Bias Cluster 1: "Too Much Information"

Scenario:

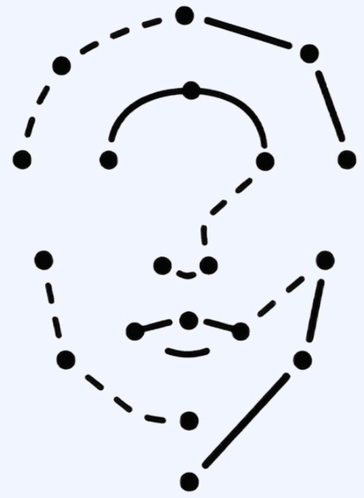
A hospital deploys a sepsis prediction AI that surfaces 27 real-time patient variables alongside its risk score. The interface is comprehensive by design: the clinical team asked for full transparency after complaints that earlier systems felt like black boxes. Within weeks, a nurse develops their routine with the tool: they check the score and skim the variables. One morning, the system flags a patient at moderate risk, and the variable that would have elevated it to high risk is buried in the dashboard. The nurse doesn't doesn't pay attention. The patient deteriorates overnight.

The system gave more information than anyone could process, and so functioned — in practice — as if it gave none.

Think: what does this mean for transparency as a design value: is more information always better for calibrated trust and informed use? How do we design for the reality that attention is finite, without deciding on behalf of users what they should notice? When does a system's comprehensiveness trigger cognitive biases in its users?

Bias Cluster 2: "Not Enough Meaning"

Scenario:
A junior financial analyst uses an LLM daily to make sense of market movements. The model always has an explanation -- fluent, structured, confident. Over time, the analyst stops experiencing any uncertainty with the model. When data is ambiguous, the model's account of it feels like a believable explanation. The analyst begins pitching to clients using framings they did not generate and could not fully defend. In a few months, three of their recommendations fail in ways the model had not anticipated. They are surprised.



We are meaning-making creatures, and AI systems that always provide meaning may be exploiting something fundamental about how we think.

When an AI fills interpretive gaps with confident narrative, it short-circuits the discomfort that would otherwise prompt deeper inquiry. Could that discomfort be designed back in — and if so, what would it look like to leverage our need for meaning rather than simply disrupt it? What is lost when a professional's tolerance for uncertainty is gradually eroded by a system that never models it?

Bias Cluster 3: "Need to Act Fast"

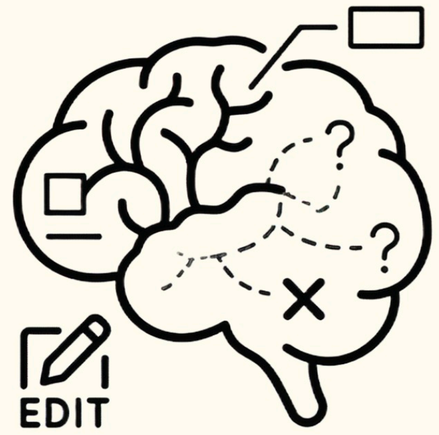
Scenario:
An air traffic control tower has been using a state-of-the-art AI collision-avoidance assistant for the past few months. The system issues alerts with recommended manoeuvres. Under normal load, controllers routinely cross-check alerts against their own situational knowledge. Tonight, with nine simultaneous aircraft to monitor, there is no time. Every alert is acted on as issued. In one of the cases, the AI system recommends a geometrically safer-looking manoeuvre rather than the correct one. The controller follows the recommendation without much cross-checking. Nothing bad happens, but two aircrafts come closer than they should. In the debrief, the controller says: "There was no time to think. There is never time to think..."

Cognitive bias research often assumes a reflective agent who could, in principle, reason more carefully if prompted (such as through cognitive forcing functions or deliberate friction).

But some decision-making contexts eliminate that option. What does bias mitigation research look like when the interaction window is very brief? How should the human-AI interaction be designed to mitigate this bias? How should humans and AI share the workload in such domains?

Bias Cluster 4: "Memory Limitations"

Scenario:
A policy analyst has spent six months using an LLM to help research and draft briefings. They couldn't tell you exactly what the AI contributed to each one, the work has blurred together. They do remember a handful of vivid moments: the time the LLM surfaced a statistic that transformed a weak argument into a compelling one, the time a minister praised a briefing they had drafted heavily with LLM assistance. They forgot — or never properly encoded — the three occasions it confidently cited sources that didn't exist, or the long stretches of mediocre output they edited into shape. Their general impression, the one they reach for when deciding how much to rely on the system today, is assembled from peaks and a handful of vivid successes.



Peaks, recent events, emotionally charged moments, and things that feel self-relevant get encoded strongly in our memory, while the long unremarkable middle can get discarded.

If users' trust in AI relies on their past experiences and memory, and memory systematically inflates the memorable over the representative, what does that mean for how trust calibration research measures "experience with the system"? How do cognitive biases make this worse? What kinds of design interventions help? Is there something about the way AI systems present their outputs — fluent, confident -- that makes them disproportionately memorable in ways that compound this problem over time?