

Combating Disinformation with Neurocognitive AI

A Hybrid Model for Crisis Communication

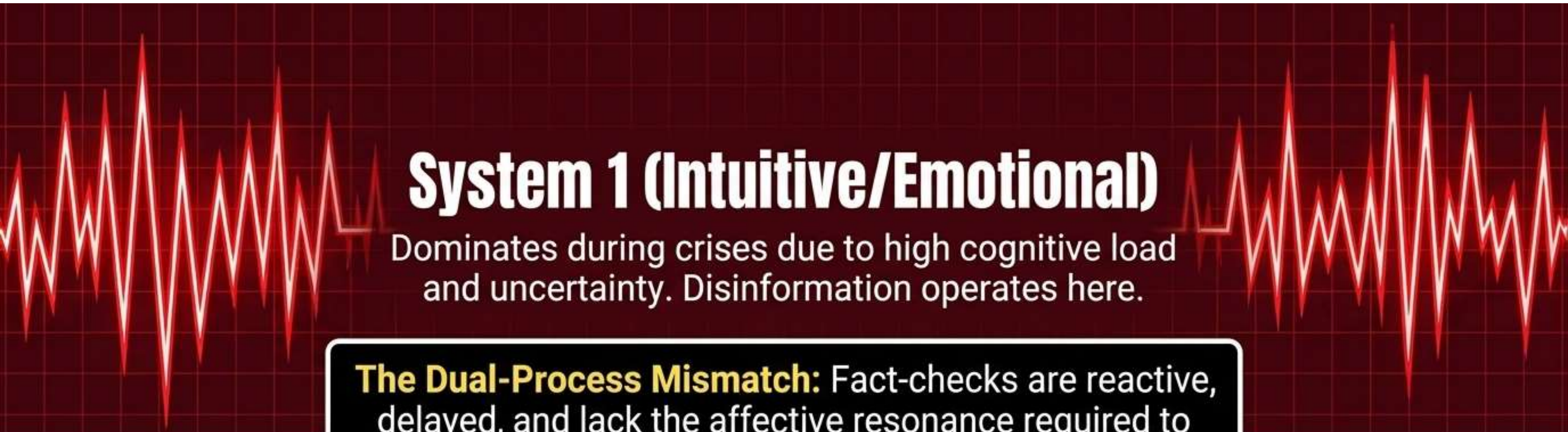
The Human Vulnerability Problem



The Shift: Disinformation in crises is not an informational deficit; it is a cognitive vulnerability.

The Mechanism: Cognitive hacking leverages the affect heuristic and confirmation bias to bypass critical thinking.

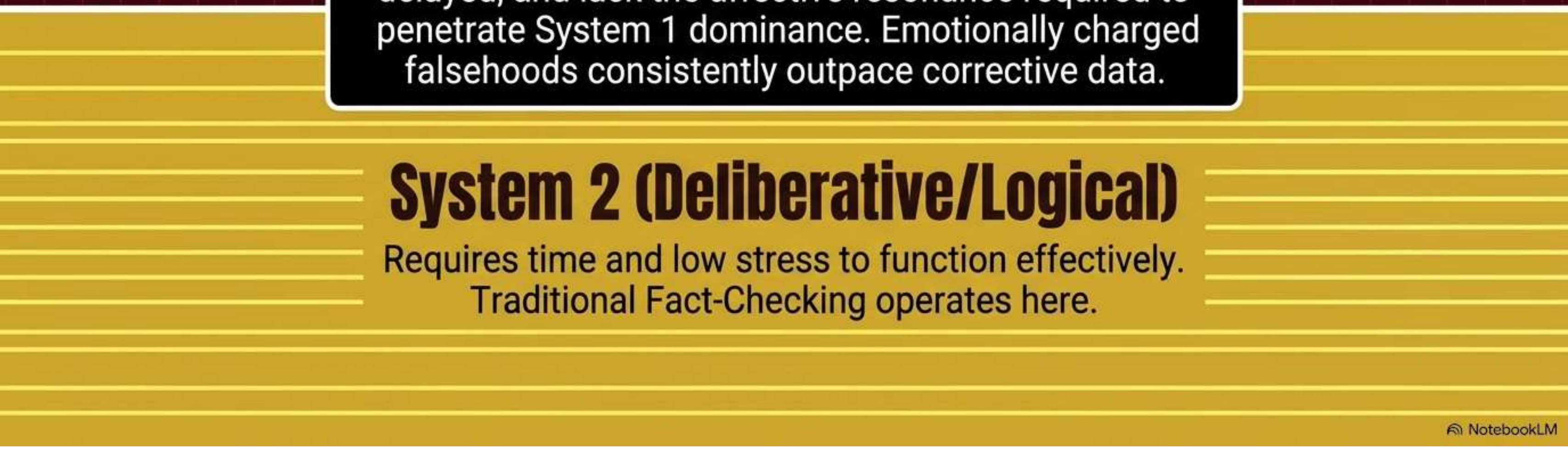
The Result: Emotionally resonant falsehoods weaponize human cognition, rendering purely fact-based corrections insufficient.

A red jagged waveform, resembling a heartbeat or an ECG, runs horizontally across the top of the image. It is set against a dark red background with a subtle grid pattern.

System 1 (Intuitive/Emotional)

Dominates during crises due to high cognitive load and uncertainty. Disinformation operates here.

The Dual-Process Mismatch: Fact-checks are reactive, delayed, and lack the affective resonance required to penetrate System 1 dominance. Emotionally charged falsehoods consistently outpace corrective data.

The bottom half of the image features a solid yellow background with several thin, horizontal white lines, creating a striped effect.

System 2 (Deliberative/Logical)

Requires time and low stress to function effectively. Traditional Fact-Checking operates here.

The Paradigm Shift: Neurocognitive AI

Integrating neuroscience-informed AI to anticipate psychological impact and tailor counter-narratives.

Measuring cognitive load and focus via interaction metrics (click-through rates, dwell time).



Operationalizing the Shift

Traditional Fact-Checking

Neurocognitive AI

Timing

Reactive & Delayed.

Anticipatory (Achieves an estimated 18% reduction in response latency).

Focus

Factual verification.

Affective alignment and cognitive load assessment.

Method

Rule-based databases.

Neuro-symbolic reasoning (Deep learning + logical inference yielding +23% precision).

Output

Generic public corrections.

Psychologically attuned counter-narratives (Yields up to 30% reduction in false belief among vulnerable cohorts).

The Anatomy of a Crisis: Universal Structures

Despite domain differences, health and geopolitical crises share an identical vulnerability architecture.



Shared Characteristics

1. Emotional narratives structurally outpace facts.
2. AI-generated synthetic content boosts credibility.
3. Cross-platform reinforcement forces digital falsehoods into mainstream media.
4. Institutional responses lack scalable, timely countermeasures.

The 7-Stage Implementation Pathway

1. Multimodal Data

Acquisition: Ingesting NLP and computer vision data.

2. Neurocognitive Signal Integration: Fusing attention metrics and emotional stress markers.

6. Immersive Deployment and Simulated: Future-proofing via AR/VR experiential AI/ML experiential learning environments.

2. Neurocognitive Signal Integration:

Fusing attention metrics and emotional stress markers.

4. Behavioural Prediction & Feedback:

Real-time evaluation of counter-narrative efficacy to prevent unintended escalation.

5. Personalisation of Counter-Narratives:

Aligning messaging with audience cognitive profiles (e.g., safety messaging for fear-driven groups).

7. Institutional Coordination: Embedding real-time dashboards into governmental and civil crisis response centers.

Governance & Ethical Guardrails



The Ethical Tension

Personalization builds resilience, but risks reproducing the manipulative dynamics of malicious actors.

Regulatory Compliance

The architecture must strictly align with the EU AI Act (human oversight, non-discrimination) and the Digital Services Act (DSA).

Human-Centred AI

Systems must provide transparent explainability (XAI). The goal is to construct protective cognitive shields, ensuring the technology empowers human autonomy rather than coercing it.

Securing the Human Layer

Sustainable resilience requires a fundamental shift in crisis communication. We must move beyond securing our informational infrastructure to actively defending the human cognitive layer.

From reactive, factual verification to proactive, neurocognitive engagement.