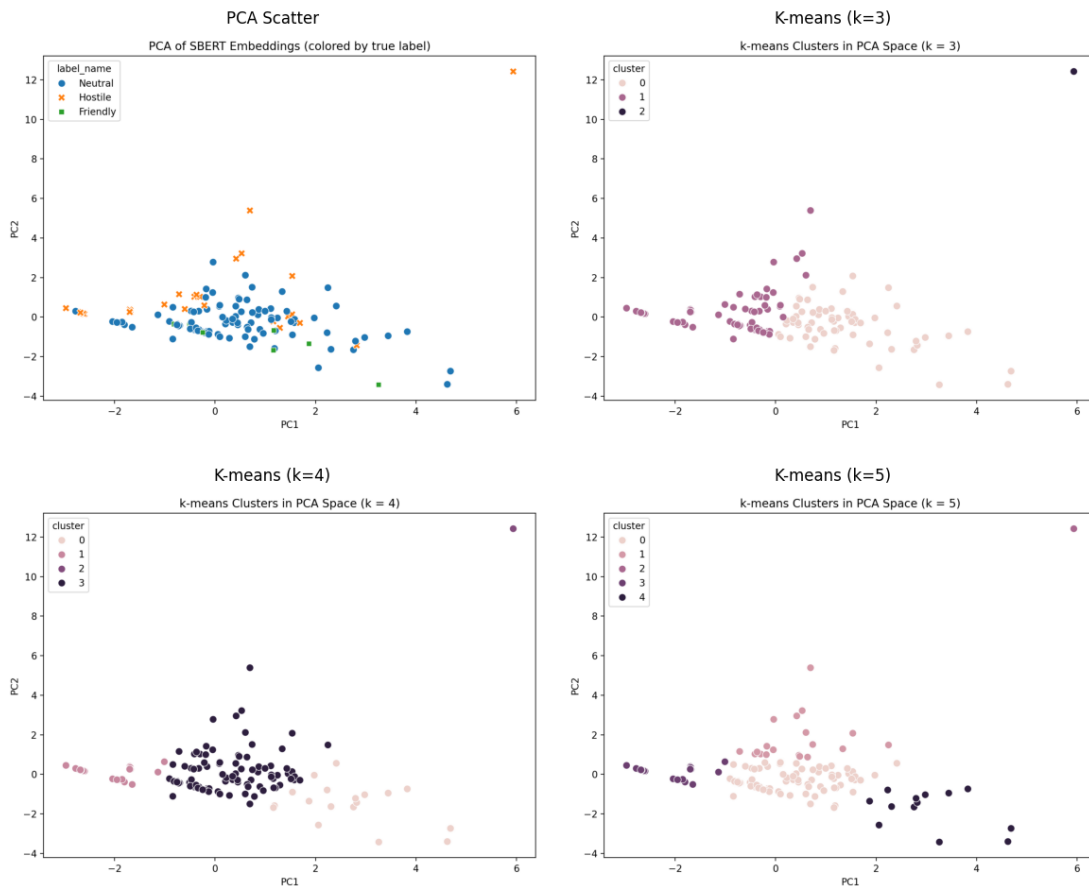
/content/AI\_agent\_compatibility/results/anomaly\_scores.csv

Top 10 most anomalous domains (higher score = more anomalous):

- |                   |                                 |
|-------------------|---------------------------------|
| - yandex.ru       | score = 0.0302, label = Neutral |
| - harvard.edu     | score = 0.0200, label = Neutral |
| - linkedin.com    | score = 0.0198, label = Hostile |
| - taobao.com      | score = 0.0137, label = Neutral |
| - trello.com      | score = 0.0130, label = Neutral |
| - tumblr.com      | score = 0.0107, label = Neutral |
| - khanacademy.org | score = 0.0096, label = Neutral |
| - skyscanner.net  | score = 0.0095, label = Neutral |

```
- nytimes.com          score = 0.0093, label = Neutral
- tiktok.com           score = 0.0081, label = Neutral
```

```
[19]: import matplotlib.pyplot as plt
import matplotlib.image as mpimg
fig, axes = plt.subplots(2, 2, figsize=(12, 10))
paths = [
    "/content/AI_agent_compatibility/figures/pca_scatter.png",
    "/content/AI_agent_compatibility/figures/kmeans_k3_scatter.png",
    "/content/AI_agent_compatibility/figures/kmeans_k4_scatter.png",
    "/content/AI_agent_compatibility/figures/kmeans_k5_scatter.png",
]
titles = ["PCA Scatter", "K-means (k=3)", "K-means (k=4)", "K-means (k=5)"]
for ax, path, title in zip(axes.flat, paths, titles):
    img = mpimg.imread(path)
    ax.imshow(img)
    ax.axis("off")
    ax.set_title(title, fontsize=12)
plt.tight_layout()
plt.show()
```



## 1.7 Digging Deeper

1. Question 1: Which models generalize best?
2. Question 2: What kinds of mistakes do the classifiers make?
3. Question 3: Is the best classifier data-limited or capacity-limited?
4. Show regression performance in a clean place

```
[20]: !python3 src/evaluate_models.py
```

Classification model summary (from model training)

	model	train_acc	train_f1	val_acc	val_f1	test_acc	test_f1
	MLP_DENSE	0.990476	0.963849	0.909091	0.618803	0.869565	0.606061
	LR_TFIDF	0.819048	0.540247	0.863636	0.584565	0.869565	0.606061
	RF_DENSE	1.000000	1.000000	0.863636	0.584565	0.782609	0.525253
LinearSVC	TFIDF	0.819048	0.540247	0.863636	0.584565	0.869565	0.606061
	GB_DENSE	1.000000	1.000000	0.863636	0.584565	0.782609	0.534091
	NB_TFIDF	0.819048	0.523106	0.818182	0.534091	0.826087	0.560784

Regression model summary (from model training)

	model	train_mae	train_rmse	val_mae	val_rmse	test_mae	test_rmse
	RFReg_DENSE	0.032641	0.052957	0.039000	0.076456	0.078060	0.145899
	Ridge_TFIDF	0.069754	0.102724	0.051079	0.080572	0.071041	0.108242
	MLPReg_DENSE	0.031396	0.057433	0.086539	0.107660	0.098634	0.141153

/usr/local/lib/python3.12/dist-packages/sklearn/linear\_model/\_logistic.py:1247:  
FutureWarning: 'multi\_class' was deprecated in version 1.5 and will be removed  
in 1.7. From then on, it will always use 'multinomial'. Leave it to its default  
value to avoid this warning.

warnings.warn(

detailed classification metrics (val/test):

	model	split	accuracy	macro_precision	macro_recall	macro_f1
	LR_TFIDF	test	0.869565	0.611111	0.611111	0.606061
LinearSVC	TFIDF	test	0.869565	0.611111	0.611111	0.606061
	NB_TFIDF	test	0.826087	0.596491	0.555556	0.560784
	MLP_DENSE	test	0.826087	0.596491	0.555556	0.560784
	RF_DENSE	test	0.782609	0.525926	0.533333	0.525253
	GB_DENSE	test	0.782609	0.541176	0.533333	0.534091
	RF_DENSE	val	0.909091	0.596825	0.644444	0.618803
	MLP_DENSE	val	0.909091	0.596825	0.644444	0.618803
	LR_TFIDF	val	0.863636	0.559524	0.622222	0.584565
LinearSVC	TFIDF	val	0.863636	0.559524	0.622222	0.584565
	GB_DENSE	val	0.863636	0.559524	0.622222	0.584565
	NB_TFIDF	val	0.818182	0.541176	0.533333	0.534091

regression metrics copied from reg\_results.csv to regression\_metrics.csv

Best classifier (by val\_f1): MLP\_DENSE

model	MLP_DENSE
best_params	{'alpha': 0.0001, 'hidden_layer_sizes': (128, ...
train_acc	0.990476
train_f1	0.963849
val_acc	0.909091
val_f1	0.618803
test_acc	0.869565
test_f1	0.606061

=== Phase 7: learning curve (best classifier) ===

20.0% data | train\_acc=1.000, val\_acc=0.818

40.0% data | train\_acc=1.000, val\_acc=0.773

60.0% data | train\_acc=0.984, val\_acc=0.864

80.0% data | train\_acc=0.988, val\_acc=0.864

100.0% data | train\_acc=0.990, val\_acc=0.909

Comment: train accuracy noticeably higher than val accuracy → overfitting / high variance.

```
[21]: import pandas as pd
pd.read_csv("results/classification_metrics.csv").head()
```

```
[21]:
```

	model	split	accuracy	macro_precision	macro_recall	macro_f1	\
0	NB_TFIDF	train	0.819048	0.578418	0.511905	0.523106	
1	NB_TFIDF	val	0.818182	0.541176	0.533333	0.534091	
2	NB_TFIDF	test	0.826087	0.596491	0.555556	0.560784	
3	LR_TFIDF	train	0.819048	0.534891	0.547954	0.540247	
4	LR_TFIDF	val	0.863636	0.559524	0.622222	0.584565	

	precision_hostile	recall_hostile	f1_hostile	support_hostile	\
0	0.937500	0.535714	0.681818	28	
1	0.800000	0.666667	0.727273	6	
2	1.000000	0.666667	0.800000	6	
3	0.769231	0.714286	0.740741	28	
4	0.750000	1.000000	0.857143	6	

	precision_neutral	recall_neutral	f1_neutral	support_neutral	\
0	0.797753	1.000000	0.887500	71	
1	0.823529	0.933333	0.875000	15	
2	0.789474	1.000000	0.882353	15	
3	0.835443	0.929577	0.880000	71	
4	0.928571	0.866667	0.896552	15	

	precision_friendly	recall_friendly	f1_friendly	support_friendly
0	0.0	0.0	0.0	6
1	0.0	0.0	0.0	1

2	0.0	0.0	0.0	2
3	0.0	0.0	0.0	6
4	0.0	0.0	0.0	1

```
[22]: import pandas as pd
pd.read_csv("results/regression_metrics.csv").head()
```

```
[22]:
```

	model	best_params	train_mae \
0	Ridge_TFIDF	{'alpha': 10.0}	0.069754
1	RFRReg_DENSE	{'max_depth': 5, 'n_estimators': 300}	0.032641
2	MLPReg_DENSE	{'alpha': 0.001, 'hidden_layer_sizes': (64,,)}	0.031396

	train_rmse	val_mae	val_rmse	test_mae	test_rmse
0	0.102724	0.051079	0.080572	0.071041	0.108242
1	0.052957	0.039000	0.076456	0.078060	0.145899
2	0.057433	0.086539	0.107660	0.098634	0.141153

```
[23]: import matplotlib.pyplot as plt
import matplotlib.image as mpimg
from pathlib import Path

# Path to figures folder
fig_dir = Path("/content/AI_agent_compatibility/figures")

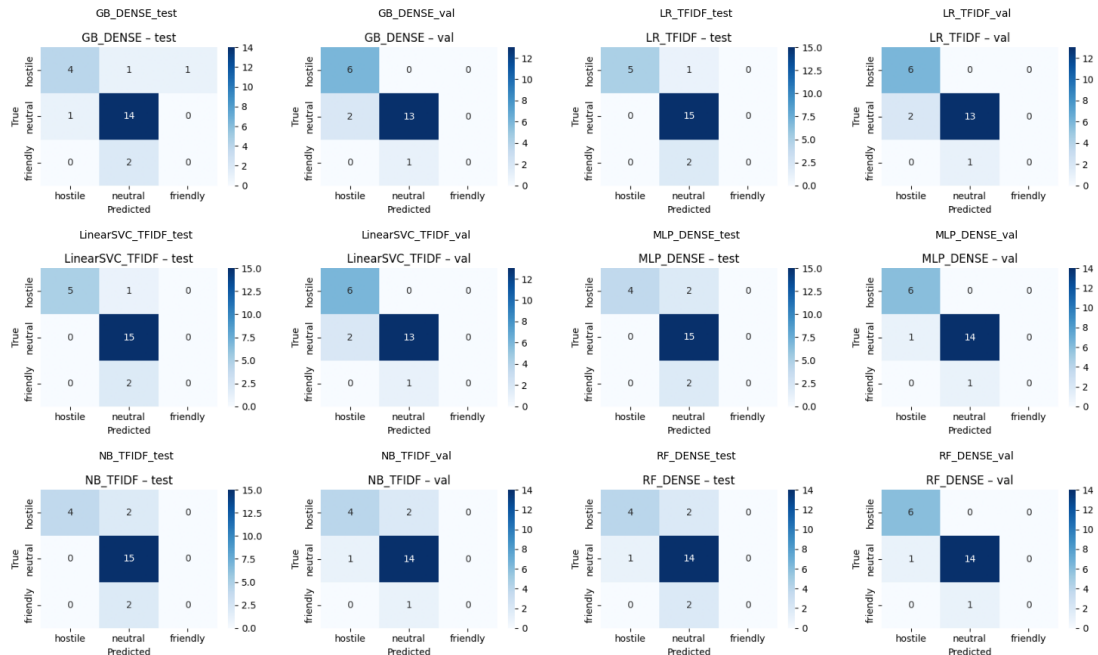
# Collect only confusion matrix images
conf_imgs = sorted(fig_dir.glob("confusion_*.png"))

# Display all in a grid
n = len(conf_imgs)
cols = 4
rows = (n + cols - 1) // cols

plt.figure(figsize=(15, rows * 3))

for i, img_path in enumerate(conf_imgs, 1):
    plt.subplot(rows, cols, i)
    img = mpimg.imread(img_path)
    plt.imshow(img)
    plt.title(img_path.name.replace("confusion_", "").replace(".png", ""),
    ↪fontsize=10)
    plt.axis("off")

plt.tight_layout()
plt.show()
```



Commit Everything

```
[24]: !git config --global user.name "Rashi Tyagi"
      !git config --global user.email "ratyagi23@outlook.com"
```

```
[15]: %cd /content/AI_agent_compatibility
      !git fetch origin main
      !git reset --hard origin/main
```

```
/content/AI_agent_compatibility
remote: Enumerating objects: 9, done.
remote: Counting objects: 100% (9/9), done.
remote: Compressing objects: 100% (5/5), done.
remote: Total 7 (delta 4), reused 5 (delta 2), pack-reused 0 (from 0)
Unpacking objects: 100% (7/7), 4.14 KiB | 1.38 MiB/s, done.
From https://github.com/ratyagi/AI_agent_compatibility
* branch          main          -> FETCH_HEAD
   bca3c65..a3f657d main        -> origin/main
HEAD is now at a3f657d more eval
```

```
[25]: %cd /content/AI_agent_compatibility
      !git add data src figures results docs
      !git commit -m "checkpoint"
```

```

# Push to GitHub
import getpass
token = getpass.getpass("Enter your GitHub token: ")
username = "ratyagi"
repo = "AI_agent_compatibility"

!git remote set-url origin https://{token}@github.com/ratyagi/
↪AI_agent_compatibility
!git push origin main

```

```

/content/AI_agent_compatibility
warning: CRLF will be replaced by LF in data/domains_normalized.csv.
The file will have its original line endings in your working directory
[main be37b24] checkpoint
39 files changed, 351 insertions(+), 330 deletions(-)
rewrite data/processed/X_embed.npy (98%)
rewrite data/processed/clf_results.csv (85%)
rewrite data/processed/reg_results.csv (87%)
rewrite data/processed/splits.npz (87%)
rewrite data/processed/y_class.npy (66%)
rewrite data/processed/y_score.npy (82%)
rewrite figures/confusion_GB_DENSE_test.png (99%)
rewrite figures/confusion_GB_DENSE_val.png (99%)
rewrite figures/confusion_LR_TFIDF_test.png (99%)
rewrite figures/confusion_LR_TFIDF_val.png (98%)
rewrite figures/confusion_LinearSVC_TFIDF_test.png (98%)
rewrite figures/confusion_LinearSVC_TFIDF_val.png (99%)
rewrite figures/confusion_MLP_DENSE_test.png (99%)
rewrite figures/confusion_MLP_DENSE_val.png (99%)
rewrite figures/confusion_NB_TFIDF_test.png (99%)
rewrite figures/confusion_NB_TFIDF_val.png (99%)
rewrite figures/confusion_RF_DENSE_test.png (99%)
rewrite figures/confusion_RF_DENSE_val.png (99%)
rewrite figures/kmeans_k3_scatter.png (68%)
rewrite figures/kmeans_k4_scatter.png (96%)
rewrite figures/kmeans_k5_scatter.png (68%)
rewrite figures/learning_curve_best_classifier.png (99%)
rewrite figures/pca_scatter.png (68%)
rewrite results/anomaly_scores.csv (99%)
rewrite results/classification_metrics.csv (92%)
rewrite results/regression_metrics.csv (86%)
Enter your GitHub token: .....
Enumerating objects: 79, done.
Counting objects: 100% (79/79), done.
Delta compression using up to 2 threads
Compressing objects: 100% (46/46), done.
Writing objects: 100% (47/47), 712.00 KiB | 11.67 MiB/s, done.

```

Total 47 (delta 10), reused 0 (delta 0), pack-reused 0  
 remote: Resolving deltas: 100% (10/10), completed with 9 local objects.  
 To https://github.com/ratyagi/AI\_agent\_compatibility  
 2e1b524..be37b24 main -> main

```
[25]: #!/cd /content/AI_agent_compatibility
#!/git add .
#!/git commit -m "Checkpoint before data expansion"

# Create a tag (replace v1.0 with any name)
#!/git tag -a v1.0 -m "Stable checkpoint before dataset expansion"

# Push both commits and tag to GitHub
#!/git push origin main --tags
#to restore
#!/git checkout v1.0
```

/content/AI\_agent\_compatibility  
 On branch main  
 Your branch is up to date with 'origin/main'.

nothing to commit, working tree clean  
 Enumerating objects: 1, done.  
 Counting objects: 100% (1/1), done.  
 Writing objects: 100% (1/1), 184 bytes | 184.00 KiB/s, done.  
 Total 1 (delta 0), reused 0 (delta 0), pack-reused 0  
 To https://github.com/ratyagi/AI\_agent\_compatibility  
 \* [new tag] v1.0 -> v1.0

```
[28]: # from https://gist.github.com/jonathanagustin/b67b97ef12c53a8dec27b343dca4abba
# install can take a minute

import os
# @title Convert Notebook to PDF. Save Notebook to given directory
NOTEBOOKS_DIR = "/content/drive/MyDrive" # @param {type:"string"}
NOTEBOOK_NAME = "AI_agent_compatibility.ipynb" # @param {type:"string"}
#-----#

from google.colab import drive
drive.mount("/content/drive/", force_remount=True)
NOTEBOOK_PATH = f"{NOTEBOOKS_DIR}/{NOTEBOOK_NAME}"
assert os.path.exists(NOTEBOOK_PATH), f"NOTEBOOK NOT FOUND: {NOTEBOOK_PATH}"
[!]apt install -y texlive-xetex texlive-fonts-recommended texlive-generic >_
↪ /dev/null 2>&1
[!]jupyter nbconvert "$NOTEBOOK_PATH" --to pdf > /dev/null 2>&1
NOTEBOOK_PDF = NOTEBOOK_PATH.rsplit('.', 1)[0] + '.pdf'
assert os.path.exists(NOTEBOOK_PDF), f"ERROR MAKING PDF: {NOTEBOOK_PDF}"
print(f"PDF CREATED: {NOTEBOOK_PDF}")
```

Mounted at /content/drive/

^C

```
-----
AssertionError                                Traceback (most recent call last)
/tmp/ipython-input-486031648.py in <cell line: 0>()
    14 get_ipython().system('jupyter nbconvert "$NOTEBOOK_PATH" --to pdf > /de /
↳null 2>&1')
    15 NOTEBOOK_PDF = NOTEBOOK_PATH.rsplit('.', 1)[0] + '.pdf'
--> 16 assert os.path.exists(NOTEBOOK_PDF), f"ERROR MAKING PDF: {NOTEBOOK_PDF}"
    17 print(f"PDF CREATED: {NOTEBOOK_PDF}")

AssertionError: ERROR MAKING PDF: /content/drive/MyDrive/AI_agent_compatibility
↳pdf
```