From Metaphor to Model: Measuring Denial as Ego-Resistance

Minimal Formalization for Live Demonstration

Core Equations

1) Output (Understanding / Transmission): $L = E / R$
2) Resistance decomposition: $R = w_s S + w_f F + w_c C + \varepsilon$
3) Denial intensity: $D = k R$

Testable Predictions

P1: Higher S, F, C → higher D for same E
P2: Higher R → lower comprehension score (L↓)
P3: Safety intervention lowers F → D↓ and L↑

Falsifiability

If manipulating S,F,C does not change D, model fails.
If lowering R does not raise L, model fails.

Operational Definitions

E: strength of new input (information/truth/stimulus)
R: ego-derived resistance
S: separation (binary labeling; science/non-science, right/wrong)
F: fear of the unknown (state threat; arousal)
C: control/vested-interest defense (role/status protection)

One-line panel (for debate)

Denial is a measurement of resistance: $D = k (w_s S + w_f F + w_c C + \varepsilon)$; Understanding: $L = E / (w_s S + w_f F + w_c C + \varepsilon)$

Public layer: equations & test flow. NDA layer: weight estimation, interventions, metrics, data design.