Transcription for Compliant Voice Dos and Don’ts

8 Dos and Don’ts When Using Transcription for Compliance

THETALAKE

Voice recording and recording review is an essential practice for highly regulated industries, because regulations concerning recording, handling, and retention have been in place for some time.

Whether to meet regulatory requirements for financial services like the CFTC's Dodd-Frank Regulation 23.202, the UK/EU MiFID II, and Canadian IDPC Rule 3804, depending on jurisdiction/location, or to adhere to internal governance and compliance standards for any regulated company, compliance recording and review needs are increasing. To meet these increasing voice recording review requirements, many organizations have turned to voice-to-text transcription, which has typically been deployed in waves in an effort to gain better insights into audio and video information.

While such use cases have been popular in regulated industries like financial services and healthcare, the relatively small gain in insights using transcription compared to the high resource and time cost to implement them hasn't yielded the right result. This is clearly the case with the first generation of transcription tools. Even in the age of evolved voice-to-text technologies that now use advanced processing and deep learning, the gains in speed and accuracy still haven't met the last-mile need for compliance context, domain relevance, and actionability.

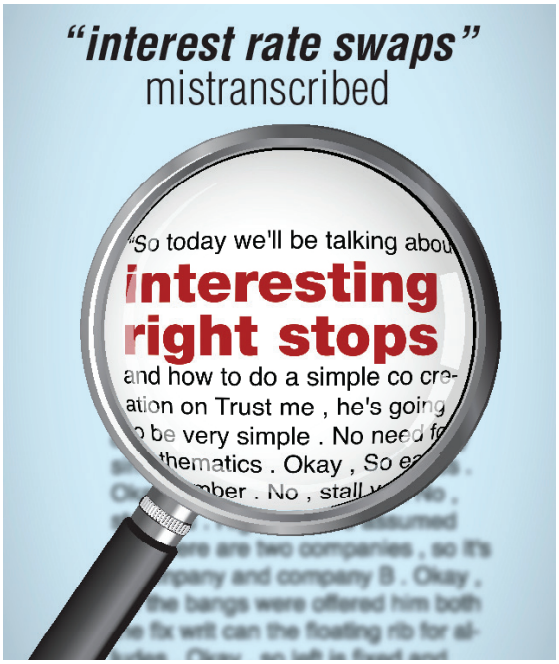
These issues are exacerbated in regulated industries, with requirements covering an ever-increasing volume of digital communications. For example, while financial institutions have historically had recording practices in place for limited operational purposes, global regulatory mandates to record telephone and some video conversations have significantly expanded since the 2008 crisis. At the same time, audit, review, and supervision of these communications remains challenging, because voice-to-text transcription and doing searches on those transcripts isn't enough to quickly and accurately find risks and address them.

Corresponding compliance requirements have also expanded, to include pre-review of audio or video-based advertising and marketing content. Using voice-to-text transcription solutions to pre-review advertising and marketing videos manually has similarly proven an extremely inefficient effort.

As you map out your audio recording program and think about Digital Communications Governance (DCG) programs, these essential dos and don'ts for transcription will serve as helpful guideposts:

1. Don't Rely on Transcription Alone

Don't build your compliance review and supervision program on a foundation that solely relies on a transcription tool. Voice-to-text transcription is not a complete solution. The required time and resource investments don't scale, and won't stand the test of time. Nor will they solve the fundamental challenge of context-based analysis.



If you've tried using a standalone transcription tool into which you have fed audio and video files, you know this presents many challenges with creating or cobbling together your own analysis of the transcript documents. It drags you into the weeds, troubleshooting potential errors in the transcription; for example, trying to gain clarity on homonyms like "for" and "four," or parsing the contextual meanings of words like "account" or "promise." Once you've spent time doing this, you then have the daunting and time-consuming task of implementing search tools to begin the error-prone and ineffective effort of using keywords to find risks. Lastly, you still need to figure out how to report and act on them.

Often, vendors will tout low "word error rates", or "WER", as evidence of more accurate and useful transcription.

However, measuring WER as the holy grail of accuracy ignores other problems with transcription pitfalls and precision. Most vendors test WER under highly controlled circumstances—it is not uncommon for a test to use studio-quality audio of a single speaker, over-articulating terms without any diction or accent nuances. Reported WER scores often neglect the fact that two or more overlapping voices are common on telephone calls or collaborative conversations. Furthermore, pristine WER tests ignore background noise and bad recording quality, which are often encountered in conversations on mobile phones, online meetings, VOIP lines, Bluetooth peripherals, and crowded trading floors. Testing WER in isolated environments leads to favorable scores, but fails to account for the real world circumstances under which conversations and recordings actually happen.

It is commonly believed that a lower WER shows superior accuracy in recognition of speech. However, at least one study has demonstrated that this may not be true, showing that a true understanding of spoken language relies on a lot more than just high word recognition for accuracy.

In short, attempting to search voice recordings to find swap-related conversation risks, collusive behavior, off-channel communications, abuse, customer complaints, and risky behavior in advisor-client discussions, is significantly more complex, less efficient, and less accurate if interpretation of the content is done only through a baseline transcript, especially if focused only on the WER of the transcript.

2. Don't "Bolt-On" Transcription to Your Legacy Compliance Tools

Relying on vendors that simply add a transcription tool into systems like legacy email-based archives for search may seem like a fast-fix approach, but it increases your manual review effort and introduces risk. Ultimately, this will inundate you with non-relevant and non-domain-specific false positives, that consume your precious review time.

While false positives create a productivity expense, this approach also introduces significant business risk from the resulting false negatives (i.e. the compliance and security risks that go undetected using this method).

3. Don't Ignore Mixed Media Communication with Voice, Text, Video and Visual Components

Don't settle for a tool that only covers the audio portion of a communication. Over time, more and more of your voice communications will be mixed in with things like SMS, meeting content from in-meeting chats to screen shares and more. For these, you'll also need to know what is typed, shown, shared, and what is said, to gain context for identifying and addressing risks in these recordings.

If the entirety of your approach can only be applied to the audio of a recording, you will impede your ability to satisfy review needs across tools and conversations with meshed aComms, eComms, and vComms components. A siloed approach also introduces a need to keep adding solutions, spend, and resources. Furthermore, you'll experience the same limitations from trying to bolt-on to your last investment, creating an inferior solution that ultimately costs time, increases risk, and increases opportunity costs.

4. Do Use a Layered Approach Including Transcription

Transcription should be considered a component of your audio, eComms, and vComms compliance infrastructure, which serves as a base layer of any compliance solution. From that layer of highly scalable and accurate transcription, you need to add a normalization filter to clean up and apply context to a transcript. Effective normalization mechanisms apply advanced techniques for analysis, including provable natural language processing (NLP) and AI technology, as well as industry-specific relevance.

For a simple example, your normalization layer should detect the words 'count dirty won for you too,' analyze the proximity and context of other words, and correct the transcript as a conversation about 'account thirty-one forty-two.'

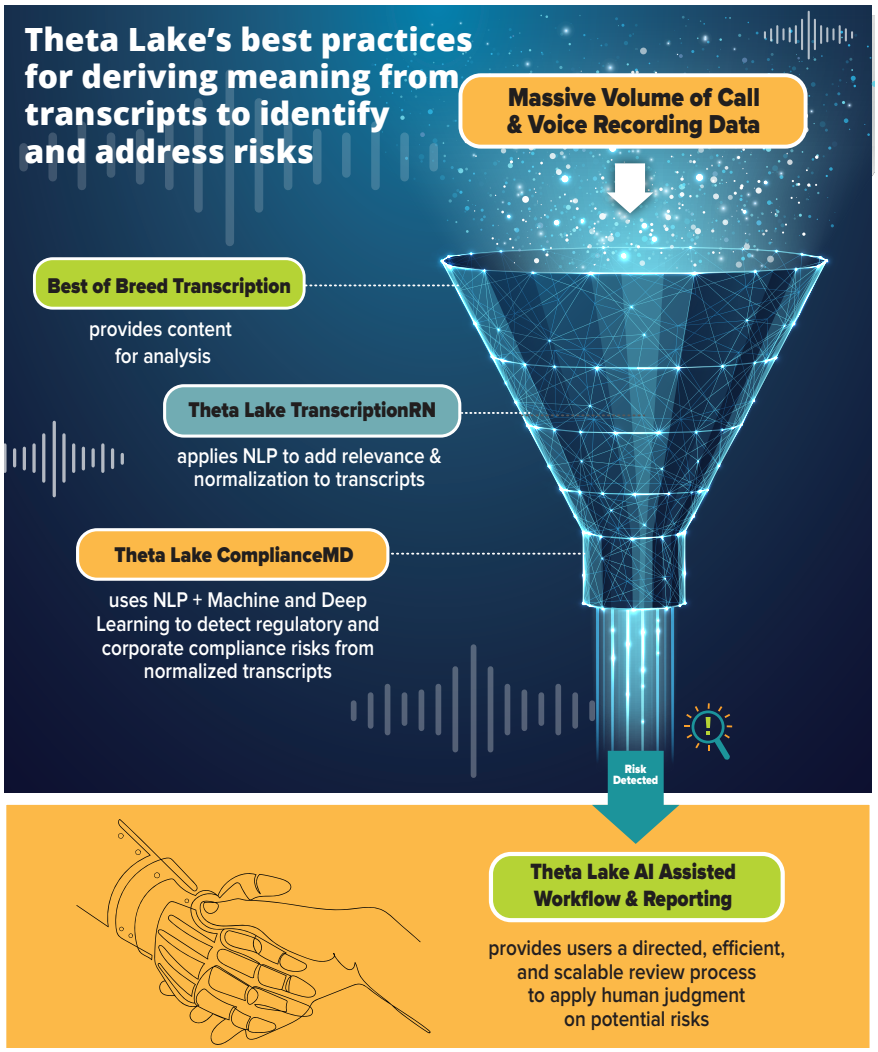
The next layer addresses your analysis methods. Instead of manual keyword searching, you need to use automated machine-learning-based detections that are trained on relevant regulatory and governance risks. Using the example from above, your deep learning system with AI should analyze your normalized transcript and trigger a PII policy detection for the presence of an account number. Then your solution should provide the ability to take that recording analysis with risk detections, and process it through a guided workflow for review and supervision of the recording.

5. Do Ensure Your Providers' Transcription Layers are up to current industry standards

Do validate that your communication compliance vendor is using a modern transcription tool that is up to the task of providing high quality transcripts, at speed and scale. Your vendor should address both the capabilities of their chosen transcription layer in terms of reliable and accurate transcription, as well as what they put on top of that layer in terms of techniques and capabilities to filter and refine results for and improved search and analysis. Hint from above: if the vendor isn't doing any transcript relevance normalization, AI, or machine learning (ML), then you are not using the right solution.

For example, your vendor should provide you with information on whether the transcription tool provides flexibility for tuning based on:

- Audio type, call center, video calls, and voice in multi-party video conferences
- Language support, with the ability to transcribe in multiple languages, ability to translate transcripts into multiple languages on-demand, ability to tune based on accents, and more
- If their transcription layer can handle and demonstrate tougher transcription problems around word error rate (WER), domain-specific models, and others
- If the solution uses targeted compliance technology to handle semantic relatedness, soundalikes, and lookalikes, with advanced analysis capabilities using fully explainable NLP and ML
- If they can support the use of other transcription tools, or transcripts from other tools, beyond storing them - and with the ability to apply compliance context to the transcript



"Ensure Regulatory Compliance with Unified Supervisory Workflows: Implement a unified supervisory workflow that reinforces defensible, risk-based supervision strategies and facilitates collaboration with regulators and external stakeholders."

6. Do Use Integrated and Automated Workflows for Review

Beyond providing recording analysis and transcripts, look for a tool that integrates automated review workflows that allow you to listen to and comment on the timeline, follow the audio with the transcript side-by-side, and report using the recording itself. Review processes based on manual review and tracking of outcomes are one of the most time-consuming efforts. Automated workflows will help you overhaul and streamline your current review processes.

Look for workflow frameworks that provide filter capabilities based on the type of recording, which can include region, participant types, device type, app type, detection types triggered, and more. Your solution should allow you to apply analysis policies and tailored workflow review steps based on each type or category of recording. For example, you should have mechanisms to specify that only calls with certain detections, certain participants, from certain regions are routed for review by a human, who must then provide review comments and obtain a secondary review before the process is closed with the record, review and audit log in place.

Given the ever-increasing use of audio and voice-enabled unified communication and collaboration tools used by employees, partners, and customers to communicate and share information, ensuring that these channels are properly captured for record keeping and supervision to meet regulatory expectations is critical.

Employing more complete solutions that use AI, ML, and NLP at every layer to extract and analyze content in communications with a unified supervisory workflow will reinforce reporting for defensible, risk-based supervision strategies, and facilitate an open and collaborative approach of interacting with regulators and others outside your organization.

7. Do Allow for Flexible Use of Any Transcription Engine

Since transcripts provide the building blocks to multiple oversight processes, organizations should have flexibility in terms of choosing the transcription vendors of their choice. Some firms may have existing trusted transcription platforms tuned to specific nuances of their environment, while others may want to use the transcription built into the UC platforms. Beyond simple generation or storing of a transcript, any modern Digital Communications Governance (DCG) platform should have a modular and flexible approach that permits firms to integrate any existing transcription technology to extend capabilities, and take advantage of pre-trained models. To validate this flexibility with your DCG provider, ask them for a review and overview of their multi-transcription test environment and examples that include at least 10 distinct transcription engines.

8. Do Use Transcripts as the Basis for LLM-Powered Summaries

As described above, transcripts must be used as the foundation for the application of AI-enabled risk detections. And, given the innovations in the area of generative AI, high quality transcripts should also be integrated into supervision processes as source material for summarization.

Along with emails, chat conversations, and other content, transcripts of voice communications should be incorporated into GAI-enabled supervisory processes that summarize user interactions across platforms over long periods of time. For example, firms must be able to aggregate and summarize a four week-long conversation between three employees that occurs across Microsoft Teams, SMS, screen shares in contact centers, Zoom meetings, and a trader voice recorded phone line. If, during the course of these interactions, the parties shared a link in Teams to an encrypted

messaging app, and subsequently discussed using the app during a recorded phone call, a summary of the conversation should contextualize this issue over time and describe its evolution across communication platforms.

Organizations must be able to compile and analyze the chat, SMS, voice transcripts, and screen-share video content to understand the complete conversation and the potential risks identified. Using transcripts to understand the context and evolution of conversations across multiple platforms is essential.

Summary

Transcription is a foundational technology for compliance and digital communications governance across unified communications. Key to the success of using that foundation is setting the right expectations, avoiding common pitfalls, and doing extended validation of DCG vendors. By following these 8 best practices, an organization can more successfully harness the value of transcription for compliance. Theta Lake is a leader in digital communications governance built to help organizations get better results from transcription and improve compliance outcomes.

About Theta Lake's Solutions for Compliance Transcription Effectiveness

Theta Lake's multi-award winning Digital Communications Governance Suite uses Theta Lake's multi-patented TranscriptionRN®. This Transcription Relevance Normalization technology includes patented capabilities to identify lookalikes, soundalikes, and semantically related terms using morphological analysis, metaphone phonetic algorithms, word formation technology to handle ASR grammar errors, and other deep technology-based approaches to drive better compliance accuracy and effectiveness with transcription.

In use across customers in North America, the United Kingdom, Europe, Asia Pacific, Latin America, and Australia & New Zealand, TranscriptionRN® is developed and continually improved by a combined team of regulatory compliance industry, data science, and engineering experts. This team leverages test harnesses developed with customers as well as internal systems that use multiple versions and variations of over 12 market leading transcription platforms. The technology and team behind TranscriptionRN® and its implementation in the Theta Lake Digital Communications Compliance Suite are built to overcome challenges with transcription Word Error Rate and increase effectiveness in reviewing audio recordings to reduce risk and maintain regulatory compliance.

Learn more

Visit our Voice Solutions Web Page