

Annotation Guide for Character Role Labeling

This project aims to create character role labels for different characters within news stories.

In narrative theory, a **character role** refers to the functional position or purpose a character occupies within a story. It is not defined by the character's individuality or psychology but by the actions, relationships, and narrative functions they perform to advance the plot or convey meaning. Character roles are thus general categories that collapse diverse character functions into recognizable and repeatable types. A good character role thus describes a character's function within a particular story and also a more general role that could apply to numerous other stories and narrative settings with similar functions.

Some examples of **character roles** applied to well-known characters:

- "Sidekick" (Robin from *Batman*, Dr. Watson from *Sherlock Holmes*)
- "Grumpy Old Man" (Carl from the Pixar movie *Up*, Ebenezer Scrooge from Charles Dickens' *A Christmas Carol*)
- "Corrupt (or Ruthless) Business Executive" (Lex Luthor from *Superman*, Miranda Priestly from *The Devil Wears Prada*)
- "Reluctant Hero" (Odysseus from Greek mythology, Luke Skywalker from *Star Wars*)
- "Trickster" (Loki from Norse mythology and Marvel, (Captain) Jack Sparrow from *Pirates of the Caribbean*)
- "Mentor" (Gandalf from *The Lord of Rings*, Morpheus from *The Matrix*)

As illustrated by the examples above, character roles are general classes of characters that possess similar traits, functions, or patterns of behaviour across different stories. In this sense, character roles are meant to be *descriptive* and *predictive*: they allow us to encapsulate a character's function in a given story by drawing on prior knowledge to infer who the character is and how they will likely behave. For example, Sidekicks can be expected to help out the main character, Reluctant Heroes will need to be cajoled into heroic action, and Tricksters will cause trouble but also find clever ways out of trouble. Overall, character roles are meant to simplify a character and abstract away many of their specific traits into a broader class.

Many character roles are quite general and apply to characters across different genres, but they can also differ in their specificity or generality. For instance, "Innocent Victim" is more specific and descriptive than "Victim", however either can be more appropriate depending on what aspects of victimhood the story emphasizes. In this particular case,

we can consider “victim” to be the primary label, and “innocent” to be a modifier of that label.

Importantly, character roles are also not mutually exclusive, and can depend on the source story or story genre. Merlin from the stories of King Arthur can be a Mentor, a Grumpy Old Man, or even a Hero depending on the specific story. Likewise, common character types for romance stories often revolve around roles in a love triangle which may not always be applicable to other stories (for example, the Hero/Heroine, the Broody Love Interest, and the Sensitive (alternative) Love Interest).

In this project, we are working with a news article dataset where the underlying character roles are not known beforehand. The goal is to identify these character role labels.

In this task, you will be presented with an article and a list of characters/entities from the text. Your goal is to write a character role label for each of the given characters based on how they are *portrayed in the text*.

Because news articles behave differently in terms of characterization than fiction, we provide a few insights around character role labeling specific to the news.

- 1) Characters in news articles can be individuals but also collective entities like corporations (Exxon), institutions (the Supreme Court), or countries (Russia). **It is important that a character role label does not repeat their entity identity.** For example, “corporation” is not a good character role label, but “corrupt corporation” is (or “Visionary Leader” instead of “Leader”). Similarly, because people in the news are often associated with their occupations **it is important that occupations not be confused with roles.** “Lawyer” is not a good role label, but “Hot-shot Lawyer” is because it gives us an idea of the type the lawyer is playing in the story. Role labels can also be more general like hero, villain, or victim if the story warrants it.
- 2) The point about labels being drawn from how characters “are portrayed in the text” is very important since the articles are drawn from real-life news and you may already be familiar with some of the characters or have preconceived notions about them. In this context, **your goal is to identify character roles based on how they are depicted by the author of the text, and not by how you view them already.** For example, if one of the characters is Justin Trudeau or Donald Trump, do your best not to let your pre-existing opinions about them influence the label you assign to that character. Either of these characters may be portrayed in a positive or a negative light depending on the source, which is what we seek to capture. Therefore for the purposes of this task, treat every

character as if you are reading about them for the first time (or only know basic things about them), and imagine the impressions the author wishes to impart on you about this character.

About the task:

- **Write 1 character label for each character listed in the CSV**
 - **The general question you are asking yourself is: what is this entity's "function" as depicted in this story? And how can I generalize that into a "type" that can reoccur in other news articles?**
 - As mentioned above, **try to ignore contextual knowledge outside the article. Only focus on how the article represents the character.**
- **Labels must all be nouns (no single adjectives, e.g. "Frightened")**
- Many labels vary in terms of specificity. E.g. "Innocent victim", "Hot-shot lawyer", "Cautious expert". To address this, we use a **two-column system to represent these labels**:

1) The primary label

- The primary label captures the general role of the character

2) An optional modifier to the role

- An adjective, or additional information, that helps make the primary label more descriptive

- Using this system:

- "Innocent Victim" -> "Victim", "Innocent"
- "Hot-shot Lawyer" -> "Lawyer", "Hot-shot"

	Primary Label	Modifier
Innocent Victim	Victim	Innocent
Hot-shot Lawyer	Lawyer	Hot-shot

- The primary role should ideally be a single word, but can also be a short phrase if it can't be split naturally because the primary label would be too general.
 - E.g. "Chosen One", "Love Interest", "Con Man", are all not split into two columns because "One", "Interest", "Man" would be too general.

	Primary Label	Modifier	
Chosen One	Chosen One		Correct
Chosen One	One	Chosen	Incorrect

- Use the word "person" when modifying a general entity with an adjective unless the gender matters.

- E.g. “Vulnerable person” (single column), “Elder woman” (single column)
- Multiple modifiers in the second column are allowed if you think it’s best. But use these sparingly and reflect on whether they are necessary:
 - E.g. “Grumpy Old Man” is a good label, but is best split as “Old man”, “Grumpy” than “Man”, “Grumpy Old”
 - “Corrupt Corporate Executive” -> “Corporate Executive”, “Corrupt” which is better than “Executive”, “Corrupt Corporate”
 - If you use multiple modifiers, do not include commas between them.
 - E.g. “Tall, Dark, and Handsome Man” -> “Handsome Man”, “Tall Dark”
- Remember the modifiers are optional, but you are encouraged to add them for greater specificity.
- No need to include “the” in your labels (“The scientist” -> “Scientist”, “The leader” -> “Leader”, etc).
- Don’t worry about capitalization.
- Write **only singular noun labels** for both singular *and* plural characters/entities. For instance, for character = Citizens of New Delhi, use a label like “Victim” and not “Victims”
- Remember that there is no one single right answer to character labelling and many roles may be applicable for any given character. **Here are some guiding thoughts about how to choose a good label.** Good labels:
 - A) *align factually* with the text; that is, the label is entailed (or implied) by the text.
 - B) Capture *important* content and *central* aspects of a character’s function in the story. An article may mention someone is a good parent, but is that the *main* focus of their role in the story? Or that just side information and there is a more salient dimension to their portrayal?
 - C) Are *descriptive* and *predictive*. They provide useful information about the character’s identity and help predict how they will behave in future situations. This feature is what discourages us from using professions as labels, since they are usually too generic to be predictive and descriptive.
 - D) Are *generalizable* to other characters in other stories (usually). As with “Innocent Victims” and “Victims” above, labels can vary in terms of generality and specificity, but most good labels are powerful because of their applicability to characters in other stories. Can you imagine your character role (not the same character) in a different story playing a similar role and would that role be informative of how they will behave?

- As a general note, you are welcome to apply any common character trope from movies/books to the characters in the articles if you think they apply (e.g. “Corrupt Business Exec”, “Underdog”, etc). In general, apply whichever role label you think best encapsulates the character, whether it is an existing trope or you come up with on your own.
- When the entities aren’t people (like companies or government entities) really try to personify them in your labels. What “character role” is that entity playing?
- Remember: No professions (unless you add a meaningful modifier)!
 - Bad labels: Executive, Manager, Minister, etc.
 - Good labels: Humble Executive, Crazy Manager, Rookie Minister

Some other notes:

- Some character names in the CSVs may look funny if the name contains accents in the text. There is no need to correct these in the CSV.
- Save your file as a CSV (not CSV UTF-8 nor CSV Mac nor CSV MS-DOS). Preferably not an Excel file (.xls) although these should be okay.
- The articles have been cleaned as best as possible, but still may contain random ads or asides (particularly near the end of some articles). If you see these, safely ignore them, but make sure the article doesn’t pick up again later in the text.
- You are not allowed to rely on GPT or other LLMs for synonyms or candidate roles.

Examples:

1) Characters: Elon Musk, OpenAI

Text: Elon Musk has unexpectedly asked a California court to withdraw a legal case against OpenAI and its boss Sam Altman, which accused them of abandoning the firm's founding mission of developing artificial intelligence (AI) for humanity's benefit.

The filing submitted by the multi-billionaire's lawyers asked for the months-old case to be dropped without offering any reason for the move.

It came just a day before the court was expected to hear the ChatGPT-developer's bid to have the case dismissed.

The latest filing asked for the case's dismissal "without prejudice", meaning Mr Musk could still reactivate it at a later stage.

The Tesla boss filed the lawsuit against OpenAI at the end of February this year, arguing the company he had helped found in 2015 had deviated from its altruistic goals to focus on making money.

OpenAI countered that Mr Musk had previously backed the idea of a for-profit structure and even suggested a merger with his electric car firm Tesla.

The feud intensified earlier this week after Apple unveiled a partnership with OpenAI to boost its Siri voice assistant and operating systems with OpenAI's ChatGPT chatbot.

After the announcement, Mr Musk posted several messages on his social media platform X, formerly known as Twitter, criticising the tie-up.

One of the posts ended with the words: "Apple has no clue what's actually going on once they hand your data over to OpenAI. They're selling you down the river."

However, investors seemed to welcome the news, as Apple's stock market value rose to a record high above \$3tn.

Mr Musk started his own AI company, called xAI, in July 2023, which he said would aim to "understand reality".

Labels

Character	Primary Label	Modifier
Elon Musk	Instigator	
OpenAI	Target	

Explanation:

- **Instigator** is applicable since the article focuses on how Elon has pursued and in some ways attacked OpenAI. Something like "Bitter Rival" may also be applicable here, and would be split as "Rival" as primary label and "Bitter" as modifier. **Tech Mogul** would also apply (and would be split as "Mogul", "Tech" to distinguish it from "Media Mogul" or "Business Mogul") from mentions that Elon Musk is the boss of Tesla, X, and xAI. The article portrays him clearly as someone influential and a key player in the tech world which is quite central to his depiction and could be used to predict future behaviour.
- **Target** is consistent with OpenAI being the intended recipient of Musk's lawsuits. It also is informative since it paints a picture of an entity that may be the subject of other attacks in the future. **Victim** might be a bit general, but pairs with Musk's Instigator label and is applicable across many different stories.

Some examples of Bad Labels

Elon Musk:

- "Tesla Boss" (true but not generalizable. This label only applies to Elon Musk and no other character in any other story or genre)
- "Businessman" (a profession which isn't very informative)
- "Conspiracy Theorist" (not entailed by the text, but a decent label in its own right)

OpenAI

- "AI Company" (similar to an occupation, not very informative)
- "Trailblazer" (not entailed or not central to the text, but a good label in its own right. It's informative and generalizable.)

- "Corrupt Corporation" (not entailed by the text, but a good label in its own right)

2) Characters: Kamala Harris

Text: A clip of Vice President Kamala Harris from 2017 calling on "everybody" to be "woke" resurfaced on social media, sparking critics and conservatives to lambaste the official Democratic nominee as a "Communist functionary" who is "too radical" to serve as president.

"We have to stay woke. Like everybody needs to be woke. And you can talk about if you're the wokeest or woker, but just stay more woke than less woke," then-Sen. Harris said in 2017 during Recode's annual Code Conference, alongside philanthropist Laurene Powell Jobs, the widow of Apple co-founder Steve Jobs, during the conference.

The conference focused on issues such as immigration and the environment, including Harris speaking out against the Trump administration's drug policies and immigration policies just moments ahead of her call on Americans to be "woke."

"I get upset when we say we're going to shut our door to refugees who are fleeing such atrocities and have been standing in line for two years to qualify for refugee status, and they arrive here, and we turn them away. Yes, I'm gonna get mad about that. I'm gonna get mad when we have an attorney general who's trying to resuscitate the war on drugs. And he thinks the greatest evil that mankind has ever seen is marijuana. ... Leave grandma's medical marijuana alone," Harris said.

Harris doubled down on her comment in a message posted to X a day later, which read: "We have to stay active. We have to stay woke."

The use of the word "woke" grew in popularity in 2017, when dictionaries such as Oxford and Merriam-Webster added the phrase to their lexicons, defining it as people who are attentive to issues such as social justice. It has since become a catch-all word for left-wing issues and agendas.

Seven years after Harris joined the 2017 conference, critics on social media are lambasting the presumptive Democratic nominee over the clip, including conservative author James Lindsay describing Harris as a "Communist functionary."

"Kamala is a low-level Communist functionary, and I'm tired of being told we have to pretend she's not," he wrote.

Labels

Character	Primary Label	Modifier
Kamala Harris	Radical	Dangerous

Explanation:

- **Dangerous Radical** applies given how Kamala Harris was introduced as a "Communist functionary" and as being "too radical". Although most of the article simply describes things Harris has done, the quote from James Lindsey cements a negative portrayal, and suggests she is something to be feared, which ties into the modifier "Dangerous". Note that "Radical" without the "Dangerous" modifier would also be acceptable.
- A label like "Extremist" may also capture that point. Another possible label may be "Threat", although it isn't as informative.

Some examples of Bad Labels

Kamala Harris:

- "Progressive" (true and consistent with the article, but it isn't salient since it doesn't capture the negative depiction by the author)
- "New Hope" (not entailed by the article)
- "Democratic nominee" (not generalizable)