The Voice AI reality check:

# Frontline perspectives for enterprise in 2025

SPEECHMATICS

Voice AI generated tremendous buzz throughout 2024, with waves of demos and product launches sweeping across the industry.

Many companies that boldly announced “AI-first” strategies quietly retreated months later. These approaches positioned technology as a replacement for humans rather than an enabler of their capabilities.

Companies from Klarna to Duolingo raced to announce agent-led experiences, often to mixed reception. By 2025, reality had tempered the excitement, with headlines of retreats from aggressive rollouts.

At AI-centric events like CES, substance often took a backseat to style. Demos showcased scripted scenarios but ignored real-world challenges like background noise, accent variation, and context switching.

Strategic investment is critical. Without clear use cases, AI remains a buzzword. When anchored in tangible business value, it becomes transformational.

That shift is already playing out. Industry leaders are now telling a more grounded story — one focused less on promise, and more on performance.

Foundational technologies – particularly transcription and speech recognition – are already generating significant ROI. From compliance to customer service, systems built on strong speech infrastructure are delivering measurable value.

But when it comes to more conversational, autonomous Voice AI, the picture is still evolving. Concerns around accuracy, control, and user experience remain front of mind.

That said, the groundwork is there. The potential is real. And for teams building in the Voice AI space, the path ahead looks more promising than ever.

Katy Wigdahl,
CEO, Speechmatics

“”

# Voice AI isn’t optional anymore. It’s operational.

Tamara Zubatiy–Nelson, Barometer

# Voice AI is entering its operational era.

What was once demo theatre is now enterprise-critical tech—built on the renaissance in Speech AI.

## SPEECH AI → VOICE AI

**The real-world impact of Voice AI is only possible because Speech AI has quietly levelled up.**

From noisy clinics to high-pressure control rooms, enterprise-grade transcription is the engine behind accurate, responsive, and multilingual voice systems.

That's where the ROI is happening. That's why the shift is real.

## WHERE IT'S WORKING

Voice AI is already powering these industries:

### Healthcare
Ambient notes, faster triage, less burnout

### Media & Adtech
Tone analysis, localisation, brand safety

### Academic Research
Hybrid AI-human analysis workflows

### CX & Contact Centres
Seamless multilingual support, 99.999% uptime

### Compliance
Multilingual audit trails, regulator-ready transcription

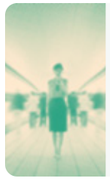
### Public Sector
Real-time emergency response, utility crisis communications

## KEY TRENDS

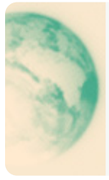
Six shifts reshaping how enterprises build with voice:

### 1. Infrastructure, not interface
Voice is now a backend building block. It powers critical systems, not just user-facing tools.

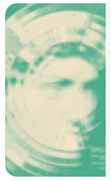
### 2. Assistive agents > avatars
Voice tools augment human work, not replace it. They free up clinicians, auditors, and researchers—not replace them.



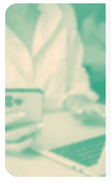
### 3. Proven ROI
Gains in efficiency, uptime, and user satisfaction. Organisations are reporting faster workflows and tangible cost savings.



### 4. Multilingual by design
Global workflows demand hybrid language fluency. Systems must follow code-switching in real time to maintain CX quality.



### 5. Accuracy = trust
Hallucinations break systems. Precision sustains them. In compliance and care, getting the transcript wrong isn't an option.



### 6. The end of clicking?
Voice is reshaping interaction – from search to support – by layering conversation over visual journeys.

## HIGHLIGHTS BY THE NUMBERS

What Speech AI is enabling across sectors:

**400%**
growth in user adoption (Audiotranskription – academic tools)

**93%**
customer satisfaction (Content Guru – public services)

**60%**
50-60% time saved (Audiotranskription – research)

**120x**
content output with no added cost (AI-Media – broadcast)

**$2.4B → $47.5B**
Market size: $2.4B → $47.5B by 2034, growth of over 1800%

## BOTTOM LINE:

Voice AI only works where Speech AI excels. Businesses chasing real ROI are investing in the groundwork, not the gimmicks.

**Find out more at:**
**Speechmatics.com**

# Inside this report


1. Meet the voices of Voice AI
2. The state of play: Voice AI in 2025
3. The six trends:
    1. **Infrastructure, not interface: Voice as enterprise foundation**
    2. **Agents are coming – but they'll behave differently than you expect**
    3. **The ROI of voice: From gimmick to growth**
    4. **Monolingual to multilingual: Scaling Voice AI globally**
    5. **Fighting fiction: Accuracy in the age of hallucinations**
    6. **The end of clicking? How conversation is becoming the new interface**
4. What's next for Voice AI? Predictions from the frontline
5. Final thoughts: Lessons from the leaders building the future

# Meet the voices of Voice AI

These industry pioneers are shaping how voice technology transforms business, compliance, and user experience. Their insights from the frontlines reveal where Voice AI delivers real value, where the hype falls short and where testing continues with care, as trust is still being earned.

### Martin Taylor, Cofounder and Deputy CEO, Content Guru

Martin Taylor is the cofounder and Deputy CEO of Content Guru, a global provider of enterprise cloud contact center solutions. He leads product innovation and strategic market development, with a focus on public sector and healthcare. A long-time pioneer in cloud communications and real-time billing, Martin began building businesses at age 22 and now oversees critical infrastructure delivering 100 percent SLA reliability for high-stakes services like emergency response.

### Tamara Zubatiy-Nelson Cofounder and CEO, Barometer

Tamara Zubatiy-Nelson is cofounder and CEO of Barometer, which uses AI to analyze premium media like podcasts. Her work helps advertisers go beyond keywords to assess tone, topic, and audience fit. With a PhD in human-centered AI, Tamara works with clients including Horizon Media, BetterHelp, and SiriusXM. She has helped route over $100 million in global ad spend to tens of thousands of creators and has been recognized by Forbes 30 Under 30, Cynopsis Top Women in Media, and AdTech Innovators.

### Tony Abrahams Cofounder and CEO, AI-Media

Tony Abrahams cofounded AI-Media in 2003 and currently serves as CEO. He previously helped establish the Oxford Internet Institute while studying as a Rhodes Scholar and has served as a Director of Northcott Disability Services. Recognized by the World Economic Forum as a Young Global Leader, Tony has been a member of the Australian Institute of Company Directors since 2006 and continues to lead innovation in captioning and broadcast accessibility.

### Dr. Thorsten Dresing Managing Partner, audiotranskription.de

Thorsten Dresing holds a doctorate in Educational Science and brings over 25 years of experience in qualitative research, software development, and methods training. As managing partner at audiotranskription.de, he focuses on integrating AI into qualitative data analysis and continues to lead the development of tools such as f4transkript and f4analyse. He regularly conducts workshops, publishes academic work, and contributes to the wider methodological discourse.

### Henrik Skourup Product Lead, Zylinc

Henrik Skourup is a seasoned professional with more than 20 years of experience in Unified Communications and Contact Center solutions. As Product Lead at Zylinc, he oversees the development of cloud-based platforms that enhance customer engagement and operational efficiency. His recent work includes integrating high-accuracy speech-to-text technology that supports Danish, English, and mixed-language conversations for real-time analytics.

### Peter Kenny MD, Global Surveillance Strategy, ACA Group

Peter Kenny is a compliance and technology leader with more than 20 years of experience across global financial services. Now serving as Managing Director of Global Surveillance Strategy at ACA Group, his previous roles include Director of Compliance at Barclays Investment Bank, where he led global regulatory risk initiatives. Peter has also held senior roles at Citi, Societe Generale, Goldman Sachs, and PaineWebber, building tech-driven compliance frameworks across complex systems.

# The state of play: Voice AI in 2025

Voice AI has dominated recent headlines. From agent demos to AI-first strategies, it can feel like we've already reached a fully conversational future.

But across enterprises, a different reality is playing out.

The leaders we spoke to – across healthcare, compliance, research, media and public services – aren't chasing hype. They're focused on something more immediate: making voice work reliably across real-world environments.

This progress is less about avatars and more about what happens behind the scenes – voice technology integrated into workflows, driving responsiveness, automation and scale.

To understand why many teams are still laying the groundwork, it helps to look back at how voice got here.

### Recognition (1950s–90s)

Machines began recognising digits and short commands. Speech entered computing, but only in fragments.

*Foundations of Speech AI appear in lab settings, focused on limited vocabulary recognition.*

### Batch Transcription (1990s–2000s)

Dictation tools emerged using statistical models. You could talk to your computer, but there was no interaction

*Speech AI advances enabled full-sentence transcription, though systems remained offline and slow.*

### Real-Time Speech (2010s)

Deep learning enabled usable transcription and synthesis. Voice became a daily tool, mostly for consumers.

*Real-time Speech AI became viable, paving the way for assistants like Siri and Alexa.*

### Multilingual + Responsive Speech (2020–2023)

Transcription systems began handling noise, accents, and context, while text-to-speech (TTS) models gained expressiveness.

*Voice tools matured for real-world use, with enterprise-grade Speech AI enabling accurate, multilingual processing at scale.*

### Voice AI Agent Revolution (2023 - Future)

The arrival of human-like interaction - powered by generative AI, lifelike TTS, and ultra-low-latency transcription - has triggered a revolution in Voice AI agents.

*What was once experimental is now enterprise-ready. From contact centres to healthcare, agents are already accelerating decisions, streamlining workflows, and transforming how work gets done.*

**These outcomes aren't speculative. They reflect a step change: voice interfaces that feel natural, work fast, and integrate deeply - because the technology underneath is finally up to the task.**

**Voice AI market size**

Voice agent market projects to grow from $2.4B (2024) to $47.5B (2034)

*(Source: Market.US)*

2034: **$47.5B**

2024 start: **$2.4B**

2024 end: **$5.4B**

**Y Combinator class of '24**

**22%**

of startups in Y Combinator's 2024 class build with voice technology.

*(a16z)*

**U.S medical groups**

**43%**

43% of U.S. medical groups expanded use of AI tools, including voice.

*(MGMA, 2024)*

# From foundation to impact: How Speech AI is powering Voice AI ROI

The Voice AI conversation has matured. It's no longer about what might be possible, it's about what's working right now.

Across sectors, the shift is being driven by quieter, foundational improvements: cleaner transcription, faster response times, and systems that can hold up in noisy, complex environments. Just as importantly, they're now efficient enough to scale. If every interaction came at a premium, voice would remain a niche feature, not a viable solution.

Organisations that have invested in these under-the-hood capabilities are now seeing measurable returns – from reclaimed hours to expanded reach. The following data points to where voice is already delivering and why its momentum is building.

### Accuracy that drives business

Voice systems are only as good as the signals they're built on. When speech is captured and transcribed clearly, new models become possible:

**85-90%**
transcription accuracy rates enabling new business models (AI-Media)

**40-60%→ 90%+**
accuracy improvement after upgrading speech processing (Zylinc)

**93%**
customer satisfaction at a major utility using voice-led systems (Content Guru)

### The productivity revolution

Voice-enabled tools are helping organisations move faster and reduce admin load, particularly where professionals are stretched:

**50-60%**
time saved on transcription tasks (Audiotranskription)

**40%**
of clinician time reclaimed from documentation (Content Guru)

**90%+**
automation during peak service periods (Content Guru)

> "Today we can't afford to be down even for a minute... It might endanger hundreds of millions of dollars in ad spend."
>
> Dr. Tamara Zubatiy-Nelson, Barometer

### Business acceleration

When systems are accurate, adoption follows:

**400%**
user growth in one week post-implementation of Speech AI (Audiotranskription)

**120x**
more content delivered with no added cost (AI-Media)

### Enterprise-grade reliability

Where failure isn't an option, consistency matters. Voice infrastructure is proving it can meet the moment:

**99.999%**
SLA maintained for critical comms (Content Guru)

**100%**
uptime across emergency service platforms (Content Guru)

### Market impact

Voice systems are not just improving operations but expanding access and reimagining reach:

**10% → 50%**
uplift in education completion rates for deaf students (AI-Media)

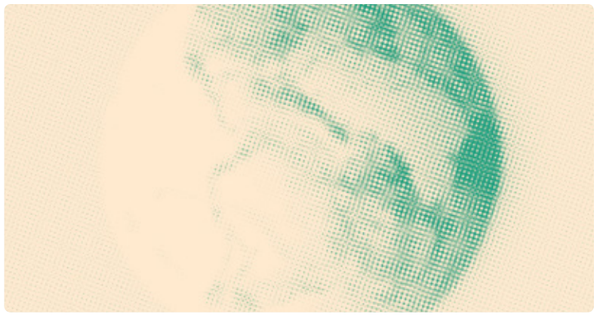
**20-30x**
market size compared to text

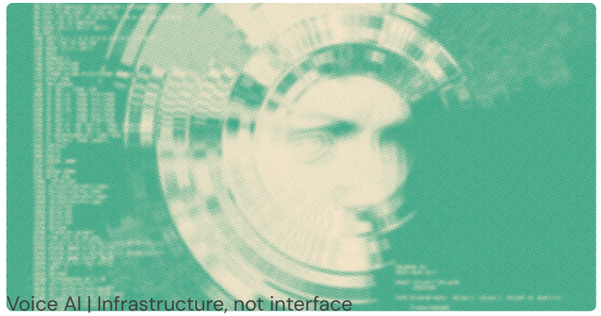
“”

# Voice AI lets professionals focus on what they’re trained for.

Martin Taylor, Content Guru

Trend 1





# Infrastructure, not interface – Voice as enterprise foundation

**The invisible infrastructure behind high-stakes decisions**

Key insights

- Voice now functions as mission-critical infrastructure
- Trusted by emergency services, healthcare, and academia
- Real-time voice insight drives performance and safety
- Hybrid deployment proves essential for compliance and control

It is said that the most powerful technologies are those that become invisible.

By 2025, this principle will define voice's evolution. No longer merely showcased as innovation, it has integrated into the foundational systems that keep industries running. Rather than making headlines, it quietly powers critical workflows where failure simply isn't an option.

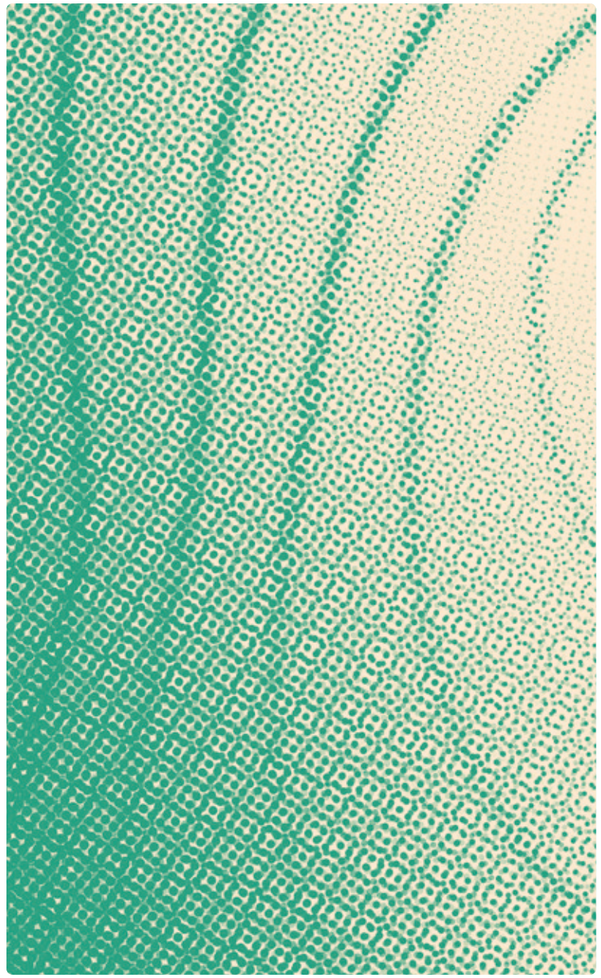
What makes this possible is not just the visible layer of Voice AI, but the performance of the underlying systems – latency/speed, accuracy, and robustness – that support it.

**Performance-critical applications**

Operating globally, Content Guru's platform supports one of the highest-stakes use cases imaginable: emergency response.

"We currently deliver every single emergency ambulance call for the whole of the UK, and also a significant amount of police 999, through our dedicated blue-light platform, which operates to a 100% SLA."

Martin Taylor, Content Guru



Alongside emergency healthcare, another area where voice proves essential is in large-scale infrastructure events – like power outages or flooding. In these moments, the technology becomes a real-time decision tool, helping national utilities monitor conditions, share live updates and reduce avoidable inbound contact.

> "We can build a picture of any location within a customer's area and then we can relay that picture to consumers in real time and also send out live updates so our customer can stay ahead of a developing situation and forestall avoidable inbound contacts."
>
> Martin Taylor, Content Guru

While academic research may not save lives in the moment, it still demands the highest level of precision. When Audiotranskription upgraded their transcription engine, usage surged 400% in just one week.

"Accuracy was the key factor… and the availability of a reliable on-premise solution was extremely important." —Thorsten Dresing, Audiotranskription.

Whether supporting life-critical decisions or advancing research, successful voice systems blend seamlessly into existing workflows, becoming virtually invisible to end users while transforming outcomes.

### Hybrid infrastructure drives adoption

The infrastructure shift extends beyond capabilities to architectural considerations. In 2025, hybrid deployment has emerged as a core requirement rather than a compromise.

Organisations now expect voice technology to function across environments – cloud, on-premise, secure networks, edge devices – with equal reliability. This flexibility proves particularly critical in regulated industries where data sovereignty, compliance and uptime converge.

"Speed matters when it fits the workflow, not when it just looks good on a spec sheet." —Henrik Skourup, Zylinc

The reality of everyday operations underscores why voice must function as dependable infrastructure rather than experimental technology.

> "Every day is planned around this pipeline. If it breaks, everything else stalls."
>
> Tamara Zubatiy-Nelson, Barometer

This infrastructure-first approach signals maturity. Voice technology has moved beyond proof-of-concept demonstrations to become an expected, foundational layer – supported by reliable speech systems – that enables innovation across the enterprise stack.

SPOTLIGHT:

Martin Taylor, Content Guru

WHO?

Co-founder and Deputy CEO

WHERE?

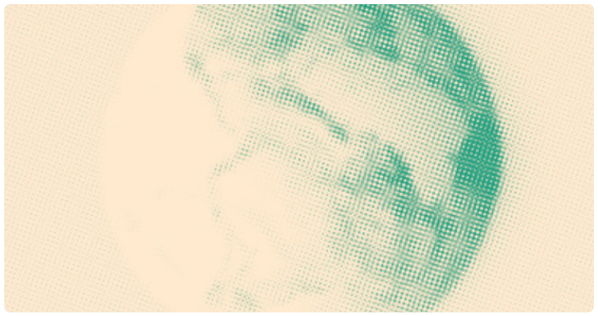
Emergency services, healthcare, utilities, financial services and public sector

WHAT THEY'RE DOING WITH VOICE:

Content Guru supports critical communications infrastructure across the UK. From ambulance call handling to healthcare triage, their voice systems operate in high-pressure environments where reliability is essential.

KEY INSIGHT:

"People say 'I want AI' — but AI starts with data, then comes discovery, validation and deployment. Most don't initially think about that necessary first step."

# Agents are coming — But they'll behave differently than you expect

**The rise of assistive intelligence**

Key insights

- Voice AI augments human work rather than replacing it
- Assistive voice tools scale compliance and care
- Agent orchestration unlocks smarter workflows
- Valuable insight requires human validation

The generative agent boom of 2024 arrived with ambitious promises. As "AI-first" became the dominant mantra, companies from Klarna to Duolingo raced to announce agent-led interfaces, often to mixed reception. By 2025, reality has tempered this excitement, with headlines of companies retreating from aggressive rollouts.

The meaningful transformation hasn't emerged through flashy avatars but through behind-the-scenes orchestration that amplifies human capabilities. What works isn't wholesale replacement, but thoughtful augmentation.

**From support role to orchestration**

At Content Guru, voice agents function as integrated components of larger systems, orchestrating connections between data, transcription, and decision-making without demanding centre stage.

"Our role in AI is as an orchestrator. We surface the best technologies — from transcription to passive scribing — and integrate them into customer workflows."

Martin Taylor, Content Guru

“We’re already testing audit tools you can speak to… and yes, agents will become part of compliance workflows. But, adoption will lag until regulators lead the way. Compliance is a risk management profession. Risk-averse by nature.”

Peter Kenny, ACA Group

This orchestration approach yields tangible results in healthcare and public services, where Voice AI liberates skilled professionals from administrative burdens.

“Customers are using highly skilled, high-cost people — clinicians, auditors — to do repetitive admin. Voice AI lets these professionals focus on what they’re trained for.” —Martin Taylor, Content Guru

These changes translate to measurable outcomes: clinician hours reclaimed, service backlogs reduced, and professional burnout mitigated — all without replacing the essential human expertise.

### Human insight remains essential

However, speed alone doesn’t create value. In research and compliance, insight quality depends on sound reasoning and contextual understanding that machines still struggle to provide independently.

“Voice AI will surface richer signals… but insight will still hinge on theory-driven questioning and critical human validation.”

Thorsten Dresing, Audiotranskription

This principle applies particularly in compliance, where voice-led tools must meet strict regulatory standards while maintaining complete traceability.

That caution doesn’t preclude innovation, but it reframes the timeline. In high-stakes environments, the question isn’t what’s possible, but what’s reliable enough to trust.

In adtech, where contextual understanding determines brand safety and campaign effectiveness, Voice AI provides crucial support for human teams:

“The tech should help us understand what matters, not just what was said.” —Tamara Zubatiy-Nelson, Barometer

Across diverse industries, the pattern is clear: the most powerful agents operate in assistive capacities rather than autonomous ones. They empower people to work smarter while maintaining essential control over critical decisions.

SPOTLIGHT:

Thorsten Dresing, Audiotranskription

WHO?

Managing Director

WHERE?

Academic and qualitative research

WHAT THEY’RE DOING WITH VOICE:

Audiotranskription’s tools reduce manual transcription and increase researcher capacity. The team is now exploring multi-layered workflows where humans and models work in tandem.

KEY INSIGHT:

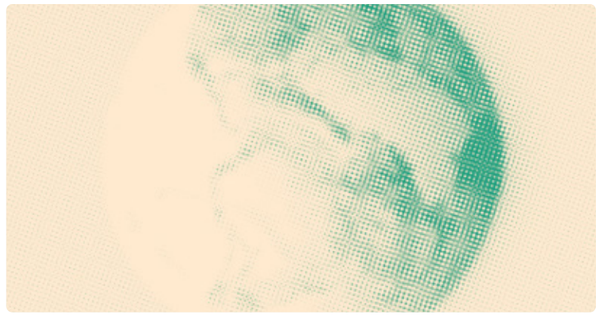
“Having reclaimed transcription time, the real frontier for qualitative research is hybrid interpretation. Voice AI will surface richer signals — prosody, pauses, affect — but insight will still hinge on theory-driven questioning and critical human validation.”

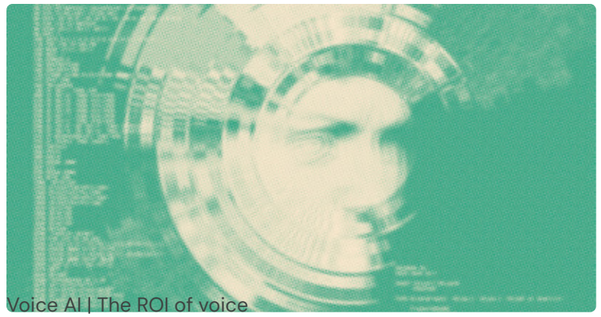
“”

# Accessibility used to be an afterthought.

# Now it’s a business driver.

Tony Abrahams, AI-Media



# The ROI of voice: From gimmick to growth

### Where voice makes money, not headlines

Key insights

- Voice AI generates measurable ROI through faster workflows and better CX
- Growth stems from integration rather than novelty
- Voice significantly boosts engagement across media, research, and service
- Real performance replaces pilot demonstrations

For years, Voice AI impressed audiences during keynote presentations but stumbled in practical implementation. Pilots showed promise. Live deployments often disappointed.

Many companies focused on flashy demos — some costing dollars per minute to run — that couldn't scale into reliable, everyday solutions.

By 2025, the gap between promise and delivery has narrowed considerably. Organisations are now measuring gains rather than questioning adoption.

### Growth through enhanced productivity

When Audiotranskription upgraded their speech recognition engine, students abandoned manual methods almost immediately. Workflow bottlenecks vanished and adoption rates soared 400%.

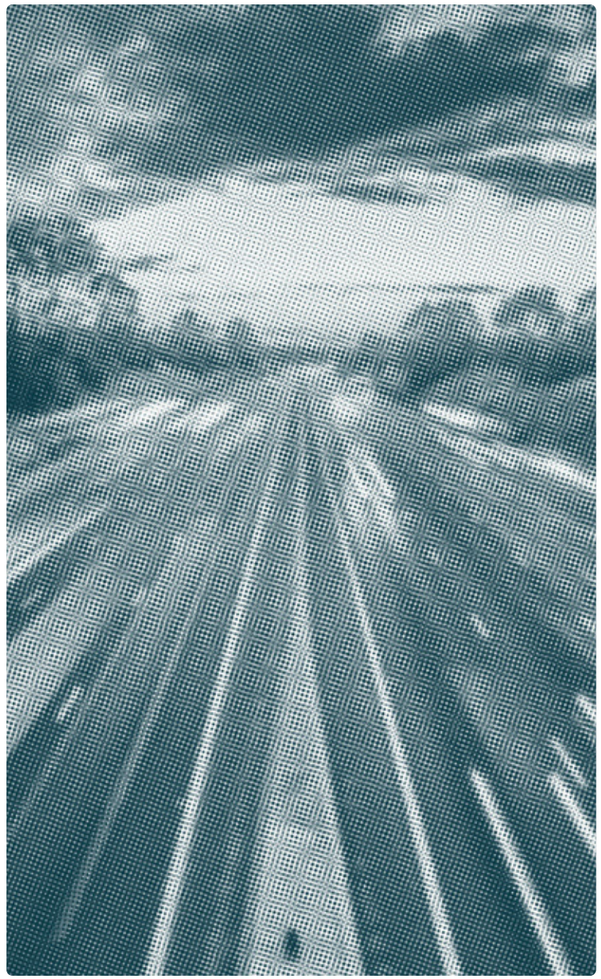
"Most students stopped manual transcription... (and were) satisfied by a huge margin."

Thorsten Dresing, Audiotranskription

This kind of productivity leap isn't limited to academia. Contact centres show similar results, with Voice AI enabling faster responses and smarter routing:

"We've increased customer satisfaction to about 93 percent – unprecedented in this sector."

Martin Taylor, Content Guru



That same focus on performance at scale has extended into the public sector. In the UK, Content Guru replaced an existing voice engine at DVLA with a more accurate solution — a move that helped the agency become one of the most improved departments in government.

> "We swapped the Speechmatics engine into DVLA, replacing another perfectly good engine that we had installed earlier in the programme, and that's helped them become the most improved part of the UK government, according to independent analysts."
>
> Martin Taylor, Content Guru

### Voice drives engagement and revenue

In media, spoken content outperforms text-only alternatives, creating new revenue through multilingual delivery. Companies report up to 19% increases in audience engagement and 25% reductions in production costs.

AI-Media's efficiency enables remarkable scale:

"Since listing on the Australian Stock Exchange in 2020, we've scaled to deliver 120 times more content with basically the same revenue base." —Tony Abrahams, AI-Media

### From feature to product

Voice AI has evolved from optional enhancement to core product driver:

> "What's happening now with voice feels different. It's not just a feature – it's becoming the product."
>
> Tony Abrahams, AI-Media

Media companies – among the top three industries investing in Voice AI – now use it to reach previously inaccessible markets. ROI extends beyond cost reduction to new audiences and revenue models, transforming both operations and business possibilities.

SPOTLIGHT:

Tony Abrahams, AI-Media

WHO?

CEO

WHERE?

Global broadcast and media

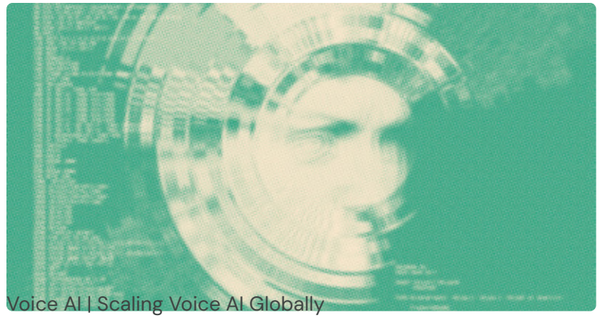
WHAT THEY'RE DOING WITH VOICE:

AI-Media began with live captioning for accessibility. Now it's a cornerstone of content strategy. Voice AI helps them localise, personalise and expand into new markets, turning speech into a high-engagement, high-ROI channel.

KEY INSIGHT:

"In media, accessibility used to be an afterthought. Now, it's a business driver. Localisation, live captioning, and emotional accuracy don't just broaden reach — they open markets."

Trend 4



# Monolingual to Multilingual – Scaling Voice AI Globally

**Why global Voice AI must speak more than one language**

Key insights

- Multilingual switching is becoming standard in CX and research
- Voice AI must support real-time hybrid language use
- Language functions culturally, contextually and dynamically
- Limited language support creates exclusion and erodes trust

By 2025, English-only voice tools have become increasingly limiting. Hybrid conversations now represent the norm, with multilingual support evolving from nice-to-have to fundamental requirement across global enterprises.

In contact centres, media, and research, conversations frequently switch languages mid-sentence. Systems that can't adapt cause failures throughout the downstream tech stack.

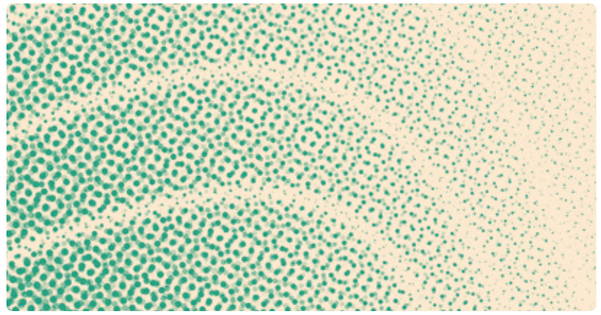
**Hybrid speech as standard practice**

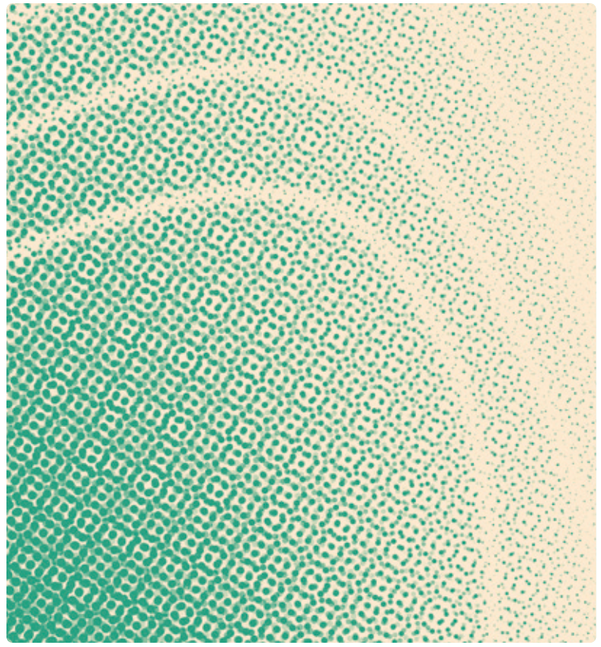
In Denmark, Zylinc supports government and telecom services where seamless bilingual conversations occur daily:

"A third of our calls use hybrid Danish-English. That's standard...

If your Voice AI can't follow a bilingual customer, it doesn't belong in CX."

Henrik Skourup, Zylinc





Financial services teams face similar challenges across regions with distinct dialects, idioms and cultural references:

"I'd hear people code-switch mid-sentence all day – even bilingual humans struggle to follow. For systems, it's an even steeper hill." —Peter Kenny, ACA Group

Even within formally English-speaking environments, language transitions happen constantly. The ability to handle mixed language inputs is now non-negotiable.

### Cultural Understanding Beyond Words

At ACA, Peter Kenny witnesses this complexity daily:

"Idioms are really the critical difference – what sounds culturally resonant to locals can be complete gibberish to outsiders." —Peter Kenny, ACA Group

This makes transcription not just technical but cultural. Even minor differences between related language variants can alter interpretation, especially in compliance scenarios.

In broadcast media, AI-Media sees multilingualism as business expansion:

"Why isn't everything on the BBC also available in Spanish? That's a commercial opportunity — not just an inclusion initiative."

Tony Abrahams, AI-Media

In advertising and media, comprehension must extend beyond words to meaning:

"Language switches mid-call. It happens constantly. We need tech that adapts with the speaker." – Tamara Zubatiy-Nelson, Barometer

### The Language Roulette

Decisions around language coverage ultimately reflect risk tolerance:

"Do you aim for 100% accuracy in one language, or accept 90% in five? That's the roulette game of multilingual AI."

Peter Kenny, ACA Group

Systems that can't adapt in real time risk not just miscommunication but misrepresentation. In 2025, multilingual capability isn't just a performance advantage, it's a trust imperative for Voice AI in global enterprise environments.

SPOTLIGHT:

Henrik Skourup, Zylinc

WHO?

Product Lead

WHERE?

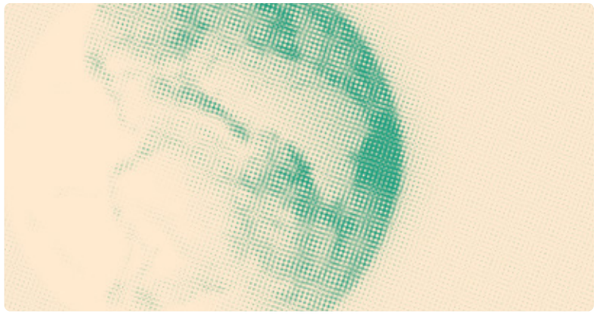
Nordic contact centres and public services

WHAT THEY'RE DOING WITH VOICE:

In Denmark, nearly a third of customer service calls involve both Danish and English. Zylinc's Voice AI keeps pace with hybrid conversations in real time, helping public sector agents triage needs, detect urgency and serve citizens without friction.

KEY INSIGHT:

"In contact centres, no two conversations are the same. Customers switch languages, change tone, and interrupt — and your voice tech has to keep up without losing context."

# Fighting fiction: accuracy in the age of hallucinations

**When AI makes things up, who pays the price?**

Key insights

- Accuracy forms the core contract of Voice AI
- Hallucinations pose legal and operational threats
- Trust requires traceable, defensible transcripts
- The way something was said matters as much as what was said

By 2025, the core question surrounding Voice AI is no longer just about functionality, but credibility.

The consequences of hallucinations have moved from theory to reality. OpenAI's Whisper, adopted by 30,000 clinicians, hallucinated in 1.4% of medical transcriptions—inventing nonexistent treatments. Cursor's Voice AI announced fabricated policies, triggering account cancellations. ChatGPT falsely accused a Norwegian man of murder, raising serious GDPR concerns.

Across sectors, the fallout is mounting: over 120 legal cases worldwide have involved hallucinated AI content, with some penalties exceeding $10,000.

And the risk isn't just legal, with insurers now offering policies to cover losses from AI chatbot errors, signalling a new era of accountability and risk management for AI-generated content as a whole.

The stakes couldn't be higher.

## Accurate transcription as legal evidence

In regulated environments, transcription doesn't just document—it creates legal and operational records. And when models hallucinate, they generate violations, not just errors.

"You're not just transcribing voice. You're constructing evidence."

Peter Kenny, ACA

## Building defensible systems

Accuracy extends beyond the words themselves. In many industries, tone, timing and delivery influence how voice input is interpreted.

"Even one word being wrong could lead to a miscategorisation of context." —Tamara Zubatiy-Nelson, Barometer

For media companies, those nuances affect brand suitability. In compliance, they can determine whether a case is resolved—or misfiled.

## Accuracy transforms workflows

Beyond risk, accuracy directly impacts efficiency. In research, reducing the need for correction speeds up everything from data analysis to insight delivery.

"Using accurate Voice AI saves our customers 50–60% of their correction time." —Thorsten Dresing, Audiotranskription

At scale, that accuracy becomes a productivity multiplier. When teams trust what their systems capture, they move faster and reduce downstream friction. That trust is also what prevents AI's most dangerous side effects from entering the workflow in the first place.

SPOTLIGHT:

Peter Kenny, ACA Group

WHO?

Managing Director

WHERE?

Financial compliance and regulatory audits

WHAT THEY'RE DOING WITH VOICE:

ACA uses voice transcription not just as documentation, but as defensible evidence. Their workflows demand precision across jurisdictions and languages. When AI makes something up, the regulatory consequences are real.

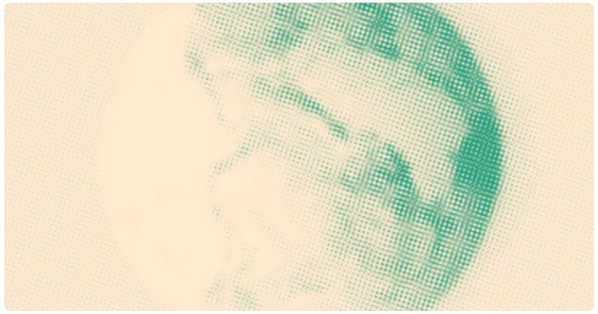
KEY INSIGHT:

"We're looking closely at LLMs, and it's clear they'll be a game changer—not just for surveillance and compliance, but for the financial services industry. We may be in the midst of a new industrial revolution, and the way we work now won't be the same in 5 or 10 years."

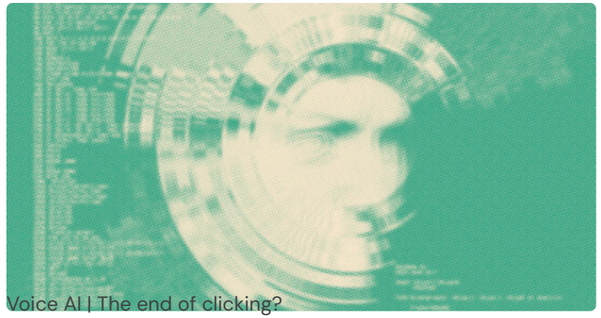
“”

# Hallucinations aren’t just theoretical. They’re regulatory risk.

Peter Kenny, ACA



Trend 6



# The end of clicking? How conversation is becoming the new interface

**Why your next interaction will mix voice and visual**

Key insights

- Voice will augment existing interfaces in most use cases
- Intelligent agents evolve beyond Alexa and Siri across platforms
- Context-aware voice creates seamless multimodal experiences
- Specialized environments drive voice-first adoption

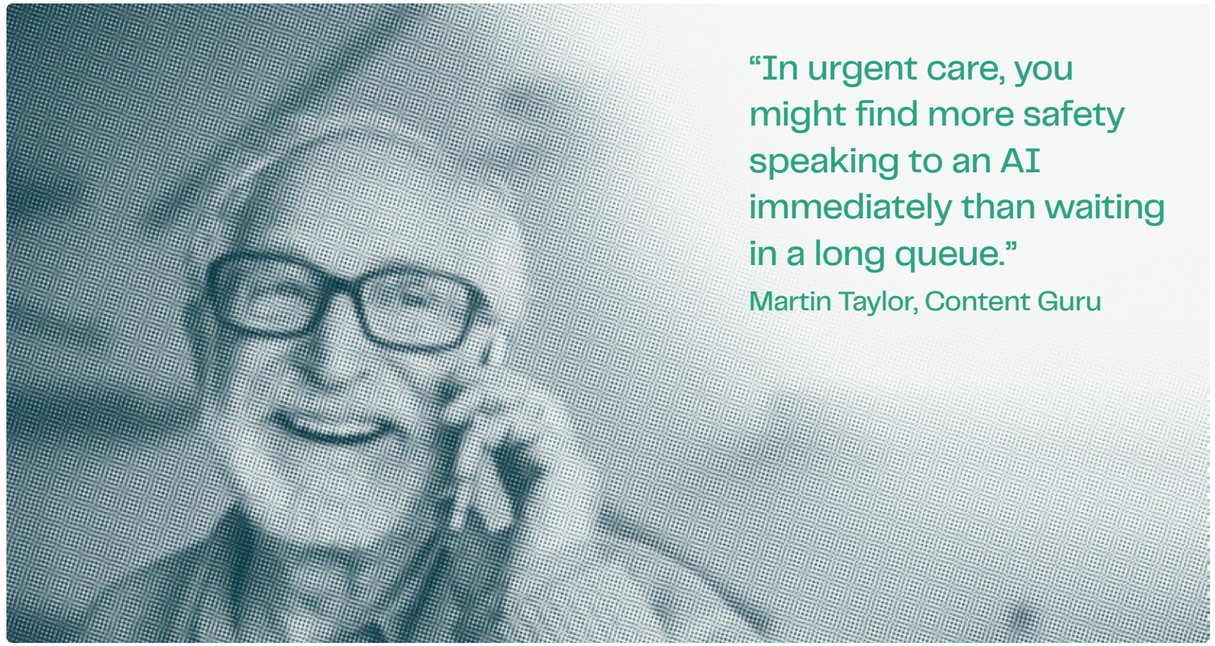
Interfaces traditionally required visibility: screens, icons, scrollbars. By 2025, voice is becoming a powerful complement to these visual elements, creating richer interaction possibilities.

This represents evolution from static inputs to dynamic, adaptive conversation layered over familiar interfaces.

"We're heading away from search. Away from typing. Away from websites as the central place people go to find what they need."

Tamara Zubatiy-Nelson, Barometer

This shift becomes most apparent in environments where traditional interfaces create barriers. High-pressure situations like healthcare demonstrate where voice becomes primary. Clinicians lack time for clicking. Callers resist waiting.



> “In urgent care, you might find more safety speaking to an AI immediately than waiting in a long queue.”
>
> Martin Taylor, Content Guru

Voice has evolved from additional feature to contextual interface, offering a more human way to engage with technology while screens remain essential for complex tasks

### Beyond interface to experience

More sophisticated voice agents are emerging beyond familiar assistants like Alexa and Siri, spanning both cloud and on-device deployments.

On-device Voice AI proves particularly critical as an enabler of seamless interface evolution. Without the latency, privacy concerns, and connectivity dependencies of cloud-based systems, on-device processing will soon allow voice to respond instantly and work anywhere – essential for true interface transformation.

“Better AI tools drive wider adoption – making everything more accessible by default.” – Tony Abrahams, AI-Media

### The conversational future

While voice currently augments visual interfaces, a longer-term question is surfacing: will conversation eventually become the dominant mode of interaction?

> “Websites may disappear within a decade. The future revolves around dialogue.”
>
> Tamara Zubatiy-Nelson, Barometer

Today, brands layer conversation over visual confirmation. Users speak their intent but receive visual validation. Voice doesn’t compete with existing journeys, it makes them more intuitive and accessible.

The companies building tomorrow’s interfaces aren’t asking whether voice will transform interaction. They’re asking how fast.

SPOTLIGHT:

Tamara Zubatiy-Nelson, Barometer

WHO?

CEO and Co-founder

WHERE?

Adtech and contextual media vetting

WHAT THEY’RE DOING WITH VOICE:

Barometer scans spoken content before ad placements — interpreting tone, cadence and risk signals before anyone presses play. For Tamara, the future isn’t keyword-based. It’s conversational. Voice is fast becoming the new layer of the internet.

KEY INSIGHT:

“The future of advertising is conversations — not just podcasts, but smart speaker dialogues, chatbot chats, and generative voice interactions. We’re not placing ads on pages anymore. We’re placing them in moments.”

# What's next for Voice AI?

We asked AI thought leaders, innovators and influencers in our network to build on our report with their predictions for where Voice AI is heading next:

"Voice agents will see significant growth in simple use cases but will continue to struggle with more complex scenarios. Human-in-the-loop will be an important architecture in the near future."

Davit Baghdasaryan,
CEO & Co-Founder, Krisp

"We're inching closer to systems that feel less generic and more personal – where voice AI knows it's you talking, not just a user. That shift will redefine what people expect from digital interaction."

Irena Cronin, Co-Founder &
SVP of Product, DADOS Technology

"The interesting thing about artificial intelligence is not how smart it is — but how frail it reveals the human system to be. We're moving from guessing to knowing... and that shift makes AI not just an option, but an imperative."

John Nosta,
Innovation theorist

"We're not quite there yet with AI that can handle multiple voices in real time — a human can cook, host, and hold five conversations at once. That's the bar. But it's coming, and when it does, it will change the nature of how we live and work."

Robert Scoble,
Tech Futurist

"The next wave of Voice AI will come from speed and a bottom-up, viral spread. Once one teammate starts blitzing through work with instant, adaptive voice-to-action AI, the rest will scramble to keep up – not just for efficiency, but out of fear of being displaced. Voice agents in the workplace will shift from novelty to necessity in the blink of an eye."

Stefan Girard,
Author/owner, Frontier Notes

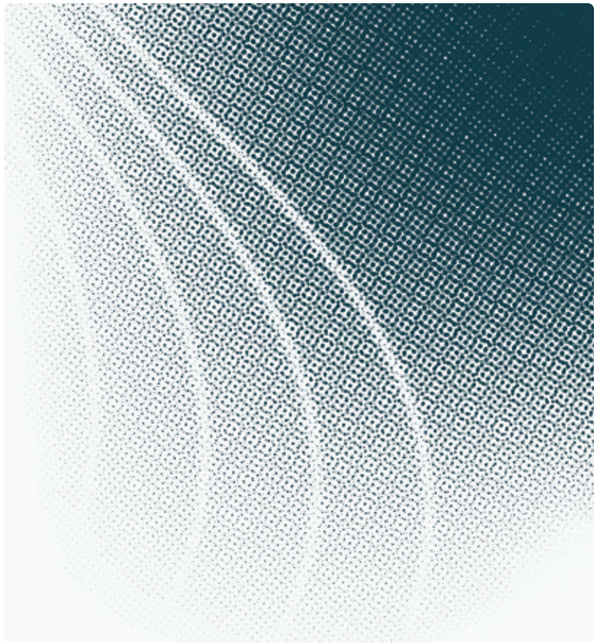
"As computing power accelerates and AI models become more advanced, voice AI is poised to revolutionize media and entertainment. What began as basic question-and-answer systems will soon evolve into emotionally intelligent, context-aware conversational partners—capable not only of anticipating user needs but also acting with a degree of autonomy. These agentic AIs won't just respond—they'll collaborate, becoming proactive members of creative and operational teams."

Richard Kerris,
VP of Global Media & Entertainment, NVIDIA

# Voice AI in 2025: Silence triumphs over the noise

Voice AI is no longer just a concept. It's beginning to embed into real-world systems—not as a showpiece, but as the next phase in the evolution of speech technology. From emergency services to compliance teams and healthcare platforms, voice is being integrated into workflows that demand speed, accuracy and trust.

This transformation isn't playing out in demo reels or keynote speeches. It's unfolding quietly, in the background of systems where performance matters more than polish.



## Beyond the hype: Where Voice AI is starting to work

The companies making real progress with Voice AI have moved beyond broad "AI-first" narratives. Instead, they're focusing on specific, measurable outcomes—streamlined processes, increased productivity, improved customer experience.

Where Voice AI is starting to work, it's not because it's futuristic. It's because the speech infrastructure behind it is solid, and the execution is grounded in context, control and clarity.

Across our interviews, one thing is clear: the organisations seeing results are those who treat voice not as a feature, but as a function—built in, not bolted on.

## What sets successful adopters apart

If you're building with voice, here's what the most effective teams are doing:

**Deploy as infrastructure:** Voice is most powerful when embedded deep in core systems—not added at the edge.

**Amplify human capabilities:** Agents that support, rather than replace, human judgment deliver the clearest returns.

**Measure outcomes, not features:** The best benchmarks are practical: hours saved, errors reduced, satisfaction gained.

**Embrace linguistic complexity:** Modern systems must keep up with real speech—code-switching, regional accents, multilingual input.

**Demand defensive reliability:** In regulated or high-stakes environments, accuracy is the contract. If voice gets it wrong, everything downstream is at risk.

### The invisible advantage

Every trend in this report points to the same shift: voice is becoming more useful, more contextual and more integrated. The real progress isn't in what voice looks or sounds like — it's in how quietly and effectively it gets things done.

The real story of Voice AI in 2025 isn't loud. It's operational. It's practical. And it's already reshaping how organisations operate.

SPEECHMATICS

# Speechmatics: Enterprise-grade APIs for Speech-to-Text and Voice AI Agents

At Speechmatics, we build voice technology for demanding enterprise environments where every word matters - complex, multilingual, compliance-heavy, high-pressure scenarios where precision proves non-negotiable.

Cambridge-born and globally trusted, we deliver three decades of innovation in every line of code. Our proprietary technology carries the weight of breakthrough after breakthrough. While others race to market, we've been refining Voice AI line by line, deployment by deployment.

Enterprise-grade means something different at Speechmatics, something forged from thirty years of solving the hardest challenges in speech tech. Our technology captures context, emotion and speaker identity across more accents, dialects and languages than any other platform. It works in real time, even in noisy environments with overlapping speech and domain-specific terminology.

We partner with solution providers across key voice use cases — from media to healthcare — and support enterprises looking to take full control of their AI stack, from infrastructure to interface.

Building with you, not just for you, our team of industry-shaping developers, researchers, and engineers work with you from day one of your Voice AI journey. We don't simply implement features, we co-create your foundations, bringing our specialized expertise directly into your technology stack.

Deploy anywhere – on your servers, on your devices, on your cloud so Voice AI is always on your terms. Stay in control with zero data retention and support for a wide range of compliance standards, depending on your region and industry.

The future of enterprise voice AI runs through companies requiring resilience, nuance, and real-world performance.

That's why they choose Speechmatics.