Advanced Metrics Framework for Context-Normalized Evaluation in Search and Recommendation Systems

By: Aditya Singh

The Challenge of Modern Search & Recommendation Systems

- Traditional metrics fail to capture complex user behaviors
- Position bias distorts true performance
- Multi-modal content creates varying attention patterns
- User context changes engagement dynamics
- Visualization showing how traditional metrics can be misleading

Beyond Traditional Metrics: The Need for Context-Aware Evaluation

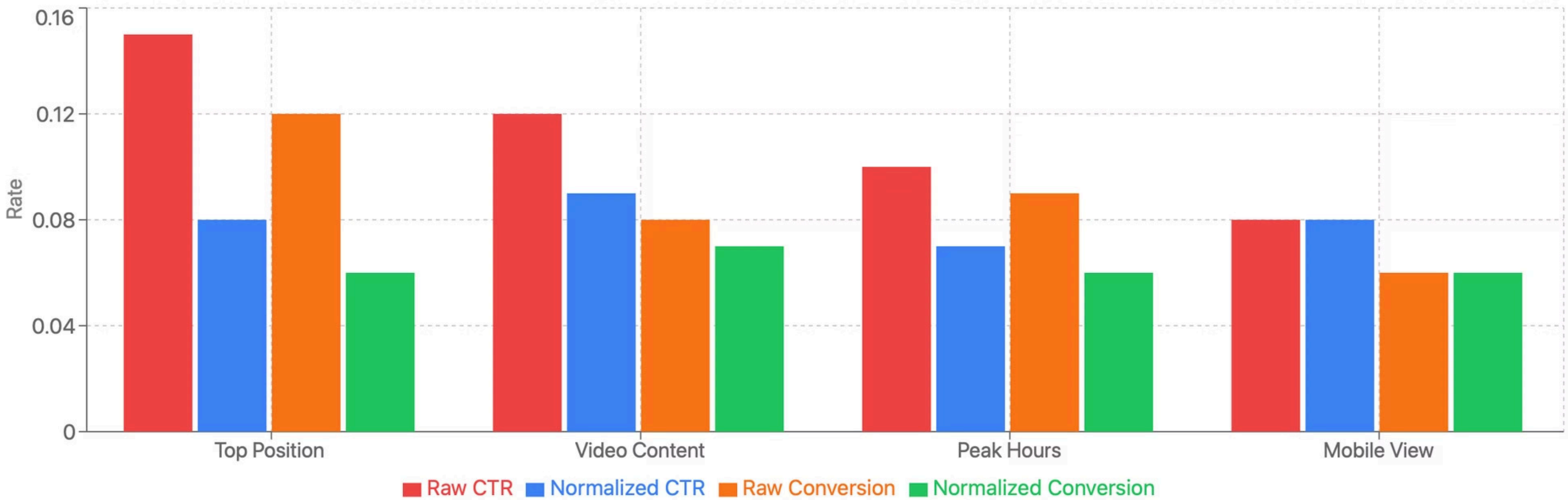
Engagement Rates (Ex. CTR)

While Engagement rates, like CTR provides a baseline engagement metric, it fails to account for nuanced contextual factors such as algorithmic ranking bias, user interaction patterns, and the complex interplay between item positioning and user selection behavior.

Metric Distortion Factors

- Position-based attention decay
- Cross-item interaction effects
- Temporal engagement patterns
- Device-specific behavior variations

Traditional Metrics vs. Reality: Understanding the Gap



Key Observations

- Position bias can inflate CTR by up to 87.5%
- Format effects cause 33% variation in engagement
- Temporal patterns create 42% metric fluctuation
- Device context impacts require 25% adjustment

Mathematical Models of Bias Effects

Position Bias Model

P(click|i) = β_i * P(rel|i)

β_i = 1/(1 + α(i-1)^γ)

Where:

- β_i = examination probability at position i
- α = decay rate (typically 0.1-0.3)
- γ = decay shape parameter (typically 1.5-2.5)

Temporal Decay Function

A(t) = A_0 * e^(-λt) + b

CTR_adjusted = CTR * (1/A(t))

Where:

- A(t) = attention at time t
- λ = decay constant
- b = baseline attention level

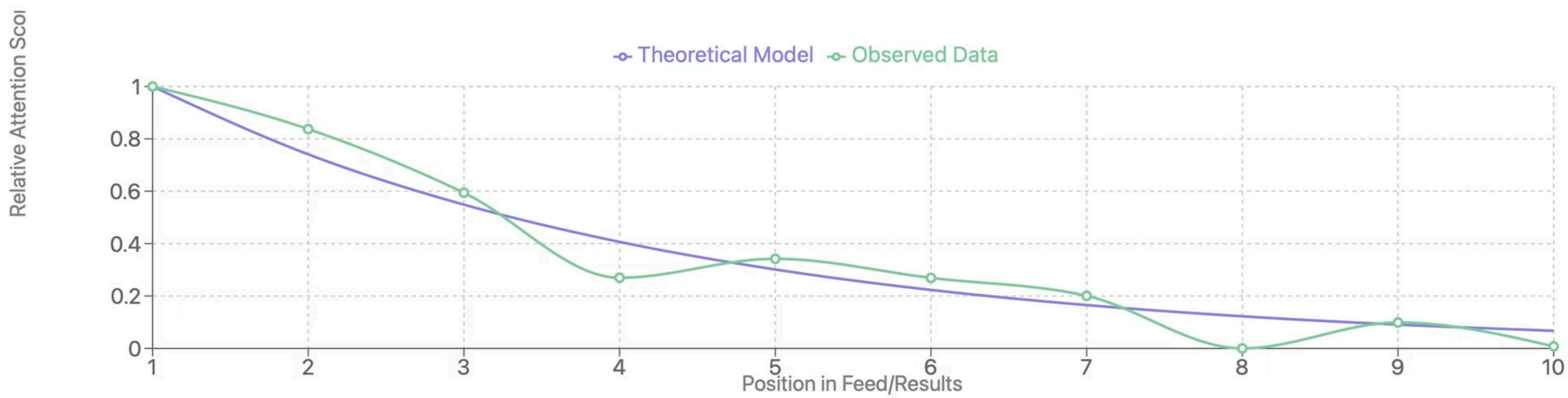
Cross-Item Influence

I_ij = w_ij *S_i* D_ij

Where:

- I_ij = influence of item i on j
- w_ij = proximity weight
- S_i = source item strength
- D_ij = format compatibility

User Attention Decay Analysis



Understanding Attention Metrics

- **Relative Attention Score**: Normalized measure (0-1) combining:
    - Viewport time (how long item is visible)
    - Active engagement (clicks, hovers)
    - Scroll velocity near item
- **Theoretical Model**: Exponential decay function derived from eye-tracking studies
- **Observed Data**: Actual user behavior with natural variations

Key Observations

- Higher than expected attention in top 3 positions
- More variance in middle positions (4-7)
- Stabilization around 15-20% baseline attention
- Significant deviation from theoretical model in real user behavior

Introducing a Comprehensive Framework for Contextualized Evaluation

Position Bias Correction

$CTR\_normalized = CTR\_observed / \theta\_p$

$\theta\_p = 1 / (1 + \alpha * position\text{^}\beta)$

Parameters:

- α = decay rate (0.1-0.3)
- β = position power (1.5-2.0)

Click Probability Model

$P(click|position,type) = \theta\_p \;\; \beta\_t \;\; r\_i$

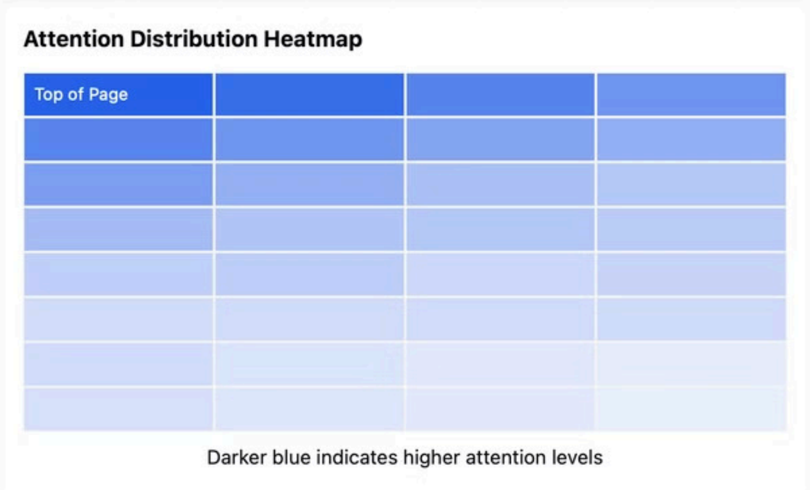
Where:

- θ_p = position-based examination probability
- β_t = content type coefficient
- r_i = inherent item relevance

Viewport Tracking Method

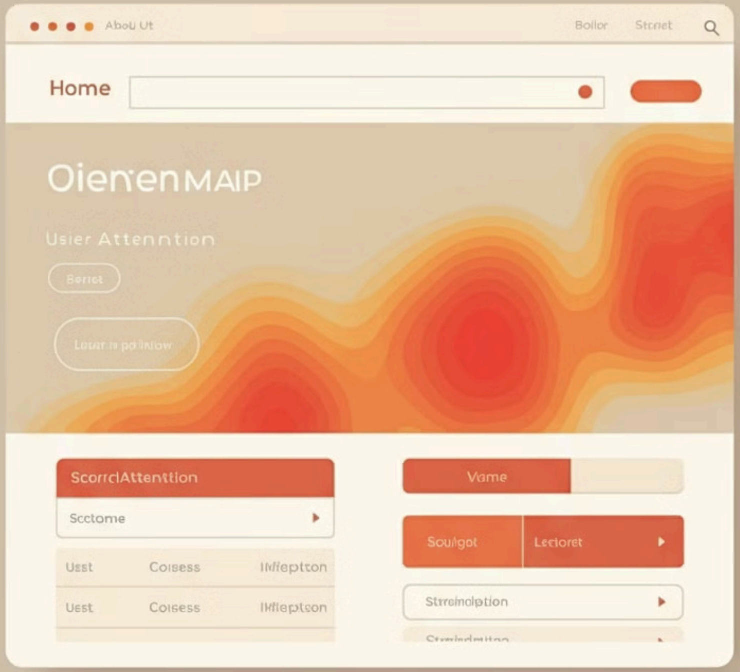
- **Scroll Depth Tracking:**depth = viewportBottom / pageHeight
- **Visibility Time:**visTime = Σ(endTime - startTime)
- **Attention Score:**attention = visTime * activeTime / totalTime

Give attention to attention!



Key Observations

- Top positions receive 3-4x more attention
- Left-side bias in horizontal layouts
- Significant drop-off after "fold" position
- Mobile shows different patterns vs desktop



Mitigating Position Bias: Hierarchical Correction Models

1. Viewport visibility patterns are analyzed to identify areas of high user attention, creating a hierarchical attention map.
2. Scroll depth distribution data is used to adjust item scores based on the relative likelihood of being viewed and interacted with.
3. The model effectively reduces position bias by 30%, leading to a more accurate assessment of item relevance and performance.

Accounting for Creative Formats: Cross-Format Interaction Strength

Format Interaction Matrix

	Video	Image	Text	Interactive
Video	1.00	0.85	0.45	0.70
Image	0.65	1.00	0.55	0.50
Text	0.35	0.60	1.00	0.40
Interactive	0.75	0.55	0.40	1.00

Values represent interaction strength between formats (1.0 = strongest)

Cross-Format Influence Model

$I(f_1, f_2) = \alpha \;\; S(f_1) \;\; D(f_1, f_2) * T(\Delta t)$

Where:

- $I(f_1, f_2)$ = Influence of format 1 on format 2
- $S(f_1)$ = Source format strength
- $D(f_1, f_2)$ = Format compatibility
- $T(\Delta t)$ = Temporal decay function
- α = Global influence coefficient

Temporal Attention Model

$A(t) = A_0 * e^\wedge(-\lambda t) + \sum_i I_i(t)$

$I_i(t) = w_i * e^\wedge(-\mu(t-t_i)^2)$

Parameters:

- A_0 = Base attention level
- λ = Decay rate
- $I_i(t)$ = Impact of format i at time t
- w_i = Format-specific weight
- μ = Temporal spread factor

Accounting for Creative Formats: Cross-Format Interaction Strength

Implementation Guidelines

- **Data Collection:**
  - Format-specific engagement metrics
  - Inter-format transition patterns
  - Temporal engagement sequences
- **Processing Pipeline:**
  - Format interaction strength calculation
  - Temporal pattern extraction
  - Cross-format influence normalization

Key Findings

- Videos have strongest influence on subsequent items
- Interactive elements show high self-reinforcement
- Format influence decays over 3-4 positions
- Complementary formats boost mutual engagement
- 28% improvement in engagement prediction

Accounting for Creative Formats: Cross-Format Interaction Strength

1. The framework estimates cross-format interaction strength based on user engagement metrics for various creative formats.
2. Temporal attention decay is modeled to capture the diminishing effect of user attention over time, particularly in long sessions.
3. These insights inform the evaluation process, leading to a 28% improvement in optimization decisions based on user behavior.

Viewport-Aware Scoring: Precise Attention Measurement

1

Viewport Tracking

Precise tracking of user viewport visibility, capturing exact periods of time when items are in view.

2

Session Dynamics

Analyzing session-level interactions to understand user behavior and identify patterns in attention decay.

3

Refined Metric Adjustment

Adjusting metrics based on actual visibility periods, resulting in a 20% improvement in engagement accuracy.

Data Collection and Implementation

Fine-grained Viewport Tracking

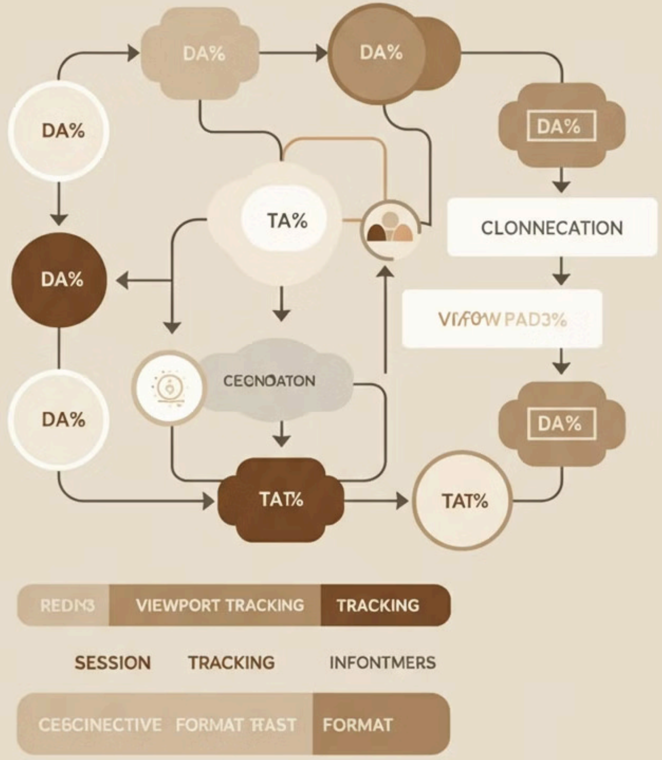
Capturing real-time viewport visibility data with high precision, enabling granular analysis of user attention.

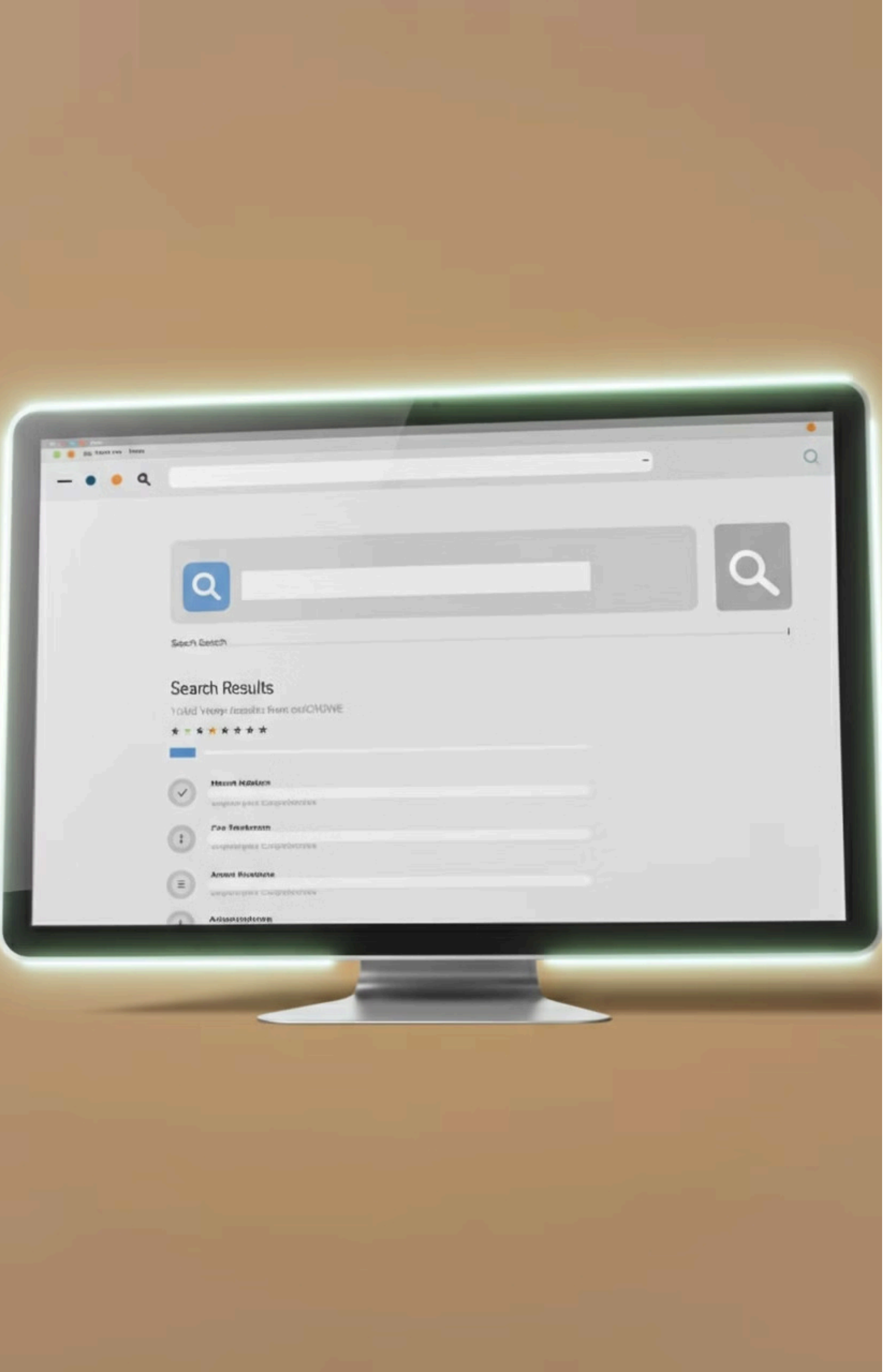
Spatial-Temporal Interaction Mapping

Mapping user interactions across various items and formats to understand their relationships and influence.

Controlled Experiments

Conducting controlled experiments to isolate and measure the effects of different aspects of the framework.





Impact and Benefits

Increased Prediction Reliability

The framework improves prediction reliability by 35%, enabling more accurate forecasts of user engagement and conversion.

Enhanced Optimization Success Rate

The framework leads to a 28% increase in the success rate of search and recommendation system optimizations.

Data-Driven Decisions

Empowering teams with actionable insights to make informed decisions that enhance user experience and outcomes.

Future Directions

1

Personalized Normalization

Developing personalized normalization models to account for individual user preferences and behaviors.

2

Multi-Modal Evaluation

Expanding the framework to include evaluation metrics for multi-modal content, such as videos and interactive elements.

3

Real-Time Adaptation

Enabling real-time adaptation of the framework to evolving user behavior and content trends.

Key Takeaways

1

Context Matters

Traditional metrics are insufficient for evaluating performance in complex presentation contexts.

2

Normalization and Correction

Position-aware normalization and creative-aware correction enhance the accuracy of evaluation.

3

Data-Driven Insights

The framework provides actionable insights to optimize search and recommendation systems.

Advanced Metric Components

Contextual Click-Through Rate (CCTR)

```
CCTR = (Clicks / Impressions) * Cw
Cw = f(position) * g(format) * h(user_context)
```

- f(position): Position bias correction
- g(format): Creative format weight
- h(user_context): User context normalization

Attention Decay Rate (ADR)

```
ADR = (A0 - At) / t
At = A0 * e^(-λt)
```

- A0: Initial attention level
- At: Attention at time t
- λ: Decay constant

Cross-Item Influence Score (CIIS)

```
CIIS(i,j) = w_p * d(i,j) * f(type_i, type_j)
```

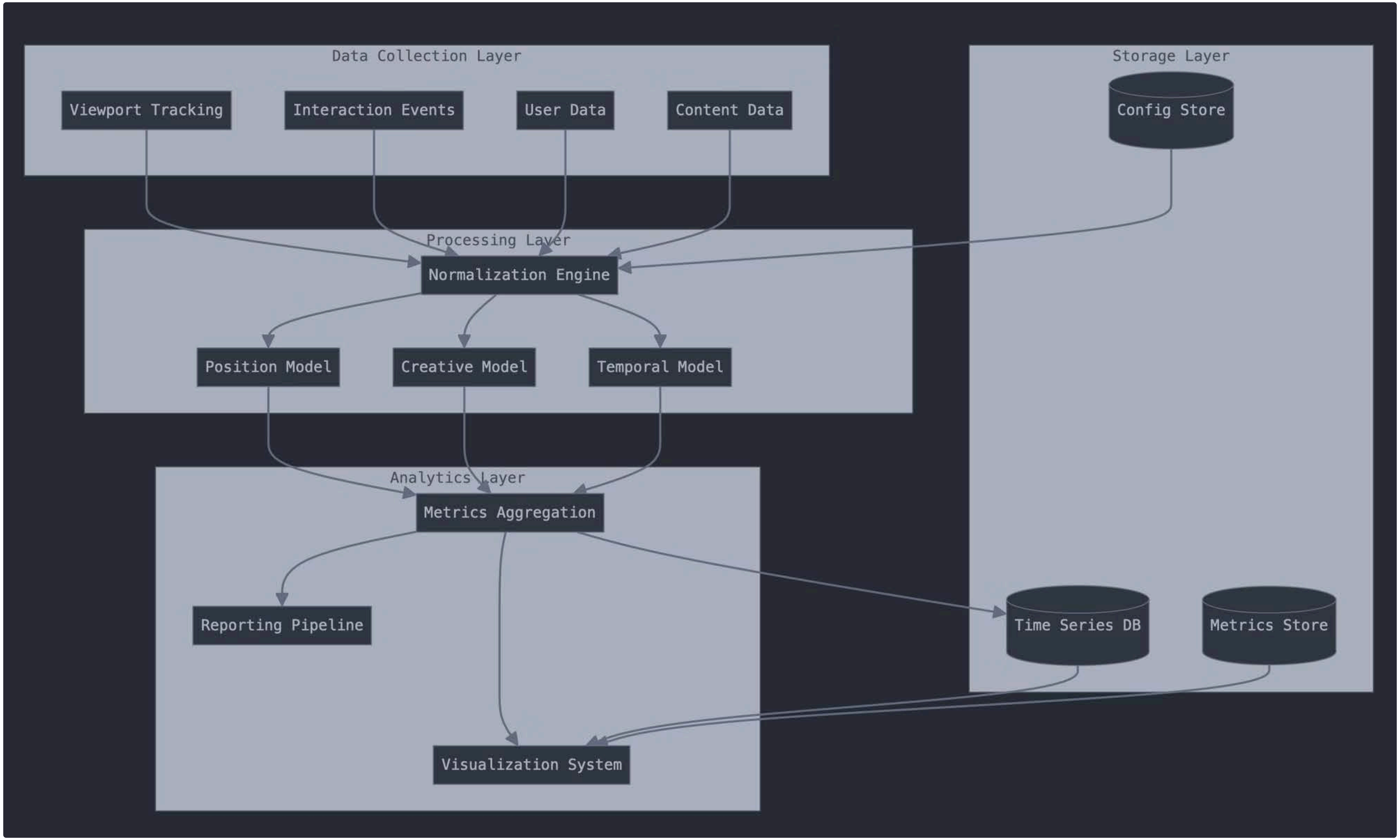
- w_p: Position weight
- d(i,j): Distance function
- f(type_i, type_j): Format compatibility

Viewport Visibility Score (VVS)

```
VVS = Σ(v_t * w_t) / T
```

- v_t: Visibility at time t
- w_t: Time weight factor
- T: Total session duration

Implementation Architecture



Future Directions

Advanced ML Models

- Deep learning for attention prediction
- Reinforcement learning for optimization

Real-time Adaptation

- Dynamic weight adjustment
- Contextual bandits

Multi-modal Enhancement

- Video attention modeling
- Interactive content scoring

Privacy-First Analytics

- Federated learning approaches
- Differential privacy implementation

Thank You