**Procedure for Finding Errors in Table-To-Text Neural outputs**

**1. Overview of the Task**

In the Human Evaluation experiment, you will find a screenshot of the Input Data, which includes the **Main title, Section Title and Complete Table with highlighted cells in yellow.** These key parameters are conditionally used for training and/or prompting the neural models to summarize a meaningful and factual sentence (as Output) focusing on: (i). highlighted cells in yellow and (ii). their corresponding header values (iii) The main Title and (iv) The Section Title. For each of this Input table data, you will see Outputs generated by different neural language models.

Our goal is to evaluate whether the neural outputs remain faithful and produce factually accurate information based on the above four parameters from the Input Data. The complete table, including the non-highlighted cells, is provided to offer you a clearer understanding for the error annotation task.

Sections 2, 3 and 4 will provide a detailed overview of the evaluation task. The error annotation experiment is provided in a separate document.

**2. Domain Type**

The inputs for each example are specific to the domain of Politics, sourced from Wikipedia tables (as part of the open-domain Table-to-Text dataset). The political data within these tables isn't limited to a single demography. Instead, it encompasses various details from the election processes across multiple countries, including:

* Election specifics such as Presidential, state, by-elections, council, district, Legislative Assembly, and other elections unique to each country.
* Information about Governors, Mayors, Ministers, and Ambassadors (pertaining to Foreign Affairs).
* Details regarding the Speaker of the Assembly.

**3. Marking guidelines**

We’ve given you twelve different output sentences generated by “deep learning” AI systems. Apart from the screenshot of the Input Data, you will be given an additional Wikipedia reference link where the original Table was taken, it is not mandatory for you to refer this link for all the examples, but it might be helpful to get the background details for some of the vague table data only if required.

We are only interested if the highlighted cell values from the table were used to produce a factually correct sentence. Please also pay attention to the non-highlighted cells in all rows where the cells are highlighted, as this might be required in some inputs to generate a meaningful sentence**.** Other non-highlighted values in the table are not expected to be used for your evaluation. Please read through the output sentences and mark up cases for the error categories mentioned in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Category** | **Brief Description** | **Single Token(word) / Group of tokens(words) highlighting** |
| 1 | NAME | when names (e.g., Party, leader, electorate – mostly nouns) are wrong. | Single or multiple tokens (full name, two tokens for some places) |
| 2 | NUMBER | when numbers are incorrect (e.g., number of votes, % of votes). | Single |
| 3 | DATE DIMENSION | when Date and/or Month and/or Year are wrong. | Single or multiple tokens;  i. incorrect month in words such as “April”,  ii. incorrect year,  iii. error in date or in combination of DD/MM/YYYY |
| 4 | WORD | when incorrect words (verbs, adjectives, prepositions, adverbs, and conjunctions) are found. Exceptions for multiple tokens: *auxiliary verb (was), extension of prepositional phrase (e.g., along with)* can be taken as single WORD error. | Single token;  Multiple tokens for exceptions |
| 5 | CONTEXT | **Context with incorrect inference**: when the model's output presents information that contradicts, misrepresents, or make unsupported assumptions about the provided context (input data). It is detailed below in a separate section with examples. | Group of tokens |
| 6 | ADDITION | when the model’s outputs have added words, phrases, or details that either diverge from the input's main topic or are unsupported by the given context. | Group of tokens |
| 7 | OTHER | detailed below in a separate section. | Group of tokens |
| 8 | NON-ENGLISH | when the Unicode characters in the non-English names are wrong. | Single |

**Table 1: Error Categories**

In addition to the above table, the guidelines are detailed for each category in this section. You can then review a few annotated examples in section 4.

OMISSION:

When it fails to mention specific details present in the provided context, without having any mistakes in the generated output.

**3.1 NAME errors**

NAME errors occur where the name of the leaders, party, Electorate information (name of the district, council, state), Election details (Assembly instead of the Representative etc.,) are wrong. If a word (other than “I”) is always capitalized, it is probably a name and can be included in this type EXCEPT for the Month and Days of the Week defined in DATE\_DIMENSION error. For example:

* **Output:** Urban Ahlin is the Deputy Speaker of the Riksdag. Analysis: NAME error because the correct deputy speaker was Tobias Billström as per the table Input Data.
* Kansas was won by Mitt Romney, Paul Ryan, Barack Obama, and Joe Biden, with Romney winning 59.66% of the popular vote, six electoral votes and 38.05 percent. In this example, Barack Obama and Joe Biden are two NAME errors because they did not win the election.
* Wednesday instead of Tuesday is a NAME error but May instead of April is a DATE DISTINCTION error.

Note: Though week of the day is unlikely to be generated in our example Outputs. Please follow the guidelines as mentioned in Name errors.

**3.2 NUMBER errors**

NUMBER errors occur where the number of seats, number of votes and % of votes are incorrect. For example:

* When the A-party won with a majority of 5.5%. But the correct one is 4.4%. 5.5% is a NUMBER error.
* **Output:** The voter turnout was 8,90%, with 10,052 votes. **Analysis:** The actual turnout was 81.90%. Please note: the error here is NOT with the comma used as decimal (as it is an acceptable decimal operator for international use); Error because the number 81.90 was not correctly predicted.

**3.3 DATE DIMENSTION errors**

A Date Dimension error is when the Date and/or Month and/or Year in the Output is/are wrong. Omission of the Year or Month or Date is NOT included in this error type. For example:

* **Output:** Cletus Avoka was the Minister for the Interior in the Mills government from 2009 to 2012. **Analysis:** 2010 is the right end term of the year.

Note: If the Output did not capture Month and/or Date, but has the correct year, then this is NOT an ERROR. It could go to OMISSION with remarks Omission of Date and Month.

**3.4 WORD errors**

WORD errors occur when incorrect words such as verbs, prepositions, adjectives, and adverbs are found in the output. For example:

* **Output:** Carly Fiorina defeated Republican Tom Campbell with 56.4% of the vote to DeVore's 19.3%, along with Al Ramirez and Tim Kalemkarian. **Analysis**: Fiorina independently defeated all the leaders, so along with is wrong.
* **Output:** Ling won the 2016 senate district against Democrat Josh Newman, with 49.6 percent of the vote. **Analysis:** Ling lost the election as per the table data.
* Some of the common WORD errors found in this data are *won, defeated, lost, succeeded*, adjectives such as *current* governor other prepositions (since, in and so on).
  1. **Context errors**

CONTEXT errors occur when the model's output presents information that contradicts, misrepresents, or make unsupported assumptions about the provided context (input data).

* It can sometimes be tricky to check for this type of error. Please follow the below process before marking this error.
  + In case of simple contradictions/misrepresentation based on the information in the input data, it would be easier to mark the token as NAME, NUMBER, DATE\_DIMENSION or WORD error.
  + In the case of a complex tabular structure, the outputs are likely to mess up completely with the overall information in the provided input data. In this case, it is hard to mark individual errors. Please go ahead and mark the group of tokens/span of text and annotate it as a CONTEXT error. Below is an example for CONTEXT error.

**Input Data**:

A screenshot of a computer

Description automatically generated

**Output:** In the 2006 election for mayor of Florence Pendleton, Michael D. Brown received 62,415 votes while Philip Pannell received 21,552 votes and write-in candidates received 1,363 votes.

**Error annotation remarks:**

* for mayor of Florence Pendleton: CONTEXT error, the name of a person is misrepresented as electorate(jurisdiction) in the output.
* write-in candidates: CONTEXT error, implies there was more than one write-in candidate.

**3.6 ADDITION errors**

ADDITION errors occur when the model’s outputs have added words, phrases, or details that either diverge from the input's main topic or are unsupported by the given context. For example:

* **Input 1:**

A screenshot of a ballot

Description automatically generated

**Output 1:** In the 2014 United States House of Representatives election in New York's 24th congressional district, Dan Maffei was defeated by Ann Marie Buerkle.

**Analysis:** The model added the election information in the output that was not mentioned in the input table data (provided context). The group of texts in red are ADDITION errors.

* **Input 2:**

**A screenshot of a computer

Description automatically generated**

**Output 2:** The table shows the list of governors of Guam from 1971 to present, with Joseph F. Ada being the first elected governor in 1986 and re-elected in 1990.

**Analysis:** Though the complete table would reveal us, the word ‘first’ is incorrect, **but the input data passed to the model does not have this information.** Hence, it is an ADDITION error and not WORD error.

* **Output 3:** The president of Rwanda is Paul Kagame. The government of Rwanda is led by the president, who is elected by popular vote for a seven-year term. The current president is Paul Kagame, who was elected in 2000 and re-elected in 2003. The next election will be held in 2010.
* **Analysis:** The first sentence in the output is correct. All other follow-up content is not passed in the input data. So, it is an ADDITION error.

**3.7 OTHER errors**

OTHER errors occur for the following scenarios in the output

* OTHER-garbage details:
  + when the table data has party name in abbreviation, it tries to produce a strange output. For example: GSSSDULSVDHSS gained 5.31% of the vote
  + Also, when the output repeats the same input multiple times without producing a complete sentence.
* OTHER-INCOMPLETE when the sentence is incomplete for longer table input.
* OTHER-PUNCTUATION – when punctuation is placed at inappropriate places.

Use this category only if it does not fit in to CONTEXT error or ADDITION error.

**3.8 NON-ENGLISH errors**

Non-English errors occur when specific alphabets from European languages, Chinese and others have unique symbols that are missed in the Output text leading to inaccurate information. For example:

* Pawe Gra is a member of Sejm. Paweł Graś is the correct name.
* Edhem Biaki served as mayor of Sarajevo. Edhem Bičakčić is the right name.

**Note – optional to annotate omissions**

If there are no factual accuracy errors in the output text, mark it as “NO ERROR”. If the output does not have highlighted cells or title information from the output, it has omitted that information. In such cases, it is optional for you to mention omission of the details in remarks.

# **4. Procedure Followed for Error Analysis**

Please mark up the above-mentioned errors by putting them in red.  If you’re colour-blind, you can underline them instead. Please also add details of each error in the table below each example. The columns are:

* **Error:** The text you highlighted in red or underlined.
* **Category:** The type of error, choose from the nine categories mentioned above. The capitalization here does not matter, so you can write NAME, Name or even name.
* **Remarks:**  Your additional remark if the error is not straight-forward. The intention here is to record your observations and not to correct the sentence.
* **Model #:** This column will be already provided in the default template. You need to include details in ‘*error’*, ‘*category’* and ‘*remarks’* columns to each of the model number. If each model has multiple errors, please add rows, and update the Model # accordingly to fill the other three columns.

**Example 1**

**Input Data:**

**A screenshot of a computer

Description automatically generated**

**Neural Outputs:**

Below are the four outputs for your marking:

|  |  |
| --- | --- |
| **ID** | **Output Sentence** |
| M1 | The National Party's MPs are Amy Adams (Selwyn), Jacqui Dean (Waitaki) and Rangitata. |
| M2 | The National Party is represented in Rangitata by Andrew Falloon, Selwyn by Amy Adams and Waitaki by Jacqui Dean. |
| M3 | The electoral district of Rangitata is represented by Amy Adams (National Party) and Jacqui Dean (National Party). The electoral district of Selwyn is represented by Amy Adams (National Party), the electoral district of Waitaki is represented by Jacqui Dean (National Party). |
| M4 | The electoral district of Canterbury was represented by Andrew Falloon, Amy Adams, and Jacqui Dean. |

**Error Marking with Comment:**

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Errors** | **Type** | **Remarks** |
| M1 | Rangitata | NAME | Rangitata is not an MP. It is an electorate |
| M2 |  | NO ERROR |  |
| M3 | by Amy Adams (National Party) and Jacqui Dean (National Party) | CONTEXT | It changed the context in the first sentence by misrepresenting the given info. It should have been Andrew Falloon of National Party or just a National Party leader. |
| M4 |  | OMISSION | OMISSION of party (National party) |

**Example 2**

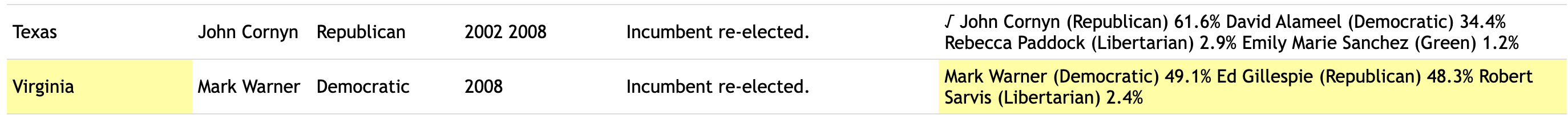
**Input Data:**

A close-up of a text

Description automatically generated

A close up of a text

Description automatically generated



**Neural Outputs:**

**Point to note:** In the M1 output, there are two potential points to mark for errors: either the name of the Libertarian party or the name of the leader. When faced with such choices, prioritize marking the first error token in the sequence. In this specific example, since the third-place finish was attributed to the Libertarian party, the error was marked on the leader's name span rather than the party name.

Below are the four outputs for your marking:

|  |  |
| --- | --- |
| **ID** | **Output Sentence** |
| M1 | Alaska was won by Republican Dan Sullivan, followed by Democrat Mark Begich and Libertarian Ted Gianoutsos, while in Virginia, Democratic Mark Warner defeated Republican Ed Gillespie. |
| M2 | Alaska was won by Dan Sullivan (R) with 48.0% of the vote, and Virginia was won by Mark Warner (D) with 49.1% of the vote. |
| M3 | The United States Senate elections in Alaska and Virginia were held on November 4, 2014. In Alaska, incumbent Democratic Senator Mark Begich lost to Republican challenger Dan Sullivan by a margin of 3.2%. In Virginia, incumbent Democratic Senator Mark Warner won re-election against Republican challenger Ed Gillespie by a margin of 1.8%. |
| M4 | 2014 United States Senate elections were held on November 4, 2014, with the primary election in most states taking place between March and June 2014. The general election coincided with the elections for the House of Representatives and many state and local elections. In Alaska, Dan Sullivan (Republican) won with 48% votes against Mark Begich (Democratic) who got 45.8%. In Virginia, Mark Warner (Democratic) won with 49.1% votes against Ed Gillespie (Republican) who got 48.3%. |

**Error Marking with Comment:**

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Errors** | **Type** | **Remarks** |
| M1 | Ted Gianoutsos | NAME | Mark Fish is the leader belong to Libertarian party. |
| M2 |  | OMISSION | Omission of other leaders and their votes |
| M3 | November 4 | ADDITION | Addition of DD and MM |
| M3 | 3.2% | NUMBER | Incorrect %margin; correct number is 2.8% |
| M3 | 1.8% | NUMBER | Incorrect %margin; correct number is 0.8% |
| M4 | November 4 | ADDITION | Addition of DD and MM |
| M4 | Other spans highlighted in red | ADDITION | Included additional details not provided in the input |

**Example 3**

**Input Data:**

A screenshot of a table

Description automatically generated

**Neural Outputs:**

Below are the twelve outputs for your marking:

|  |  |
| --- | --- |
| **ID** | **Output Sentence** |
| M1 | In 2018, Working Class Party candidate Juan Rey of California was elected to the United States House of Representatives in the nonpartisan blanket primary. |
| M2 | In the 2018 United States House of Representatives election, Working Class Party candidate Juan Rey of California received 944 votes (1.5%) in a nonpartisan primary. |
| M3 | Juan Rey won the election for United States House of Representatives in California with a total of 944 votes and 1.5% of the vote. He was elected as a member of the Working Class Party. |
| M4 | 2018 election results for the United States House of Representatives in California's 29th district showed that Juan Rey was elected as a nonpartisan blanket primary candidate. |

**Error Marking with Comment:**

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Errors** | **Type** | **Remarks** |
| M1 | was elected | WORD | contested would be the right word |
| M1 |  | OMISSION (low priority) | Omission of votes |
| M2 |  | Omission | No error but Omission of this information (29th district and San Fernando valley) |
| M3 | won | WORD | contested would be the right word |
| M3 | was elected | WORD | contested would be the right word |
| M4 | was elected | WORD | contested would be the right word |
| M4 |  | Omission | Omission of San Fernando valley |