Los Angeles | September 17, 2025

# AWS SUMMIT

DEV203

# Crafting Real-Time Personalized Game Highlights using AI with AWS

**Gaganjot Kaur Kang**
(she/hers)
Staff Software Engineer
Sony Interactive Entertainment

# Agenda

- The Content Creation Bottleneck
- The AI-powered Highlight Factory with AWS
- High-Level Architecture
- Deep dive
- Results & Benefits

# The Content Creation Bottleneck

- The best gaming moments are sometimes hidden in 5+ hour long gameplay, at exactly 4:12.
- Viewers are drowning in gameplay but starving for highlights.


**Creator Economy Market projected to rise from $212.3B in 2024 to $894.8B by 2032 [Source: Yahoo]**

# The Content Creation Bottleneck

Why 99% of gaming moments disappear forever?

## Opportunity

- Game streamers generate hours of content daily.
- Viewers want engaging clips right now.

## Challenge

- Streams are long, highlight discovery is manual.
- Viewers have short attention spans.

## What if?

- AI could instantly turn raw gameplay into personalized highlights tailored to each viewer’s interests?

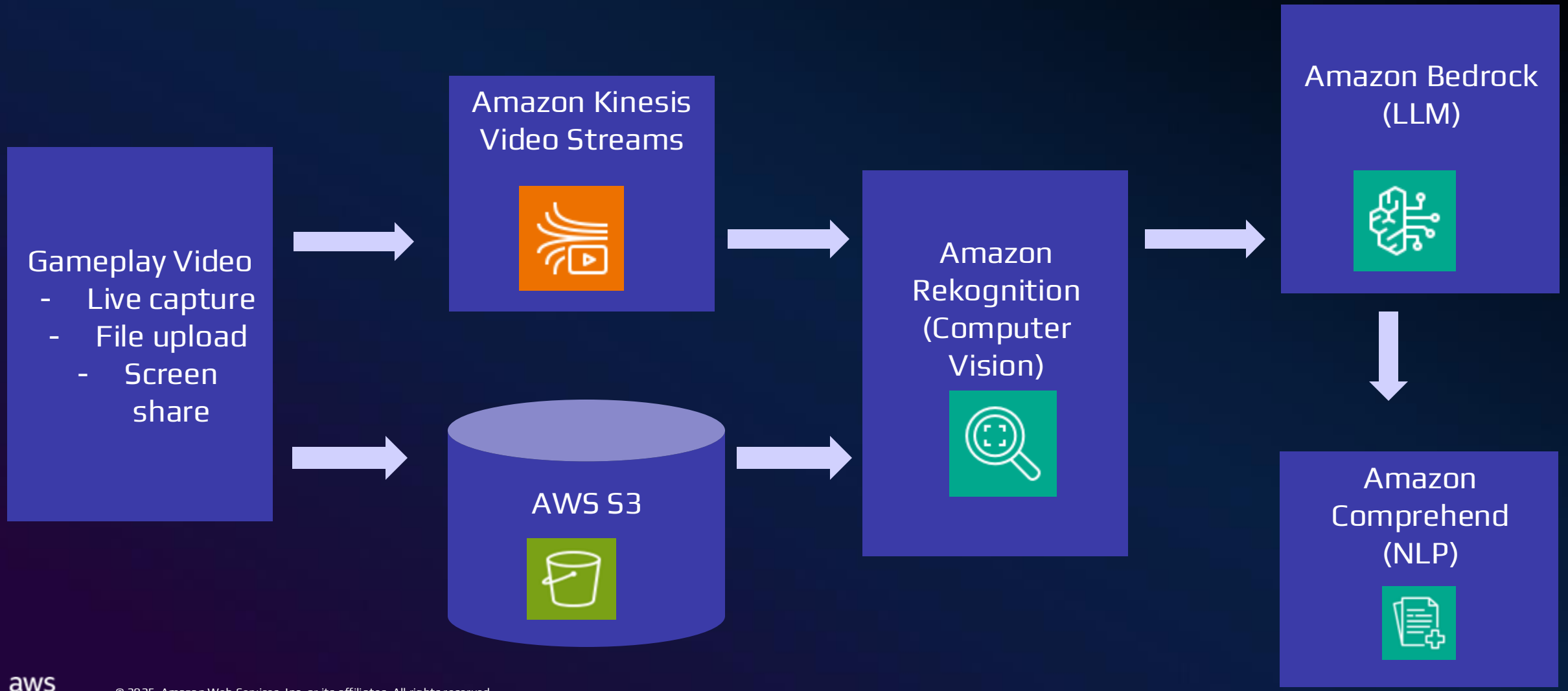
# The AI-powered Highlight Factory using AWS

| | |
|---|---|
| **Ingestion** | Ingest video stream data to Amazon Kinesis Video Streams |
| **Analysis** | Analyze the streams with Amazon Rekognition (event/face/object detection), Amazon Comprehend (contextual metadata) |
| **Decision** | Summarise the clips using Amazon Bedrock (LLM for summaries) |
| **Personalization** | Personalize clips based on past user data using Amazon Personalize (User preference, smart scoring, diversity) |
| **Delivery** | Deliver using Amazon CloudFront, Websocket, Real-time API |

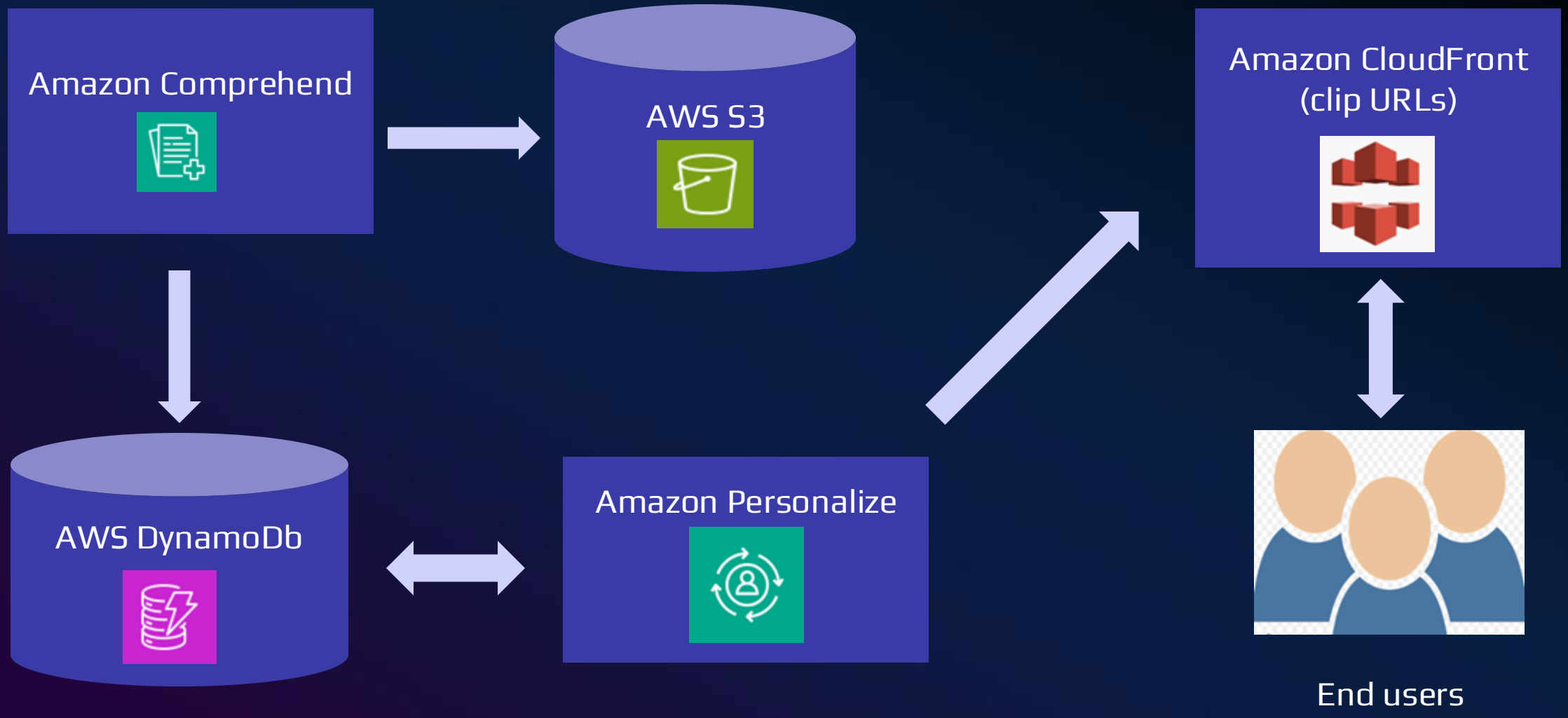
# Key Architectural Principles

- **Event-Driven**: S3 uploads trigger Lambda functions
- **Cost-effective**: Pay only for actual processing time
- **Serverless scale:** Handle 1 or 1000s seamlessly
- **AI-Enhanced**: Every highlight enriched with metadata
- **Real-Time**: WebSocket updates for live notifications

# High Level Architecture



# High Level Architecture



# High Level Architecture

AWS Lambda

- The video analysis Lambda function orchestrates the entire AI pipeline: it calls Rekognition for computer vision, waits for the results, then sequentially invokes Bedrock for contextual enhancement and Comprehend for sentiment analysis, all within the same function execution.

# Event detection that understands the game story

## Amazon Rekognition + Bedrock

- Rekognition analyzes entire video and performs label detection, person tracking.
- The video analysis algorithm finds 5-second highlight windows with 3+ interesting events.
- Amazon Bedrock transforms raw computer vision data into meaningful, contextual insights using Claude 3 Haiku; turning simple object detection into rich gaming narratives.
- Processing 10-30 seconds per highlight batch vs. minutes for larger models
    - Excitement Level: AI rates content engagement (1-10)
    - Play Classification: Understands gaming context
    - Natural Titles: Human-readable descriptions

Without Bedrock:

- Generic labels: "Ball", "Person", "Sports"

With Bedrock:

- Contextual insights: "Goal celebration with crowd reaction"

# Context understanding with Personalization

## Amazon Comprehend and Personalize

- Comprehend analyzes the AI-generated text from Bedrock.
- It is not analyzing the video directly; it is understanding the language descriptions.
- This creates a feedback loop where AI-generated content gets further AI analysis.
- Sentiment analysis tells us if a highlight title sounds exciting, neutral, or negative.
  - For example: 'Spectacular Goal!' gets POSITIVE sentiment with 95% confidence.
- Entity extraction identifies specific gaming terms, team names, and player references.
  - From 'Manchester United scores amazing goal' it extracts 'Manchester United' as ORGANIZATION.
- We automatically classify content into emotional tones and gameplay types.
- This creates personalization dimensions beyond basic sport categories.
- Users can be matched by emotional preference - some prefer 'exciting', others 'intense'.

# Production Readiness

- **Observability:** CloudWatch metrics/alarms; distributed tracing on Lambdas.
- **Cost controls:** tiered storage (S3 IA/Glacier for older clips), batch Transcribe for non-urgent ASR.
- **Governance:** IAM least privilege; human-in-the-loop review list for “borderline” clips; content filters.

# PoC Link

- GitHub Repo: https://github.com/gkang050/game-spotlights-ai
- Programming Language: Node.js, JavaScript
- IDE used: Amazon Kiro


**LinkedIn**: https://www.linkedin.com/in/gaganjotkaurkang/
Github Handle: gkang050