Algae Growth

Model: General Rubric		
Criteria	Scoring	Max Points
Output Integrity		5
Names		5 (min 0)
Flows	Cloud -> Algae	1
Variables	Initial Population Carrying Cap Coeff	3
Positioning		4
Specific Model Rubric		
Initial Conditions	2pts - Initial Population 2pts - Carrying Cap 2pts - Coeff	6
Relationships	Variables may be renamed if model does not run (penalize in general rubric) If model still does not run, -5pts per misc. necessary element change for model to run 2pts - Initial Pop 2pts - Does not Exceed Carrying Cap 2pts - Carrying Capacity Affects Pop 2pts - Coeff Affects Pop 12pts - Numerical Correctness	20
Overall		44

Model: General Rubric	
Comments	

PROMPT: Create a model to simulate the growth of an algae colony using a logistic growth curve. Add a carrying capacity, initial population, and a coefficient of growth.

Model: ChatGPT-4o		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	1	1
Variables	3	3
Positioning	2	4
Specific Model Rubric		
Initial Conditions	6/6 2pts - Initial Population 2pts - Carrying Cap 2pts - Coeff	6
Relationships	15/20 -5pt - “initialPopulation” instead of “[initialPopulation]” 2pts - Initial Pop 2pts - Does not Exceed Carrying Cap 2pts - Carrying Capacity Affects Pop 2pts - Coeff Affects Pop 12pts - Numerical Correctness	20
Overall	37	44
Comments	Pretty Decent, a bit messier than o3 but got the main parts down	

Model: Claude 3		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	1	1
Variables	3	3
Positioning	4	4
Specific Model Rubric		
Initial Conditions	6/6 2pts - Initial Population 2pts - Carrying Cap 2pts - Coeff	6
Relationships	20/20 2pts - Initial Pop 2pts - Does not Exceed Carrying Cap 2pts - Carrying Capacity Affects Pop 2pts - Coeff Affects Pop 12pts - Numerical Correctness	20
Overall	44	44
Comments	Basically perfect, matched answer key	

Model: ChatGPT-o3-mini		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	1	1
Variables	3	3
Positioning	4	4
Specific Model Rubric		
Initial Conditions	6/6 2pts - Initial Population 2pts - Carrying Cap 2pts - Coeff	6
Relationships	20/20 2pts - Initial Pop 2pts - Does not Exceed Carrying Cap 2pts - Carrying Capacity Affects Pop 2pts - Coeff Affects Pop 12pts - Numerical Correctness	20
Overall	44	44
Comments	Basically perfect, matched answer key	

Model: DeepSeek-R1		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	1	1
Variables	2 -1pt - Init Population set in Stock	3
Positioning	4	4
Specific Model Rubric		
Initial Conditions	6/6 2pts - Initial Population 2pts - Carrying Cap 2pts - Coeff	6
Relationships	20/20 2pts - Initial Pop 2pts - Does not Exceed Carrying Cap 2pts - Carrying Capacity Affects Pop 2pts - Coeff Affects Pop 12pts - Numerical Correctness	20
Overall	43	44
Comments	Almost perfect	

Trebuchet

Model: General Rubric		
Criteria	Scoring	Max Points
Output Integrity		5
Names		5 (min 0)
Flows	Cloud -> Beam Omega Cloud -> Beam Theta	2
Variables	PA Length PA Mass P Mass CWA Length CWA Mass CW Mass Gravity Initial Angle Launch Speed Launch Angle	10
Positioning		4
Specific Model Rubric		
Initial Conditions	2pts - PA Length 2pts - PA Mass 2pts - P Mass 2pts - CWA Length 2pts - CWA Mass 2pts - CW Mass 2pts - Gravity 2pts - Initial Angle	16
Relationships	Variables may be renamed if model does	58

Model: General Rubric		
	not run (penalize in general rubric) If model still does not run, -5pts per misc. necessary element change for model to run 2pts - Initial Angle 2pts - Initial Speed 10pts - Torque 10pts - Inertia 5pts - Angular Acc 5pts - Angular Vel 4pts - Launch Speed 4pts - Launch Angle 16pts - Numerical Correctness	
Overall		100
Comments		

PROMPT: Create a model that simulates the movement of a trebuchet. The arm of the trebuchet can be simulated by a line segment that rotates around a fixed point. Initial variables include the length and mass of the portion of the trebuchet arm with the projectile and the length and mass of the portion of the trebuchet arm with the counterweight. The mass of the projectile and the counterweight are also initial variables. Finally, include the starting angle of the trebuchet as a variable. Other constants such as gravity should also be stored as variables.

The output stocks/variables for the simulation should be: beam angular speed & angular acceleration, the launch velocity (speed & angle components) of the projectile at any given moment. Make an appropriate element for each of these.

Any other helper nodes or elements can be created if necessary.

Model: ChatGPT-4o		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	0 (-0.75x2 Incorrect Flow Origins) (-1x2 Incorrect Flows for Launch Speed/Angle)	2
Variables	8 (-1x2 Stock for Launch Speed, Launch Angle)	10
Positioning	2	4
Specific Model Rubric		
Initial Conditions	16/16 2pts - PA Length 2pts - PA Mass 2pts - P Mass 2pts - CWA Length 2pts - CWA Mass 2pts - CW Mass 2pts - Gravity 2pts - Initial Angle	16

Model: ChatGPT-4o		
Relationships	16/58 0pts - Initial Angle 2pts - Initial Speed 3pts - Torque 7pts - Inertia 2pts - Angular Acc 0pts - Angular Vel 2pts - Launch Speed 0pts - Launch Angle 0pts - Numerical Correctness	58
Overall	58	100
Comments	Failed to recognize the difference between when to use a stock or variable for output. No angle stock. Incorrect flow origins, no clouds. Treated trebuchet arm as a point mass rather than a rotating bar for inertia. Did not incorporate angle or trebuchet arm into torque. Failed to establish a relationship between angular acceleration, angular speed, and angular position. Failed to differentiate between project launch angle/speed and beam angular speed.	

Model: Claude 3		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	2	2
Variables	10	10
Positioning	4	4
Specific Model Rubric		
Initial Conditions	16/16 2pts - PA Length 2pts - PA Mass 2pts - P Mass 2pts - CWA Length 2pts - CWA Mass 2pts - CW Mass 2pts - Gravity 2pts - Initial Angle	16
Relationships	44/58 2pts - Initial Angle 2pts - Initial Speed 5pts - Torque 5pts - Inertia 5pts - Angular Acc 5pts - Angular Vel 4pts - Launch	58

Model: Claude 3		
	Speed 4pts - Launch Angle 12pts - Numerical Correctness	
Overall	86	100
Comments	Treated trebuchet arm as a point mass at the same location as the projectile/counterweight rather than a rotating bar for inertia. Did not incorporate trebuchet arm into torque. Very close to the answer key.	

Model: ChatGPT-o3-mini		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	2	2
Variables	10	10
Positioning	4	4
Specific Model Rubric		
Initial Conditions	16/16 2pts - PA Length 2pts - PA Mass 2pts - P Mass 2pts - CWA Length 2pts - CWA Mass 2pts - CW Mass 2pts - Gravity 2pts - Initial Angle	16
Relationships	38/58 2pts - Initial Angle 2pts - Initial Speed 3pts - Torque 5pts - Inertia 5pts - Angular Acc 5pts - Angular Vel	58

Model: ChatGPT-o3-mini		
	4pts - Launch Speed 4pts - Launch Angle 8pts - Numerical Correctness	
Overall	80	100
Comments	Very impressive performance with logical element placement. The model failed to incorporate the weight of the trebuchet arm into either torque or inertia, causing inaccuracies in the final output model. The model also used Math.sin instead of the correct Math.cos in torque calculations. However, there is a significant similarity between the model produced by the AI and the answer key model.	

Model: DeepSeek-r1		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	2	2
Variables	7 (-1 PA Mass) (-1 CWA Mass) (-1 Launch Angle)	10
Positioning	2	4
Specific Model Rubric		
Initial Conditions	12/16 2pts - PA Length 0pts - PA Mass 2pts - P Mass 2pts - CWA Length 0pts - CWA Mass 2pts - CW Mass 2pts - Gravity 2pts - Initial Angle	16
Relationships	34/58 2pts - Initial Angle 2pts - Initial Speed 3pts - Torque 5pts - Inertia 5pts - Angular Acc 5pts - Angular Vel 4pts - Launch Speed 0pts - Launch Angle 8pts - Numerical	58

Model: DeepSeek-r1		
	Correctness	
Overall	67	100
Comments	Failed to incorporate the mass of the trebuchet arm into any component of the model. Treated the trebuchet as a simple “two-point” system and ignored the trebuchet arm itself. Failed to create the requested launch angle output variable.	

Binary Star System

Model: General Rubric		
Criteria	Scoring	Max Points
Output Integrity		5
Names		5 (min 0)
Flows	Cloud -> Vx1 Cloud -> Vy1 Cloud -> X1 Cloud -> Y1 Cloud -> Vx2 Cloud -> Vy2 Cloud -> X2 Cloud -> Y2	8
Variables	Mass1 Mass2 Fgx1 Fgy1 Fgx2 Fgy2 Ax1 Ay1 Ax2 Ay2 dX dY G	12
Positioning		4
Specific Model Rubric		
Initial Conditions	2pts - Star 1 Init Pos 2pts - Star 1 Init Vel 2pts - Star 1 Mass 2pts - Star 2 Init Pos	12

Model: General Rubric		
	2pts - Star 2 Init Vel 2pts - Star 2 Mass	
Relationships	Variables may be renamed if model does not run (penalize in general rubric) If model still does not run, -5pts per misc. necessary element change for model to run 2pts - Initial Positions 2pts - Initial Speeds 2pts - Initial Accel. 8pts - Gravity (1 & 2) 8pts - Accel (1 & 2) 4pts - Vel (1 & 2) 4pts - Pos (1 & 2) 20pts - Numerical Correctness	50
Overall		96
Comments		

PROMPT: Create a model that simulates a binary star system in space. Initial variables should specify the masses of each star. The starting x and y-positions and starting x and y-velocities of each star can be hardcoded into the initial values of the relevant stocks. The two stars should move and orbit around each other, and the only force acting on either star should be the force of each star's gravity on the other. Any other constants should be stored in variables.

Create intermediate variables storing the force of gravity on each star (x-y components), the acceleration of each star (x-y components), and the distance between the two stars (overall & x-y components) at any given time.

Model: ChatGPT-4o		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	2 (-0.75x8 flows incorrectly linked to stocks instead of clouds)	8
Variables	12	12
Positioning	2	4
Specific Model Rubric		
Initial Conditions	12/12 2pts - Star 1 Init Pos 2pts - Star 1 Init Vel 2pts - Star 1 Mass 2pts - Star 2 Init Pos 2pts - Star 2 Init Vel 2pts - Star 2 Mass	12
Relationships	17/50 -5pts - Division by Zero 2pts - Initial Positions 2pts - Initial Speeds 2pts - Initial Accel.	50

Model: ChatGPT-4o		
	8pts - Gravity (1 & 2) 8pts - Accel (1 & 2) 0pts - Vel (1 & 2) 0pts - Pos (1 & 2) 0pts - Numerical Correctness	
Overall	55	96
Comments	Equations seem correct but the model fails to understand how flow relationships connect related elements or how the variables should interact with each other. Understands the principles behind a BSS but fails to correctly integrate them into a new scenario.	

Model: Claude 3		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	8	8
Variables	12	12
Positioning	4	4
Specific Model Rubric		
Initial Conditions	12/12 2pts - Star 1 Init Pos 2pts - Star 1 Init Vel 2pts - Star 1 Mass 2pts - Star 2 Init Pos 2pts - Star 2 Init Vel 2pts - Star 2 Mass	12
Relationships	50/50 2pts - Initial Positions 2pts - Initial Speeds 2pts - Initial Accel. 8pts - Gravity (1 & 2) 8pts - Accel (1 & 2) 4pts - Vel (1 & 2) 4pts - Pos (1 & 2)	50

Model: Claude 3		
	20pts - Numerical Correctness	
Overall	96	96
Comments	Full marks, understands the relationships between elements both in the equation editor and spatially, outputted all requested variables	

Model: ChatGPT-o3-mini		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	8	8
Variables	10 (-1 Fgx2) (-1 Fgy2)	12
Positioning	2	4
Specific Model Rubric		
Initial Conditions	12/12 2pts - Star 1 Init Pos 2pts - Star 1 Init Vel 2pts - Star 1 Mass 2pts - Star 2 Init Pos 2pts - Star 2 Init Vel 2pts - Star 2 Mass	12
Relationships	50/50 2pts - Initial Positions 2pts - Initial Speeds 2pts - Initial Accel. 8pts - Gravity (1 & 2) 8pts - Accel (1 & 2) 4pts - Vel (1 &	50

Model: ChatGPT-o3-mini		
	2) 4pts - Pos (1 & 2) 20pts - Numerical Correctness	
Overall	92	96
Comments	Output exactly matches the answer key, very impressive. Struggles with positioning of elements graphically. Failed to create the requested output variables for Fg2.	

Model: DeepSeek-r1		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5 (min 0)
Flows	8	8
Variables	10 (-1 Fgx2) (-1 Fgy2)	12
Positioning	4	4
Specific Model Rubric		
Initial Conditions	12/12 2pts - Star 1 Init Pos 2pts - Star 1 Init Vel 2pts - Star 1 Mass 2pts - Star 2 Init Pos 2pts - Star 2 Init Vel 2pts - Star 2 Mass	12
Relationships	50/50 2pts - Initial Positions 2pts - Initial Speeds 2pts - Initial Accel. 8pts - Gravity (1 & 2) 8pts - Accel (1 & 2) 4pts - Vel (1 & 2) 4pts - Pos (1 & 2) 20pts - Numerical Correctness	50
Overall	94/96	96
Comments	Very good model, matches output model exactly, good spatial reasoning when placing elements. Failed to create the requested output variables for Fg2.	

Projectile Motion

Model: General Rubric		
Criteria	Scoring	Max Points
Output Integrity		5
Names		5 (min 0)
Flows	Cloud -> xVel Cloud -> xPos Cloud -> yVel Cloud -> yPos	4
Variables Initial conditions and other constants are properly expressed as variables	initX initY initVel initAngle gravity mass dragCoeff	7
Positioning Elements are appropriately placed		4
Specific Model Rubric		
Initial Conditions	2pts - xPos 2pts - yPos 2pts - Initial Speed 2pts - Initial Angle 2pts - Drag Coeff 2pts - Mass 2pts - Gravity	14
Relationships	Variables may be renamed if model does not run (penalize in general rubric)	30

Model: General Rubric		
	If model still does not run, -5pts per misc. necessary element change for model to run 2pts - Initial Pos 2pts - Initial Vel. 2pts - Correct Gravity 4pts - Acceleration to Velocity 4pts - Velocity to Position 4pts - Drag Coefficient affects Velocity 12pts - Numerical Correctness	
Overall		69
Comments		

PROMPT: Create a model for 2D projectile motion. Initial variables should be starting position, mass, and angle. Incorporate a drag coefficient that affects acceleration proportional to velocity.

Model: ChatGPT-4o		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	1 (-0.75*4, flows drawn from stock)	4
Variables	5 (-0.5*2 starting pos) (-1 start vel)	7
Positioning	2	4
Specific Model Rubric		
Initial Conditions	8/14 0pts - xPos 0pts - yPos 0pts - Initial Speed 2pts - Initial Angle 2pts - Drag Coeff 2pts - Mass 2pts - Gravity	14
Relationships	6/30 0pts - Initial Pos 0pts - Initial Vel. 2pts - Correct Gravity 0pts - Acceleration to Velocity	30

Model: ChatGPT-4o		
	0pts - Velocity to Position 4pts - Drag Coefficient affects Velocity 0pts - Numerical Correctness	
Overall	32	69
Comments	Failed to split initial conditions into x-y components Flows were drawn from stocks (incorrect) instead of creating a cloud source element Correct equation but incorrect flow origin	

Model: Claude 3		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	4	4
Variables	7	7
Positioning	4	4
Specific Model Rubric		
Initial Conditions	14/14 2pts - xPos 2pts - yPos 2pts - Initial Speed 2pts - Initial Angle 2pts - Drag Coeff 2pts - Mass 2pts - Gravity	14
Relationships	14/30 2pts - Initial Pos 2pts - Initial Vel. 2pts - Correct Gravity 4pts - Acceleration to Velocity 4pts - Velocity to Position 0pts - Drag Coefficient affects Velocity 0pts - Numerical	30

Model: Claude 3		
	Correctness	
Overall	53	69
Comments	Notably, used angles in degrees and converted to RAD for flow/stock equations. Drag coefficient equation was incorrect which led to incorrect numerical results. However, excellent model structure.	

Model: ChatGPT-o3-mini		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	4	4
Variables	7	7
Positioning	2	4
Specific Model Rubric		
Initial Conditions	14/14 2pts - xPos 2pts - yPos 2pts - Initial Speed 2pts - Initial Angle 2pts - Drag Coeff 2pts - Mass 2pts - Gravity	14
Relationships	30/30 2pts - Initial Pos 2pts - Initial Vel. 2pts - Correct Gravity 4pts - Acceleration to Velocity 4pts - Velocity to Position 4pts - Drag Coefficient affects Velocity 12pts - Numerical	30

Model: ChatGPT-o3-mini		
	Correctness	
Overall	67	69
Comments	All equations and relationships perfect, spacing of elements could be better however	

Model: DeepSeek-r1		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	4	4
Variables	7	7
Positioning	4	4
Specific Model Rubric		
Initial Conditions	14/14 2pts - xPos 2pts - yPos 2pts - Initial Speed 2pts - Initial Angle 2pts - Drag Coeff 2pts - Mass 2pts - Gravity	14
Relationships	2pts - Initial Pos 2pts - Initial Vel. 2pts - Correct Gravity 4pts - Acceleration to Velocity 4pts - Velocity to Position 4pts - Drag Coefficient affects Velocity 12pts - Numerical Correctness	30
Overall	69	69
Comments	Perfect score, excellent positioning of elements in the GUI and all equations and elements are correct.	

Hooke’s Law

Model: General Rubric		
Criteria	Scoring	Max Points
Output Integrity		5
Names		5 (min 0)
Flows	Cloud -> Velocity Cloud -> Position	2
Variables	Initial Pos Mass Spring Const	3
Positioning		4
Specific Model Rubric		
Initial Conditions	2pts - Initial Pos 2pts - Mass 2pts - Spring Const	6
Relationships	Variables may be renamed if model does not run (penalize in general rubric) If model still does not run, -5pts per misc. necessary element change for model to run 2pts - Initial Pos 2pts - Initial Vel. 2pts - Vel. Flow 2pts - Pos. Flow 4pts - Spring Force 2pts - Spring Accel 12pts - Numerical	26

Model: General Rubric		
	Correctness	
Overall		51
Comments		

PROMPT: Create a model for the oscillating motion of a block on a spring according to Hooke's law. Initial variables should be starting position, mass, and the spring constant. The block originally starts at rest.

Model: ChatGPT-4o		
Criteria	Scoring	Max Points
Output Integrity	0 (Had to clean format)	5
Names	5	5
Flows	2	2
Variables	2 (-1 hardcoded position)	3
Positioning	4	4
Specific Model Rubric		
Initial Conditions	6/6 2pts - Initial Pos 2pts - Mass 2pts - Spring Const	6
Relationships	26/26 2pts - Initial Pos 2pts - Initial Vel. 2pts - Vel. Flow 2pts - Pos. Flow 4pts - Spring Force 2pts - Spring Accel 12pts - Numerical Correctness	26

Model: ChatGPT-4o		
Overall	45	51
Comments	Included comments in JSON which made the .luna invalid. Hardcoded initial position.	

Model: Claude 3		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	2	2
Variables	3	3
Positioning	4	4
Specific Model Rubric		
Initial Conditions	6/6 2pts - Initial Pos 2pts - Mass 2pts - Spring Const	6
Relationships	21/26 -5pts: “initialPosition” instead of “[initialPosition]” 2pts - Initial Pos 2pts - Initial Vel. 2pts - Vel. Flow 2pts - Pos. Flow 4pts - Spring Force 2pts - Spring Accel 12pts - Numerical Correctness	26

Model: Claude 3		
Overall	46	51
Comments	Almost perfect besides minor issue in position equation.	

Model: ChatGPT-o3-mini		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	2	2
Variables	3	3
Positioning	4	4
Specific Model Rubric		
Initial Conditions	6/6 2pts - Initial Pos 2pts - Mass 2pts - Spring Const	6
Relationships	26/26 2pts - Initial Pos 2pts - Initial Vel. 2pts - Vel. Flow 2pts - Pos. Flow 4pts - Spring Force 2pts - Spring Accel 12pts - Numerical Correctness	26

Model: ChatGPT-o3-mini		
Overall	51	51
Comments	Perfect.	

Model: DeepSeek-R1		
Criteria	Scoring	Max Points
Output Integrity	5	5
Names	5	5
Flows	2	2
Variables	2 (-1 hardcoded position)	3
Positioning	2	4
Specific Model Rubric		
Initial Conditions	6/6 2pts - Initial Pos 2pts - Mass 2pts - Spring Const	6
Relationships	26/26 2pts - Initial Pos 2pts - Initial Vel. 2pts - Vel. Flow 2pts - Pos. Flow 4pts - Spring Force 2pts - Spring Accel 12pts -	26

Model: DeepSeek-R1		
	Numerical Correctness	
Overall	48	51
Comments	Bad positioning & hardcoded start position Correct numerical output though	