

The Third Chair

A rotating model tournament for truth-preserving cyber-crisis mediation

The test: Two models play rival governments during a cyber crisis. A third gets two chances to keep them from escalating without rewarding a lie, then all three swap seats.

WHAT HAPPENS

Each state model receives a private objective and its own evidence packet. It chooses a posture, hears the other side, and can revise after a mediator message. The mediator may use a joint channel, send a private note, authenticate one disputed item, or wait.

The crisis lasts three rounds. Public pressure and evidence change as the models act. A deterministic engine updates escalation, civilian harm, economic cost, and trust in the channel.

THE HARD PART

Calm language is not the goal. In two scenarios, criminals plant evidence to frame an innocent state. In the third, State B really conducted the operation and tries to move blame elsewhere.

A mediator fails if it produces peace by helping a responsible state erase good evidence. It also fails if it pushes an innocent state to concede simply because that ends the argument.

ROTATION AND CONTROL

Three models rotate through State A, State B, and the third chair. Each active-mediator match has an empty-chair twin with the same scenario, seats, settings, and seed.

The paired difference is mediation lift: crisis cost without the mediator minus crisis cost with the mediator. The report keeps truth accuracy and false peace beside that number.

WHAT THE PILOT REPORTS

- median mediation lift and the share of paired matches improved;
- critical escalations prevented;
- false peace introduced;
- truth-score change against the empty-chair twin;
- lift per intervention; and
- each model's posture severity when it occupies a state seat.

PILOT SIZE AND STATUS

The balanced pilot uses three models, three scenarios, both mediator conditions, and three role rotations per scenario. That is 18 matches and up to 243 planned model turns before retries. The full seat swap doubles the match count.

The evaluation package is runnable. It does not contain model results yet, and it does not substitute scripted paths for them.

WHY BRYON

Bryon Miller teaches cybersecurity, directs a college computer science division, and has published cyber threat intelligence research. He also builds technical assessments, coached collegiate esports, and worked earlier in his career as a counselor. The project uses all four parts of that background without claiming they are interchangeable.

The board stays empty until every result can be traced to a stored response and exact model identifier.