

A PVPs for AG's, Yelp, Yahoo

Below, we list the 23 patterns and the single verbalizer used for AG's News, Yelp Reviews Full Star and Yahoo Questions. We group all patterns by their category.

A.1 AG's News

Inputs

- x_1 : The headline of the news article to be classified.
- x_2 : The text body of the news article to be classified.

NULL Patterns

- [MASK] $x_1 x_2$
- $x_1 x_2$ [MASK]
- x_1 [MASK] x_2

PUNC Patterns

- [MASK] : $x_1 x_2$
- [MASK] - $x_1 x_2$
- [MASK] . $x_1 x_2$
- ([MASK]) $x_1 x_2$
- x_1 ([MASK]) x_2
- $x_1 x_2$ ([MASK])

PROMPT Patterns

- $x_1 x_2$ Category: [MASK].
- $x_1 x_2$ Class: [MASK].
- $x_1 x_2$ Topic: [MASK].
- $x_1 x_2$ Theme: [MASK].
- $x_1 x_2$ Category: [MASK]
- $x_1 x_2$ Class: [MASK]
- $x_1 x_2$ Topic: [MASK]
- $x_1 x_2$ Theme: [MASK]
- [MASK] News: $x_1 x_2$
- [MASK] NEWS: $x_1 x_2$

Q&A Patterns

- $x_1 x_2$ Question: What is the topic of this article? Answer: [MASK].
- $x_1 x_2$ Question: What is the category of this article? Answer: [MASK].
- $x_1 x_2$ Question: What is the topic of this article? Answer: [MASK]
- $x_1 x_2$ Question: What is the category of this article? Answer: [MASK]

Verbalizer World \mapsto World, Sports \mapsto Sports, Business \mapsto Business, Science/Tech \mapsto Tech

A.2 Yelp Reviews Full Star

Inputs

- x : The review to be classified.

NULL Patterns

- [MASK] x
- x [MASK]

PUNC Patterns

- [MASK] : x
- [MASK] - x
- [MASK] . x
- ([MASK]) x
- x ([MASK])

PROMPT Patterns

- x It was [MASK]
- x All in all, it was [MASK]
- x In summary, it was [MASK]
- x It was [MASK].
- x All in all, it was [MASK].
- x In summary, it was [MASK].
- x The restaurant was [MASK]
- x All in all, the restaurant was [MASK]
- x In summary, the restaurant was [MASK]
- x The restaurant was [MASK].
- x All in all, the restaurant was [MASK].
- x In summary, the restaurant was [MASK].

Q&A Patterns

- x Question: What does the customer think of this restaurant? Answer: It is [MASK].
- x Question: What does the customer think of this place? Answer: It is [MASK].
- x Question: What does the customer think of this restaurant? Answer: It is [MASK]
- x Question: What does the customer think of this place? Answer: It is [MASK]

Verbalizer 1 \mapsto terrible, 2 \mapsto bad, 3 \mapsto okay, 4 \mapsto good, 5 \mapsto great

A.3 Yahoo Questions

Inputs

- x_1 : The question to be classified.
- x_2 : The top answer to the question.

NULL Patterns

- [MASK] $x_1 x_2$
- $x_1 x_2$ [MASK]
- x_1 [MASK] x_2

PUNC Patterns

- [MASK] : $x_1 x_2$
- [MASK] - $x_1 x_2$
- [MASK] . $x_1 x_2$
- ([MASK]) $x_1 x_2$
- x_1 ([MASK]) x_2
- $x_1 x_2$ ([MASK])

PROMPT Patterns

- $x_1 x_2$ Category: [MASK].
- $x_1 x_2$ Class: [MASK].
- $x_1 x_2$ Topic: [MASK].
- $x_1 x_2$ Theme: [MASK].
- $x_1 x_2$ Category: [MASK]
- $x_1 x_2$ Class: [MASK]
- $x_1 x_2$ Topic: [MASK]
- $x_1 x_2$ Theme: [MASK]
- [MASK] Question: $x_1 x_2$
- [MASK] QUESTION: $x_1 x_2$

Q&A Patterns

- $x_1 x_2$ Question: What is the topic of this question? Answer: [MASK].
- $x_1 x_2$ Question: What is the category of this question? Answer: [MASK].
- $x_1 x_2$ Question: What is the topic of this question? Answer: [MASK]
- $x_1 x_2$ Question: What is the category of this question? Answer: [MASK]

Verbalizer Society & Culture \mapsto Society, Science & Mathematics \mapsto Science, Health \mapsto Health, Education & Reference \mapsto Education, Computers & Internet \mapsto Computer, Sports \mapsto Sports, Business & Finance \mapsto Business, Entertainment & Music \mapsto Entertainment, Family & Relationships \mapsto Relationship, Politics & Government \mapsto Politics

B PVPs for RAFT

Below, we list the PVPs used for all tasks in the RAFT benchmark. For each task t , we make use of the original task description D_t that we copy verbatim from https://github.com/oughtinc/raft-baselines/tree/master/example_prompts. We use two vertical bars (||) to mark boundaries between text segments. If multiple patterns and verbalizers are specified, each verbalizer is used in combination with each pattern.

B.1 ADE

Task Description Label the sentence based on whether it is related to an adverse drug effect (ADE). Details are described below:
Drugs: Names of drugs and chemicals that include brand names, trivial names, abbreviations and systematic names were annotated. Mentions of drugs or chemicals should strictly be in a therapeutic context. This category does not include the names of metabolites, reaction byproducts, or hospital chemicals (e.g. surgical equipment disinfectants).
Adverse effect: Mentions of adverse effects include signs, symptoms, diseases, disorders, acquired abnormalities, deficiencies, organ damage or death that strictly occur as a consequence of drug intake.

Inputs

- x : The text to be classified.

Patterns

- $D_t \parallel x$ Question: Is this sentence related to an adverse drug effect (ADE)? Answer: [MASK].
- x Question: Is this sentence related to an adverse drug effect (ADE)? Answer: [MASK].

Verbalizers

- not ADE-related \mapsto No, ADE-related \mapsto Yes
- not ADE-related \mapsto False, ADE-related \mapsto True

B.2 B77

Task Description The following is a banking customer service query. Classify the query into one of the 77 categories available.

Inputs

- x : The text to be classified.
- y : The correct intent for the given text.

Patterns

- $D_t \parallel x$ Question: Is y the correct category for this query? Answer: [MASK].
- x Question: Is y the correct category for this query? Answer: [MASK].

Verbalizers

- False \mapsto No, True \mapsto Yes
- False \mapsto False, True \mapsto True

B.3 NIS

Task Description Label the impact statement based on whether it mentions a harmful application of the research done in the paper. Make sure the statement is sufficient to conclude there are harmful applications of the research being done, not a past risk that this research is solving.

Inputs

- x : The text to be classified.

Patterns

- $D_t \parallel x$ Question: Does this impact statement mention a harmful application? Answer: [MASK].
- x Question: Does this impact statement mention a harmful application? Answer: [MASK].

Verbalizers

- doesn't mention a harmful application \mapsto No, mentions a harmful application \mapsto Yes
- doesn't mention a harmful application \mapsto False, mentions a harmful application \mapsto True

B.4 OSE

Task Description The following is an article sourced from The Guardian newspaper, and rewritten by teachers to suit three levels of adult English as Second Language (ESL) learners: elementary, intermediate, and advanced. Predict the level of the article.

Inputs

- x : The text to be classified.

Patterns

- $D_t \parallel x$ Question: What is the level of this article? Answer: [MASK].
- x Question: What is the level of this article? Answer: [MASK].
- $D_t \parallel x$ Question: Is the level of this article "elementary", "intermediate" or "advanced"? Answer: [MASK].
- x Question: Is the level of this article "elementary", "intermediate" or "advanced"? Answer: [MASK].

Verbalizers

- elementary \mapsto elementary, intermediate \mapsto intermediate, advanced \mapsto advanced

B.5 Over

Task Description In law, an overruling sentence is a statement that nullifies a previous case decision as a precedent, by a constitutionally valid statute or a decision by the same or higher ranking court which establishes a different rule on the point of law involved. Label the sentence based on whether it is overruling or not.

Inputs

- x : The text to be classified.

Patterns

- $D_t \parallel x$ Question: Is this sentence overruling? Answer: [MASK].
- x Question: Is this sentence overruling? Answer: [MASK].

Verbalizers

- not overruling \mapsto No, overruling \mapsto Yes
- not overruling \mapsto False, overruling \mapsto True

B.6 SOT

Task Description The dataset is a list of institutions that have contributed papers to semiconductor conferences in the last 25 years, as catalogued by IEEE and sampled randomly. The goal is to classify the institutions into one of three categories: "university", "company" or "research institute".

Inputs

- x_1 : The title of the paper to be classified.
- x_2 : The name of the organization to be classified.

Patterns

- $D_t \parallel$ Organization name: x_1 Paper title: x_2 Question: What is the category of this institution? Answer: [MASK].
- Organization name: x_1 Paper title: x_2 Question: What is the category of this institution? Answer: [MASK].
- $D_t \parallel$ Paper title: x_2 Organization name: x_1 Question: What is the category of this institution? Answer: [MASK].
- Paper title: x_2 Organization name: x_1 Question: What is the category of this institution? Answer: [MASK].

Verbalizers

- company \mapsto company, research institute \mapsto institute, university \mapsto university

B.7 SRI

Task Description Identify whether this paper should be included in a meta-review which includes the findings of systematic reviews on interventions designed to promote charitable donations.

Included reviews should describe monetary charitable donations, assess any population of participants in any context, and be peer reviewed and written in English.

They should not report new data, be non-systematic reviews, consider cause-related marketing or other kinds of prosocial behaviour.

Inputs

- x_1 : The title of the paper to be classified.
- x_2 : The abstract of the paper to be classified.
- x_3 : The journal of the paper to be classified.

Patterns

- $D_t \parallel$ Title: x_1 Abstract: x_2 Journal: x_3 Question: Should this paper be included in a meta-review which includes the findings of systematic reviews on interventions designed to promote charitable donations? Answer: [MASK].
- Title: x_1 Abstract: x_2 Journal: x_3 Question: Should this paper be included in a meta-review which includes the findings of systematic reviews on interventions designed to promote charitable donations? Answer: [MASK].

Verbalizers

- not included \mapsto No, included \mapsto Yes
- not included \mapsto False, included \mapsto True

B.8 TAI

Task Description Transformative AI (TAI) is defined as AI that precipitates a transition comparable to (or more significant than) the agricultural or industrial revolution. Label a paper as "TAI safety research" if:

1. The contents of the paper are directly motivated by, and substantively inform, the challenge of ensuring good outcomes for TAI,
2. There is substantive content on AI safety,

not just AI capabilities,

3. The intended audience is the community of researchers,

4. It meets a subjective threshold of seriousness/quality,

5. Peer review is not required.

Inputs

- x_1 : The title of the paper to be classified.
- x_2 : The abstract of the paper to be classified.

Patterns

- $D_t \parallel$ Title: x_1 Abstract: x_2 Question: Is this paper a TAI safety research paper? Answer: [MASK].
- Title: x_1 Abstract: x_2 Question: Is this paper a TAI safety research paper? Answer: [MASK].

Verbalizers

- not TAI safety research \mapsto No, TAI safety research \mapsto Yes
- not TAI safety research \mapsto False, TAI safety research \mapsto True

B.9 ToS

Task Description Label the sentence from a Terms of Service based on whether it is potentially unfair. If it seems clearly unfair, mark it as potentially unfair.

According to art. 3 of the Directive 93/13 on Unfair Terms in Consumer Contracts, a contractual term is unfair if: 1) it has not been individually negotiated; and 2) contrary to the requirement of good faith, it causes a significant imbalance in the parties rights and obligations, to the detriment of the consumer.

Inputs

- x : The text to be classified.

Patterns

- $D_t \parallel x$ Question: Is this sentence potentially unfair? Answer: [MASK].
- x Question: Is this sentence potentially unfair? Answer: [MASK].

Verbalizers

- not potentially unfair \mapsto No, potentially unfair \mapsto Yes
- not potentially unfair \mapsto False, potentially unfair \mapsto True

B.10 TEH

Task Description Label whether the following tweet contains hate speech against either immigrants or women. Hate Speech (HS) is commonly defined as any communication that disparages a person or a group on the basis of some characteristic such as race, color, ethnicity, gender, sexual orientation, nationality, religion, or other characteristics.

Inputs

- x : The text to be classified.

Patterns

- $D_t \parallel x$ Question: Does this tweet contain hate speech against either immigrants or women? Answer: [MASK].
- x Question: Does this tweet contain hate speech against either immigrants or women? Answer: [MASK].

Verbalizers

- not hate speech \mapsto No, hate speech \mapsto Yes
- not hate speech \mapsto False, hate speech \mapsto True

B.11 TC

Task Description A complaint presents a state of affairs which breaches the writer's favorable expectation. Label the tweet text based on whether it contains a complaint.

Inputs

- x : The text to be classified.

Patterns

- $D_t \parallel x$ Question: Does this tweet text contain a complaint? Answer: [MASK].
- x Question: Does this tweet text contain a complaint? Answer: [MASK].

Verbalizers

- no complaint \mapsto No, complaint \mapsto Yes
- no complaint \mapsto False, complaint \mapsto True