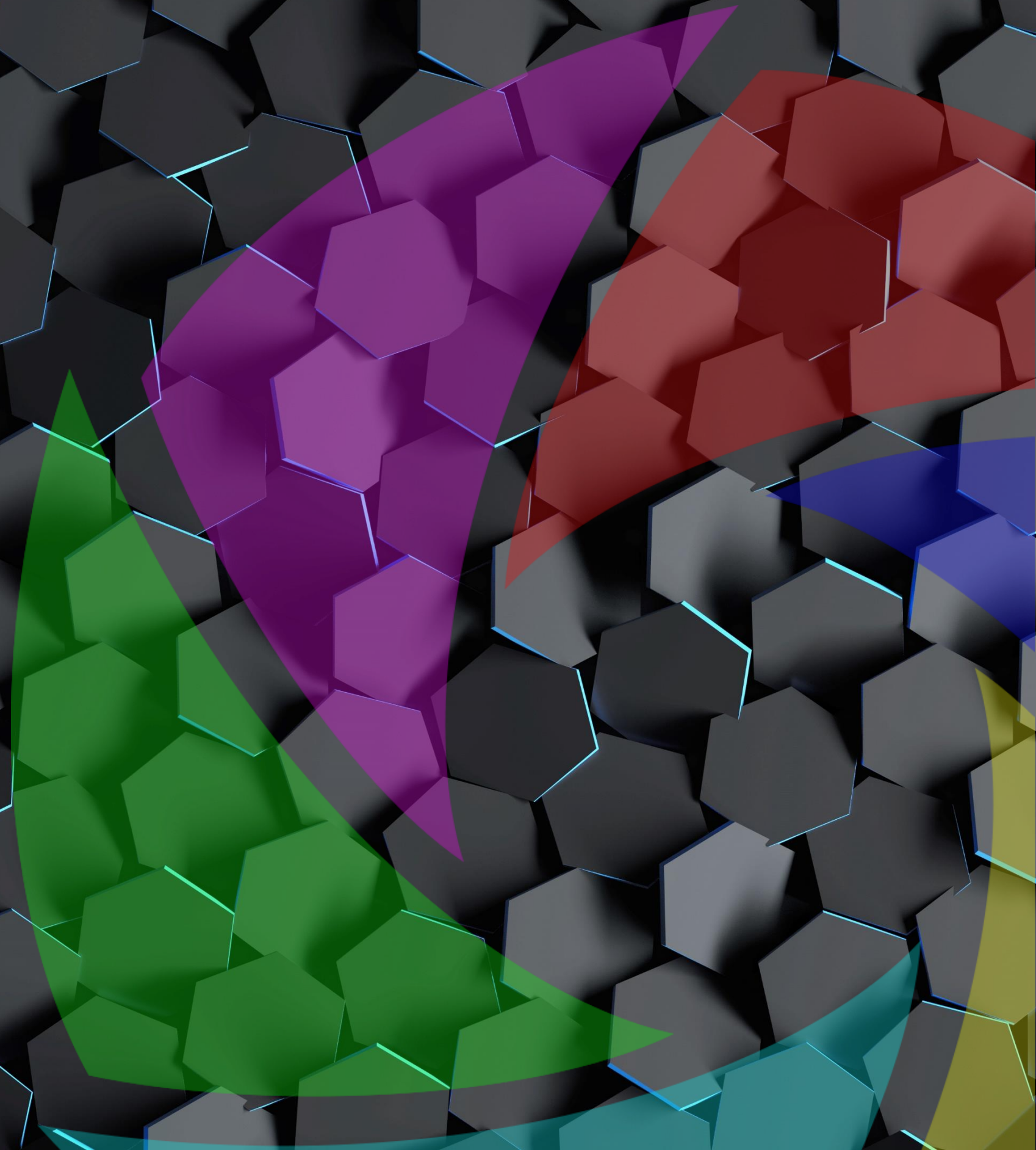




# BITS BY THE BAY



SMPTE  
BITS BY  
THE BAY

2026

# AI in Live Production

---

Presenter: Jim Doyle, VP of Creative Services – Ross Video

# VoiceControl

Voice Control is a breakthrough voice-control technology that lets directors, operators, and talent control video production systems using natural speech.

From switching cameras to triggering graphics, you simply say the command, and Voice Control handles the rest.

# VoiceControl | NAB 2024 Switcher Control



# VoiceControl | NAB 2025 Weather Control



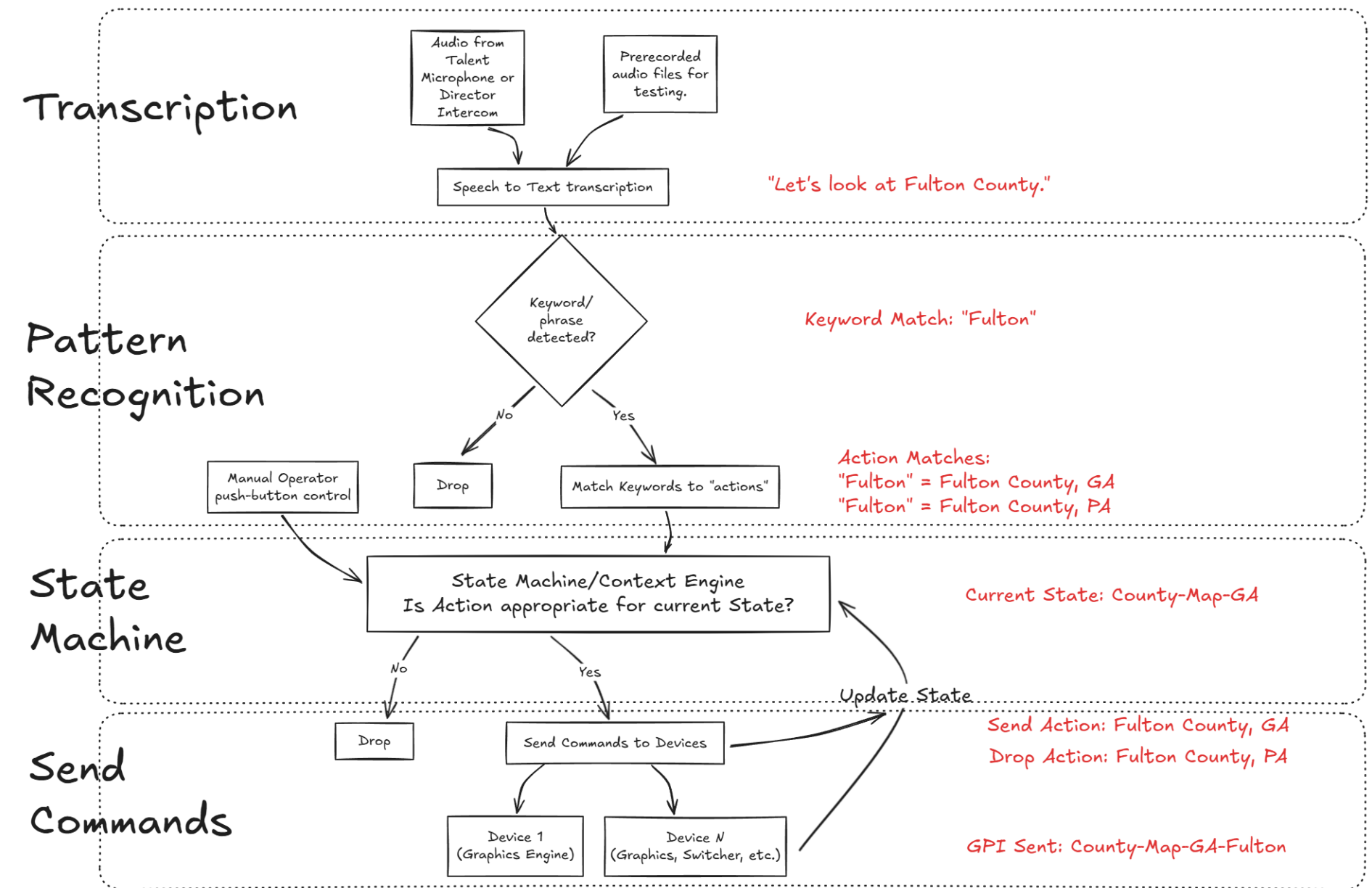
# Background & Industry Context

- Voice UIs common in consumer tech but not suited to live broadcast.
- Broadcast workflows now too complex for manual control alone.
- Advances in ASR/NLP enable reliable, domain-specific voice control.
- Voice Control: built for live TV — fast, redundant, context-aware, operator-safe.
- Talent drives visuals by voice → faster reactions & richer storytelling.
- Next: architecture, field results, and future impact.



# System Architecture

- Four subsystems
  - Transcription - Whisper ASR, optimized for noisy environments
  - Pattern Recognition - Phrase parsing + probabilistic matching
  - Context Evaluation - State-aware disambiguation
  - Command Dispatch - RossTalk, DashBoard, WebSocket
  
- End-to-end latency: ~400 ms



# System Design

- Fully on-prem
- Containerized for security
- Native Ross DashBoard integration
- Operator override always available
- Grafana telemetry confidence monitoring



# **VoiceControl** | Election Night Deployment

**PENNSYLVANIA 2024** ★ ELECTORAL VOTES **19**

PHILADELPHIA IN: 78%

VOTE COUNT  
DEM REP

PITTSBURGH PHILADELPHIA

**FOX 4** IN: 98% ★ **MISSISSIPPI SENATE** ★

	<b>PINKINS</b> 388,304	36.9%		<b>WICKER*</b> 663,217	63.1%
--	---------------------------	-------	--	---------------------------	-------

★ **PRESIDENT** ★

<b>HARRIS</b> 216	<b>TRUMP</b> 232
----------------------	---------------------

NEEDED TO WIN: 270

★ ★ ★ ★

	388,304	36.9%		663,217	63.1%
--	---------	-------	--	---------	-------

IN: 98% ★ **MISSISSIPPI SENATE** ★

<b>HARRIS</b> 216	<b>TRUMP</b> 232
----------------------	---------------------

★ **PRESIDENT** ★

NEEDED TO WIN: 270

# The Results

- 8 hours continuous live broadcast
- 351 phrases detected → 42 valid triggers
- Zero false positives, only 2 misses
- Reliability under high-pressure conditions



# Key Advantages

- Reduced latency between intent and on-air execution
- Lower operator workload, freeing staff for higher-level tasks
- Greater freedom for on-air talent to explore interactively
- Fluid storytelling during unscripted moments
- Hands-free control
- Stronger viewer engagement



# Benefits of Voice-Controlled Workflows

- Collapsed workflow steps → fewer producer/operator handoffs
- Voice-triggered graphics & transitions keep conversation flowing
- Streamlined cadence: talent moves freely, improvises, reacts instantly
- Operators refocused on oversight → fewer errors, better collaboration
- Modular architecture makes it adaptable



# Voice Control Challenges & Solutions

## Challenges

- Phrase ambiguity → vague/indirect commands
- Variable microphone quality
- Speech overlap in multi-host environments
- 2 false negatives during election broadcast

## Solutions

- Expanded pattern recognition → semantic intent + phonetic scoring
- Improved mic handling & audio robustness
- Speaker identification for multi-host setups
- Separate contextual models per on-air talent

# Roadmap

- Multilingual input support
- Deeper automation integrations
- Cloud-hybrid model for remote/distributed control
- Predictive control → preload graphics from partial commands
- Feedback cues (visual/audio) reduce cognitive load
- Voice as a core interface for next-gen media systems



# Broader Applications

- Realtime Open & Closed Captioning
  - Automated Speech Recognition model produces incredibly accurate captions at low latency.
  - Upload a “custom dictionary” to tune the ASR for your event.
- Audio Concealment



# **Video**Tracking | AI in Studios & Sports

# AI Face & Body Tracking | Vision[Ai]ry



# AI Player Tracking | Piero



# Conclusion

## VoiceControl

- Natural language triggers graphics & automation
- Empowers talent as visual storytellers; reduces operator workload
- Flexible architecture for news, sports, weather, & live events

## VideoTracking

- Highly efficient
- Enhanced speed, accuracy, & predictability
- Operational 'game changer'



SMPTE  
BITS BY  
THE BAY

2026

**Thank you**