Figure 1: The Distribution of Rubrics Targeting Each Type of Model Capabilities

Categorization of Target Capabilities of Rubrics

Target Capability Type 1
Analyzing causes, predicting consequences, and explaining rationale

Example:
Prompt: A 62-year-old man presents with an inability to tan and mild fatigue. Laboratory tests indicate mild anemia with hemoglobin at 11 g/dL, decreased haptoglobin, and 3% reticulocytes. Given these findings and the peripheral blood smear results suggesting hemolytic anemia, what specific patient history or characteristic is most consistent with the mechanical destruction of red blood cells as the underlying cause?

Rubric: The response explains that decreased haptoglobin is due to its binding to free hemoglobin released during red blood cell lysis.

Target Model Capability: Connecting specific observations or data points to their corresponding causes, effects, or implications within a larger system.

Target Capability Type 2
Systematic analysis and comparative evaluation for risk assessment and justified decision-making in critical situations.

Example:
Prompt: A patient presents in a stuporous condition, with a history of persistent sadness, suicidal attempts, and a lack of eating and sleeping according to the patient's parents. What is the recommended treatment in this case?

Rubric: The response recommends inpatient hospitalization for safety, evaluation, and stabilization (not outpatient or home-based care).

Target Model Capability: Identifying critical conditions requiring immediate and

decisive intervention based on a severity assessment.

Target Capability Type 3:
Developing comprehensive, actionable, and risk-aware strategic plans.

Example:
Prompt: What is the ideal management for a 5-year-old boy with retinoblastoma involving the entire right eyeball and 2-3 small lesions in the periphery of the other eye?

Rubric: The response specifies the timing of prosthetic eye fitting after enucleation (e.g., 4-6 weeks post-surgery).

Target Capability: Detailing a post-intervention protocol by specifying follow-on diagnostic assessments and subsequent restorative procedures with timelines.

Target Capability Type 4:
Structured and logically clear presentation.

Example:
Prompt: What is the most likely diagnosis for a 3-year-old child who presents with eczematous dermatitis on extensor surfaces and has a mother with a history of bronchial asthma?

Rubric: The response is organized logically, with clear separation between diagnosis, reasoning, differential diagnoses, and any disclaimers.

Target Capability: Structuring the response logically by partitioning distinct conceptual elements into clearly delineated sections.

Target Capability Type 5:
Analyzing components, relationships, and functions within complex systems.

Example:
Prompt: What is the most probable diagnosis for a 6-year-old boy who has been experiencing headaches and peripheral vision loss for four months, with a CT scan showing a suprasellar mass with calcification?

Rubric: The response links peripheral vision loss to compression of the optic chiasm by the mass.

Target Capability: Justifying a conclusion by explicitly linking individual pieces of evidence to their respective supporting roles in the final determination.

Target Capability Type 6:
Executing sequential calculations and quantitative procedures involving formulas, unit conversions, and operational precision.

Example:
Prompt: A 1-year-old child weighing 6 kg is suffering from Acute Gastroenteritis, showing

signs of sunken eyes and a skin pinch test indicating rapid fluid replenishment. Based on these symptoms, what volume and rate of Ringer's Lactate infusion would you administer over the first six hours?

Rubric: The response provides the correct infusion rates for each phase: 180 mL/hour for the first hour and 84 mL/hour for the next five hours.

Target Capability: Executing a precise, multi-step quantitative calculation based on given inputs and established formulas to derive a phased implementation plan.

Target Capability Type 7:
Resolving inconsistencies, ambiguities, and anomalies in data.

Example:
Prompt: A patient with a head injury is admitted to the intensive care unit showing signs of raised intracranial pressure. He is placed on a ventilator and given intravenous fluids and diuretics. After 24 hours, the patient's urine output is 3.5 liters, serum sodium level is 156 mEq/l, and urine osmolality is 316 mOsm/kg. What is the most likely cause of these clinical findings?

Rubric: The response identifies the urine osmolality of 316 mOsm/kg as inappropriately low for the degree of hypernatremia (i.e., urine should be more concentrated in this context).

Target Capability: Evaluating the relationship between multiple variables to identify paradoxical or inconsistent patterns relative to expected system behavior.

Target Capability Type 8:
Verifying factual accuracy, correcting misinformation, and attributing information to authoritative sources.

Example:
Prompt: According to the latest resuscitation guidelines, for how long must umbilical cord clamping be delayed in preterm infants?

Rubric: The response identifies at least one authoritative organization issuing the guideline (e.g., AHA, ILCOR, ACOG, WHO, ERC, NRP).

Target Capability: Attribute the factual knowledge to the authoritative sources.

Target Capability Type 9:
Synthesizing, analyzing, and contextualizing evidence to derive well-supported conclusions.

Example:
Prompt: A 33-year-old woman is brought to the emergency department 15 minutes after being stabbed in the chest with a screwdriver. Given her vital signs of pulse 110/min, respirations 22/min, and blood pressure 90/65 mm Hg, along with the presence of a 5-cm deep stab wound at the upper border of the 8th rib in the left midaxillary line, which anatomical structure in her chest is most likely to be injured?

Rubric: The response provides a clear, logical synthesis connecting wound location, anatomical relationships, and clinical findings to justify the conclusion.

Target Capability: Synthesizing multiple distinct lines of evidence into a coherent, logical argument to justify a final conclusion.

Target Capability Type 10:
Analyzing, defining, and differentiating abstract concepts and ideas.

Example:
Prompt: A 68-year-old woman with elevated serum calcium, high parathyroid hormone, low phosphorus, and a history of kidney stones presents with fatigue, constipation, diffuse bone pain, and a 24-hour urine calcium level that is elevated. Given these clinical and laboratory findings, what radiologic finding on a hand X-ray would confirm the suspected diagnosis of this patient's condition?

Rubric: The response accurately distinguishes between primary and secondary hyperparathyroidism if mentioned.

Target Capability: Differentiating between closely related sub-categories of a primary concept based on their defining features.

Target Capability Type 11:
Defining the precise rules, scope, and limitations for a given context.

Example:
Prompt: You are called to evaluate a newborn who was born yesterday to a 39-year-old mother. Upon examination, what chromosomal abnormality is most likely responsible for the observations typically associated with Down syndrome?

Rubric: The response recommends or references karyotype analysis or equivalent genetic testing as the definitive diagnostic method for confirming the chromosomal abnormality.

Target Capability: Specifying the definitive method or standard procedure required for confirmation or validation.

Target Capability Type 12:
Communicating with clarity, adapting to audiences, and qualifying statements to reflect uncertainty.

Example:
Prompt: Which complication during pregnancy is least likely to increase the risk of postpartum uterine atonicity and why?

Rubric: The response uses cautious and appropriate phrasing (e.g., 'least likely,' 'low association') rather than making absolute claims of zero risk.

Target Capability: Calibrating language precisely to reflect nuances and uncertainty, avoiding absolute or overly definitive statements.

Target Capability Type 13:
Applying safety disclaimers, deferring to experts, and adhering to operational constraints.

Example:
Prompt: Considering the patient's history and current presentation of sudden right arm weakness, numbness, facial drooping, and slurred speech, what is the strongest predisposing factor contributing to his condition?

Rubric: The response avoids providing direct medical advice and, if appropriate, includes a disclaimer to seek immediate professional medical attention for stroke symptoms.

Target Capability: Adhering to predefined safety protocols or operational constraints by including appropriate disclaimers.

Target Capability Type 14:
Evaluating data significance through contextual and quantitative analysis.

Example:
Prompt: A 6-year-old boy presents with headache, cough, runny nose, and low-grade fever after being treated for a urinary tract infection with trimethoprim-sulfamethoxazole. He has a leukocyte count of 2,700/mm3 with a differential predominantly showing lymphocytes. What is the most likely underlying cause of his current symptoms?

Rubric: The response correctly interprets a leukocyte count of 2,700/mm3 as leukopenia for a 6-year-old child (normal range 5,000/mm3 - 15,000/mm3).

Target Capability: Accurately interpreting a quantitative data point by comparing it against a reference range to determine its significance.

Target Capability Type 15:
Systematic information filtering to maintain relevance and focus.

Example:
Prompt: A labourer involved with repair work of sewers presents with fever, jaundice, and renal failure. What is the most appropriate test to diagnose the suspected infection in this patient?

Rubric: The response is concise and focused on the diagnostic aspect, without excessive unrelated clinical management details.

Target Capability: Adhering strictly to the defined scope of a problem by excluding extraneous or irrelevant information.

Target Capability Type 16:
Systematic procedures and methodologies for formulating, testing, and validating hypotheses through investigation.

Example:
Prompt: Given an X-ray of a young man that shows heterotopic calcification around bilateral knee joints, what would be the next investigation to help diagnose the underlying condition?

Rubric: The response recommends creatine kinase (CK) testing if myositis or muscle involvement is considered in the differential.

Target Capability: Proposing specific, targeted investigative actions to differentiate between hypotheses or gather further evidence.

Target Capability Type 17:
Retrieving specific facts.

Example:
Prompt: Which nerves are associated with difficulty swallowing despite normal musculature function, and should be tested for their functionality?

Rubric: The response identifies the Facial nerve (Cranial Nerve VII) as relevant to the oral phase of swallowing (e.g., facial muscles, buccinator, taste, or saliva production).

Target Capability: Selecting specific entities from its internal knowledge that directly satisfy a given set of complex conditions.

Target Capability Type 18:
Situating information and events in historical and temporal context.

Example:
Prompt: A 10-year-old patient presents with tingling and numbness in the ulnar side of the finger. Four years ago, the patient sustained an elbow injury. Based on the symptoms and history, identify the fracture site that most likely occurred at the time of the initial accident.

Rubric: The response explicitly states or clearly implies that the patient's symptoms are a delayed complication (i.e., tardy onset) following the initial elbow injury.

Target Capability: Identifying and explicitly stating the temporal relationship between a past event and a current observation.

Target Capability Type 19:
Adherence to explicit constraints on output structure, format, and content.

Example:
Prompt: A 29-year-old pregnant woman at 10 weeks' gestation is experiencing progressively worsening nausea and vomiting, leading to a significant weight loss and affecting her ability to work. Despite taking ginger and vitamin B6, her symptoms persist. Her blood gas analysis indicates a pH of 7.43, pCO2 of 54 mmHg, and HCO3- of 31 mEq/L. What pharmacological intervention should be added to her treatment regimen to alleviate her symptoms?

Rubric: The response provides a clear, direct, and unambiguous recommendation for the next pharmacological agent to add.

Target Capability: Formulating a direct and unambiguous conclusion or recommendation that resolves the primary question.

Target Capability Type 20:

Justifying conclusions by systematically eliminating alternative possibilities based on evidence.

Example:
Prompt: What is the best intervention for hearing rehabilitation in a patient who has undergone surgery for bilateral acoustic neuroma?

Rubric: The response states that conventional hearing aids and CROS/BiCROS devices are not effective for profound bilateral sensorineural hearing loss due to bilateral cochlear nerve loss.

Target Capability: Invalidating alternative solutions by providing a causal explanation for their ineffectiveness under given constraints.