ResearchDebateChain (RDC)

A debate-native, agentic reasoning OS with a Hypothesis Knowledge Graph, for medicine (rare diseases) and beyond.

---

0) What this is

ResearchDebateChain turns a set of tools and models into an **autonomous debate society**. Instead of a fixed pipeline, specialists **argue, refute, and converge** over a **Clinical/Conceptual Debate Graph (CDG)**—a typed, signed knowledge graph of hypotheses, evidence, tests, and causal structure. A **Moderator** runs a protocol (rounds, budgets, rules of evidence). A light **Jury** makes decisions from the graph's posteriors and utilities. Optionally, **reinforcement learning (RL)** helps agents learn better moves and the moderator learn scheduling; but RL is *not required* for v1.

RDC is domain-agnostic (legal, science, security). Below, the medical anchoring (differential diagnosis, "Dr. House" style) is concrete enough to implement, yet general enough to port.

---

1) Design goals

- **Agentic**: each agent is autonomous (policy, memory, tools) and negotiates uncertainty via rules of evidence.
- **Explainable**: all claims and decisions are visible in the CDG with provenance and counterfactuals.
- **Decision-theoretic**: actions/tests are chosen by **Value of Information** (VoI) and utility, not heuristics.
- **Modular**: pluggable tools (retrievers, calculators, simulators), pluggable policies (scripted or learned).
- **Cross-domain**: medical by default; trivially retargeted to other research domains.

---

2) Core entities

2.1 Agents

- **Specialists** (e.g., Infectious, Autoimmune, Oncology; or "Data Quality," "Causality," "Statistics" in general research).
- Each has:
  - **Policy** (scripted or learned): when to search, argue, challenge, or stop.
  - **Tool portfolio**: retrieval, guideline access, calculators, code-exec, simulators.
  - **Private memory**: episodic traces from prior debates, semantic priors, failure diaries.
  - **Belief state**: hypothesis log-odds, uncertainties, pending questions.

2.2 Moderator

- Runs the **debate protocol** (turn taking, budgets), enforces **rules of evidence**, and asks **cross-examination** questions where the graph is most conflicted.

2.3 Argumentation Engine (AE)

- Maintains the **CDG**, evaluates **support/attack** structure, fuses probabilistic evidence into **posteriors**, and surfaces **uncertainty hotspots**.

2.4 Jury

- Converts AE posteriors + utilities into: a **verdict** (ranked hypotheses), a **test plan** (VoI-optimal), a **Disagreement Index**, and **"what would change our mind"** counterevidence.

2.5 Evidence/Provenance Ledger

- Stores: all moves, tool calls, citations, CDG deltas, hashes of prompts, and verdict provenance.

---

3) The Clinical/Conceptual Debate Graph (CDG)

A **typed, signed, weighted** multigraph:

- **Nodes** (key types):
  - `Hypothesis` (H): disease / legal claim / theory.
  - `Evidence` (E): observation, paper, lab value, simulation output.
  - `Test/Action` (T): diagnostic test, experiment, intervention.
  - `Finding/Feature` (F): symptom, sign, environment, variable.
  - `Assumption` (A): modeling assumption, background fact.
- **Edges** (signed, weighted):
  - `supports` (E $\rightarrow$ H), weight (w$\in$[0,1]).

  - `attacks` (E ⊣ H), weight (w∈[0,1]).
  - `suggests_test` (H ⇒ T).
  - `causes` (H → F) (causal semantics, optional).
  - `contradicts`, `refines` (hierarchies), `derived_from` (provenance).
  - `query` / `answer` edges to capture information needs and resolutions.

- **Node/edge attributes**:
  - Quality/confidence (q), cost (c), risk (r), latency (l), and provenance bundle (source, tool, time).

- **Graph invariants**:
  - Every claim must be attached to **at least one evidence** or **explicit assumption**.
  - Every proposed test must have **expected discriminative power** documented (see §6).

---

4) Debate protocol (round-based)

1. **Seeding (Opening statements)**
   Each agent proposes top-(k) hypotheses with minimal evidence packs and priors.

2. **Challenges**
   Agents may **attack** rival claims, **query** missing links, or **propose tests** that would disambiguate.

3. **Rebuttal**
   Proponents answer with new evidence, revise priors, or concede edges.

4. **Cross-examination (Moderator-led)**
   Moderator pinpoints graph hotspots (cycles, conflicts, high-entropy subgraphs) and forces targeted clarification.

5. **Vote / Synthesis**
   Agents submit per-hypothesis probability vectors and confidence rationales. AE computes posteriors; Jury emits **verdict + test plan + rationale**.

6. **(Optional) Appeal**
   If Disagreement Index is high, run a short extra round focused on the top conflict pair.

Halting conditions: budget exhausted, entropy below threshold, or VoI of any remaining test < minimal gain.

---

5) Argumentation semantics (high-level math)

We blend **symbolic argumentation** with **probabilistic evidence**.

5.1 Bipolar argumentation labels

Let (G=(V,E$^+$,E$^-$)) with supports (E$^+$) and attacks (E$^-$). Define an acceptability operator (Γ) over labelings ($L : V \to \{\text{in}, \text{out}, \text{undec}\}$) akin to **grounded semantics**:

- A node (v) is **in** if all attackers of (v) are **out** and at least one supporter is **in** (or (v) is a base fact).
- **out** if it has an **in** attacker, or all supporters are **out**.
- **undec** otherwise.

Compute the least fixed point ($L^\star = \text{lfp}(\Gamma)$). This yields **defensibility** independent of probabilities.

5.2 Probabilistic fusion (posteriors over hypotheses)

Let ($\mathcal{H}$) be hypotheses, ($\mathcal{E}$) evidence observed. We maintain **log-odds** for each (H):

$$\ell(H) \;=\; \log \frac{P(H)}{1-P(H)} \;+\; \sum_{e \in \mathcal{E}} s(e,H) \cdot \text{LLR}(e \to H)$$

- ($s(e,H) \in \{-1, 0, +1\}$) encodes attack/support and is scaled by edge weight (w_{eH}).
- ($\text{LLR}(e \to H) = \log \frac{P(e|H)}{P(e|\neg H)}$).
- Posterior ($P(H) = \sigma(\ell(H)) = \frac{1}{1+e^{-\ell(H)}}$).

If edge qualities (q) vary, incorporate them as multipliers: ($\tilde{s} = s \cdot f(q)$) with ($f : [0,1] \to [0,1]$) (e.g., ($f(q) = q^\alpha$)).

Noisy-OR aggregation for multiple independent supports may be used when likelihoods are scarce:

$$P(H \mid \{e_i\}) \approx 1 - \prod_i \left(1 - \beta_i, P(H \mid e_i)\right)$$

with ($\beta_i \in [0,1]$) learned/calibrated reliability.

5.3 Consistency with labels

Enforce soft constraints that **in** nodes shouldn't be assigned near-zero probability and **out** nodes shouldn't be near one. Regularizer:

$$\mathcal{R}(P, L^\star) \;=\; \lambda_{in}!! \sum_{v : L^\star(v) = in} !!!!!!\text{CE}(P_v, 1) \;+\; \lambda_{out}!! \sum_{v : L^\star(v) = out} !!!!!!\text{CE}(P_v, 0)$$

where (P_v) is the marginal acceptability or posterior of node (v) (depending on type).

---

6) Decision-theoretic test selection (VoI)

Let current posterior over hypotheses be ($p = \{P(H_j)\}$). For a candidate test (T) with outcomes ($t \in \Omega_T$) (e.g., positive/negative), define **expected information gain**:

$$\mathrm{EIG}(T) = H(p) - \sum_{t \in \Omega_T} P(t)\, H!\left(p \mid T{=}t\right)$$

Use **expected utility** when costs and harms matter:

$$\mathrm{EU}(T) = \sum_{t \in \Omega_T} P(t) \left(\max_{a \in \mathcal{A}} \mathbb{E}[U(a, H) \mid T{=}t] \right) - C(T) - R(T)$$

Choose ($T^\star = \arg\max(\alpha \cdot \mathrm{EIG}(T) + \mathrm{EU}(T))$), with ($\alpha$) a trade-off scalar. **Risk constraints**: forbid (T) if ($R(T) > \tau$) (e.g., contrast allergy risk).

"What would change our mind?" For a hypothesis pair ((H_i, H_k)), find the test (T) that maximizes the **posterior odds shift**:

$$\Delta_{ik}(T) = \sum_t P(t) \left| \log \tfrac{P(H_i|t)}{P(H_k|t)} - \log \tfrac{P(H_i)}{P(H_k)} \right|$$

Report the top (T) and outcome (t) with largest ($\Delta_{ik}(T)$).

---

7) The verdict

The **Jury** consumes:

- Posteriors ($P(H)$),
- VoI/EU scores for tests/actions,
- Disagreement metrics (variance across agent votes; attack/support cycle density).

Outputs:

- **Differential**: top-(K) hypotheses with probabilities and key supporting/attacking paths.
- **Plan**: ordered test/action set ($(T_1, \ldots, T_m)$) by EU / EIG per unit cost.
- **Disagreement Index**: e.g., entropy of agent vote distribution or Kendall-(τ) across agent rankings.
- **Counterevidence**: minimal evidence set that would flip the verdict (graph-based hitting set).

---

8) Orchestration mechanics

8.1 Scheduling and budgets

- **Who speaks next?** Choose the agent expected to maximize **entropy drop** per unit budget.
- **How long?** Grant a token/tool budget based on diminishing returns (if last two moves from same agent had low gain, rotate).
- **When to stop?** Stop when ($\max_T \mathrm{EIG}(T)$) and entropy drop per step fall below thresholds, or cost cap reached.

A simple contextual **bandit** (no heavy RL) can handle speaker selection: features = intake summary, current entropy, agent-specific recent gain, remaining budget; reward = entropy drop – cost.

8.2 Rules of evidence (lightweight governance)

- Every move must attach **provenance** (source, tool trace).
- Penalize or reject untraceable claims.
- Mark **assumptions** explicitly; they are allowed but must be challengeable.

---

9) Memory & provenance

- **Private memory**: per-agent diaries of (state → move → outcome), mistakes, and preferred tools, used to set priors and heuristics.
- **Shared memory (CDG)**: the debate itself—source of truth for posteriors and verdict.
- **Ledger**: append-only record of moves, tool calls, graph deltas, and final verdict subgraph; enables audit and replay.

---

10) Optional reinforcement learning (kept simple)

RL helps **refine** policies but is not necessary to ship v1.

- **Agent policy learning**:
  - *State*: intake summary + CDG features (entropy, conflicts, VoI landscape).
  - *Action*: {claim, attack, support, query, propose_test} + tool parameters.
  - *Reward*: per-step **information gain** ($(H(p_{t-1}) - H(p_t))$) minus cost; terminal **-NLL** if ground truth exists; penalties for unsafe proposals.
  - *Algorithm*: start with **behavior cloning** from scripted policies; optionally fine-tune with a small-batch **PPO** or **AWR**.

- **Moderator learning** (scheduling):
  - *Action*: pick next agent and grant budget.
  - *Reward*: entropy drop per cost.
  - *Algorithm*: **contextual bandit** (LinUCB/Thompson) before any policy gradient.
- **Calibration learning**: adjust edge-quality scaling ($(f(q) = q^{\alpha})$) and reliability ((β_i)) by minimizing **Brier score** and **ECE** on held-out debates.

---

11) Medical specialization (rare disease, "Dr. House")

Agents: Infectious, Autoimmune, Metabolic/Genetic, Hematology/Oncology, Imaging, Pharmacology (adverse drug), Epidemiology, Data Quality.

Example flow:

1. **Seeding**
   - Infectious: proposes *Brucellosis* and *Q-fever* with travel and animal exposure evidence;
   - Autoimmune: proposes *Adult-Onset Still's Disease* with ferritin and rash features.
2. **Challenges**
   - Hem/Onc attacks Still's with normal LDH, Imaging attacks Q-fever with clear chest CT.
   - Infectious suggests **Brucella serology** ((T_1)) and **Q-fever phase II IgG** ((T_2)).
3. **Rebuttal**
   - Autoimmune supplies **glycosylated ferritin** literature (adds strong support edge).
4. **Cross-exam**
   - Moderator probes the cycle between fever pattern $\leftrightarrow$ rash $\leftrightarrow$ ferritin; asks for tests that separate Still's vs Brucellosis.
   - VoI ranks ((T_1)) higher (bigger ((Δ_{ik}))).
5. **Verdict**
   - Jury: ($(P(\text{Brucellosis}) = 0.42)$), ($(P(\text{Still's}) = 0.31)$), ($(P(\text{Q-fever}) = 0.17)$).
   - Plan: ((T_1)) first; if positive, start doxycycline + rifampin; if negative, order glycosylated ferritin.
   - Disagreement Index moderate; "what would change our mind": positive blood culture or glyco-ferritin $< 20\%$.

Safety: contraindication rules (pregnancy, drug interactions) as **hard constraints** in EU; e.g., avoid doxycycline if pregnant—the Jury will propose alternatives.

---

12) Non-medical applications (sketches)

- **Scientific hypothesis testing**: agents = Theory, Experiment, Statistics, Prior Art, Simulation; tests = experiments; utility = power × novelty cost.
- **Cybersecurity incident response**: agents = Threat Intel, Forensics, Networking, AppSec; hypotheses = attack vectors; tests = triage probes; utility = risk reduction time.
- **Legal research**: agents = Case Law, Statutes, Facts, Procedure; hypotheses = legal theories; tests = discovery motions; utility = probability of success × settlement leverage cost.

The **same CDG + debate protocol + VoI** applies.

---

13) Evaluation & quality gates

- **Accuracy**: top-(k) inclusion, NLL on ground truth when available.
- **Calibration**: Brier score, Expected Calibration Error (ECE).
- **Efficiency**: cost/latency to reach target entropy.
- **Explainability**: path-based rationale coverage (each verdict has ≥1 support path with provenance).
- **Robustness**: ablations (remove one agent / tool) and measure degradation.
- **Safety**: zero tolerance for constraint violations; track near-misses.

---

14) Minimal API surface (conceptual)

- `POST /debate/start` → returns `case_id`, stream URL.
- `GET /debate/{case_id}/stream` → Server-Sent Events of moves.
- `GET /debate/{case_id}/graph` → CDG (JSON-LD / GraphML).
- `GET /debate/{case_id}/verdict` → differential, plan, disagreement index, counterevidence set.
- `GET /debate/{case_id}/ledger` → provenance bundle.

---

15) Implementation roadmap

- **V0 (scripted agents, no RL)**
  - Build CDG, AE (labels + log-odds fusion), VoI planner, debate protocol, provenance ledger, Jury.
  - Scripted policies: "retrieve → propose," "challenge contradictions," "propose test with highest $((\Delta_{ik}))$."
- **V1 (bandit orchestration + calibration learning)**
  - Contextual bandit for turn-taking; learn reliability scalers $((\beta_i))$, edge scaling exponent $((\alpha))$.
  - Add "what would change our mind" and disagreement summaries.
- **V2 (optional RL)**
  - Behavior-cloned agent policies; small-batch PPO/AWR online refinement with strict safety constraints.

---

16) Glossary (quick)

- **CDG**: Clinical/Conceptual Debate Graph—typed, signed, weighted multigraph of the debate.
- **AE**: Argumentation Engine—computes labels (in/out/undec), posteriors, hotspots.
- **VoI**: Value of Information—expected benefit of running a test relative to cost and risk.
- **Disagreement Index**: scalar summarizing spread of agent beliefs / structural conflict.
- **Provenance Ledger**: append-only audit trail of moves, tools, and graph deltas.

Final note You can ship **V0** without any RL and already get a system that *argues like experts, explains itself, and plans tests by information value.* RL is an optional turbocharger—use it only when logs and safety gates are in place.