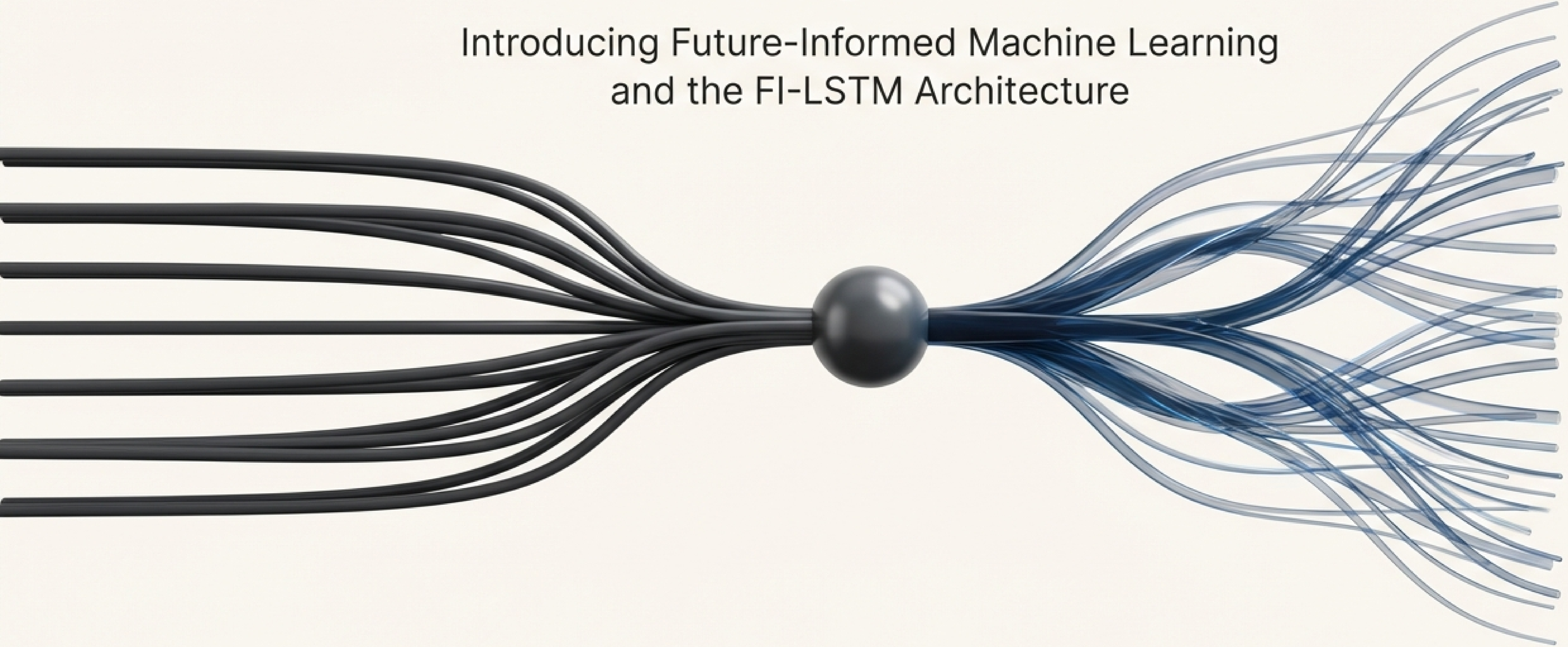
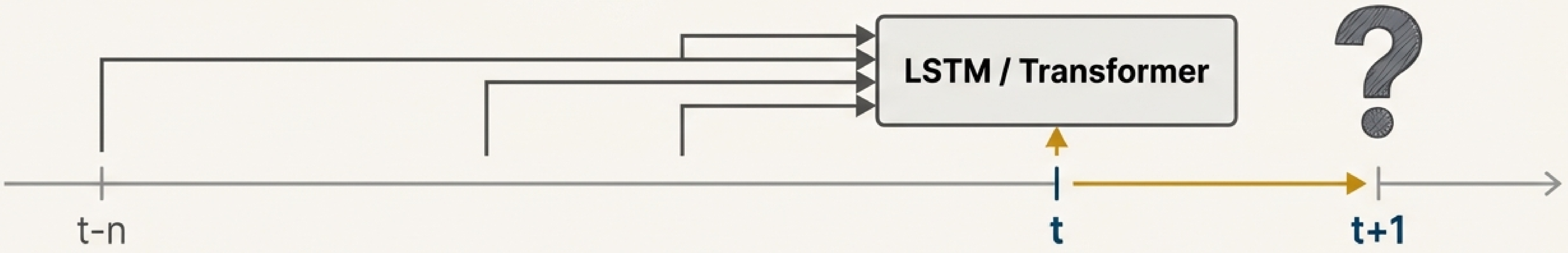
Beyond Memory: AI That Learns by Simulating the Future

Introducing Future-Informed Machine Learning and the FI-LSTM Architecture



Current Sequence Models Are Powerful, Yet Myopic



Dominant architectures like LSTMs and Transformers excel at learning from historical data.

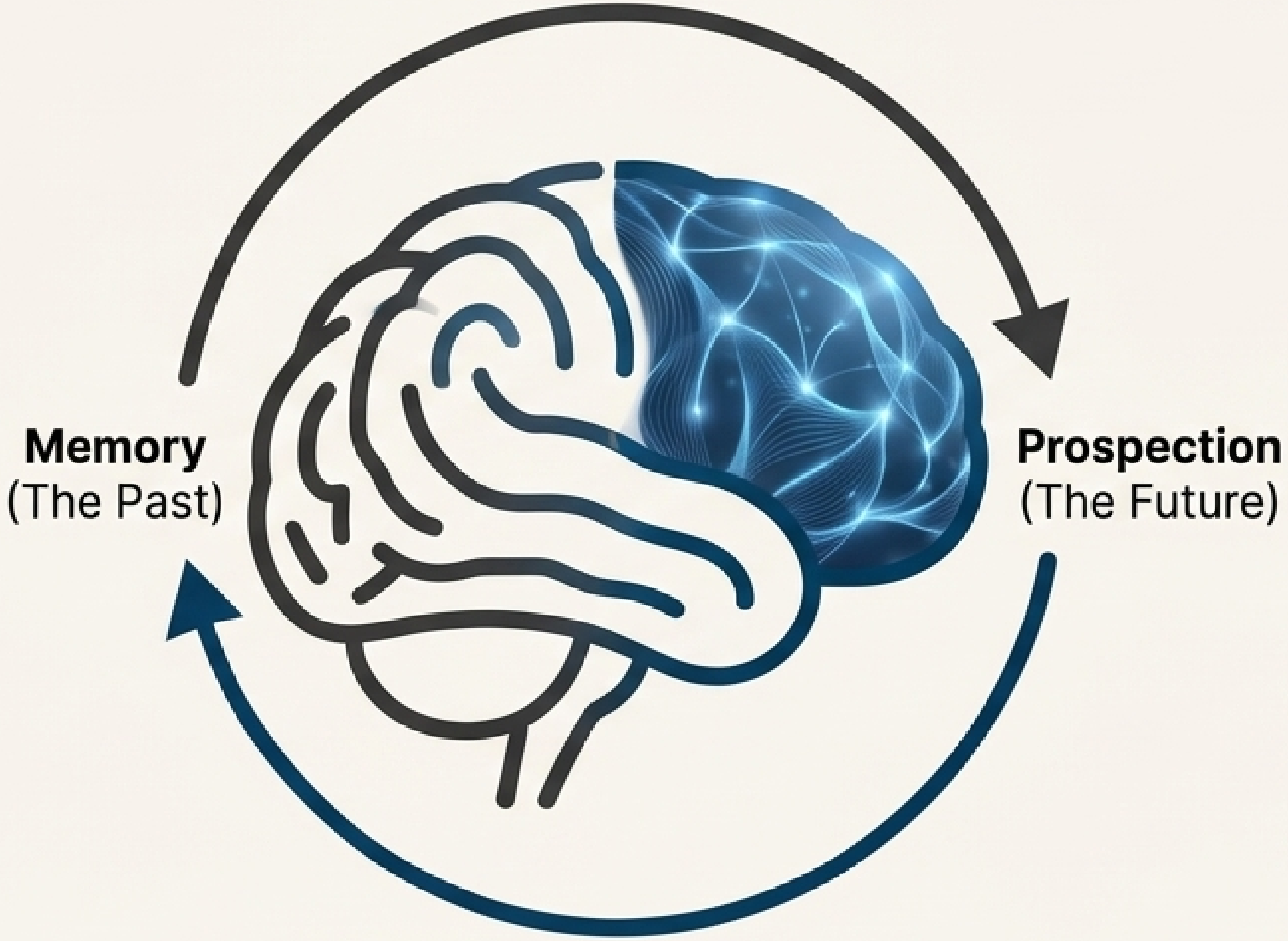
However, they operate under the assumption that the past contains sufficient information to determine the present. This is

often suboptimal in volatile or ambiguous real-world scenarios.

Even when Transformers or bidirectional RNNs use 'future' tokens, this access is syntactic and retrospective—they require the entire sequence to be known.

They do not engage in imaginative forecasting of an *unknown* future.

Human Intelligence Fuses Memory of the Past with Simulations of the Future



- Cognitive science reveals that human intelligence relies heavily on "**prospective cognition**" or "**episodic future thinking**." We constantly run mental simulations to anticipate outcomes and plan.
- This ability is not separate from memory; research shows that brain regions like the hippocampus and prefrontal cortex are active in both remembering the past and imagining the future.
- We can embed this principle into our models, treating simulated futures as structured, dynamic '**soft priors**' that guide decision-making in the present.

We Propose a New Paradigm: Future-Informed Machine Learning (FIML)

We extend Knowledge-Informed Machine Learning (KIML), which uses external facts and rules, to include a new kind of knowledge: dynamically forecasted futures. This creates a clear conceptual path from a broad paradigm to a specific, novel architecture.

KIML (Knowledge-Informed Machine Learning)

↓

→ **FIML** (Future-Informed Machine Learning)

↓

→ **FINN** (Future-Informed Neural Networks)

↓

→ **FI-RNN** (Future-Informed Recurrent Neural Networks)

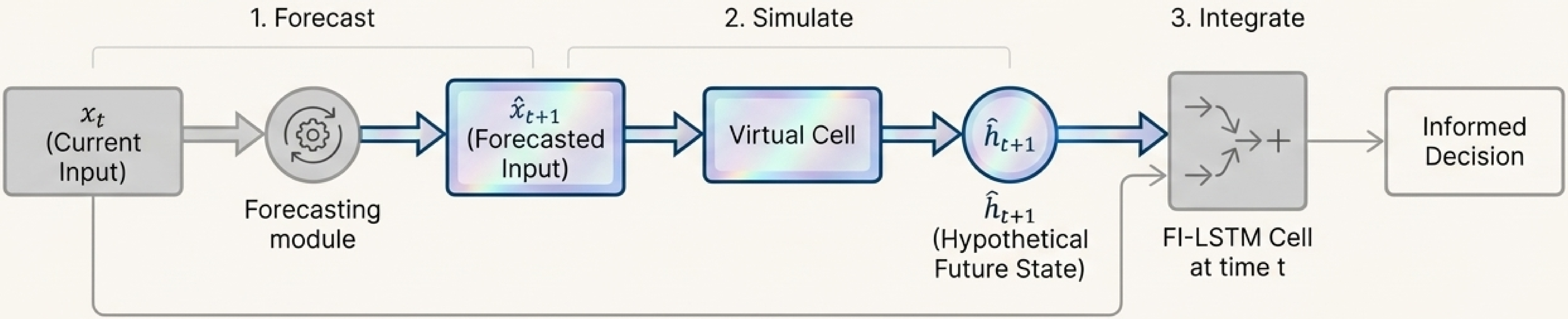
↓

→ **FI-LSTM** (Future-Informed LSTM)

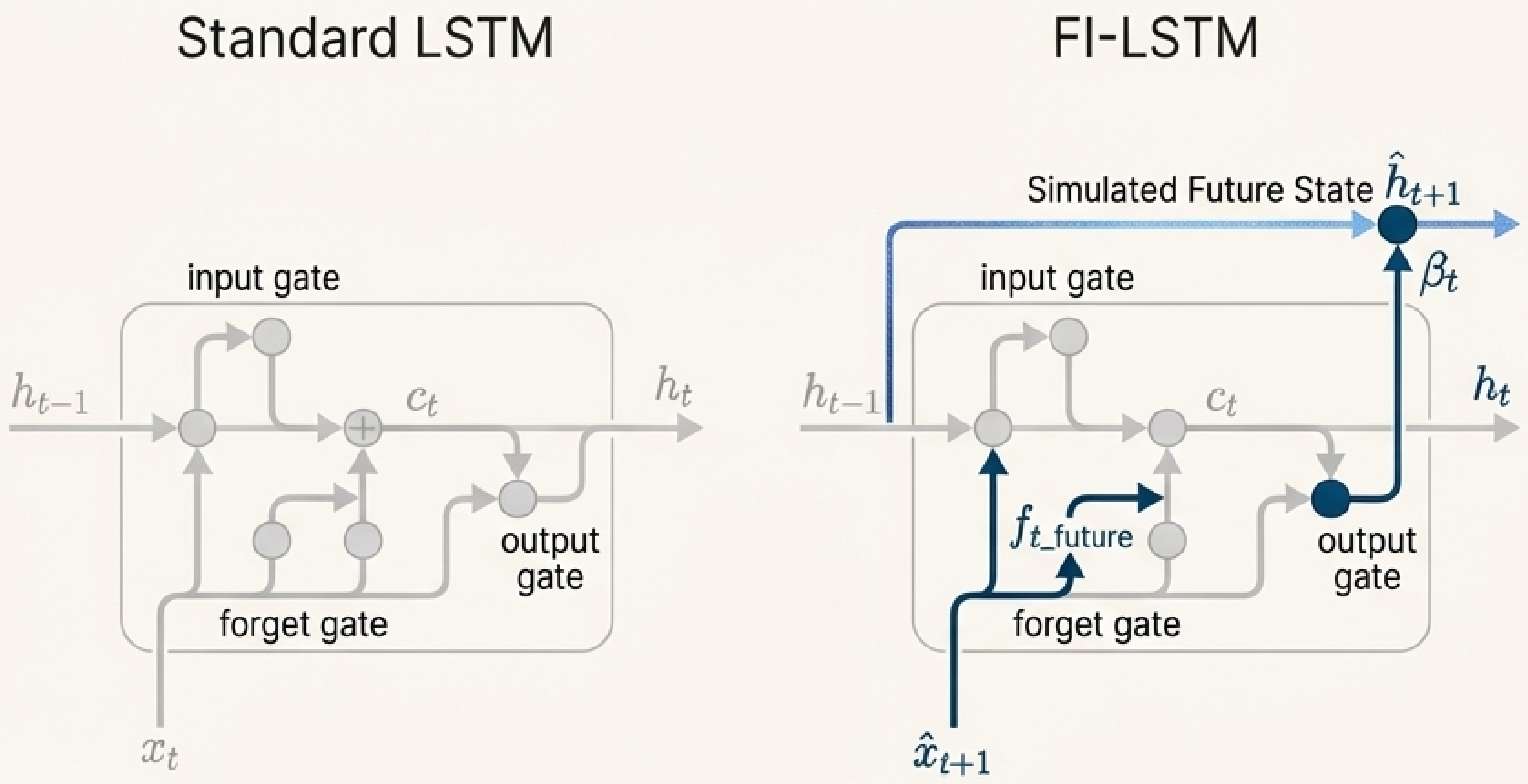
Side note: 'FINNs' is a term that not only reflects their predictive spirit but also celebrates the origins of this work – Finland.

The FI–LSTM: An Architecture That “Peeks Ahead” by Simulating the Next Step

The model’s logic is “Forecast-Next-Input-Then-Decide.”



Inside the FI-LSTM: New Gates to Manage Imagined Futures



New Components

$\hat{x}_{t+1}$: Forecasted Input
The model's best guess for the next input in the sequence.

$\hat{h}_{t+1}$: Simulated Future State
A hypothetical hidden state based on the forecasted input.

f_{t_future}: Forget Future Gate
A learned gate that controls how much influence the simulated future has. This creates a functional symmetry with the standard forget gate.

β_t: Blend Gate
A learned gate that dynamically balances the contributions of the present-focused state and the future-informed state.

NotebookLM

The Computational Pipeline Blends Past, Present, Present, and Future

$$\hat{x}_{t+1} = F(x_t, h_{t\_\text{temp}})$$

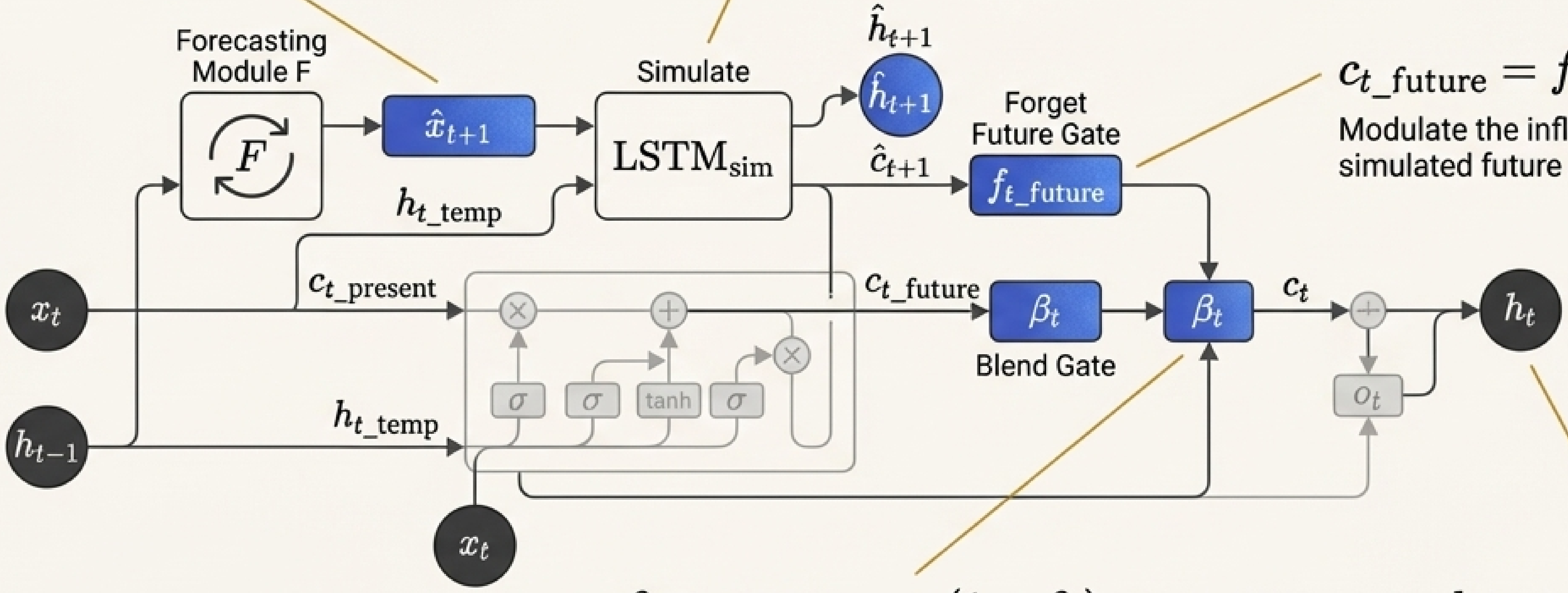
Predict the next input using a forecasting module F.

$$(\hat{h}_{t+1}, \hat{c}_{t+1}) = \text{LSTM}_{\text{sim}}(\hat{x}_{t+1}, h_{t\_\text{temp}}, c_{t\_\text{present}})$$

Compute a hypothetical future state by processing the forecasted input.

$$c_{t\_\text{future}} = f_{t\_\text{future}} \odot \hat{c}_{t+1}$$

Modulate the influence of the simulated future memory.



$$c_t = \beta_t \odot c_{t\_\text{present}} + (1 - \beta_t) \odot c_{t\_\text{future}}$$

Dynamically interpolate between the present memory and the future-informed memory. The model learns how much to trust each source.

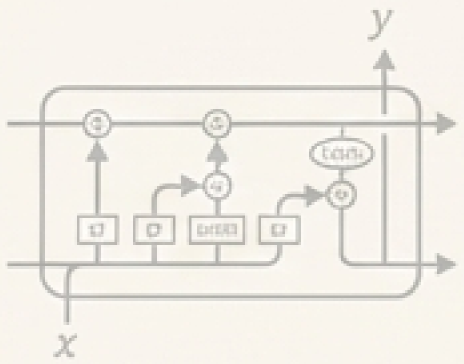
$$h_t = o_t \odot \tanh(c_t)$$

Compute the final hidden state, now informed by both past and simulated future.

FI-LSTM’s Advantage Emerges When Past and Present Signals Are Weak

In experiments, FI-LSTM significantly outperforms standard LSTM in scenarios where the correlation between the input sequence $X(t)$ and the target $Y(t)$ is low.

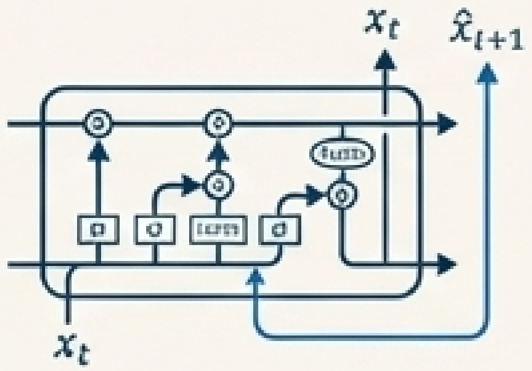
High Correlation



Performs well.

The past of X is a strong predictor for Y.

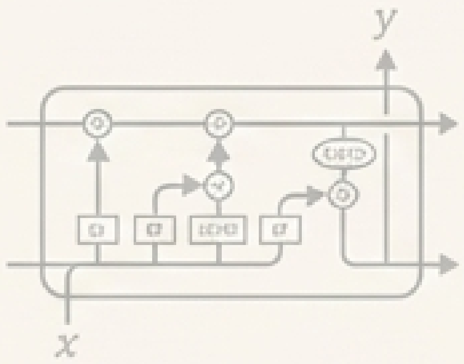
High Correlation



Performs similarly.

Extra complexity offers little benefit.

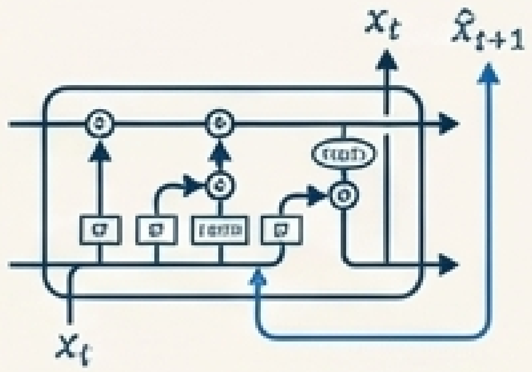
Low Correlation



Struggles.

The past of X is a weak signal.

Low Correlation



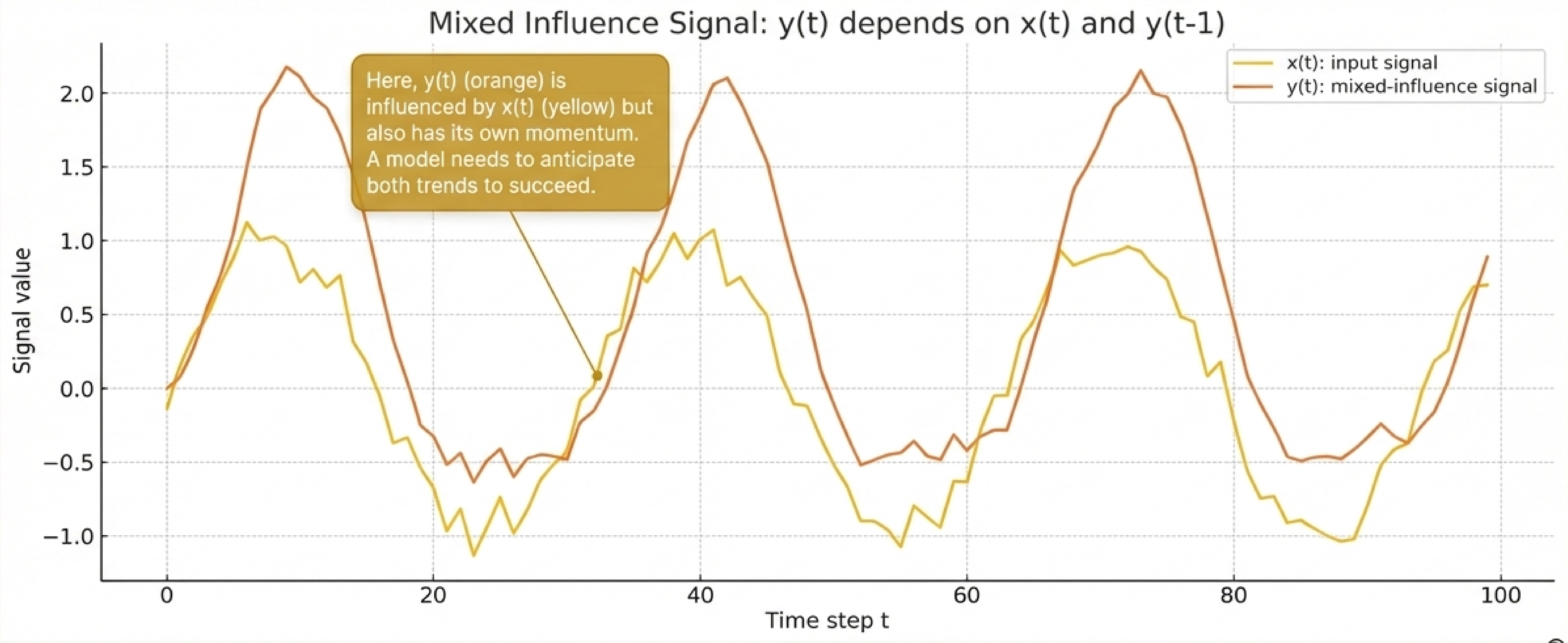
Excels.

Auxiliary signal from forecasting $\hat{x}_{t+1}$ provides crucial context.

Visualizing Performance on a Mixed-Influence Signal

Consider a target signal y(t) that depends on both an external input x(t) and its own past values (inertia). $y(t) = \alpha \cdot y(t-1) + \beta \cdot f(x(t)) + \ldots$

Predicting y(t) requires understanding both the external driver x and the internal momentum of y. This is precisely the scenario where forecasting both the future input $\hat{x}_{t+1}$ and the target's own trend $\hat{y}_t$ provides a decisive advantage.

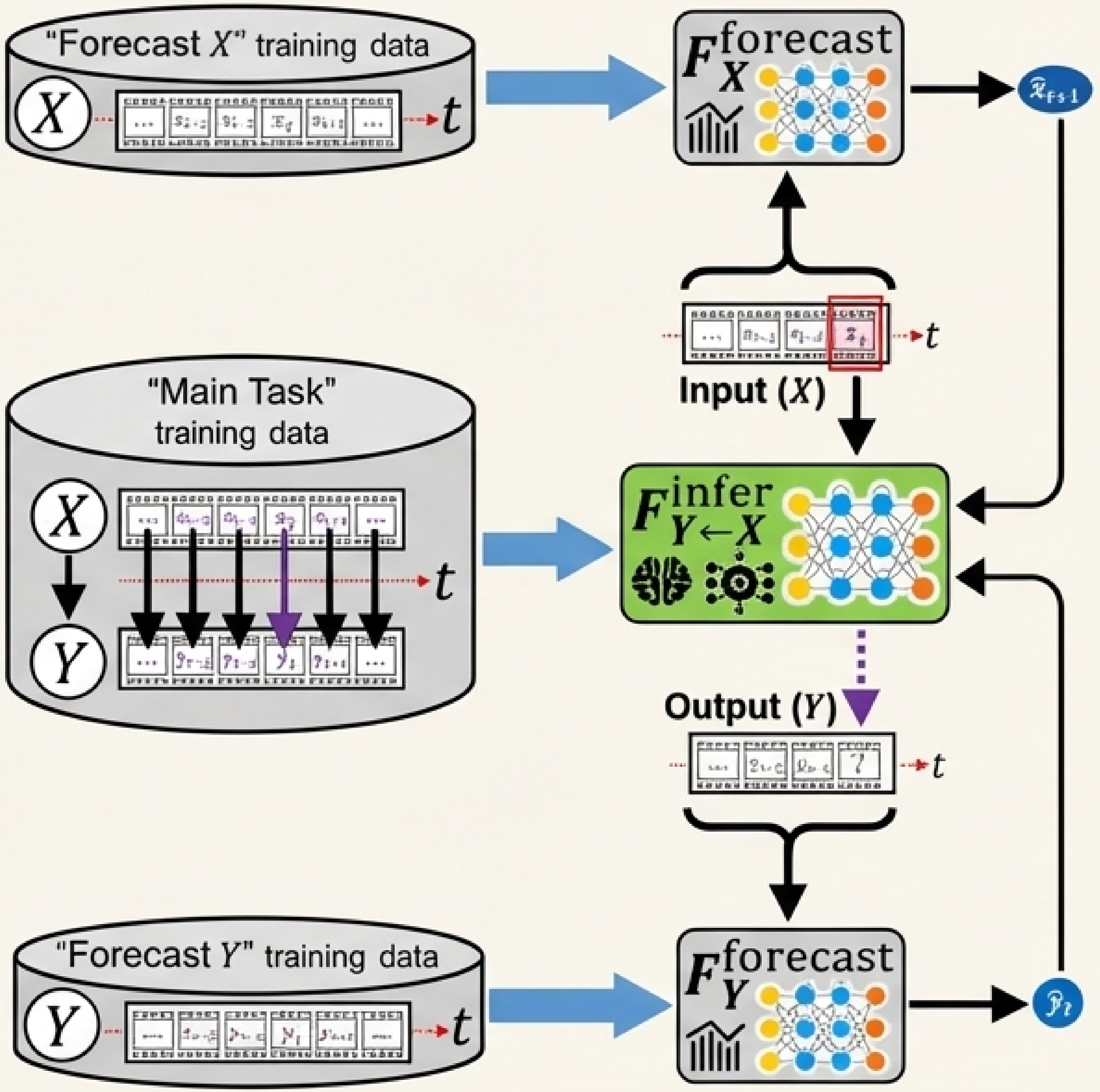


The Next Step: A Dual-FI Model That Forecasts Both Input and Output

We can enhance the architecture by providing it with two streams of future simulation:

1. **Forecasted Input ($\hat{x}_{t+1}$):** Simulates where the *environment* or external signal is going.
2. **Forecasted Output ($\hat{y}_t$):** Simulates the target's own *inertia* or internal dynamics, providing a 'second opinion' based on its own trajectory.

Analogy: This is cognitively natural. Humans make decisions based on what we observe (x_t), our expectations about the world ($\hat{x}_{t+1}$), and **our own habits** or momentum ($\hat{y}_t$).

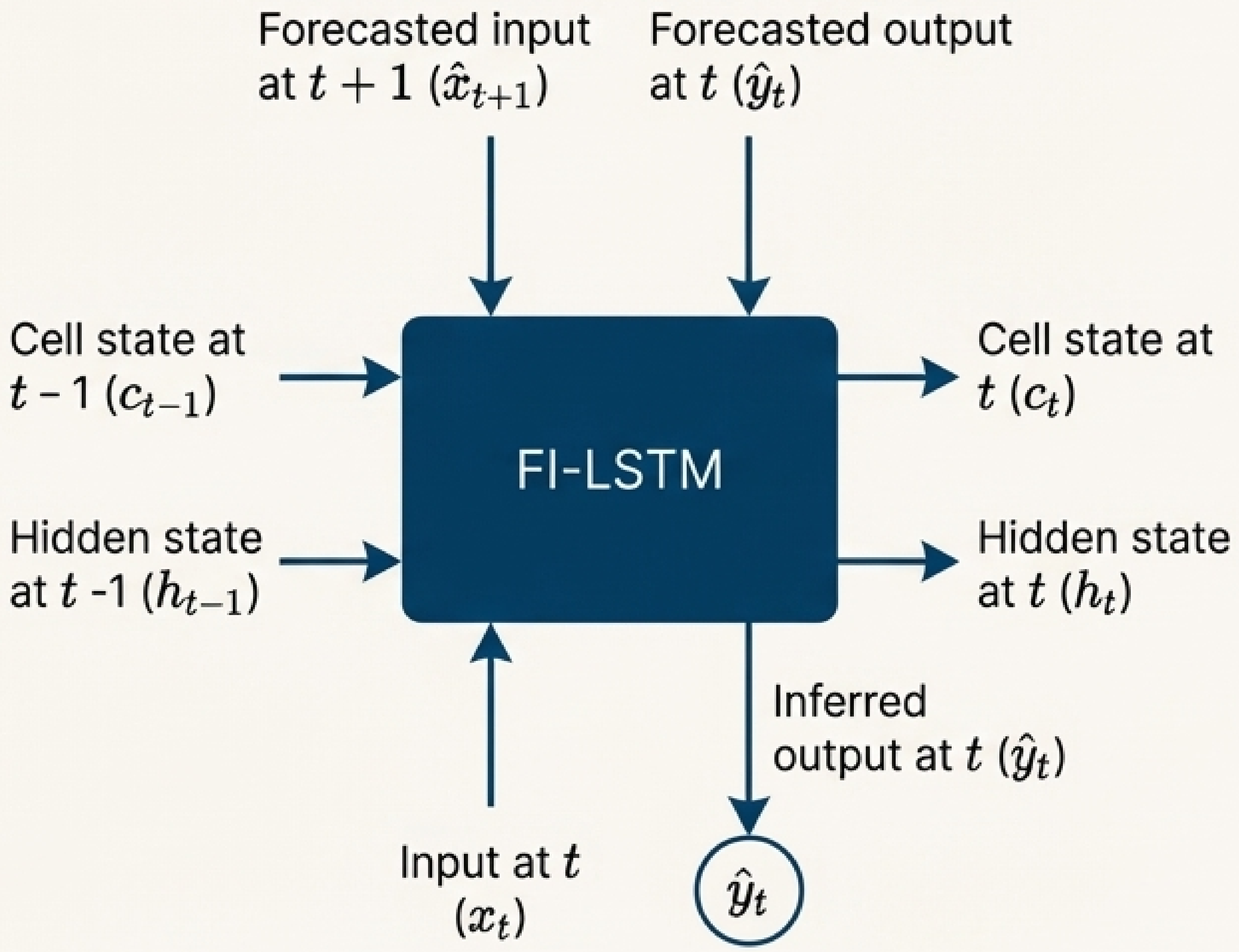


Integrating Dual Forecasts into the FI-LSTM Cell

The FI-LSTM cell is modified to accept three inputs at each time step t:

- The current input x_t.
- The forecasted *next* input $\hat{x}_{t+1}$.
- The forecasted *current* output $\hat{y}_t$.

Modified Computation: The gate activations now become functions of all three signals, e.g., $f_{t+1} = \sigma(W_f \cdot [h_{t\_temp}, \hat{x}_{t+1}, \hat{y}_t] + b_f)$. This allows the network to learn complex interactions between external drivers and internal dynamics.

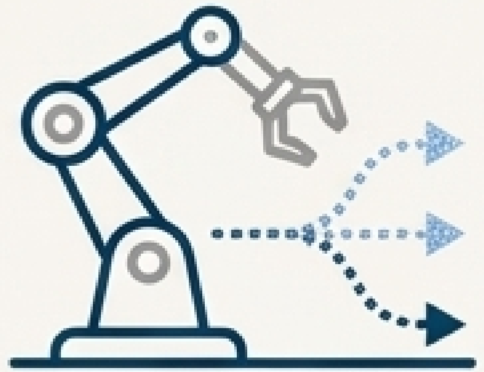


FI-LSTM’s Real-Time Foresight Is Fundamentally Different from Existing Methods

Model	How it Uses the Future	Limitation
Bidirectional RNN	Accesses the actual future when the full sequence is already known.	Retrospective. Cannot operate in online scenarios.
Transformer	Attends to the full sequence (if not causally masked). Can ‘look ahead’ at known data.	Cannot imagine or simulate a truly unknown future.
FI-LSTM	**Operates in truly online settings. Actively forecasts and simulates the unknown future on the fly.**	**Integrates imagined possibilities, not just observed future data points.**

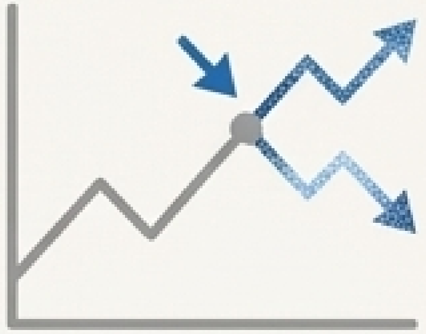
FI-LSTM Is a Concrete Step Toward Proactive, Anticipatory AI

By embedding future-aware reasoning into the core of sequential modeling, modeling, this paradigm opens new possibilities for systems that must operate under uncertainty.



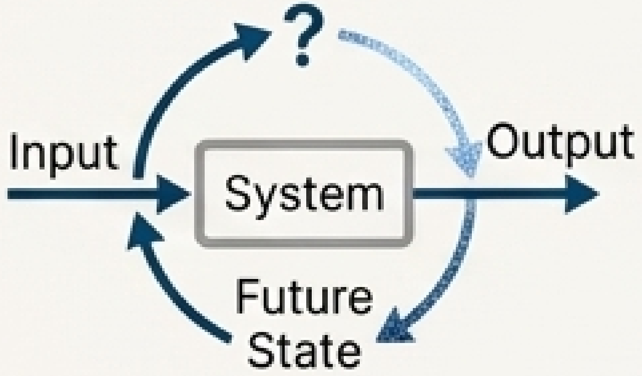
Robotics

Agents that simulate hypothetical futures before acting.



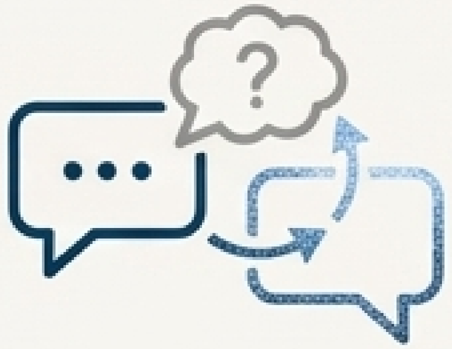
Financial Forecasting

Models that anticipate market volatility instead of just reacting to it.



Control Systems

Real-time decision-making that accounts for likely future states.

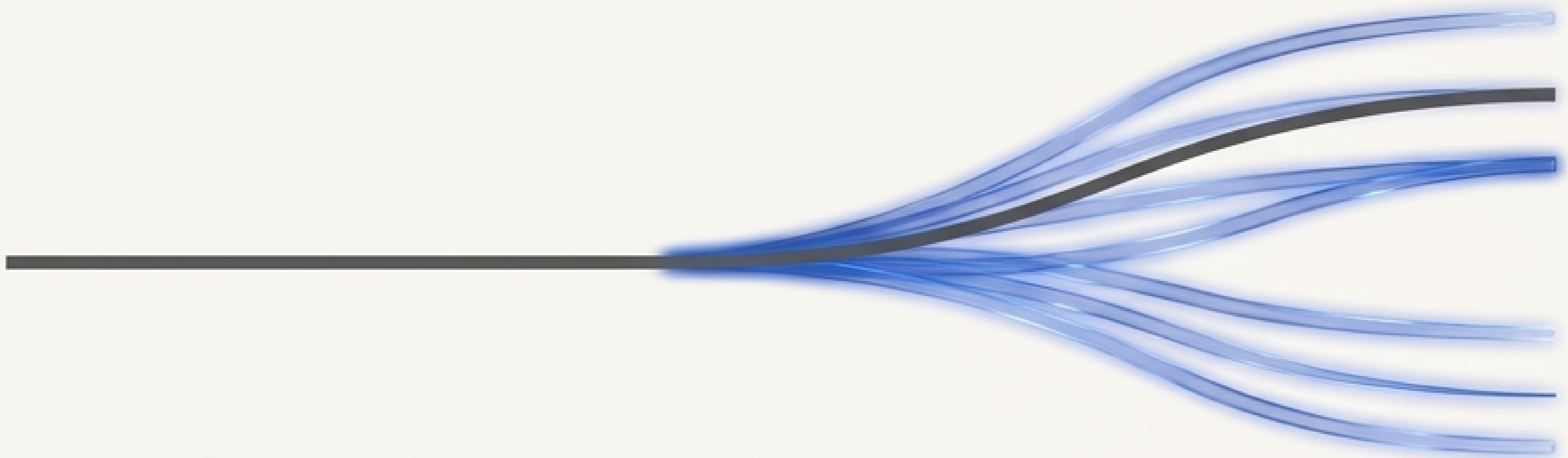


Dialogue Systems

Planning conversational turns by anticipating the user's response.

"This allows models to 'imagine the future and be shaped by it' in every moment."

The Future of Learning Is Not Just About Memory, But Expectation



FIML elevates the role of **future prediction from an auxiliary task to a core structural principle of model design**.

This invites a **fundamental shift in how we build intelligent systems**: moving from models that simply react to the past to agents that are actively guided by **credible simulations** of what lies ahead.