**Aggregate Metrics Across All Domains (30 Submissions)**

* **Total Correct Predictions:** 11/30
* **Total Wrong Predictions:** 19/30
* **False Negative Rate (AI → Human/Hybrid):** 10/10
* **False Positive Rate (Human → AI/Hybrid):** 3/10
* **Hybrid Misclassification Rate:** 7/10

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Domain** | **Submissions** | **Correct Predictions** | **Wrong Predictions** | **Key Observations** |
| **Psychology** | 6 | 5 | 1 | AI submissions mostly detected correctly. Human & Hybrid texts largely identified accurately. Subtle hybrid text caused one misclassification. |
| **Teaching** | 6 | 2 | 4 | Strong false negative bias: AI submissions misclassified as Human. Hybrid texts collapsed into Human. Human texts mostly correct. |
| **Engineering** | 6 | 2 | 4 | AI texts misclassified as Human. Hybrid submissions failed detection. Detector struggles with technical, structured academic style. |
| **IT** | 6 | 1 | 5 | Both AI submissions misclassified as Human. Hybrid texts collapsed into Human. Mixed handling of Human submissions. Strong false negative bias. |
| **Accounting** | 6 | 1 | 5 | Detector struggled across AI, Human, and Hybrid. Formal, technical writing style made AI signals harder to detect. |

**Cross-Domain Observations**

1. **Strong False Negative Bias:**
   * AI submissions across all domains (Teaching, Engineering, IT, Accounting) are frequently misclassified as Human.
   * Polished, structured, formal academic writing confuses the detector.
2. **Hybrid Text Challenges:**
   * Hybrid submissions are often misclassified, either collapsed into Human or AI.
   * Detector struggles to balance subtle AI assistance with human-written content.
3. **Domain Sensitivity:**
   * Psychology domain performed best. Likely due to narrative style, use of examples, and less formal/technical language.
   * Technical/structured domains (Accounting, IT, Engineering, Teaching) show poor detection.
4. **Human Text Detection:**
   * Generally reliable in Psychology and Teaching; less so in Accounting and IT due to over-polished, academic writing mimicking AI style.
5. **Pattern Recognition Issue:**
   * Detector relies heavily on structural, formal, and repetitive cues.
   * Subtlety in AI use (Hybrid) or highly formal student writing often leads to misclassification.

**Conclusion**

* Current AI detection tool performs **well on narrative, less formal domains** (e.g., Psychology).
* Performance **drops sharply in technical, structured, or formal domains** (Accounting, IT, Engineering, Teaching).
* Major weaknesses:
  1. Fails to detect AI in highly polished academic text (False Negatives).
  2. Struggles with subtle Hybrid texts (misclassifications).
  3. Collapses nuanced human writing into Hybrid/AI incorrectly.

**Recommendation:** For multi-domain detection, the model needs **domain-adaptive calibration** and better differentiation between human polish vs. AI polish, especially in technical/academic writing.