

Clontract: A Multi-Agent System for Spotting Risks in Voice Acting Contracts and Negotiating Control Over Voice Data

Yihao Zhou^{*}, Farhad Hossain^{*}, Ayae Ide, Tory Park, Tanusree Sharma[†]

Pennsylvania State University, State College, PA, USA

{yihaozhou, mjh7856, ayaeide, vcp5105, tanusree.sharma}@psu.edu

Abstract

Early large-scale audio datasets, such as LibriSpeech, were built with hundreds of individual contributors whose voices were instrumental in the development of speech technologies. Yet, a decade later, these same contributions and their regular voice work now expose actors to unauthorized cloning, impersonation, and long-term contractual exploitation. While existing ethical frameworks emphasize Consent, Credit, and Compensation (C³), they do not adequately address the emergent risks involving vocal identities that are increasingly decoupled from context, authorship, and control.

To close this gap, (a) we first introduce **PRAC³ (Privacy, Reputation, Accountability, Consent, Credit, Compensation)** framework in assessing risks that emerge over time and beyond contractual boundaries; (b) we then introduce **Clontract**, a multi-agent system that delivers context-aware contract assessment and negotiation support for voice actors. Clontract enhances context understanding and reasoning in voice actor contract analysis, leveraging PRAC³, NIST, and OWASP standards to identify vague terms on voice copyright, voiceprint reuse, and Text-to-Speech / AI training or use. A Preliminary evaluation on real-world voice acting contracts (Amazon ACX), Clontract provides comprehensive and actionable negotiation advice. By pairing an empirically grounded risk taxonomy with an agentic analysis pipeline, Clontract lays a foundation for automated, domain-aware protections that support voice actors to make informed decisions during contract negotiation to manage the risk of the voice data.

Keywords

AI Risk Assessment, Voice Data, Voice Professionals, Agentic AI

1 Introduction

Data sharing has long been a contested domain between individual contributors, professionals, and data controllers. Individuals or groups contribute data either deliberatively, whether in pursuit of social value, to receive financial compensation, or as part of their primary profession [6, 11, 19, 22]. Among these contributors and professionals, voice actors are one of the pragmatic contributors and professionals who played a foundational role in the development of modern speech technologies [26, 27, 29]. A notable innovation was the early large-scale audio datasets, LibriSpeech, derived from thousands of contributions to LibriVox and other public domain audiobook platforms, underpinned early breakthroughs in automatic speech recognition and the voice assistants we use today [17, 23, 25, 30]. These contributions, originally made in the spirit of open knowledge and accessibility, have since been repurposed

into commercial AI pipelines often without consent, attribution, or safeguards [7]. A decade later, these same contributions have exposed voice actors to a range of harms and may devalue or displace the very actors who created them.

Unlike textual or visual data, voice is not only expressive but also biometric, and it is uniquely identifiable to a person [5]. Thus, voice contributors are prone to a wide range of harms, including unauthorized cloning, impersonation, reputational damage, and identity theft [10, 13]; however, these risks have received little systematic attention and tools to support voice actors. Moreover, voice actors face risks throughout their professional voice work lifecycle. They work across a range of sectors such as, commercials & advertising, followed by audiobooks, animation & cartoons, and E-Learning & educational content, video games, and podcasting & audio dramas, dubbing & localization, live performance & theatrical productions, each showing high demand across the industry. As a part of their primary profession, typically following a structured workflow: **(a) discovery of the work; (b) Audition ; (c) Contracting; (d) Recording and File sharing.** Each phase comes with certain risks.

Due to the limited to no literature addressing risks faced by voice professionals, we began by investigating real-world incidents and associated threat models throughout the voice actor lifecycle (Figure 1, detail empirical results are in Appendix A) throughout the voice lifecycle by **interviewing 20 voice actors**. Drawing on these risk scenarios and threat models, we developed a framework that extends the existing C³ model (Consent, Credit, Compensation) into PRAC³ which incorporates Privacy, Reputation, and Accountability. To the best of our knowledge, this is the first study to examine the experiences of voice professionals through the lens of risk assessment, rather than solely focusing on the general harms associated with creative labor.

We then focused on Contract Negotiation, a critical phase that governs the use of voice data and associated risks over time. Contracts contain intricate terms related to credit, consent, and compensation, which are often misunderstood by voice actors. Particularly, when many voice actors lack union support or legal representation to assess contracts and enforce AI-related clauses (Figure 2) in the era of Generative AI. Text-to-speech model training further adds complexity to these risks, where obscure AI clauses often expose voice actors to unforeseen risks and biometric data rights. Our interviews point out incidents of how contracts have been and can be exploited against voice actors. An instance from **the original male voice of a voice assistant tool of major tech recounted how a one-time session and a yearly non-compete fee evolved into widespread, unauthorized use of his voice.**

“It was released 5 or 6 years back. I regret not having a lawyer review the contract, which included broad

^{*}Yihao Zhou and Farhad Hossain contributed equally to this work.

[†]Corresponding author: Tanusree Sharma (tanusree.sharma@psu.edu).

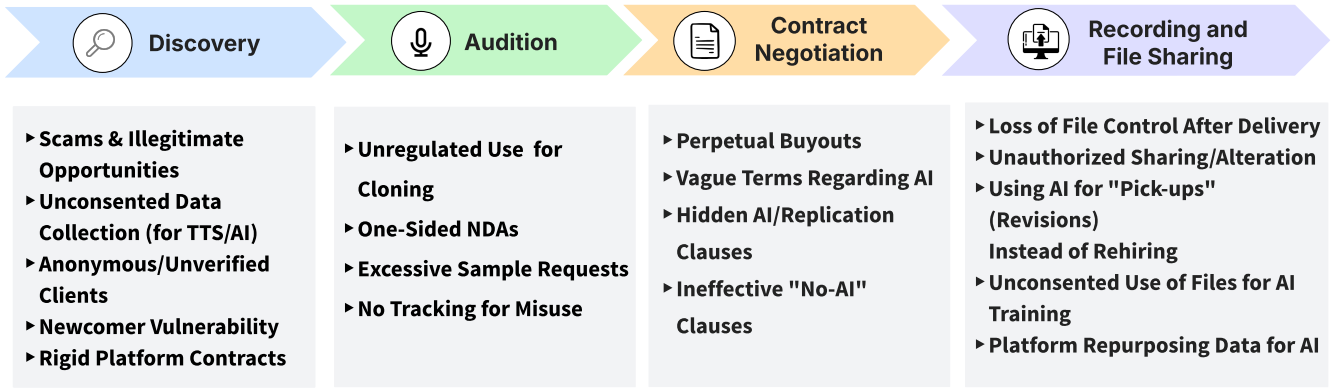


Figure 1: Risks and AI-related threats in different stages of voice acting work, including discovery, audition, contracting, recording and file sharing.

terms like 'in any form or technology now known or unknown, in perpetuity.' I later found out my voice is rented to Y and Z companies. Interestingly got to know that from my daughter and friend. That didn't feel good. I hadn't understood how my voice would be used, but for a while, people kept asking if I did a hotel ad in Berlin or other projects. My voice ended up in explainer videos, commercials, and even a video game chatbox without additional pay. Since then, I've renegotiated for higher compensation. Still, the original deal locked me into a much lower rate especially compared to the female voice, who reportedly earns around \$250k a year."

Similar trends emerged in the case of Bev Standing, who took legal action against TikTok for unauthorized use of her voice in its text-to-speech feature [7]. This highlights the mismatch between contractual terms and long-term value in the AI era. Unlike other creative professionals and data contributors, voice actors are unique because the shared content is **voiceprint**, which is both a personal and professional tool.

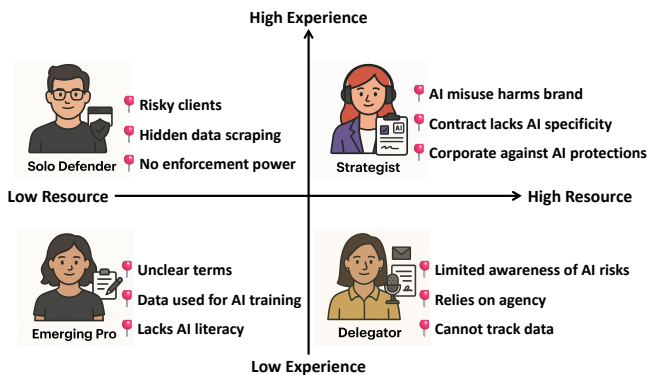


Figure 2: Personas of Voice Actors

Building on this **empirical** understanding of voice actors, we focus on **Contract Assessment and Negotiation** as a key intervention point. We also drew motivation from recent advances in AI-driven agentic systems, enabling automatic parsing and interpretation of contractual terms, potentially mitigating risks by clearly identifying ambiguous language and hidden implications [12, 20, 21, 28]. However, automated contract understanding tools may lack a deep understanding of the context and risks unique to the acting, such as clauses involving voice copyright, AI voice duplication risks, and biometric data protection.

Thus, we build a proof-of-concept multi-agent system – "**Clontract**" to help voice actors enhance their understanding of the contract context and risks. It helps voice actors better understand contract language by identifying ambiguous clauses and AI-related risks. We also incorporated industry specific risk frameworks, as as, NIST, OWASP and our newly created PRAC³ (voice actor-specific) framework for a comprehensive assessment of contract input from voice actors. Clontract uses task decomposition and collaborative agents to perform deep contextual analysis of contract text. It incorporates industry-standard frameworks like NIST and OWASP, along with our domain-specific PRAC³ framework, for comprehensive risk assessment unique to the voice acting profession.

In summary, our contributions are the following:

- (1) An Extended Risk framework **PRAC³** for voice Actors (built from empirical evidence);
- (2) A proof-of-concept tool, **Clontract**, – A multi-agent collaborative architecture to enhance the understanding of the unique context in voice actor contracts;
- (3) A preliminary system evaluation on voice acting platform contract examples (Amazon ACX), which proves the reliability of Clontract in automatic contract analysis and risk assessment.

We hope that Clontract will lay the foundation for future research on automated and context-aware contract analysis.

2 Background

2.1 Ethical Frameworks: From C³ to PRAC³

Our work broadens the discussion of ethical AI data use by expanding the "C³" (Consent, Credit, Compensation) to PRAC³, adding

Privacy, Reputation, and Accountability, which emerged through our findings and are important dimensions for long-term risk assessment. Prior work centered on creators' consent to their data and receive attribution and payment [8, 18]. PRAC³ model can capture context-transcending risks posed by generative AI, for instance, how voice actors' "vocal identities" can become decoupled from context, authorship, and control in AI systems. Our findings reveal that voice, as a unique identifier, can be misused by clients or downstream users, causing harm to contributors' personal and professional identities. PRAC³ thus reframes voice actors as stakeholders, not just content sources, to offer a comprehensive model for assessing risks.

Privacy, as a pillar, encourage rethinking voice data not merely as creative output but as biometric personal data. Voiceprints which is central to voice actors' identity, are often scraped or shared without consent, echoing Zuboff's "surveillance capitalism," where human experience becomes unconsented raw material [31]. Our findings present that voice actors' sign a contract for their voice performance, not the voiceprint. Despite growing legal recognition (e.g., Illinois' BIPA [9], EU AI Act [2], CCPA [14]), our findings reveal widespread misuse, particularly in privacy, security, and safety, due to a lack of provenance. Once voice data is embedded in models and spread across platforms, it's nearly impossible to trace or retract. Unlike image watermarking, to the best of our knowledge, robust voice provenance tools remain undeveloped [15, 24]. Legal protections lag, with gaps illustrated by the TikTok text-to-speech case, where a voice actor's work was repurposed without her knowledge [16]. Further, we found voice data reused in controversial memes, raising unresolved questions of accountability regarding whether to attribute the harm to secondary content creators who used the voice sample or the original voice actors whose voice been used. This indicated reciprocal reputational harm for voice actors. By positioning voice data as personal data tied to privacy, reputation, and accountability, our work advances frameworks for voice data governance in AI.

2.2 Automated Contract Review

Contract review has various stages of complexity: at the foundational level, referred to as "contract analysis," the task involves identifying and extracting risky clauses; whereas at the more advanced "counseling" level, professionals must contextually interpret risks and recommend tailored solutions [12]. In the context of voice acting, the counseling phase requires deeper contextual understanding of industry practices, voice-actor-specific risks (such as biometric data misuse or unauthorized voice cloning), and individual risk tolerance, which poses a big challenge for automated systems.

Traditionally, contract review is a specialized and high-value task that demands considerable time from legal professionals when performed manually [12]. Large voice-acting agencies and platforms have to handle extensive amount of contracts. Meanwhile, individual voice actors and smaller agencies frequently face barriers to professional legal assistance due to limited resources, potentially leading to unfavorable contract terms or exposure to hidden clauses.

Research into automated contract review, especially in a specific domain, remains scarce, although some related efforts have

emerged. In 2020, a study introduced ALeaseBERT [20], a BERT-based language model designed for identifying problematic clauses in lease agreements. Another strategy is Retrieval-Augmented Generation (RAG), proposed in [21], leveraging external knowledge retrieval alongside generative models (e.g., GPT-4) for clause comparison and contextual reasoning. With the advance of LLM, recent studies have explored prompt engineering for contract review [28]. Despite these advances, the literature clearly indicates a gap in automated contract review tailored explicitly to creative workers, or more specifically in our case, voice actors.

3 PRAC³: Conceptualized Threat Model to Assess Risks

As the voice industry intersects increasingly with generative AI, voice actors face distinct and compound risks to their identity, labor, and safety. These risks are structural, embedded in how digital labor is extracted, synthesized, and monetized.

Based on the risk indicators through different phases of voice actors' interaction to digital platforms as well as their experienced and perceived risks (Appendix section A.1 and section A.2), we proposed a PRAC³ framework. This offer a conceptual tool for threat modeling these long-tailed risks, especially in assessing harms that emerge over time and beyond contractual boundaries. PRAC³ stands for "**Privacy, Reputation, Accountability, Consent, Credit, Compensation**". Each dimension represents a critical vector of exposure or harm for voice actors in the AI data economy. **Consent, Credit, Compensation** presents foundational rights which often overlooked or bypassed in AI data pipelines. newly added components from voice actor's experience: **Privacy** which presents breaches of biometric identity through cloning or surveillance; **Reputation**, which represents harm from voice misuse in misaligned, offensive, or deceptive contexts and finally; **Accountability** which present legal and technical gaps in traceability and recourse when voice actors data is misused by adversarial actors and harm general users.

Voice actors experience **three archetypal threat scenarios** that encapsulate both direct and downstream risks. These scenarios highlight how harm is not limited to the moment of data creation but often arises through redistribution, secondary use, and platform-driven commodification.

(a) **Voluntary, non-monetary contributors:** Actors donate voice data for public good, only to have it later surface in unauthorized commercial tools.

(b) **Monetized contractual contributors:** Initial legal agreements include ambiguous language often enabling resale, transfer, or indefinite reuse of voice data, especially following corporate changes.

(c) **Secondary, informal misuse:** Legally recorded voices leak into meme culture, satire, or political propaganda via AI tools, distorting public perception and damaging actors' professional standing.

Across all scenarios, key assets are voice recordings with identifiable voice features, voiceprint which is a unique vocal fingerprint capable of identification or cloning, reputational credibility, and contractual protections. When a voice actor performs, they manipulate multiple acoustic and articulatory signals to create different

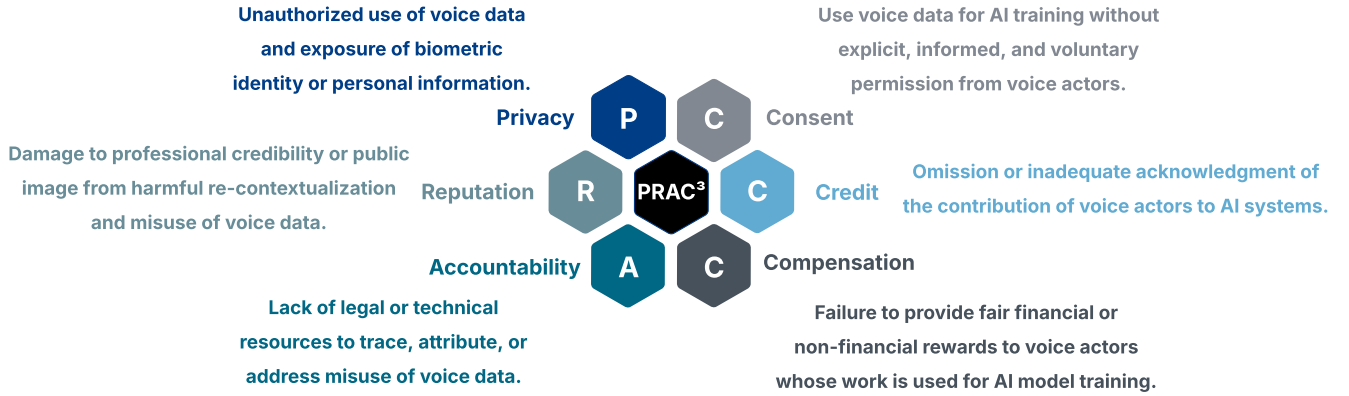


Figure 3: The PRAC³ framework, including six risk dimensions in the use of voice actors’ data in the context of generative AI: Privacy, Reputation, Accountability, Consent, Compensation, and Credit.

characters, emotions, or identities. These changes affect the perceived voice, but the underlying biometric voice signature often remains partially detectable by machines (e.g., AI voice recognition); even if data is anonymized before being shared, advanced analytics or cross-referencing with other datasets could re-identify contributors.

Table 2 illustrates PRAC³ framework by mapping real-world incidents shared by voice actors to the six dimensions. Each case illustrates how risks unfold across time and contexts: P7’s incident where a modder used AI to generate explicit content using a recognizable voice from a game voice character a violation of Reputation, Consent, and Privacy.

4 Tool Development & Evaluation

In this section, we will introduce Clontract, a multi-agent agentic system for helping voice actors understand and negotiate control over contracts and voice data. We begin by formulating the problem in Section 4.1; Next, in Section 4.2, we define the tasks of contract understanding and elaborate on the system architecture, agent design, and operational logic that enable Clontract to deliver accurate and robust solutions. Finally, in Section 4.3, we detail the specialized tools integrated into the system.

4.1 Problem Formulation

Understanding voice-actor contracts involves different tasks, such as case search, contract review, and risk assessment. Completing these tasks requires reasoning across both textual content and specialized domain knowledge. To address these challenges, we propose an agentic framework with multiple collaborative agents. The distribution of tasks ensures accurate context interpretation, adaptability, and clarity in different tasks.

Formally, our system’s core processing can be represented as: $r = \mathcal{A}(q; \mathcal{T})$. Here, \mathcal{A} denotes our agentic system, consisting of two primary components: \mathcal{A}_{plan} , which determines a set of tasks required to analyze the given contract-related query, and \mathcal{A}_{exec} , which executes this subset of tasks.

The input query (q) represents specific contract-related questions or instructions, while each agent is equipped with a specialized

toolkit $\mathcal{T} = \{t_1, t_2, \dots, t_n\}$, where each t_i denotes a distinct tool for functions such as document analysis or knowledge retrieval. Each agent performs its assigned task based on provided instructions (s) given by \mathcal{A}_{plan} , and produces an output $\tau = t(s)$. Finally, the system generates a comprehensive and context-aware response (r) by synthesizing and reasoning over these outputs.

4.2 Clontract

Given the confidentiality of the contract, we considered both performance and accessibility when designing our system. Therefore, we prioritize integration with open-source tools and models whenever feasible. To meet this need, we leverage Agno [4]—an open-source and lightweight agent-centric library, as our main framework. We choose Agno as it provides several essential functions in designing an agentic system: (i) a unified API for instantiating heterogeneous agents (chat, retrieval or heuristic agents); (ii) support for knowledge stores by vector database; and (iii) a packed coordinate agent team mode that automates multi-agent collaboration through the internally supplied reasoning and planning toolkits.

Clontract has **four** functions: **legal research, contract review, risk assessment, and custom query**. To implement these functions, we designed four different agents, as shown in Table 1. Specifically, the **execution of the task consists of two main stages: task planning and contextual reasoning**. In each stage, specialized agents driven by LLM collaborate to achieve different goals and integrate tools such as DuckDuckGo web search, Qdrant vector database, PDF reader, and text embedder to optimize the results. This structured workflow starts with parsing and understanding the contract, then collecting and organizing relevant regulations, cases, risk assessment framework, and conducting iterative collaborative discussions to finally achieve comprehensive report generation.

4.2.1 Task and Agent Design. Contract’s agent design is aligned with its tasks: from the initial understanding of the contract to a complete risk assessment, different agents contribute in distinct ways.

Table 1: Task-specialist agent design.

Agent	Main Role	Dedicated Tools	Core Instruction
Coordinator	Assign tasks & coordinate discussions	Knowledge Retrieval	“Coordinate analysis between team members...”
Researcher	Retrieve statute & precedents	Web Search + Knowledge Retrieval	“Find and cite relevant cases...”
Analyst	Clause extraction, detailed analysis & risk spotting	Knowledge Retrieval	“Identify ambiguous AI clauses...”
Strategist	Synthesis & negotiation advice	Knowledge Retrieval	“Draft mitigation strategies and negotiation tips...”

For **Contract Review**, the task is to review the contract and identify key terms, obligations, and potential issues. For **Legal Research**, the task is to search relevant cases and precedents related to this contract. The goal of these two tasks is to give the user a detailed overview of the contract. The **corresponding agents (Analyst, Researcher)** carry out their tasks independently, then submit the results to the Coordinator for confirmation before presenting them to the user. The **output is delivered in three modules: “Analysis”, “Key Points”, and “Negotiation”**, to reflect layered results that progress from comprehensive to concise and actionable as per the expectation of voice actors.

For Risk Assessment, the task is to analyze potential legal risks and liabilities in the uploaded contract based on common risk assessment frameworks. All team members work on this collaboratively. The Researcher retrieves relevant cases, regulations, and frameworks, ensuring the other members have solid references; the Analyst, equipped with sufficient background knowledge, leads the contract analysis; and the Strategist offers user-oriented advice based on the findings. The entire process is not strictly linear but iterates under the Coordinator’s supervision. For example, the Analyst first conducts a preliminary analysis, then refines it using the frameworks and cases supplied by the Legal Researcher. The results are provided in two parts: “Risk Review” and “Scoring”, where the former is a comprehensive report and the latter gives risk scores (0–100) based on the referenced frameworks, to provide the user with an intuitive understanding of the results.

For Custom Query, all agents collaborate to analyze the query and report the results, using the same output format as in Legal Research and Contract Review.

4.2.2 Task and Tool Chain Planning. As illustrated in Figure 4, task planning is the initial phase upon receiving a query. Given a specific query or task (q), the coordinator \mathcal{A}_{plan} will systematically decompose it into a subset of queries. Specifically, this is achieved through reasoning about the query’s requirements and the capabilities of the agents available in the team.

4.2.3 Contextual Reasoning. Once the task is assigned, the execution agents \mathcal{A}_{exec} will process iteratively based on the decomposed query and the accumulated execution history. Each step in the process is history-aware under the guidance of the coordinator \mathcal{A}_{plan} , which enables the agent to adaptively determine appropriate inputs for each tool. At the i th step, the coordinator \mathcal{A}_{plan} will generate the instruction input (s_i) based on the query and the accumulated execution history \mathcal{H}_i , which is defined as: $\mathcal{H}_i = \{(t_1, s_1, \tau_1), \dots, (t_{i-1}, s_{i-1}, \tau_{i-1})\}$, with $\mathcal{H}_0 = \emptyset$.

The instruction input (s_i) and the tools (t_i) in the toolbox \mathcal{T} are then used to generate the output (τ_i) as follows: $\tau_i = \mathbf{t}_i(s_i)$, $s_i =$

$\mathcal{A}_{exec}(q, \mathcal{H}_{i-1}; t_i)$. This iterative process ensures that each agent can fully understand the context of query and current analyzing progress when reasoning the next step.

Specifically, for the Researcher, we integrate a pre-built dataset that contains contract-related policies and documents from many voice acting platforms, as well as a risk assessment framework (PRAC³) specifically for voice actors, into its knowledge base. This dataset enables Clontract to better understand the context when analyzing voice acting-related documents and write targeted reports.

4.2.4 Output Generation. When the feedback from the execution agents \mathcal{A}_{exec} is evaluated as sufficient and appropriate by the coordinator \mathcal{A}_{plan} , it will generate the final response (\mathbf{r}) in a structured format. For example, in the risk assessment task, the output for risk review will be in markdown format with section-based bullet points of risks in different categories, and the output for scoring will contain a comprehensive table with corresponding risk scores in the categories.

4.3 Toolbox

Clontract integrates five specialized tools. These tools are used to guide the decision-making and operational processes within our system. As illustrated in Figure 3, the tools are detailed as follows:

4.3.1 Web Search. This tool facilitates real-time retrieval of relevant information from the internet and enable timely access to updated legal guidelines, standard contract practices, and current industry standards for voice actors. In our system, we use DuckDuckGo as the search engine as it focuses on user privacy and can avoid filter bubbles caused by personalized searches.

4.3.2 Knowledge Retrieval. In our system, the documents uploaded by the user are stored as knowledge in a vector database, and the agents will use it for RAG or dynamic few-shot learning. Agno agents use Agentic RAG by default, which means they will search the knowledge base for the specific information they need to complete a task. Specifically, the agent will call language models to generate a set of keywords, and then call the knowledge retrieval tool to retrieve the relevant information or few-shot examples.

4.3.3 PDF Reader. The PDF reader is implemented to handle and parse contract documents directly. This tool extracts textual content from PDF files and segments it into several small chunks for the following semantic analysis and clause interpretation process within the agentic framework.

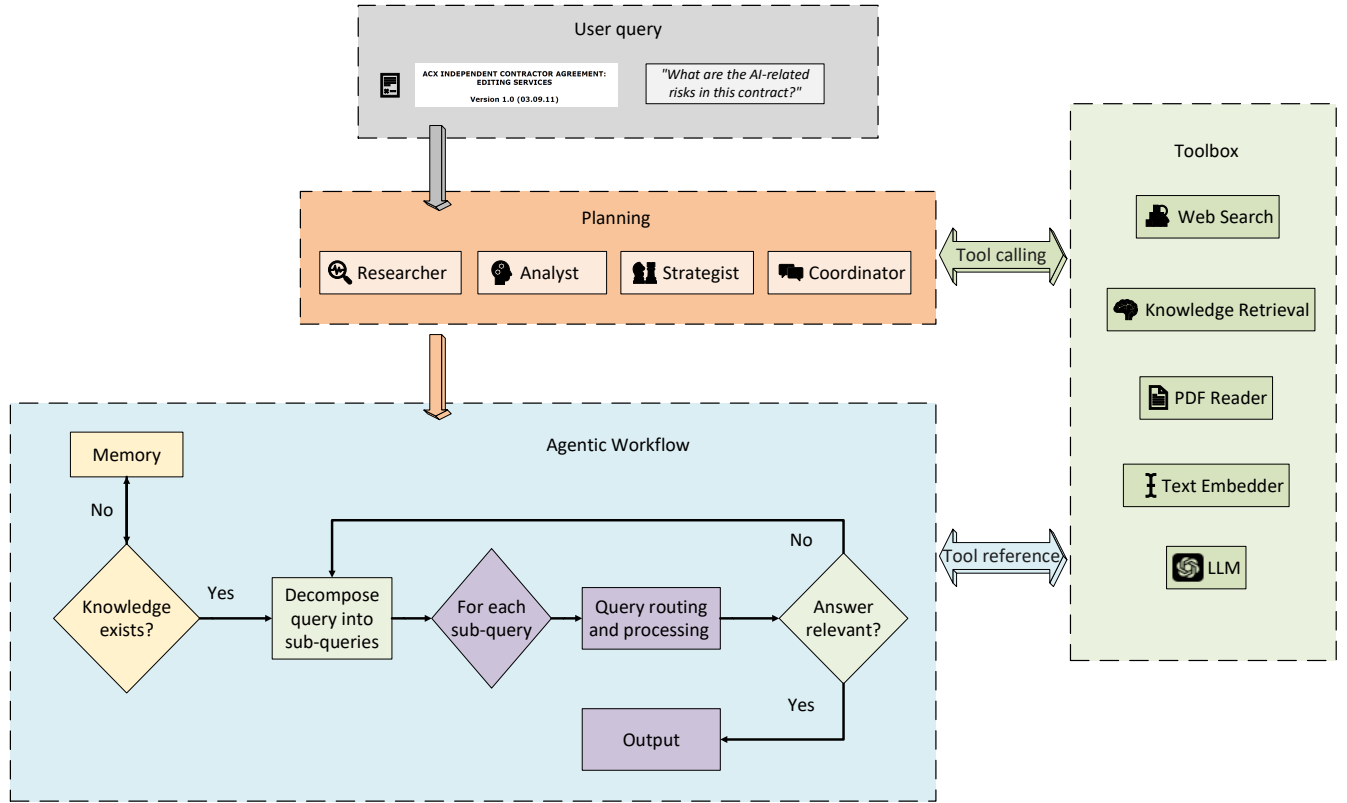


Figure 4: Clontract Architecture Overview. We design an agentic multi-agent system to decompose the contract-related query and utilize task-specialist agents with different tools to generate comprehensive responses.

4.3.4 Text Embedder. Before processing the contract, our system utilizes text embedder to convert textual data into numerical embeddings for efficient and accurate semantic searching, similarity comparison, and contextual reasoning tasks.

4.3.5 Large Language Model. As a powerful, general-purpose language processing engine, Large Language Models (LLMs) are implemented in our system to handle different natural-language tasks, including interpreting clauses, identifying potential risks, and generating clear, user-friendly explanations and recommendations.

text-embedding-3-small model and indexed in a Qdrant vector database. The team is built in "coordinate" mode of Agno, with reasoning tools for explicit chain-of-thought and DuckDuckGoTools available to the Legal Researcher for supplementary web search. All agent queries sample with temperature 0.4, return up to 1024 tokens, and retrieve the top-5 nearest chunks from the vector database; these settings keep inference costs predictable while ensuring sufficient contextual coverage for contract analysis.

5 Experiments

We start from the experiment setup in Section 5.1. Then in Section 5.2, we describe the dataset we collected for improving the context understanding ability of Clontract. Finally, in Section 5.3, we use real contract examples publicly available from Amazon ACX to conduct a qualitative analysis of our system.

5.1 Experiment Setup

All experiments run on the same agent team (Researcher, Analyst, and Strategist). Each agent is powered by GPT-4.1 and configured to log every tool invocation. Documents are chunked into 1,000-token pieces with 200-token overlap before being embedded by OpenAI's

5.2 Dataset Curation

To support context understanding in contract analysis and risk assessment for voice actors, we collected 76 primary-source legal and policy documents from 21 leading voice-over and creative-service platforms including audiobook production (e.g., ACX), freelance marketplaces (e.g., Upwork, Fiverr, Freelancer), and dedicated voice-talent exchanges (e.g., Voices, Voice123, VOPlanet). For each platform we captured its public contract and legal-related policies and documents, such as Terms of Service/User Agreements, privacy policies, content and community guidelines, royalty-payment and escrow instructions, independent-contractor or production agreements, licensing and distribution terms, and specialized clauses on AI, intellectual-property, or biometric data.

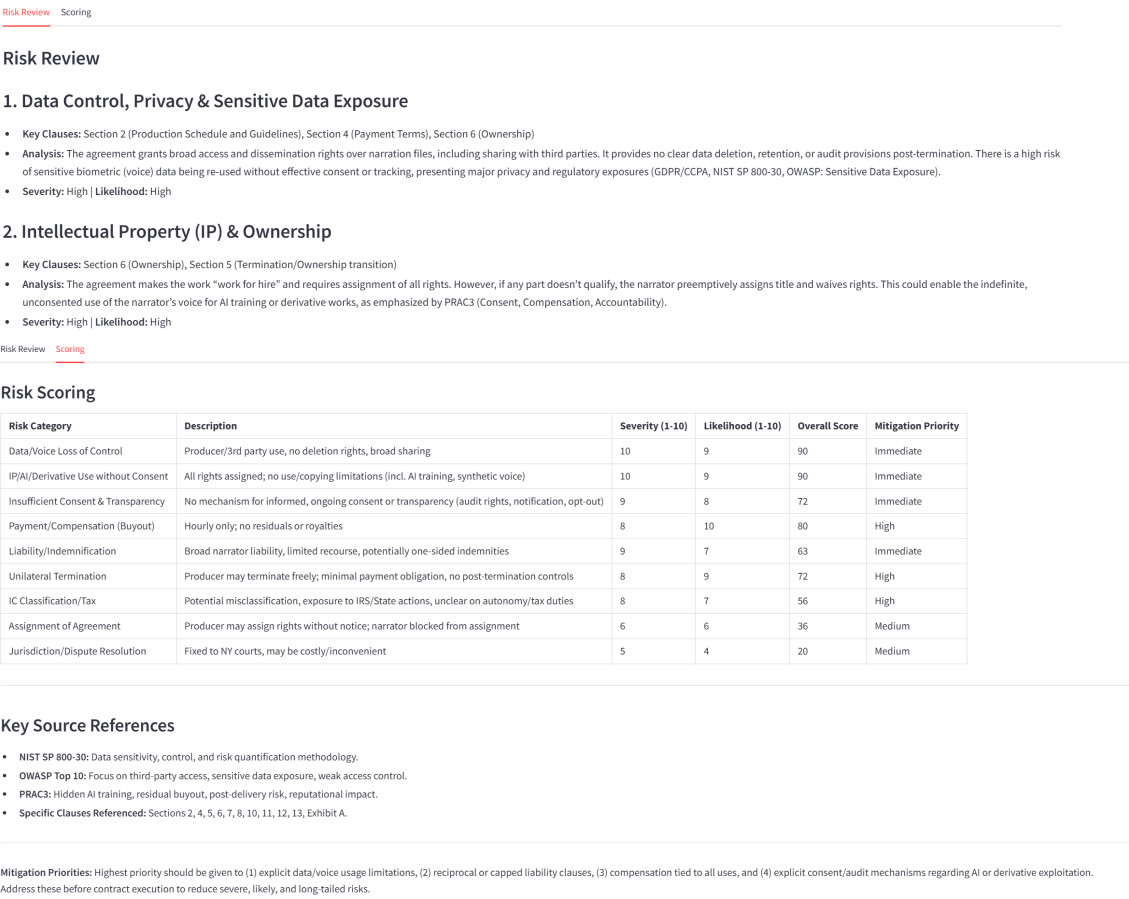


Figure 5: Task Performance Results. We use risk assessment as an example to show the structured output of Clontract.

Further more, as described in Section 3, voice data has its unique vulnerability. Therefore, we propose a novel risk assessment framework, PRAC³ - an expansion of C³ that foregrounds Privacy, Reputation, Accountability, Consent, Credit, and Compensation as interdependent pillars of data used in the synthetic voice economy. This framework captures how privacy risks are amplified through non-consensual training, how reputational harm arises from de-contextualized deployment, and how accountability can be modified in AI Data ecosystems. When performing risk assessment, the PRAC³ framework, which is stored in the knowledge base, will be cross-referenced with other common frameworks such as NIST and OWASP.

5.3 Task Performance

We use ACX Independent Contractor Agreements [3] from Amazon ACX as our sample contract for evaluating our system. In Figure 6, we show the results of Clontract performing the risk assessment task. It can be observed that both the risk review and scoring parts provide structured outputs. The former not only provides a review report by category, but also references common risk assessment frameworks, such as NIST and OWASP, as well as our own PRAC³ framework. In the scoring part, Clontract provides intuitive output

in the form of a table and indicates which section in the contract it corresponds to. This example demonstrates Clontract’s ability to understand the context of the contract and its ability to use domain-specific knowledge.

6 Discussion

Using real world voice acting contract as a test case, we find that Clontract shows the ability to perform context-based contract analysis and risk assessment.

6.1 Social Impact and Industry Standardization.

and We believe that the Clontract adapted PRAC3 framework can significantly contribute to social impact by establishing industry standards for voice data in AI economy. Voice actors, especially freelancers and early-career careers, do not have access to legal support, union protections, or well-defined industry standards. Consequently, they often encounter a continuous influx of intricate, legally complicated contracts under time constraints, and are frequently required to relinquish long-term rights, particularly concerning AI utilization, data licensing, and voice repurposing, without sufficient understanding or options for recourse. Although there are some informal, community-driven safety measures like the Nava

Analysis
Key Points
Negotiations

Negotiations

Negotiation strategies and positions

Here are key, actionable negotiation strategies and priority points for the most significant sections of the "ACX Independent Contractor Agreement: Mastering Engineering Services," directly grounded in the contract analysis and supported by relevant case law and industry practice:

Section 5: Termination Provisions and Fees

Negotiation Points and Strategies:

- **Clarity on Termination Fees:**
 - Ensure all potential termination fees (amount, trigger events) are transparent and capped.
 - Tactic: Request a schedule of fees/examples or negotiate a pro-rata formula instead of flat fees.
 - Use Case Law: Cite *Kavounas v. United Aluminum* to confirm that courts demand clear, non-unconscionable rules.
- **Mutuality:**
 - Negotiate for reciprocal termination rights and fees if only one side's remedies are described.
 - Tactic: Point out "freedom of contract" includes mutual remedies.

Section 6: Work Made for Hire / Ownership

Negotiation Points and Strategies:

- **Assignment Backup:**
 - Ensure the "assignment" clause fully covers any scenario where "work for hire" might not apply, and clarify compensation for unanticipated uses.
 - Tactic: Propose explicit language covering all ancillary materials, and reserve moral rights where feasible.
 - Use Case Law: Reference *CCNV v. Reid* to argue for clear assignment and full payment.
- **Scope of Assignment:**
 - Limit assignment to works specifically commissioned under the agreement, not broader creations or portfolios.

Figure 6: Outcome of Negotiation Tips- structured output of Clontract.

AI Rider [1], which certain voice actors request their clients to incorporate in contracts, such practices are not widespread. We hope that Clontract can serve as a pathfinder to get the ball rolling in creating context-based solutions to govern AI-related risks in voice data. Looking forward, we plan to improve the system and expand our contribution from multiple aspects:

6.2 Future Direction

This work is not comprehensive. We plan to continue this work to the following direction.

System Evaluation and User-Centered Design. We have not yet performed a quantitative performance evaluation. Currently, in voice professions, there is no existing benchmark systems, rather, contract is typically handled on a case-by-case basis by a legal representative for those who have access. We plan to conduct a two-fold, human-centered evaluation involving both voice actors and legal experts to assess Clontract's performance.

Expanding the Dataset. Current system includes contracts and policy documents from over 21 major voice platforms (e.g., Amazon ACX, Fiverr, Voice123), covering licensing terms, AI and data usage clauses, NDAs, and voice-over agreements. Moving forward, we aim to curate a more diverse set of contract samples from voice actors engaged in non-platform work such as individual projects,

collaborations with agents, or contracts from production houses to enrich the knowledge base.

Improving System Adaptability. The current Clontract prototype is deployed based on the lightweight framework Agno and OpenAI Chatgpt. In the future, we will work to adapt the system to more popular agent frameworks (such as Autogen) and language models (such as Google Gemini). At the same time, platforms such as Auth0 are used to implement authentication and authorization services to ensure user privacy.

In the long term, we will build an AI Risk Observatory specifically for voice actors with incidents to build anticipatory risk modeling of voice data in the evolving AI ecosystem.

7 Conclusion

In this work, we proposed an agentic multi-agent system, Clontract, that helps voice actors assess risks in contracts and provide negotiation assistance. Our experiment shows that Clontract can generate comprehensive and structured responses to contract-related tasks. Built on a pre-built dataset of legal and policy documents from leading voice-over and creative-service platforms, as well as a risk assessment framework specifically designed for voice actors, Clontract can provide contextually situated, robust analysis results and negotiation suggestions.

References

- [1] [n. d.]. SYNTHETIC VOICE/ARTIFICIAL INTELLIGENCE RIDER - NAVA — navavoices.org. <https://navavoices.org/2023/01/23/artificial-intelligence-rider/>. [Accessed 01-06-2025].
- [2] EU AI Act. 2023. EU AI Act: first regulation on artificial intelligence | Topics | European Parliament — europarl.europa.eu. <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. [Accessed 23-05-2025].
- [3] ACX. 2025. ACX Independent Contractor Agreements. <https://help.acx.com/s/article/acx-independent-contractor-agreements>. Accessed: 2025-05-30.
- [4] Agno-AGI. 2025. Agno: A lightweight, high-performance library for building Agents. <https://github.com/agno-agi/agno>. Accessed: 2025-05-30.
- [5] Petar S Aleksic and Aggelos K Katsaggelos. 2006. Audio-visual biometrics. *Proc. IEEE* 94, 11 (2006), 2025–2044.
- [6] Anita L Allen. 1999. Privacy-as-data control: Conceptual, practical, and moral limits of the paradigm. *Conn. L. Rev.* 32 (1999), 861.
- [7] BBC. 2021. Actor sues TikTok for using her voice in viral tool. <https://www.bbc.com/news/technology-57063087>. [Accessed 23-05-2025].
- [8] Allie Blaising and Laura Dabbish. 2022. Managing the transition to online freelance platforms: self-directed socialization. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–26.
- [9] Andrew Cook. 2024. Illinois BIPA: A Litigation Nightmare for Employers. *UIC Law Review* 57, 2 (2024), 5.
- [10] Brenda Leong Ekene Chuku-Okeke, Ade Adetunji. 2023. Voice actors and generative AI: Legal challenges and emerging protection. <https://iapp.org/news/a/voice-actors-and-generative-ai-legal-challenges-and-emerging-protections>. [Accessed 23-05-2025].
- [11] Béatrice Godard, Jörg Schmidtke, Jean-Jacques Cassiman, and Ségolène Aymé. 2003. Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective. *European journal of human genetics* 11, 2 (2003), S88–S122.
- [12] Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review. *arXiv preprint arXiv:2103.06268* (2021).
- [13] Wiebke Hutiri, Orestis Papakyriakopoulos, and Alice Xiang. 2024. Not my voice! a taxonomy of ethical and safety harms of speech generators. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 359–376.
- [14] Odia Kagan. 2020. CCPA Regulations: Are Audio Recordings Personal Information? <https://dataprivacy.foxrothschild.com/2020/06/articles/california-consumer-privacy-act/ccpa-regulations-audio-recordings/>. [Accessed 23-05-2025].
- [15] Daniel Kang, Tatsunori Hashimoto, Ion Stoica, and Yi Sun. 2022. Zk-img: Attested images via zero-knowledge proofs to fight disinformation. *arXiv preprint arXiv:2211.04775* (2022).
- [16] Jacob Kastrenakes. 2021. TikTok settles lawsuit with actress over its original text-to-speech voice. <https://www.theverge.com/2021/9/29/22701167/bev-standing-tiktok-lawsuit-settles-text-to-speech-voice>. [Accessed 22-05-2025].
- [17] Jodi Kearns. 2014. Librivox: Free public domain audiobooks. *Reference Reviews* 28, 1 (2014), 7–8.
- [18] Lin Kyi, Amruta Mahuli, M Six Silberman, Reuben Binns, Jun Zhao, and Asia J Biega. 2025. Governance of Generative AI in Creative Work: Consent, Credit, Compensation, and Beyond. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [19] Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum. 2014. *Privacy, big data, and the public good: Frameworks for engagement*. Cambridge University Press.
- [20] Spyretta Leivaditi, Julien Rossi, and Evangelos Kanoulas. 2020. A benchmark for lease contract review. *arXiv preprint arXiv:2010.10386* (2020).
- [21] Savinay Narendra, Kaushal Shetty, and Adwait Ratnaparkhi. 2024. Enhancing Contract Negotiations with LLM-Based Legal Document Comparison. In *Proceedings of the Natural Legal Language Processing Workshop 2024*. 143–153.
- [22] Board on Health Sciences Policy and Committee on Strategies for Responsible Sharing of Clinical Trial Data. 2015. Sharing clinical trial data: maximizing benefits, minimizing risk. (2015).
- [23] Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. 2015. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 5206–5210.
- [24] Pavel Pantiukhov, Dmitrii Koriakov, Tatiana Petrova, Jeovane Honorio Alves, Vijay K Gurbani, and Radu State. 2024. Enhanced DeFi Security on XRPL with Zero-Knowledge Proofs and Speaker Verification. In *2024 IEEE International Conference and Expo on Real Time Communications at IIT (RTC)*. IEEE, 23–30.
- [25] Kishore Prahallad. 2010. *Automatic building of synthetic voices from audio books*. Carnegie Mellon University.
- [26] Kishore Prahallad, E Veera Raghavendra, and Alan W Black. 2010. Learning speaker-specific phrase breaks for text-to-speech systems. In *SSW*. 162–166.
- [27] Kishore Prahallad, E Veera Raghavendra, and Alan W Black. 2010. Semi-supervised learning of acoustic driven prosodic phrase breaks for text-to-speech systems. In *Proceedings of 5th International Conference on Speech Prosody (Speech Prosody 2010)*, Chicago, Illinois.
- [28] Adam Roegiest, Radha Chitta, Jonathan Donnelly, Maya Lash, Alexandra Vtyurina, and Francois Longtin. 2023. Questions about contracts: Prompt templates for structured answer generation. In *Proceedings of the Natural Legal Language Processing Workshop 2023*. 62–72.
- [29] Joshua Ian Tauberer. 2010. *Learning [voice]*. University of Pennsylvania.
- [30] Royal Van Horn. 2007. Online books and audiobooks. *Phi Delta Kappan* 89, 2 (2007), 154.
- [31] Shoshana Zuboff. 2023. The age of surveillance capitalism. In *Social theory re-wired*. Routledge, 203–213.

A Appendix

A.1 Indication of Risk Through Interaction

In their routine professional activities, voice actors work across a range of sectors such as, commercials & advertising, followed by audiobooks, animation & cartoons, and E-Learning & educational content, video games, and podcasting & audio dramas, dubbing & localization, live performance & theatrical productions, each showing high demand across the industry. We found a reservation of voice actors to work on Text-to-Speech (TTS) & AI Voice Training project due to both normative and practical concerns. For all participants, voice acting was their primary profession, typically following a structured workflow: (a) discovery of the work; (b) Audition ; (c) Contracting; (d) Recording and File sharing. In this section, we lay out the risks in each stage of their interaction, in particular risks pertaining to with advanced AI landscape, as shown in Figure 1.

Discovery. The first phase involves voice actors seeking projects aligned with their skills, interests, and availability. Participants identified three primary channels for finding work: commercial platforms (e.g., Voice123, Amazon ACX), social media job postings (e.g., Facebook, Twitter, LinkedIn, Bluesky, Discord), and agent representation.

In this phase, participants explained a key concern, which was the difficulty in distinguishing legitimate opportunities, particularly when platforms allow anonymous clients. Thus, determining if the project is legit or a means to collect voice data for unconsented use in Text-to-Speech (TTS) applications, often challenging for voice actors. Participants expressed a preference for working with agents or clients who engage in direct, identifiable communication, allowing for dialogue and verification in contrast. However, a very small number of participants had access and the means for agents or direct connections with publishing houses and individual clients, which largely contingent on the actor’s experience and the professional network they had developed over time. This highlights a level of gatekeeping and structural advantage not available to less experienced voice actors, leading to risks of their voiceprint ending with bad actors.

Audition. Auditions serve as a gateway for voice actors to secure roles, yet this crucial stage remains largely unregulated in terms of how submitted samples may be used beyond the selection process. Participants mentioned producers or client often shielded by non-disclosure agreements (NDAs) signed by voice actors, where the actors themselves typically operate without reciprocal legal protection. The current industry norm relies on informal trust. We encountered incident where audition sample were indeed used, but later remediate by booking the artist, as P11 recounted,

"It's just an unspoken agreement, you have to sort of trust... I've only had once-client used my audition for the job. But they booked me later, so it was fine."

Participants consistently identified this auditioning phase as one of the most vulnerable to misuse particularly regarding unauthorized AI training or voice cloning. To mitigate these risks, voice actors rely on informal, ad-hoc strategies, such as, mark as read flag if (a) requests for unusually long audition samples (b) lack of communication from intermediaries (e.g. agent, client, casting director). As P18 said

"Usually when you do a read like that, it's between a minute to 5 minutes long based on the project types like, commercial might need longer one where audiobook should not need more than 60s. If its longer, there is something off with the client."

Some participants mentioned technical deterrents, such as inserting beeps into their samples to prevent unauthorized use. However, most rejected this approach due to concerns that it compromised the quality and may jeopardize job opportunities. Notably, several participants suspected that their auditions had indeed been misused, yet felt powerless due to no means of tracking or legal recourse.

A.2 Current & Long Term Risks Perception

Our analysis with professional voice actors revealed awareness of both current vulnerabilities and future threats, beyond the three Cs, particularly around privacy, reputation, and accountability.

Security and Identity Concerns Participants expressed growing concern over the biometric nature of voice data and the ease with which it can now be cloned and reused without authorization. Several actors identified the potential for fraud and impersonation particularly, in financial or emergency contexts. For instance, P16 noted

"Scammers can now... call you and say 'Mommy, I'm being hurt' using your kid's voice. And you don't know if it's real. Its really frightening. My voice is out there more than an average users."

Some reported concerns on voice authentication in banking. Meanwhile, actors like P16 pointed to the existential challenge of deepfakes, describing it as *"Not being able to verify your own voice because someone has stolen it.. next-level voice theft."* P8 explained -

"If financial institutions use voices... that's not a good idea considering how easy it is to duplicate. I also sometime wonder- banks that ask for voice verification... Is it being used to train something else?"

One participant with a cybersecurity background (P6) emphasized that some deepfake uses cross into serious crime, noting incidents where AI-cloned voices were used for "swatting" (calling in fake threats) and other dangerous hoaxes. These concerns underscore the shift from theoretical risk to practical harm, particularly for security and safety of voice actors in their personal life.

Reputational and Ethical Risks Voice actors also raised serious concerns about their voices being used in ways that contradict their values and can often damage their personal standing. We found scenarios where some participants found their voice being misused to create AI-generated voice content in controversial

media such as, political, controversial media. One actor recalled a case where P4 mentioned-

"I initially worked on a anime character which was normal. then they made that character do AI-generated porn... that reflects badly on me, which was never consented."

Some also feared their voices could be embedded in propaganda or defamatory content, with no clear mechanism for recourse or correction. P17 described an unsettling experience of hearing accidentally a TV commercial on political agenda in gender issues which sounded like her own voice which she never recorded. This lack of control over one's digital likeness raises questions about the professional and personal boundaries in the age of generative AI.

Accountability and Legal Uncertainty. Participants expressed frustration over the lack of enforceable rights and mechanisms to trace, remove, or contest the misuse of their voice. For example, P17 described a situation in which a TikTok user initially perceived as a fan used a voice sample from her website to create a reel video:

"At first, I hear my voice in the background, it seemed benign. Then I realized there was AI to clone certain words I never said. If those memes become more extreme, who is accountable- me, the person who cloned, or the TikToker?"

Beyond the concerns of accountability, some participants added concern of professional and economical reputation. P17 highlighted how their voice association with low-quality productions or cloned by individuals could damage his credibility, as audiences might conflate the synthetic performance with the original artist. Similarly, P3 explained the opacity of content distribution chains and the inadequacy of existing legal measures:

"I don't doubt one day some content's gonna feature my voice ... and I'm very much scared for that day to navigate legal world... more scared when legitimate companies and criminals alike, now a temptation to 'rip off everybody' by harvesting voices, and our legal system is only starting to grapple with it."

These difficulties were particularly severe for non-union actors, who frequently did not have the financial or institutional backing necessary to explore abuses or seek redress. With many intermediaries, such as, casting agents, platforms, production studios for a project, standing between them, identifying source of harms, and tracing accountability becomes a near-impossible task.

Table 2: Reported data-misuse and AI-related incidents affecting professional voice actors.

ID	Scenario	Incident (Participant)	Analysis using PRAC ³ Framework
1	Audition sample reused in national commercial	P17 discovered her voice in an ad she never recorded (P17)	<p>PRAC³ Domain: Consent, Compensation, Accountability</p> <p>Threat Agent: Client/Studio</p> <p>Asset at Risk: Voice data, creative labor</p> <p>Potential Impact: Unauthorized commercial use; loss of income; reputational risk</p> <p>Mitigation Status: None – discovered post-facto</p>
2	Voice used in AI-generated adult content	Game mod used AI to create pornographic scenes with actor's voice (P7)	<p>PRAC³ Domain: Reputation, Consent, Accountability</p> <p>Threat Agent: Third-party modders</p> <p>Asset at Risk: Public persona, moral integrity</p> <p>Potential Impact: Defamation; emotional distress</p> <p>Mitigation Status: Unreported; no recourse</p>
3	Exhibit A clause allows post-production cloning	Audiobook contract allowed voice replication without notice (P4)	<p>PRAC³ Domain: Consent, Compensation, Accountability</p> <p>Threat Agent: Publisher</p> <p>Asset at Risk: Voice likeness; residual earnings</p> <p>Potential Impact: Job displacement; IP erosion</p> <p>Mitigation Status: Discovered post-signing</p>
4	AI voice scam using child's cloned voice	Scam calls using cloned voice of loved one (P16)	<p>PRAC³ Domain: Privacy, Identity, Accountability</p> <p>Threat Agent: Cybercriminals</p> <p>Asset at Risk: Biometric identity</p> <p>Potential Impact: Financial fraud; emotional harm</p> <p>Mitigation Status: Hypothetical/precautionary</p>
5	Podcast platform AI-translates and clones voice	Company [X] translated podcaster's voice without opt-out (P19)	<p>PRAC³ Domain: Consent, Privacy, Accountability</p> <p>Threat Agent: Platform provider</p> <p>Asset at Risk: Voice data; linguistic identity</p> <p>Potential Impact: Unconsented speech generation</p> <p>Mitigation Status: Actor manually obstructed usage</p>
6	No disclosure of voice reuse for AI training	P4 reported clause only found post-distribution	<p>PRAC³ Domain: Consent, Privacy, Compensation</p> <p>Threat Agent: Client</p> <p>Asset at Risk: Voice training data</p> <p>Potential Impact: Unpaid AI training use</p> <p>Mitigation Status: No consent captured</p>
7	AI-generated voice used in foreign language translation	Company [Y] used AI to translate podcaster's voice without clear opt-in (P16)	<p>PRAC³ Domain: Privacy, Consent, Accountability</p> <p>Threat Agent: Platform</p> <p>Asset at Risk: Voice identity; language authenticity</p> <p>Potential Impact: Loss of control over voice use, misrepresentation</p> <p>Mitigation Status: Voice actor manually obstructed feature with background audio</p>
8	Audition samples used without hiring actor	Actors heard their audition voices in released work (P14, P16, P17, P18, P20)	<p>PRAC³ Domain: Consent, Compensation, Credit</p> <p>Threat Agent: Client/Producer</p> <p>Asset at Risk: Audition recordings; performance data</p> <p>Potential Impact: Unpaid labor; reputational confusion</p> <p>Mitigation Status: Typically undiscovered until after release</p>
9	Voice used in modded game porn content	AI-generated adult content using voice actors' characters (P7)	<p>PRAC³ Domain: Reputation, Privacy, Accountability</p> <p>Threat Agent: Third-party users</p> <p>Asset at Risk: Character alignment; public image</p> <p>Potential Impact: Moral distress; brand damage</p> <p>Mitigation Status: No action taken; actors unaware until fans reported</p>

Continued on next page

(Continued from previous page)

ID	Scenario	Incident (Participant)	Details
10	Hidden AI training clauses in audiobook contracts	Exhibit A allowed voice replication post-recording (P4)	PRAC³ Domain: Consent, Accountability, Compensation Threat Agent: Publisher Asset at Risk: Creative control; residuals Potential Impact: Job replacement by AI; under-compensation Mitigation Status: Clause discovered only post-facto
11	Client reuses voice clip across projects without permission	P17's voice reused in ad without consent	PRAC³ Domain: Consent, Accountability, Credit Threat Agent: Client Asset at Risk: Vocal performance; authorship Potential Impact: Unauthorized branding; reputational risk Mitigation Status: No prior notification; discovered incidentally
12	Scam calls using AI voice cloning of relatives	Actors fear scammers using their voice for fraud (P3, P16)	PRAC³ Domain: Privacy, Identity, Accountability Threat Agent: Cybercriminals Asset at Risk: Biometric voice identity Potential Impact: Financial scams; family trauma Mitigation Status: No technical prevention mechanisms
13	AI contracts lack explicit voice usage limitations	Contracts omit AI voice use clauses (P14, P1)	PRAC³ Domain: Consent, Privacy, Accountability Threat Agent: Clients/Platforms Asset at Risk: Legal rights over voice data Potential Impact: Non-consensual reuse or AI training Mitigation Status: Actors often overlook contract language
14	Perpetual license buried in email agreements	Clients assume full rights from email threads (P10, P18)	PRAC³ Domain: Consent, Compensation, Credit Threat Agent: Clients Asset at Risk: Work ownership; royalties Potential Impact: Lack of residuals; misappropriation Mitigation Status: No formal legal review of communication
15	Replacement by AI for minor roles or demo work	Lost work for minor roles to AI-generated voices (P14)	PRAC³ Domain: Compensation, Reputation, Accountability Threat Agent: Clients Asset at Risk: Job opportunities; creative career pathways Potential Impact: Job displacement Mitigation Status: Community advocacy; union action (no technical protection)
16	Voice licensed and mass redistributed via third-party	Large tech Company [Z] licensed actor's voice to third-party platforms (P12)	PRAC³ Domain: Consent, Compensation, Accountability, Privacy Threat Agent: Clients Asset at Risk: Voice data; public image Potential Impact: Ongoing uncompensated use; loss of control; reputational risk Mitigation Status: Attempted renegotiation failed