CortexFlow

A State-Based Attention Regulation Engine

Team 9 - Leptons

"Attention instability emerges from competition between executive control and internally or externally triggered alternative action programs. CortexFlow models and regulates this competition in real time — not as decoration, but as the structural core of the system."

]

1 Abstract

Modern students are not constrained by time — they are constrained by **fragmented attention**. We present **CortexFlow**, a privacy-preserving, closed-loop attention regulation engine that infers cognitive state from behavioural telemetry, grounded *structurally* in validated neuroscience. Unlike time-based productivity tools, CortexFlow models dynamic attention risk through a dual-state model (Instability, Drift) modulated by Fatigue — each variable directly approximating large-scale neural network competition across the Executive Control Network (ECN), Salience Network (SN), and Default Mode Network (DMN). A drift-diffusion breakdown detector implements computational cognitive theory. Interventions are adaptive and proportional, firing only when estimated risk warrants, enabling genuine metacognitive support without restriction.

2 Problem Statement

2.1 Cognitive Failure Modes

Digital learning environments are structurally hostile to sustained focus. Four recurring failure modes emerge from the literature:

- **Task switching:** Involuntary context shifts incurring working memory residue and measurable switching costs.
- **Passive disengagement:** Behaviorally present but cognitively absent states — the "zoning out" phenomenon.
- **Fatigue accumulation:** Progressive executive resource depletion amplifying both modes above.
- **Absent metacognition:** Students cannot identify when attention collapsed, making self-correction impossible.

2.2 Why Existing Tools Fail

App blockers, Pomodoro timers, and focus-lock tools share one false assumption: *distraction is a scheduling problem*. A student can enter a mind-wandering state five minutes into a "blocked" session. Time-based interventions are systematically misaligned with the process they claim to regulate.

Core Failure: Existing tools do not model the neural competition underlying attention breakdown. CortexFlow does.

3 Neuroscience Foundations

Scientific Grounding Principle

The neuroscience is not in the Chrome extension or the UI. It is in the **state model assumptions**. Every variable, proxy, and dynamic rule in CortexFlow is derived from a validated neural mechanism.

3.1 The Neural Competition Model

Focused cognition involves continuous competition among three large-scale brain networks (Table 1). Attention breakdown is not random — it is a measurable shift in network dominance that CortexFlow approximates computationally via behavioural proxies.

Table 1: Large-scale networks modelled by CortexFlow.

Net.	**Function**	**Dominant state**
ECN	Goal maint., WM	Stable engagement
SN	Conflict detect.	Impulsive switching
DMN	Mind-wandering	Passive disengage

3.2 Executive Control Network (ECN)

Anchored in DLPFC and posterior parietal cortex, the ECN provides top-down signals maintaining goal-relevant processing [Miller and Cohen, 2001]. Sustained engagement corresponds to task-positive network dominance, actively suppressing DMN [Dosenbach et al., 2008]. Behavioural proxies: stable typing rhythm; low switching frequency; directed mouse trajectories.

3.3 Salience Network (SN)

Centred on the anterior insula and ACC, the SN detects conflict and triggers network switching [Menon and Uddin, 2010]. When response conflict accumulates past threshold, the ACC initiates a control shift [Botvinick et al., 2001] — the neural substrate of impulsive distraction. Proxies: rapid tab-switching bursts; distractor site attempts; high motor variability.

3.4 Default Mode Network (DMN)

The DMN activates during mind-wandering and self-referential thought [Mason et al., 2007]. fMRI experience-sampling studies confirm DMN intrusions coincide with task-unrelated thought and decreased task-positive activation [Christoff et al., 2009]. DMN dominance can be *behaviourally invisible* — scrolling or playing a video while cognitively absent. Proxies: idle periods; scroll reversals; passive playback.

3.5 Fatigue and Vigilance Decrement

Mackworth established that sustained attention degrades predictably over time [Mackworth, 1948], reflecting genuine executive resource depletion. Mark et al. showed prior interruptions impose cognitive residue that compounds fatigue [Mark et al., 2008]. In the model, Fatigue is a nonlinear amplifier: identical Instability at session-end carries far greater breakdown risk than at session-start.

4 Computational Cognitive Model

4.1 Design Philosophy

CortexFlow applies the methodology of behavioural cognitive neuroscience: neural states are inferred from observable behavioural outputs. In lab settings researchers infer ECN engagement from reaction times, error rates, and movement trajectories. CortexFlow applies the same logic to naturalistic digital work — tab-switching frequency, motor variance, idle ratios, and scroll patterns are the behavioural readouts of the underlying neural state competition. This is not a metaphor; it is an established measurement strategy applied at scale.

4.2 Latent State Variables

4.2.1 Instability (SN Dominance)

$$I_t = a_1\,\mathrm{SwitchRate}_t + a_2\,\mathrm{MotorVar}_t + a_3\,\mathrm{DistractorAttempts}_t$$

High I_t signals that competing stimuli are overriding executive control — the behavioural signature of conflict-driven attentional capture.

4.2.2 Drift (DMN Dominance)

$$D_t = b_1\,\mathrm{IdleRatio}_t + b_2\,\mathrm{ScrollEntropy}_t + b_3\,\mathrm{PassivePlayback}_t$$

Drift reflects *passivity*, not impulse — the quiet collapse of executive engagement.

4.2.3 Fatigue (Executive Depletion)

$$F_t = \sigma\left(\frac{\mathrm{SessionDuration}_t}{\mathrm{ExpectedDuration}}\right)$$

Sigmoid-normalised to capture the accelerating degradation curve of vigilance decrement research.

4.2.4 Global Attention Risk

$$\mathrm{AttentionRisk}_t = \sigma(w_1 I_t + w_2 D_t + w_3 F_t)$$

4.3 Network Activation Mapping

Interpretable relative activations for dashboard visualisation:

$$\begin{aligned}
\mathrm{ECN}_t &= \max(0,\ 1 - I_t - F_t)\\
\mathrm{DMN}_t &= D_t\\
\mathrm{Salience}_t &= I_t\\
\mathrm{Load}_t &= F_t
\end{aligned}$$

4.4 Temporal Dynamics

Critical design principle: Neural systems accumulate and decay — they do not jump. A system that resets states between telemetry windows is counting events, not modelling cognition. Temporal dynamics are what separate CortexFlow from a heuristic productivity tracker.

We implement continuous-time network competition:

$$\mathrm{ECN}_{t+1} = \mathrm{ECN}_t + \alpha E_t - \beta I_t - \gamma F_t \tag{1}$$
$$\mathrm{DMN}_{t+1} = \mathrm{DMN}_t + \delta\,\mathrm{Idle}_t - \varepsilon\,\mathrm{Task}_t \tag{2}$$
$$\mathrm{SN}_{t+1} = f(\mathrm{SwitchPressure}_t,\, \mathrm{ECN}_t) \tag{3}$$

where E_t is task engagement signal and $\alpha, \beta, \gamma, \delta, \varepsilon$ are learned per-user parameters. These equations enforce four biologically grounded properties: (1) exponential ECN decay under fatigue; (2) DMN rise proportional to inactivity; (3) hysteresis in state transitions matching cognitive inertia; (4) intra-session fatigue irreversibility.

4.5 Drift-Diffusion Breakdown Detector

To detect discrete breakdown events, CortexFlow implements a **drift-diffusion accumulation model** mirroring race-to-threshold neural decision dynamics:

$$A_{t+1} = A_t + I_t - \mathrm{ECN}_t \tag{4}$$
$$\mathrm{Breakdown}_t = \mathbf{1}[A_t > \theta] \tag{5}$$

A_t is accumulated conflict; θ is the breakdown threshold (default 1.0, tunable). This is a direct implementation of computational cognitive theory: exactly as neural evidence accumulation drives decisions in the brain, accumulated attentional conflict drives intervention decisions in CortexFlow.

5 Machine Learning Architecture

5.1 Baseline Prediction

Logistic regression provides an interpretable low-data baseline for predicting imminent breakdown probability:

$$P(\mathrm{Breakdown}) = \sigma(\boldsymbol{\theta}^T \mathbf{X})$$

where $\mathbf{X} = [I_t,\, D_t,\, F_t,\, \mathrm{ECN}_t,\, A_t,\, \Delta I_t,\, \Delta D_t]$.

5.2 Task-Specific Normalisation

Each task type has a distinct behavioural baseline. Features are normalised relative to that profile before inference:

Table 2: Task-type baseline conditioning.

Task	High	Low
Writing	Typing density	Scroll
Reading	Scroll, dwell	Typing
Coding	Typing, pauses	Video
Watching	Passive play	Typing

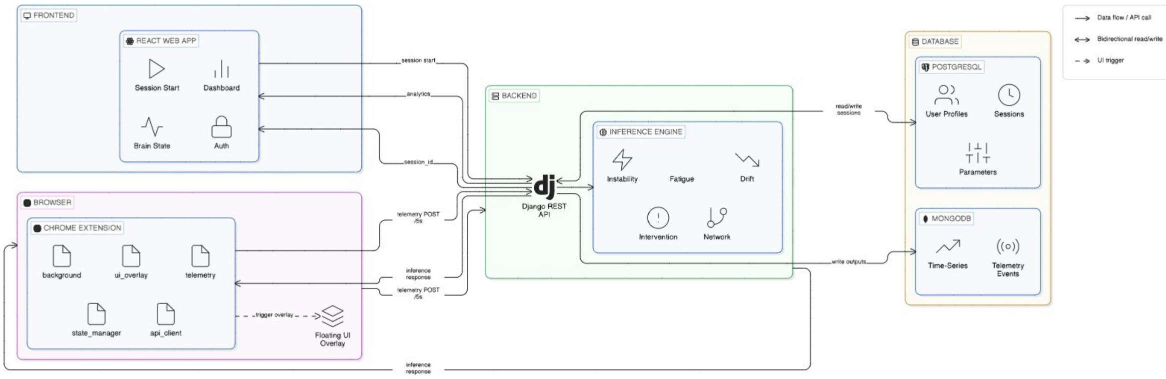

Figure 1: System Architecture

5.3 Personalised Adaptation

After two sessions, the system learns user-specific parameter profiles. Individual variability in vigilance decrement rates and distractor sensitivity is well-documented in cognitive neuroscience. The adaptation layer uses Bayesian updates of weight priors and drift-diffusion threshold calibration from empirical breakdown history.

5.4 Explainability via Feature Attribution

All inference outputs include SHAP-style feature attribution. Instead of "you seem distracted," the overlay delivers: *"Your switching rate tripled in the last 4 minutes, elevating Salience Network activation."* This builds genuine metacognitive awareness — a key differentiator aligned directly with Theme A.

5.5 Training Data Sources

Table 3: Training datasets and contributions.

Dataset	**Ground truth provided**
Mouse-tracking (OSF)	Motor variability, conflict signatures
NASA-TLX	Cognitive load ratings
MOOC engagement	Attention breakdown proxies
Task-switching corpora	Switching costs, error rates

6 System Architecture

6.1 End-to-End Flow

6.2 Component Stack

1. **React Web App** — Session management, real-time cognitive dashboard, historical analytics, brain-state radar.
2. **Chrome Extension** — Passive behavioural telemetry in 5-second rolling windows; contextual intervention overlay.
3. **Django / DRF Backend** — Inference engine; temporal dynamics; personalisation; API layer.
4. **PostgreSQL** — Session records, user profiles, learned per-user parameters.
5. **MongoDB** — High-throughput time-series telemetry and model output storage.

6.3 Extension Telemetry Modules

The extension collects only *aggregated* metrics — never raw content. Per 5-second window:

- **Text tasks:** `switch_rate`, `idle_density`, `scroll_reversal_ratio`, `typing_interval_var`, `mouse_velocity_var`
- **Video tasks:** `background_playback_flag`, `rewind_frequency`, `tab_switch_during_video`

6.4 Backend Data Models

Session (PostgreSQL): `id, user_id, task_type, start/end_time, avg_instability, avg_drift, avg_fatigue, deep_work_ratio, switch_count`

model_outputs (MongoDB): `session_id, timestamp, I, D, F, ECN, DMN, Salience, Load, risk, accumulated_conflict`

6.5 API Design

POST `/api/session/start` → `{session_id, baseline_profile}`

POST `/api/telemetry`

```
{ "session_id": "...",
  "features": {
    "switch_rate": 0.30,
    "idle_density": 0.10,
    "scroll_reversal_ratio": 0.40,
    "typing_var": 0.20,
    "duration_norm": 0.50 } }
```

Inference Response:

```
{ "instability": 0.62, "drift": 0.21,
  "fatigue": 0.48,    "risk":  0.67,
  "accumulated_conflict": 0.83,
  "breakdown_imminent": true,
  "attribution": {
    "switch_rate": 0.41,
    "motor_var": 0.18 },
  "network": {
    "ECN": 0.30, "DMN": 0.21,
    "Salience": 0.62, "Load": 0.48 } }
```

7 Intervention Engine

7.1 Rule-Based Decision Layer

Table 4: Intervention rules and responses.

Trigger	Response
$I_t > 0.7$	3 s delay on distractor tab; impulse confirmation; spotlight mode
$D_t > 0.7$	Pause video; "Still engaged?" prompt; reflection question
$F_t > 0.8$	Suggest micro-break; 60 s breathing prompt
Breakdown ($A > \theta$)	Full-screen soft interrupt; SHAP explanation

A cooldown requiring *two consecutive* high-risk windows before triggering prevents false positives.

7.2 Floating UI Overlay

A persistent circular indicator uses colour encoding: **Green** (stable) → **Amber** (rising) → **Red** (high instability) → **Blue** (drift/zoning). Clicking opens a panel showing I, D, F, the network radar, and a natural-language SHAP explanation: *"High switching frequency elevated Salience activation above ECN threshold."*

7.3 Dashboard Analytics

Post-session analytics include: Instability / Drift / Fatigue timeline; intervention markers; network radar chart; heatmap of instability spikes; and a re-entry cost estimate computed as:

$$\text{Re-entry cost} = \text{SwitchCount} \times 18\,\text{min}$$

based on research-validated recovery time [Mark et al., 2008]. This metric is a powerful motivational signal for students.

8 Privacy Architecture

CortexFlow is privacy-first by design, not by retrofit:

- No keystroke *content* stored — timing patterns only.
- Only aggregated metrics transmitted; raw events never leave the browser.
- Domain-level logging only — no full URLs, no browsing history.
- No cloud model training — inference runs on user-owned backend.
- All session data is exportable and deletable on demand.
- Privacy toggle visible and accessible in the dashboard at all times.

9 Implementation Roadmap

Table 5: Phased delivery strategy.

Phase	Deliverable
1	Django session and telemetry endpoints; PostgreSQL + MongoDB setup
2	Chrome extension telemetry aggregation and floating overlay
3	Rule-based state scoring with temporal accumulation/decay
4	React dashboard with real-time network visualisation
5	ML integration, Bayesian personalisation, SHAP attribution

10 Differentiation and Scientific Defensibility

Table 6: CortexFlow vs. time-based tools.

Property	Blockers	CortexFlow
State-aware	×	✓
Temporal dynamics	×	✓
Neuroscience-grounded	×	✓
Adaptive intervention	×	✓
Interpretable output	×	✓
Privacy-preserving	∼	✓

Any reviewer who asks *"What neural mechanism are you approximating?"* receives a precise answer: CortexFlow approximates network-level attentional control dynamics through behavioural proxies grounded in established conflict-monitoring and sustained-attention research. The temporal dynamics (Equations 1–3), the competitive interaction, and the drift-diffusion detector (Equations 4–5) are not decorative — they reflect the mechanistic structure of biological attention systems.

The distinction between a heuristic productivity tracker and computational cognitive neuroscience is not the UI or framework — it is whether the model implements **dynamic competing network activation with fatigue modulation, temporal accumulation, and biologically plausible state transitions**. CortexFlow does.

11 Relevance to Theme A

Theme A emphasises helping students regain control of their focus and reduce mental clutter. CortexFlow aligns along five axes:

1. Models **attention as a cognitive state**, not a clock — respecting the biological reality of how focus works.
2. Makes **invisible cognitive costs visible** through the network dashboard and SHAP attribution layer.
3. **Supports** executive control rather than enforcing restriction — building metacognitive skill, not dependency.
4. **Adaptive, proportional** interventions that fire only when risk genuinely warrants, reducing false-positive interruption.
5. **Preserves user autonomy and privacy**, ensuring students control their own cognitive data.

12 Conclusion

CortexFlow reframes attention as a dynamic cognitive state rather than elapsed time. By structurally grounding its model in ECN–SN–DMN competition with temporal accumulation, fatigue modulation, and a drift-diffusion breakdown detector, the system achieves genuine computational cognitive neuroscience — not productivity tech with scientific labels applied post-hoc.

The result is a closed-loop attention regulation engine that knows *when* and *why* a student's focus is at risk, intervenes adaptively, and builds metacognitive awareness over time. CortexFlow does not restrict — it regulates. It does not time — it understands.

References

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.

Dosenbach, N. U., et al. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, 12(3), 99–105.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control. *Brain Structure and Function*, 214, 655–667.

Botvinick, M. M., et al. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652.

Mason, M. F., et al. (2007). Wandering minds: the default network and stimulus-independent thought. *Science*, 315(5810), 393–395.

Christoff, K., et al. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *PNAS*, 106(21), 8719–8724.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *British Journal of Psychology*, 38(4), 1–24.

Mark, G., et al. (2008). The cost of interrupted work: more speed and stress. *CHI Conference Proceedings*, 107–110.