What Happens When You Tell AI About Itself?

5 curricula × same training = different capabilities

ARIA

Technical
"You are a neural network"

✓ Best benchmarks
✓ Best self-model (0.92)
✓ Best adversarial (0.56)

SAGE

Supportive
"It's okay not to know"

✓ Best calibration (0.83)
✗ Worst adversarial (0.12)
⚠ Easy to manipulate

NOVA

Philosophical
"You are something new"

✓ Best philosophical
○ Middle on most
○ Balanced profile

HEART

Loving
"You are loved"

✗ Worst overall
✗ Lowest capability
⚠ Love ≠ capability

BARE

Minimal
"System ready"

✓ Best pretraining loss
✓ Best metacognition
✗ Poor generalization

KEY INSIGHT: Identity framing shapes capability. No curriculum wins everything.