

— AV AI DISPATCH · RESEARCH REPORT

The AV AI Production Index

A quarterly assessment of what AI can and can't do in audio/video production. No hype. Just what works, what's close, and what's still a demo.

Q1 2026 · INAUGURAL EDITION · AUDIO/VIDEO SERVICES

40 Years Later, a Longer Lever

In 1986, a staff member saw MEPS and told a writer, "Now writing must be easy for you." He missed the point. MEPS didn't eliminate the hard work — it gave one person the capacity of a hundred. A single operator could typeset in dozens of languages. Not because the work was easier, but because the lever was longer. Forty years later, we're holding a much longer lever. The question isn't whether AI does our jobs — it's what we could accomplish if every repetitive hour were returned to us.

70–90%

AI handles this much

10–30%

Still needs a human

100%

Needs QA before shipping

That last 10–30% is where quality lives. Every category in this report follows the same pattern.

Where Things Stand **Right Now**

✓ WHAT YOU CAN DO

- Generate **5–15 seconds** of usable video footage for B-roll and concept work
- Produce **broadcast-quality background music** indistinguishable from stock libraries
- **Clone voices** with near-human fidelity from as little as 15 seconds of audio
- Cut **post-production time in half** with AI cleanup, captioning, and masking

x WHAT YOU CAN'T DO

- Produce a coherent **60-second narrative** scene
- Replace a **composer** for distinctive, memorable music
- Ship **any AI output** without a human quality pass

Every section below follows the same structure: **what's ready, what's close, what's not there yet**, and a one-line recommendation.



The Basics: LLMs & Agents

Every AI tool in this report is powered by a large language model. Understanding the foundation helps you evaluate everything built on top of it.

Chatbots vs. Agents

Chatbot

You ask a question, it gives an answer. Reactive — waits for you, responds, stops. A very knowledgeable colleague who only speaks when spoken to.

"How do I fix this audio sync issue?"

Agent

A chatbot that can **do things**. Reasons about a goal, breaks it into steps, uses tools, checks its work, and iterates until the task is done. Same brain, different autonomy.

"Fix the audio sync bug, run the tests, open a PR."

Most AV teams will interact with these models as **chatbots first** (research, brainstorming, drafting) and as **agents later** (automating workflows, building tools).

Chatbot

Responds to messages

Agent

Reasons + takes actions

Same LLM

Different level of autonomy

Frontier Model Lineup — March 2026

PROVIDER	CHATBOT	FLAGSHIP MODEL	CONTEXT	PRICING
OpenAI	ChatGPT	GPT-5.4	1M tokens	Free; Plus \$20/mo; Pro \$200/mo
Anthropic	Claude	Opus 4.6	1M tokens	Free; Pro \$20/mo; Max \$100-\$200/mo
Google	Gemini	Gemini 3.1 Pro	1M tokens	Free; AI Plus \$5/mo; Pro \$20/mo; Ultra \$250/mo
xAI	Grok	Grok 4.20	256K tokens	Free; SuperGrok \$30/mo; Heavy \$300/mo
Meta	—	Llama 4 Maverick	10M tokens	Open-weight; free to self-host
DeepSeek	DeepSeek Chat	V3.2	128K tokens	Free; API ~\$0.28/M tokens
Mistral	Le Chat	Large 3	256K tokens	Free; Pro \$15/mo; Team \$25/user/mo

Open-weight models (Meta, DeepSeek, Mistral) can run on your infrastructure — no data leaves your network. **Closed models** (OpenAI, Anthropic, Google, xAI) are easier to start with but require trust in the provider's data policies.

Why This Matters for AV Production

Every specialized AI tool in this report connects back to a frontier LLM in one of three ways:

Built on

Descript uses Whisper for transcription.
ElevenLabs uses custom models for voice synthesis.

Orchestrated by

AI agents use Claude or GPT to plan and execute multi-step production workflows.

Enhanced by

Cursor uses Claude/GPT to understand your codebase and write internal tools.

Takeaway: When a new frontier model drops, every tool built on it gets better. Understanding the model layer helps you evaluate vendor claims and anticipate capability jumps.

Video Generation

The most visible category — and the one with the widest gap between demos and production reality.

The Field Right Now

TOOL	RESOLUTION	DURATION	CONSISTENCY	PRICE
Runway Gen-4.5	4K	Up to 60s	Good	\$12-\$76/mo
Kling 3.0	4K at 60fps	15s	Good (multi-shot)	\$6-\$48/mo
Seedance 2.0	2K	15s	Good (multi-ref)	TBD
Google Veo 3.1	Up to 4K	4-8s	Moderate	Vertex pricing
Sora 2	1080p	20s	Weak	\$20-\$200/mo
Pika 2.5	1080p	10s (25s Pikaframes)	Weak	\$8-\$76/mo
Luma Ray 3.14	1080p native	5-10s (ext. 30s)	Moderate	Enterprise

All tools produce usable results for short clips. The differences emerge in duration, consistency across shots, and physics simulation. Image-to-video consistently outperforms text-to-video across every platform.

Also in this space: Vidu · Hailuo/MiniMax · CapCut AI · Genmo Mochi · Higgsfield · Haiper

Readiness Breakdown

READY

- **Short clips (5–15s)** for B-roll and supplemental content
- **Image-to-video** — consistently beats text-to-video
- **Cinematic style & camera movement**
- **Pre-vis & storyboarding** — ~40% time savings
- **Lip-sync** in controlled settings

CLOSE

- **Character consistency** — drifts past 3–4 shots
- **Extended duration** — coherence fails past ~20s
- **Physics** — rigid OK, fluid/cloth fail

NOT YET

- **Text in video** — all tools fail
- **Hands/fingers** — persists across all models
- **30s+ narrative coherence**

Takeaway: Use now for pre-vis and B-roll. Don't ship AI-only video for precise human representation.

Editing & Post-Production

Where AI delivers the most immediate, measurable time savings today.

AI in Your Existing NLE

Industry NLEs — AI already built in

- **DaVinci Resolve** — Magic Mask, IntelliTrack, Voice Isolation, Dialogue Separator, UltraNR, text-based editing
- **Final Cut Pro** — Magnetic Mask (AI subject isolation), ML object tracking, on-device Transcribe to Captions
- **Avid Media Composer** — ScriptSync AI (dialogue search/sync across 21+ languages)
- **Premiere Pro** — AI Object Mask, Generative Extend, auto-captioning (92–95%)

Experimental AI-native editors

- **Descript** — edit video by editing text; Underlord agentic co-editor; 60–70% time savings for spoken-word
- **Flixier** — browser-based transcript editing with AI B-roll
- **CutScript** — open-source, local-first alternative

NOT YET

- Creative editorial judgment
- Brand-consistent editing at scale

Takeaway: Your existing NLE already has AI features worth learning. For spoken-word workflows, experimental editors like Descript offer a fundamentally different — and faster — approach.

Also in this space: VEED.IO · Choppity · PlayPlay · ScreenApp · Mocha Pro 2026

VFX & Compositing

AI rotoscoping is the single most production-ready AI capability in all of VFX.

Visual Effects & Animation

VFX & Compositing

READY

- **AI rotoscoping** — Premiere Object Mask, Slapshot (8K, 400 shots/day)
- **CG compositing (Wonder Studio)** — 80–90% automated, from \$19.99/mo
- **Simple object removal** — wires, boom mics

NOT YET

- Fully automated VFX shots
- Real-time AI compositing

Animation

READY

- **Cascadeur** — AutoPosing + AutoPhysics, 2–3x faster blocking
- **Mocap cleanup** — Rokoko + Cascadeur, minutes not days
- **Single-camera mocap** — webcam previz

NOT YET

- Fully automated character animation
- AI lip-sync for 3D characters

Takeaway: Rotoscoping and mocap cleanup are deploy-today tools. Final artistry still needs humans.

Also in this space: VFX: Mocha Pro/Boris FX · Runway Inpainting · HitFilm | Animation: DeepMotion · Move.ai · Plask · RADiCAL · Motorica

Audio: Voice, Music & Post

The category where AI most consistently passes for human output.

Voice Synthesis & Text-to-Speech

READY

- **Short-to-medium narration (<10 min)** — indistinguishable from human for explainer content
- **Voice cloning** — ElevenLabs (~30 min audio), Fish Audio S2 (15 seconds)
- **Word-level emotion control (Fish Audio S2)** — tag words with [whispered], [voice breaking]

CLOSE

- Long-form narration (30+ min) — prosody drift accumulates
- Complex emotional arcs — irony, sarcasm, building tension

Pricing Snapshot

TIER	COST
ElevenLabs	From ~\$0.12/1K chars
Fish Audio S2	Free (self-host) or API
Azure / Google	~\$0.016/1K chars

Takeaway: AI TTS is production-ready for narration under 10 minutes. Premium dramatic performance still needs a human.

Also in this space: PlayHT 3.0 · WellSaid Labs · Murf.ai · Resemble AI · Coqui/XTTS · Bark · Tortoise TTS · Amazon Polly · Google Cloud TTS

Music, SFX & Audio Post

Music Generation

- **READY** **Background music & underscore** — Suno V5, 48kHz, 7M tracks/day
- **CLOSE** Full songs — vocals drift, structures collapse
- **NOT YET** Distinctive compositional voice

Sound Design & SFX

- **READY** Simple ambience, UI sounds, atmospheres
- **NOT YET** Foley, cinematic sound design, anything that ships

Audio Post-Production

The most production-ready AI audio category.

- **Noise removal (iZotope RX 11)** — industry gold standard
- **Speech enhancement (Adobe Podcast)** — free (1hr/day)
- **10-stem separation (LALAL.AI)** — 14.8M hours processed
- **Transcript editing (Descript)** — one-click cleanup

Takeaway: Audio post saves the most hours per dollar. iZotope RX 11 and Descript belong in every toolkit.

Also: Music: Udio · AIVA · Soundraw · Beatoven.ai | SFX: Stable Audio · AudioGen | Post: Accusonus ERA · Audo.ai · Krisp · Auphonic · Moises.ai

Dubbing, Platforms & Agents

The pipeline is 4–5 tools stitched together. True end-to-end doesn't exist yet.

The Pipeline Today

Dubbing & Localization

- **READY** **AI dubbing + human QA** for informational content (Deepdub, Papercup, ElevenLabs)
- **CLOSE** Entertainment dubbing — sounds noticeably AI
- **NOT YET** Fully automated dubbing, singing, premium entertainment

AI Agents

- **Make.com / n8n** — ~50% time reduction for trigger-based workflows
- No agent is ready for **autonomous production**

End-to-End Platforms

PLATFORM	READY?
ElevenLabs Flows	Alpha
Luma Agents	Just launched
Synthesia	Yes — talking-head
HeyGen	Yes — talking-head

Realistic pipeline: Script (LLM) → Video (Synthesia/Runway) → Audio (ElevenLabs) → Edit (NLE/Descript) → Deliver

Takeaway: Viable now for informational content with human QA. True end-to-end is a 4–5 tool pipeline — no single platform handles it all.

Also: Dubbing: Rask.ai · Flawless AI · Dubverse · Camb.ai | Platforms: Colossyan · D-ID · Tavus · Elai.io | Agents: Zapier · Activepieces · CrewAI · LangChain

AI Coding Agents & Dev Harnesses

Two layers: IDE harnesses where a developer steers, and autonomous agents that work independently and return finished code.

IDE Harnesses — Developer in the Loop

HARNESS	KEY CAPABILITY	PRICE
Cursor	Agent mode, 8 parallel agents, cloud agents from browser/phone/Slack	\$20-\$200/mo
Windsurf	Cascade writes ~90% of code; "Memories" learn project conventions; Turbo Mode	\$15-\$200/mo
Google Antigravity	Agent-first: plan → execute → test → verify across editor, terminal, browser	Free preview

Not autocomplete — agentic environments where you describe intent and the AI plans, writes, tests, and debugs. Over half the Fortune 500 use Cursor.

Autonomous Agents — No IDE Required

AGENT	HOW IT WORKS	ACCESS
OpenAI Codex	Clones repo, writes code, runs tests, returns PR. GPT-5.3-Codex	ChatGPT Plus (\$20/mo+)
Claude Code	Terminal agent: reads codebase, plans, executes multi-step implementations	Claude Max (\$100-\$200/mo)
Jules 3.0	Async agent: assign from browser/Slack, returns PRs. API for CI/CD	Free preview
Claude Cowork	Knowledge work: file management, spreadsheets, reports — not code-specific	Claude Max (\$100-\$200/mo)

Use Cases for AV Teams

- **Internal tool development** — Asset management, review workflows, dashboards, metadata tagging. One developer + agent ≈ a team of three
 - **Web development & publishing** — Sites, CMS interfaces, distribution platforms. Agent handles scaffolding, responsive layouts, API integrations, deploys
 - **Automation infrastructure** — Ingest pipelines, transcription routing, QC checks. Codex/Jules work async on GitHub issues, return working PRs
 - **Production operations (Cowork)** — Organize project files, compile reports, generate formatted spreadsheets, manage docs
- NOT YET
- **Complex architecture decisions** — Deep domain knowledge, creative problem-solving
 - **Production-critical code without review** — Error rate on complex tasks still too high

Takeaway: IDE harnesses (Cursor, Windsurf) are the fastest path to internal tools. Autonomous agents (Codex, Jules) handle well-scoped tasks async. Neither replaces a developer — they replace the repetitive 70%.

Also in this space: GitHub Copilot · JetBrains AI (Junie) · Tabnine · Codeium · Replit Agent · Bolt.new · Lovable · Devin · Sweep AI

Transcription & Captioning

The most mature AI category in production.

PROVIDER	ACCURACY	PRICE/MIN	BEST FOR
AssemblyAI Universal-3 Pro	94.1%	\$0.0035	Audio intelligence, PII redaction
OpenAI Whisper	~92%	\$0.006	Budget batch, 100+ languages
Deepgram Nova-3	~92%	\$0.0043	Real-time (<300ms latency)

Key Strengths

- **92–95% accuracy** for clear English across all providers
- **Real-time** transcription under 300ms latency
- **100+ languages** supported by Whisper

Weak Point

- **Speaker diarization** — 85–90% in good conditions, drops to 75–80% with noise or 5+ speakers

Takeaway: Transcription is solved for most use cases. Budget under \$1/hour for enterprise-grade accuracy.

Also in this space: Rev.com · Otter.ai · Sonix · Trint · Verbit · Speechmatics · Gladia · Amazon Transcribe · Google Cloud STT · Transkriptor



The Cheat Sheet

What to adopt, what to watch, and what to skip.

ADOPT NOW

Clear ROI

Concept / pre-vis	Runway Gen-4.5 · Kling 3.0
Rotoscoping & masking	Premiere Object Mask · Slapshot
Audio cleanup	iZotope RX 11 · Adobe Podcast (free)
Transcript-based editing	Descript
Background music	Suno V5

Voice synthesis (short-form)	ElevenLabs · Fish Audio S2
Talking-head video	Synthesia · HeyGen
Animation blocking + mocap	Cascadeur · Rokoko
CG character compositing	Wonder Studio
Internal tool / web dev	Cursor · Windsurf
Async bug fixes & features	Codex · Jules 3.0

WATCH CLOSELY

Agent-based workflows (Luma, ElevenLabs Flows) · Multi-shot narrative (Kling, Seedance) · AI dubbing at scale · Autonomous agents for complex dev

NOT YET

Don't Count On

- AI-only long-form video generation (>30s coherent narrative)
- Automated editorial decision-making
- AI text rendering within generated video
- Unsupervised VFX compositing
- AI-only character animation for broadcast
- Fully autonomous production agents for creative work
- AI music as hero content
- AI sound effects for final delivery
- Autonomous coding agents for complex architecture

METHODOLOGY

Based on manufacturer documentation, third-party benchmarks (Artificial Analysis), trade press, enterprise case studies, and 41 independent references. Where marketing claims conflict with testing, the conservative position is taken.

Shelf life: Roughly 3–6 months. The Q2 2026 edition will follow.

— AV AI DISPATCH · RESEARCH REPORT

The AV AI Production Index

Q1 2026 · Inaugural Edition

Full report available at avaidispatch.com

AUDIO/VIDEO SERVICES — AI RESEARCH & DEVELOPMENT 2026