Coding Manual: Toulmin Model Analysis of Student Arguments on AI

This manual provides guidelines for analyzing high school students' essays on Artificial Intelligence (AI) based on Stephen Toulmin's model of argumentation. The objective of the coding process is to identify argumentative elements, assess the quality of the arguments, and uncover underlying argumentation patterns.

The Coding Framework

The six elements of the Toulmin model in the context of AI:

- **Claim (C):** The student's main opinion or thesis about AI (e.g., AI is useful/dangerous/important).
- **Data (D):** Factual information or examples mentioned by the student.
- **Warrant (W):** The implicit or explicit link between the data and the claim.
- **Backing (B):** Mention of additional evidence, sources, or authorities.
- **Qualifier (Q):** Phrases indicating the certainty of the claim (e.g., probably, in some cases).
- **Rebuttal (R):** Mention of counter-arguments, limitations, or exceptions.

Example:

„I think artificial intelligence will be very useful in the future (C), because it is already used in many fields, such as medicine and transportation (D). These applications help people (W), as has been shown in several studies (B). However, it will probably (Q) also cause problems with job displacement (R).”

Coding:

- **Claim (C):** „artificial intelligence will be very useful in the future”
- **Data (D):** „it is already used in many fields, such as medicine and transportation”
- **Warrant (W):** „These applications help people”
- **Backing (B):** „as has been shown in several studies”
- **Qualifier (Q):** „probably”
- **Rebuttal (R):** „also cause problems with job displacement”

Complexity: High (3) - contains a rebuttal.
Explicitness: High (3) - six elements are explicitly present.

The Coding Rounds

The coding is conducted in four rounds:

1. **First Round:** Independent coding of the first 50 essays by two coders.
2. **Calibration:** The inter-rater reliability (Cohen's Kappa) is calculated, and discrepancies are discussed by the coders to align understanding.
3. **Second Round:** The entire sample is coded based on the calibrated guidelines.
4. **Third Round:** The reliability metric (Cohen's Kappa) is calculated for the full sample, disagreements are discussed, and conflicts are resolved where possible.
5. **Fourth Round:** A third coder resolves any remaining disagreements.

The Elements of the Toulmin Model and Their Coding

1. Claim (C)

- **Definition:** The central thesis statement formulated by the student regarding AI.
- **Coding:**
  - **Presence:** 1 = claim is present, 0 = no identifiable claim.
  - **Explicitness:** 1 = explicit claim, 0 = implicit claim.
- **Examples:**
  - *Explicit:* „Artificial intelligence is a useful tool for humanity.”
  - *Explicit:* „AI poses a threat to the job market.”

    - *Implicit:* „AI is appearing in more and more fields." (Implicit claim could be: the spread of AI is happening).

2. Data (D)

- **Definition:** Facts, examples, or personal experiences used by the student to support their claim.
- **Coding:**
    - **Presence:** 1 = data is present, 0 = no data.
    - **Explicitness:** 1 = explicit data, 0 = implicit data.
- **Examples:**
    - „ChatGPT can already write essays."
    - „Self-driving cars have driven millions of miles without an accident."
    - „The quality of machine translation has improved significantly in recent years."

3. Warrant (W)

- **Definition:** The bridge connecting the data to the claim; it explains how the data leads to the conclusion.
- **Coding:**
    - **Presence:** 1 = warrant is present, 0 = no warrant.
    - **Explicitness:** 1 = explicit warrant, 0 = implicit warrant.
- **Examples:**
    - „Since ChatGPT can imitate human text, we can assume that AI will soon replace some human jobs."
    - „The accident-free statistics of self-driving cars prove that AI can be safer than humans."

4. Backing (B)

- **Definition:** Additional evidence, authorities, or sources mentioned to support the warrant.
- **Coding:**
    - **Presence:** 1 = backing is present, 0 = no backing.
    - **Explicitness:** 1 = explicit backing, 0 = implicit backing.
- **Examples:**
    - „According to scientists, the development of AI is exponential."
    - „A study by Google showed that..."
    - „Our teacher also told us that..."

5. Qualifier (Q)

- **Definition:** Indicates the certainty or limitations of the claim; it shows how confident or conditional the statement is.
- **Coding:**
    - **Presence:** 1 = qualifier is present, 0 = no qualifier.
    - **Explicitness:** 1 = explicit qualifier, 0 = implicit qualifier.
- **Examples:**
    - „Probably," „presumably," „in some cases," „mostly."
    - „Not in all cases," „not necessarily."

6. Rebuttal (R)

- **Definition:** Acknowledgment of opposing viewpoints, exceptions, or limitations of the claim.
- **Coding:**
    - **Presence:** 1 = rebuttal is present, 0 = no rebuttal.
    - **Explicitness:** 1 = explicit rebuttal, 0 = implicit rebuttal.
- **Examples:**
    - „Although AI is useful, it also hides dangers."

- „Many people are afraid that AI will take their jobs, but..."
- „Undoubtedly, there are ethical concerns related to AI."

Assessing the Quality of Argumentation

IMPORTANT: Neither Argument Complexity nor Explicitness needs to be coded manually. These variables are automatically derived from the coding of the six base categories (C, D, W, B, Q, R) and their explicitness. This will be handled by the lead researcher (T.M.) during data preparation.

1. Argument Complexity

- **Low (1):** Only Claim and Data are present (C+D).
- **Medium (2):** Claim, Data, and either a Warrant, Backing, or Qualifier are present (C+D+W/B/Q).
- **High (3):** The argument contains a Rebuttal (R) AND at least one of the basic elements (C or D).
  - *Reasoning for the „AND" condition:* In the Toulmin model, a rebuttal indicates a more complex argument as it shows the student considers other perspectives. However, a rebuttal alone is not a high-level argument; the rest of the structure matters. The model's logic implies a rebuttal typically responds to an existing argumentative structure. If this base structure is missing, the rebuttal does not fully serve its function.

2. Argument Explicitness

- **Low (1):** One element is explicit OR only implicit elements are present.
- **Medium (2):** Two elements are explicit.
- **High (3):** Three or more elements are explicit.

Coding Procedure Steps

1. Read the entire essay once to get a general impression.
2. Identify and code the Claim (C).
3. Identify and code the explicitness of the Claim (C_explicit).
4. Identify and code the Data (D).
5. Identify and code the explicitness of the Data (D_explicit).
6. Identify and code the Warrant (W).
7. Identify and code the explicitness of the Warrant (W_explicit).
8. Identify and code the Backing (B).
9. Identify and code the explicitness of the Backing (B_explicit).
10. Identify and code the Qualifiers (Q).
11. Identify and code the explicitness of the Qualifiers (Q_explicit).
12. Identify and code the Rebuttals (R).
13. Identify and code the explicitness of the Rebuttals (R_explicit).

Difficult Cases and Guidelines

- **Implicit Elements:** If an element is not explicitly stated but can be clearly inferred from the text, it is coded as present but implicit (0).
- **Multiple Claims:** If there are multiple claims, code the main claim of the argument.
- **Contradictory Arguments:** If the essay contains a self-contradiction, code one instance but add a note (e.g., by coloring the cell). A final decision will be made during consultation between coders and the lead researcher.
- **Incomprehensible Essays:** If an essay is incomprehensible or irrelevant, mark it as Not Codable (NC).

Coding Sheet

The Excel sheet contains the following columns:

ID Essay C C_explicit D D_explicit W W_explicit B B_explicit Q Q_explicit R R_explicit