

Article

When AI Speaks in a Person's Voice: Contextual Identifiability and Technical Control in Chinese Law

Yike He^{1,*}

¹ School of Law, Tiangong University, Tianjin, China

* Correspondence: Yike He, School of Law, Tiangong University, Tianjin, China

Abstract: Voice cloning technology converts a situated act of speaking into a reusable capacity to make a person appear to speak in settings that person did not choose, raising profound questions about identity, autonomy, and legal protection in the age of generative artificial intelligence. Existing rules centred on copied recordings capture only part of that wrong, leaving significant gaps in how the law addresses synthetic speech. Drawing on Article 1023(2) of the Chinese Civil Code and the Beijing Internet Court's 2024 Yin Mouzhen judgment, this article develops a sequential account of synthetic voice as relational digital identity. The protected interest is identified as personality autonomy: a bounded authority to determine whether, by whom, and in which contexts an identity-bearing voice may represent the person, rather than ownership of accent, timbre, or style. Contextual identifiability supplies the threshold test by combining the protected identity connection, the relevant audience, unprompted recognition, acoustic evidence, and contextual attribution, while remaining analytically distinct from authorisation, justification, fault, causation, and remedy. Responsibility should then follow technical control, understood as the auditable capacity to prevent, trace, constrain, or terminate a defined risk at a particular system node. Chinese rules on personal information, deep synthesis, and AI-content labelling demonstrate how provenance, scoped consent, model documentation, access controls, logs, labels, and complaint mechanisms can translate legal duties into technical architecture. The resulting framework offers a portable but limited technolaw method that restores control where defendants can practically act while preserving commentary, parody, art, accessibility, and innovation.

Keywords: voice cloning; personality rights; artificial intelligence; chinese law; digital identity; technical control

1. Introduction

Synthetic speech transforms the focus of legal attention. A conventional voice-related matter typically arises from a discrete audio recording that has been copied, edited, broadcast, or used in advertising. Voice cloning technology, however, enables the extraction of identity-bearing vocal characteristics from limited source material [1]. A brief sample may be used to construct a system capable of generating speech the individual never uttered—and sustaining that output across new interactions, digital platforms, and social environments. The subject of legal consideration thus extends beyond a static audio file to encompass a technical capability for projecting a person's vocal identity into contexts not chosen or authorized by that individual.

This shift reveals an ethical concern that goes beyond issues of misrepresentation or economic substitution [2]. Even an output that is factually accurate, technically proficient, and clearly labeled as synthetic may still cause a natural person to appear as the speaker. The core concern lies in the transformation of a finite, context-specific communicative act into an ongoing, externally managed representational function. It affects confidence in speaker attribution, but more fundamentally, it raises questions about authority over how one's vocal identity is digitally represented. Consent must therefore serve as an ongoing mechanism for defining boundaries—specifying purpose, technological scope, output

Received: 11 August 2026

Revised: 04 October 2026

Accepted: 16 October 2026

Published: 21 October 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

formats, authorized users, time limits, and conditions for further use—rather than functioning as a singular, unconditional permission tied to a recording.

The central legal challenge is not merely whether an audio file is inaccurate. Rather, it concerns whether a technical system has converted vocal features linked to a specific individual into a persistent capacity for attribution. Research on synthetic media highlights risks such as misrepresentation, misuse, and diminished public trust, while also cautioning against overly broad restrictions that could hinder legitimate applications [3]. A framework grounded in personality interests must therefore be precisely calibrated: it should identify the protected link between vocal characteristics and personal identity, maintain distinct pathways for lawful expression, and assign responsibilities to those actors who possess meaningful control over relevant technical risks.

Legal provisions in certain jurisdictions provide a valuable reference point for analyzing this issue. Judicial interpretation has affirmed that vocal characteristics may be afforded protection analogous to visual likeness when members of the public—or individuals within a relevant professional domain—can reliably associate timbre, intonation, and articulation patterns with a particular person [2]. A publicly documented judicial decision further clarified that rights associated with a sound recording—or permissions granted for its use—do not inherently extend to authorizing AI-driven processing of that voice for text-to-speech synthesis.

That decision offers a foundational reference, but does not resolve all dimensions of the issue. First, the threshold for protection must distinguish a verifiable connection to identity from general auditory similarity [1]. Second, identifiability alone does not determine legal responsibility: authorization status, permissible usage, expressive context, fault, and causal linkage remain distinct analytical considerations. Third, distributed development and deployment complicate accountability, as different entities—including infrastructure providers, model developers, application interfaces, platform operators, and end users—exercise control over distinct stages where risks may arise. Finally, appropriate remedies must address both economic displacement and the loss of practical oversight over enduring generative capabilities, without presuming that every deployed model can be fully removed or its learned patterns entirely reversed.

The analysis follows a coherent progression [4]. Vocal characteristics can constitute a form of digital identity; the principle of personal integrity supports bounded protection for that identity link; contextual identifiability establishes the threshold for protection; technical control determines how duties and evidentiary responsibilities are allocated across the system; integrated technolaw measures embed those duties through provenance tracking, consent documentation, system transparency, access governance, activity logging, standardized labeling, and mechanisms for redress; and remedies aim to restore functional control at points where responsible parties retain actionable influence. This structure deliberately avoids three oversimplifications: voice data are not uniformly classified as sensitive biometric information; recognizability does not, by itself, establish legal violation; and noncompliance with regulatory requirements does not automatically define the full scope of civil accountability.

The approach is doctrinal and normative, drawing primarily on domestic legislative frameworks and judicial reasoning. Comparative references to other legal traditions serve only to highlight potential pitfalls—such as reducing identity solely to commercial value or separating personality interests from their practical implications in digital systems. The analysis begins with voice as relational digital identity and persistent generative capacity, proceeds through judicial interpretation and the interface between personality interests and information governance, and then examines contextual identifiability, technical accountability, remedial design, and the framework's broader applicability and inherent limitations. Such comparison provides a shared conceptual vocabulary while respecting jurisdiction-specific legal foundations, available defenses, and available remedies [5].

2. Voice, Personhood and Persistent Generative Capacity

2.1. Voice as Relational Digital Identity

Voice is protected under Chinese legal frameworks because it serves as an identifier of natural persons. Characteristics such as timbre, cadence, intonation, accent, and habitual pronunciation—individually or collectively—can associate spoken output with a specific individual [6]. The protected interest lies not in sound as an abstract physical phenomenon, nor in isolated vocal traits, but in the auditory representation that fulfils an identifying function. Academic discussion situates voice protection within the broader category of personality markers, while perspectives differ on whether its economic dimension should be treated as a separate interest.

Article 1023(2) establishes protection through functional reference rather than by establishing an independent 'voice right'. Its structural placement is significant [2]. Provisions governing portraits prohibit distortion, defacement, or forgery using information technology and generally require consent for creating, using, or publishing a portrait, subject to exceptions for reasonable uses. Applying these provisions by analogy acknowledges a functional parallel: portraits enable visual identification, whereas voices enable auditory identification. This approach does not eliminate distinctions between sensory modalities. Visual recognition may occur nearly instantaneously, while voice recognition often depends on duration, linguistic content, listener familiarity, and contextual cues. Consequently, rules adapted from portrait regulation must be carefully calibrated—not mechanically transposed—to suit the auditory domain.

This framework integrates personal and economic considerations within a unified structure. A person may object when synthetic speech creates misleading associations, undermines personal dignity, or places an identity marker in contexts inconsistent with their autonomy. The same use may also involve the market value associated with a professional or widely recognized voice [5]. Some analyses emphasize the personality-based nature of voice interests, even where commercial applications are present; others highlight the increasing economic significance of voice data and the importance of acknowledging both dimensions. A balanced interpretation avoids rigid dichotomies. Economic value arises precisely because the voice refers to a person; conversely, personality protection must not render every phonetic similarity actionable. The requirement of a demonstrable identity connection thus serves both as the rationale for protection and as a necessary limitation on its scope.

Digital identity is not inherently a singular, transferable digital replica of a person [5]. Rather, it frequently consists of context-dependent identifiers whose meaning is shaped by the technical system and audience in which they operate. Voice aligns with this plural and relational understanding. The same vocal output may serve as a reliable identifier for a performer within a dubbing industry context, yet remain unidentifiable to the general public, and acquire stronger associative force when paired with a name, image, or product description. Describing voice as a digital identity therefore captures a functional role—not the existence of a discrete, owner-controlled digital asset.

Comparative approaches to publicity rights illustrate the risks of reifying functional identification into proprietary control [7]. Analyses have shown how personality-based claims can evolve into broadly assignable rights that constrain both individual autonomy and public expression; similarly, image-based claims may overshadow the expressive content in which the identity appears. Case law confirms that deliberate commercial imitation of a distinctive voice cannot be excused solely on the grounds that no original recording was reproduced. The appropriate focus remains on the person-linked auditory representation—its scope defined by the necessity of identification and the legitimacy of the interest asserted, rather than by exclusive control over acoustic features.

2.2. Persistent Generative Capacity and Personality Autonomy

With voice cloning, the legally salient act changes. In a recording case, the evidential path is comparatively direct: identify the protected performance or recording, establish copying or unauthorised use, and assess the defendant's conduct. In a synthesis case, the claimant's audio may serve as training input, while the output consists of newly generated speech. The output can differ waveform by waveform from every source sample yet retain

a recognisable vocal identity [8]. Treating physical identity tied to a recording as a threshold would allow a technical transformation to undermine the personality interest protected under relevant legal provisions.

The transformation is not merely qualitative; it is enduring. A copied recording is a finite object, even if it can be duplicated [9]. A trained voice capability, by contrast, can generate an indefinite range of content. The relevant use may begin with sample acquisition, continue through feature extraction and model training, and culminate in repeated calls to a model. Chinese commentary has therefore described natural-person voice as both an identity interest and a data resource in AI systems. The legal inquiry must remain disciplined. 'Data use' does not automatically constitute infringement of voice interests. A claimant must still establish that the process or output engages the protected identity connection, that the use lacks a legal basis or exceeds consent, and that the defendant satisfies the applicable conditions of liability.

Consent in this setting depends on purpose as much as form. An actor may authorise a studio to record a performance, a publisher to distribute an audiobook, or a producer to edit a particular work. None of those permissions necessarily authorises extraction of acoustic features for a general-purpose model. The difference is not merely quantitative. Model training can transform a bounded performance licence into an open-ended ability to generate new speech, including speech outside the original professional and reputational context. A sound authorisation analysis should therefore examine the object of consent, the contemplated technology, the permitted outputs, duration, users, markets, and onward transfer. Generic language about 'use of recordings' should not implicitly extend to a fundamentally different use of personality [10].

The ethical distinction between the two transactions is qualitative [11]. Permission to capture a situated performance does not by itself authorise a system to place the speaker in new roles, markets, or relationships. The original recording remains bounded by its text and production context; a modelled voice can become an enduring capability whose future utterances, audiences, and commercial associations are unknown at the time of sample collection. This is why the relevant interest concerns the conditions under which another party may acquire and exercise the capacity to represent the person, rather than a blanket restriction on all acoustically similar sounds.

Ongoing consent is not equivalent to perpetual revocability in every transaction. It is a governance principle requiring the permitted purpose, synthesis technique, model or product, output categories, duration, market, recipients, and onward transfer to remain inspectable and enforceable. Withdrawal may lead to different consequences depending on contractual terms, personal information regulations, and technical feasibility, but a system must preserve the distinction between a narrowly licensed recording and an indefinitely reusable identity model. A consent record must therefore carry provenance limitations rather than function as a detached yes-or-no token [12].

This account treats consent as contextual authority rather than a comprehensive theory of privacy. A contextual-integrity framework explains why information transfer can be inappropriate even when the same information was legitimately available in another setting. For synthetic voice, the critical transition is from a bounded performance relationship to a capability that alters speakers, texts, audiences, and purposes. The inquiry remains grounded in positive law, but the contextual insight helps clarify why a recording licence should not implicitly resolve questions concerning model training or deployment.

3. The Yin Case and the Personality–Information Interface

3.1. Facts, Holding and Doctrinal Limits

The official source is the People's Courts Case Database entry *Yin Mouzhen v Beijing Intelligent Technology Company and Others*, entry no 2025-07-2-474-001. The Beijing Internet Court decided the first-instance case, (2023) Jing 0491 Min Chu No 12142, on 23 April 2024; neither side appealed and the judgment became effective. The entry identifies

Yin as a voice actor who discovered works using a voice associated with her on several applications. Tracing the output led to a text-to-speech product offered through a platform operated by an intelligent-technology company [13].

In 2019 Yin had recorded a work commissioned by a cultural-media company, which held rights in the sound recording. That company later supplied the recording data to a software company under a data-use agreement and produced a data authorisation that Yin had not signed. The software company used only Yin's recording as material to develop the disputed text-to-speech product. The product was sold through a cloud platform and was subsequently procured and called through an application programming interface by downstream companies. The official entry records the platform-displayed play count but also shows that the product was removed during proceedings; because the displayed figure was contested in the underlying materials, this article does not use it as a judicially established measure of actual listening [14].

At an in-court inspection, output generated by the product was highly consistent with Yin's timbre, intonation and pronunciation style. The court reasoned that AI-processed voice falls within article 1023(2) when the general public or a public in the relevant field can connect those features to the natural person [15]. It further held that rights in the recording did not include authority to permit AI use of Yin's voice. Because the purported data authorisation was not signed by Yin, the cultural-media company and software company lacked a lawful basis for the AI use.

The remedial result was mixed and actor-specific. The cultural-media company and software company were jointly and severally ordered to pay RMB 250,000 in economic loss. The intelligent-technology company and software company were ordered to provide Yin with a written apology [11]. The remaining claims were dismissed. The three downstream defendants were found not to be at fault for damages, but it is therefore inaccurate to say that all three simply escaped every form of responsibility. The court also held that Yin had not demonstrated serious harm to her mental interests and rejected the claim for mental-distress damages.

The decision establishes three propositions that can safely bear doctrinal weight. AI transformation does not remove voice interests when the resulting output remains identifiable. A general or relevant-field public can supply the recognition audience. And a right to use a recording does not automatically encompass the natural person's consent to AI processing and a TTS product. These propositions move the inquiry beyond literal recording identity while keeping it tied to a specific person.

The entry does not establish that any recognisable output is necessarily unlawful. Its concise case summary links identifiability with lack of authority on the facts before the court; a general framework must still preserve reasonable uses, expressive justification, fault, causation and actor-specific remedies. Nor does the decision adopt personality control as an independent basis for monetary damages. The court's rejection of mental-distress compensation is part of the holding [1]. The argument below—that serious loss of practical personality control may justify disabling, corrective or, where statutory requirements are met, non-economic relief—is a normative extension rather than a description of the Yin result.

3.2. Personality Rights and Personal Information

Personality rights and personal information protection address overlapping factual scenarios through distinct legal frameworks. A claim grounded in personality rights examines whether an output or use appropriates a voice that carries identity-signifying features tied to the individual within their social environment. A claim grounded in personal information protection assesses whether recorded or derived data can identify a natural person—either independently or when combined with other data—and whether its processing rests on a lawful basis. A voice may give rise to one type of claim without satisfying the criteria for the other at a given stage. For instance, a synthetic output that is socially identifiable may raise concerns about individual integrity even if no conventional voiceprint is retained; conversely, a stored speaker embedding may qualify as identifying

personal information from a technical standpoint, even if no publicly recognizable output has yet been generated [9].

Four distinct layers must therefore be analytically separated. Ordinary vocal expression refers to the transient act of speaking [13]. A recording captures speech and may constitute personal information when linked to an identifiable individual. A voiceprint or speaker embedding isolates distinctive acoustic features for identification purposes and may fall within the category of biometric identifiers. A generative model—or a fine-tuned component thereof—embodies a functional capability whose relationship to any specific training sample depends on architectural design and operational context. Conflating these four layers would mischaracterize everyday speech as inherently sensitive while failing to capture the persistent regulatory and ethical implications associated with reusable modeling capabilities.

Regulatory provisions governing biometric identifiers impose heightened requirements for purpose specification, demonstrable necessity, and robust safeguards, along with explicit consent and clear communication regarding potential impacts on individual rights. These standards apply directly where systems extract or deploy voiceprints or comparable biometric features. However, they do not support the broad assertion that every audio recording of human speech automatically qualifies as sensitive biometric information [10]. Classification must be determined by the nature of the information actually processed, its capacity to enable identification, and the purpose for which it is used.

Remedial and evidentiary frameworks also differ across these domains. Claims based on personality rights typically rely on foundational civil law principles, including considerations of fault, causal connection, and the threshold severity required for non-material harm compensation. In contrast, statutory provisions assign a presumption of responsibility to entities handling personal information when harm arises from improper processing; such entities bear the burden of demonstrating absence of fault, and compensation is calibrated to actual loss or benefit gained, with judicial discretion applied where quantification proves challenging. The same conduct may satisfy conditions for both types of claim, but adjudicative analysis should explicitly distinguish the protected interest, evidentiary standard, and applicable remedy rather than applying procedural assumptions from one domain to the other without justification.

Regulatory obligations complement private-law analysis without supplanting it. Certain administrative measures require lawful sourcing of training data and mandate consent or another valid legal basis when personal information is involved. Other provisions assign service providers responsibilities akin to those of content originators and personal information handlers, safeguard user-submitted inputs and interaction logs, and establish formal channels for raising concerns. Additional guidance specifies that separate notice and consent are required when services modify biometric identifiers such as voice. While such requirements help define reasonable technical and organizational precautions and may serve as evidence of system governance, compliance or noncompliance with them does not, by itself, determine all elements of a civil liability assessment [14].

These two frameworks intersect at three key operational stages. First, initial collection may engage both the speaker's interest in personal integrity and the rules governing processing of identifiable recordings. Second, feature extraction may yield biometric identifiers used for speaker recognition, thereby activating stricter controls applicable to sensitive categories. Third, deployment may result in outputs that become socially identifiable—even if underlying representations remain internal and undisclosed [12]. Courts should therefore trace each operation to its specific object and responsible party, rather than treating 'the voice' as a monolithic entity passing unchanged through the technical pipeline.

The frameworks may also operate independently [2]. A general-purpose model trained on lawfully sourced speech may generate outputs that acquire identifiability only through downstream contextual cues—such as association with a name, image, or

targeted communication—thereby implicating integrity interests without involving claimant-specific biometric processing. Conversely, an internal database of voice features may infringe personal information protections without ever producing audible public output. Maintaining this analytical distinction ensures that protection is not contingent solely on public recognizability in information-related contexts, nor is technical identifiability alone treated as determinative for integrity-based claims.

4. Contextual Identifiability as a Bounded Legal Threshold

Defining the Protected Identity Connection: Once literal copying ceases to be the threshold, the boundary problem arises. Voices share common linguistic and acoustic features. Actors imitate styles; comedians parody public figures; performers adopt genre conventions; assistive systems may offer generic synthetic voices. If a claimant could establish control merely by showing resemblance, personality law would create a broad monopoly over sound. Identifiability must therefore be distinguished from distinctiveness and from unlawfulness. A voice may be distinctive but not recognised as a particular person by the relevant audience. It may be identifiable yet used lawfully for news, criticism, art or another justified purpose. Conversely, a voice that is not widely known may be highly identifiable within the specific market where the use occurs.

The more defensible move is from recording identity to contextual identity, not from copying to unbounded association. Scholarship on portrait recognisability is useful by analogy because it treats identification as a social judgment informed by features and context rather than a demand for exact duplication. Recent research in China on AI-generated voice similarly proposes a combined inquiry into listener recognition, technical evidence and use context. That approach becomes more compelling when applied to a professional voice actor [10]. The audience relevant to an audiobook service or dubbing market may recognise a professional vocal identity even if the general population does not. The legal task is to define that audience without reducing it to either 'everyone' or a handful of intimates.

The first step is therefore to identify the legally protected connection. The voice in question must materially contribute to attribution to a particular natural person; generic timbre, accent, emotion, genre convention and vocal style remain outside the scope of protection by themselves. The inquiry should ask what feature or combination makes the output function as that person's auditory representation, whether the link persists across more than a prompted comparison, and whether the claimant's argument depends on the voice itself or only on unrelated contextual information [10, 14]. This definition protects transformation-based identity appropriation without creating property rights in the fundamental elements of speech.

4.1. Selecting the Relevant Audience

A contextual identifiability inquiry has three connected stages, beginning with recognition by the relevant audience [2]. The court should define the audience by reference to the defendant's actual or reasonably foreseeable use, then ask whether an appreciable part of that group would connect the output with the claimant under ordinary reception conditions. For a public figure, the group may be broad. For a dubbing actor, it may consist of listeners, clients and professionals in the relevant field. For a private person, the audience may be narrower where the deployment itself targets a workplace, family or local community. The law need not require majority recognition, but it should require more than an idiosyncratic reaction from a primed witness.

Recognition evidence should be designed to reduce suggestion [10]. Relevant questions include whether listeners identified the claimant without being given a name; whether the group was familiar with the claimant's voice before the matter arose; whether recognition persisted across different generated sentences; and whether contemporaneous comments or customer communications associated the voice with the claimant. Courts should distinguish evidence that the voice "sounds similar" from evidence that it identifies a person. Analyses of portrait recognisability support this

socially grounded approach, while foundational work on voice-specific personality rights supplies its theoretical basis.

The general public or relevant-field public framework should be applied functionally. Courts should identify the audience exposed to the challenged use or the market in which substitution or attribution occurs. Close relatives or colleagues may supply corroboration, especially in a targeted private deployment, but recognition after being told the claimant's name will ordinarily be too suggestive to establish the threshold by itself. The audience inquiry prevents celebrity from becoming a condition of protection while also preventing every intimate recognition from externalising a broad right against the public.

4.2. Triangulating Recognition, Acoustic and Contextual Evidence

Objective acoustic evidence provides a second layer of analysis. Voiceprint analysis, spectrographic comparison, and expert evaluation can assess whether generated output preserves stable features associated with the source. These methods are particularly valuable when listener-based evidence is contested or when the output has undergone modification. Their function is corroborative. Experts must clearly articulate the comparison material used, recording conditions, underlying model assumptions, reported error rates, and inherent limitations of any similarity metric. A court should not treat a vendor's proprietary threshold as a binding legal standard. Where competing technical reports yield divergent conclusions, the court must evaluate methodological rigor and evidentiary context rather than numerically averaging results.

Contextual attribution completes the inquiry. Names, images, job titles, signature phrases, visual presentation, product labels, and the functional purpose of the output may either reinforce or undermine an identity connection. Commercial intent is significant: a service explicitly marketed as reproducing a known individual—or deployed within the professional domain where that individual operates—employs resemblance in a distinct manner compared to a neutral illustration of speech synthesis capability. Context must not artificially construct an identification interest where the audio itself contributes no identifying information; however, the law must also recognize the cumulative effect of integrated representations. Chinese courts should require that the voice, evaluated alongside lawful contextual evidence, makes a material and discernible contribution to identification.

Evidence should be triangulated rather than subjected to rigid hierarchical ranking. Unprompted recognition by a properly defined audience tests social attribution. Acoustic or voiceprint analysis can verify whether stable features support that attribution, provided experts fully disclose the comparison material, recording conditions, methodology, and analytical limitations. Names, images, roles, product descriptions, and marketing materials reveal how the output was positioned and framed. No single category is invariably necessary or sufficient. The court must articulate how the totality of evidence substantiates a stable identity connection, while discounting cognitive priming effects, low-fidelity recordings, selectively curated examples, and contextual framing that obscures the minimal identifying value contributed by the voice itself.

4.3. Separating Identifiability from Infringement

Taken together, these stages establish only identifiability. The claimant must still demonstrate an act that engages the protected voice interest and the absence of consent or other lawful justification. Courts should assess whether the defendant generated, used, disclosed, sold, or enabled use of a voice linked to the claimant, and whether such conduct fell within any applicable authorization. They should then weigh reasonable uses and competing interests, including news reporting, criticism, scholarship, parody, artistic creation, and accessibility. Commercial purpose is relevant but not determinative: a paid documentary may be justified, whereas a non-profit deceptive impersonation may not be [9].

Cases from other jurisdictions illustrate similar analytical considerations under differing doctrinal frameworks. One involved deliberate commercial imitation of a distinctive singer following her refusal of a commercial request, making appropriation

particularly evident. Another relied heavily on contextual evocation to establish identity linkage. Neither case should be treated as establishing a rule that any associative connection is inherently actionable. Their significance lies in confirming that literal replication is not required to raise voice-related concerns, while also underscoring the need for doctrinal boundaries—particularly to prevent overextension of control into stylistic elements, symbolic associations, or broader cultural references. For Chinese law, the analogy to portrait rights should preserve its inherent limiting structure, including enumerated reasonable uses, rather than functioning as an unqualified expansion of control.

Such boundaries are especially important where expressive activity is involved. Parody, criticism, historical representation, and artistic performance may intentionally invoke a real person's voice precisely because the identity connection conveys meaning. A framework treating mere recognisability as conclusive infringement would undermine that expressive function. Chinese law's reference to portrait protections should incorporate their established structure of reasonable-use limitations and interest balancing. Courts should evaluate purpose, degree of transformation, necessity of the identity linkage, likely audience understanding, potential for commercial substitution, and safeguards against deception. Disclosing that a voice is synthetic may mitigate confusion but does not, by itself, validate unauthorised commercial appropriation; conversely, the absence of such disclosure does not, standing alone, establish a personality-right violation.

The full analytical sequence is therefore as follows: identification of the protected object, specification of the relevant audience, and triangulated evidence collectively establish contextual identifiability; a second tier of assessment then examines whether the defendant used or enabled use of the identity connection, whether authority existed and extended to the specific technology and outputs employed, whether a statutory or expressive justification applies, and whether the defendant satisfies fault, causation, and remedial requirements. This conceptual separation is essential to safeguard criticism, parody, artistic performance, historical representation, scholarship, news reporting, accessibility applications, and generic speech systems. Disclosure that a voice is synthetic may reduce confusion and support a justification, but it neither automatically validates unauthorised commercial identity appropriation nor, in its absence, suffices to establish a personality-right tort.

5. Technical Control as Human-Centred Technolaw

5.1. Risk-Bearing Nodes and Differentiated Duties

Responsibility becomes harder to place once production is fragmented. Legal frameworks generally require fault-based attribution, while provisions addressing joint involvement apply where multiple parties contribute to an outcome. A conventional account links conduct, fault, causation and damage. Voice synthesis distributes each element across the value chain. The supplier controls access to recordings and the scope of any licence. The developer controls feature extraction and model training. An interface provider may control authentication, logging and use restrictions [11]. A platform controls listing and removal. A commercial user chooses the content and deployment. None necessarily controls the entire process.

The chain should be reconstructed around operations rather than corporate labels. Source acquisition determines whether recordings enter the process and under what conditions [9]. Feature extraction and training determine the conversion into a reusable capability. Model or API access governs who can generate outputs, at what scale and with what audit trail. Product presentation and distribution govern contextual attribution and continued availability. A downstream deployment governs text selection, audience reach and implied association. The same enterprise may occupy several nodes, while several enterprises may share one. Technical control is legally relevant only when tied to a defined risk, time and practicable measure.

Technical control offers an organising principle, though it does not displace the ordinary elements of responsibility [4]. It means the practical capacity to govern the feature or process through which the infringement risk arises. Relevant forms include the capacity to verify authority over recordings, decide whether features are extracted, select training data, determine model architecture and fine-tuning, label or restrict a voice, authenticate users, preserve logs, restrict an interface, suspend deployment and remove generated products. Control is specific to an act and a time. A company may control model training but not a later user's choice of text; a platform may lack control at training yet acquire control over continued distribution after notice.

The model developer controls the conversion from recordings into a reusable capability. Its duty should include risk-based provenance review, separation of copyright or recording permissions from personality-related consent, records linking models to data sources, testing for output associated with specific individuals, access restrictions and a process for restricting a voice under question. The developer need not guarantee the legality of every input in all circumstances [3]. Greater scrutiny is justified, however, where a dataset is built around a named individual, contains professional recordings or is intended to reproduce a particular identity. The distinction between generic speech research and a marketable voice representation is material to foreseeable risk.

Interface and platform operators require functional analysis. An operator that selects named voices, describes them as substitutes for professionals, prices access and can restrict a model exercises substantial control even if it did not train the model [6]. A neutral intermediary with no reasonable means to detect output associated with specific individuals may owe a more limited duty before notice. After a credible complaint, its control over logging, preservation, suspension and removal becomes salient. Regulatory provisions support duties relating to consent, assessment, labelling and logs in covered services, but civil liability should still depend on the actor's function and fault.

Looking at the chain node by node makes proportionate liability possible. Relevant factors are: the indispensability of the actor's contribution; its knowledge or reason to know; its ability to prevent or stop the harm; the benefit obtained; the duration of control; compliance with documentation, logging and notice procedures; and whether another actor's action was extraordinary or foreseeable. Where actors knowingly cooperate, jointly commercialise the same representation or produce an indivisible injury, joint liability may be appropriate. Where independent acts make separable contributions, apportioned liability better reflects risk contribution. Equal division should not be the default.

Technical control must not become enterprise liability by another name. Financial size, technical sophistication or proximity to AI does not by itself prove fault. Nor should the concept impose impossible monitoring of all generated speech [10]. Its value is evidential and normative: it asks which precaution was available to which actor at the relevant node. A duty is strongest where a party both controls the risk and benefits from its creation or continuation. It weakens where detection is infeasible, the use is remote and the actor responds reasonably when control becomes possible.

5.2. Provenance, Scoped Consent, Logging, Labelling and Contestability

Design matters because digital systems structure what users and intermediaries can perceive and do. Legal protection can and should extend to the design of technologies rather than treating architecture as neutral; standards, networks, and institutional design are integral to the legal construction of informational power. Technical control translates this insight into a narrower liability inquiry, focusing on which actor could take which precaution against which voice-related risk at a specified time.

Provenance should record the origin of a voice sample, the identity or authority of the transferor, permitted processing purposes, restrictions on synthesis and onward transfer, and the connection between those records and the dataset or model component actually used [14]. It is not sufficient to retain an assertion that data are 'licensed'. Rights in a recording and authority over an identity-bearing voice must be separately described. Inconsistencies, unverified permissions, or unsigned authorisations should trigger review

proportionate to claimant-specific risk, especially where a dataset is built around a named performer or a product markets a recognisable substitute.

Consent records should be machine-actionable only to the extent that their legal meaning remains reviewable. A scoped record can express permitted model, purpose, output class, market, duration, recipients, sublicensing, and withdrawal or termination conditions. It should travel with downstream restrictions and retain a human-readable source. Treating consent as an unqualified binary flag would undermine the nuanced approach required by regulatory guidance. The Deep Synthesis Provisions' article 14 requirement for notice and separate consent in voice-biometric editing supplies a regulatory foundation for this design, while the Personal Information Protection Law determines when sensitive-information rules apply [12].

Logging provides temporal evidence of access, generation, model version, user identity where lawfully collected, complaints, suspension, and remedial action. It can distinguish initial creation from subsequent distribution after notice and make causation more granular [6]. Logging must itself comply with data-minimisation and security obligations; the Interim Measures prohibit unnecessary collection and unlawful retention of identifiable user inputs and use records. The design task is therefore purpose-bound traceability, not indefinite surveillance.

Labelling is the clearest current Chinese example of law entering technical architecture. The 2022 Deep Synthesis Provisions require technical marks and logs, conspicuous labels for synthetic or imitative voice likely to cause confusion, and protection against removal or concealment. The 2025 Measures for Labelling AI-Generated Synthetic Content and mandatory national standard GB 45438-2025, effective from 1 September 2025, distinguish perceptible explicit labels from implicit labels in file metadata and add verification responsibilities for content-distribution services. These mechanisms can reduce deception, support provenance, and reveal platform control. They cannot authorise the underlying appropriation of a person's vocal identity.

Contestability closes the loop. A claimant needs an accessible channel to identify the output, preserve relevant evidence, obtain a reasoned response, and secure proportionate interim action. Credible notice can activate control that did not exist at training: a platform may preserve logs, suspend a named voice, restrict access, or remove a product while authority is checked. The Interim Measures and Deep Synthesis Provisions already require complaint or reporting mechanisms. Civil doctrine should treat the design and operation of those mechanisms as evidence of both practicable precaution and post-notice fault.

The architecture must also remain contestable. Recent scholarship reframes human oversight as a site of contestation rather than a purely supervisory label. In this article's narrower setting, contestability requires a channel through which the affected person can identify an output, preserve evidence, obtain a reasoned response, and trigger proportionate interim measures. It is not a promise that a human will review every generation. It is an institutional capability to make control visible and answerable when a claimant-specific risk emerges.

These artefacts also discipline remedies. Provenance disclosure can identify the node at which authority failed. Access controls and suspension can stop future calls. Segregation can isolate a claimant-specific component. Machine unlearning may sometimes remove the influence of specified training data, but techniques can be exact or approximate and their feasibility and verification depend on the model. Courts should therefore require evidence before ordering or accepting unlearning, and should prefer the least over- or under-inclusive measure that restores control. NIST's Generative AI Profile likewise supports lifecycle risk management and content-provenance practices as governance tools, not as substitutes for legal judgment.

6. Control-Restoring Remedies

For economic harm, courts should begin with the measure best supported by the record: proven loss, attributable profit, or a reasonable licensing fee. For professional voice

use, a hypothetical licence is often the most workable baseline when the other measures cannot be isolated. The question is not what the claimant charged for an ordinary recording session. It is what reasonable parties would have agreed for the actual rights taken: source-data use, feature extraction, training, number and type of outputs, scope of application, duration, exclusivity, sublicensing, revocability and association with particular products.

The Yin result anchors the positive-law boundary. It awarded RMB 250,000 in economic loss against the cultural-media and software companies, ordered a written apology from the intelligent-technology and software companies, and rejected the claim for mental-distress damages because serious harm to mental interests was not proved. The case therefore supports neither an automatic personality-based award nor the proposition that a licence measure fully resolves the identity wrong. It shows instead that monetary and non-monetary remedies can attach to different defendants and interests [12].

A reasonable licence remains a compensatory counterfactual, not compelled consent. Its scope must reflect the rights actually taken: training, generation, outputs, duration, exclusivity, sublicensing and association. Where future generation remains possible, restoration may require suspension, access restriction, segregation, traceability, correction or technically feasible deletion or unlearning. A court should identify the defendant's control node and require evidence about efficacy, cost, collateral effects and verification rather than assume that complete model deletion is always possible.

The law should also recognise that personality harms do not always arrive as measurable market loss. Legal systems can under-protect privacy when harm is reduced to tangible financial or physical injury. Synthetic voice sharpens that problem because a claimant may lose practical authority over future representation before reputational or economic consequences can be proved [11]. Chinese positive law still controls the availability of damages, including the seriousness threshold for mental-distress compensation. The analytical response is therefore to distinguish the wrong from the remedy and to use non-monetary control-restoring orders where their legal conditions are met.

By contrast, non-economic relief should focus on a serious loss of personality control. Relevant factors include whether the use places the claimant in speech or roles they would reasonably refuse; creates confusion about authorship or endorsement; reaches a substantial or sensitive audience; persists through an open-ended model; resists effective withdrawal; involves personal or intimate contexts; or continues after objection. Distress is evidence, but the wrong should not depend exclusively on a claimant's ability to prove psychiatric or reputational injury. A professional may experience a serious loss of control even where the generated content is competent and non-defamatory [3].

Remedial design must also guard against double recovery. A licence measure compensates the economic value of the unauthorised use; non-economic damages address serious impairment of dignity, identity or personal control. The same fact, such as prolonged deployment, may be relevant to both but for different reasons. Courts should explain the distinction. They should also consider non-monetary orders first where those orders directly reduce ongoing loss. This is particularly important for model-based harm, because an award that leaves the voice capability active does little to restore control.

7. International Significance and Limits

The framework's international significance lies in its logical sequence, not in proposing a jurisdiction-specific legal remedy. Across diverse legal systems, synthetic media cases commonly engage concerns related to misrepresentation, personal privacy, data handling, publicity rights, and intellectual property. These areas of law safeguard distinct interests and entail differing evidentiary and procedural requirements. Relational digital identity provides a unified conceptual description of the underlying phenomenon, while contextual identifiability clarifies when a synthesized voice functions as a reference to a specific individual rather than merely evoking a stylistic resemblance.

Technical control offers a transferable mechanism for aligning legal expectations with distributed system design. It aligns with design-integrated approaches to privacy and oversight, yet remains more circumscribed than broad assertions that technical architecture alone governs behavior. The analysis remains focused on particular actors, concrete risks, and defined temporal contexts. It can support obligations regarding origin transparency, user notification, and practical remediation—without predetermining the applicable legal basis or equating technical capability with automatic responsibility [2].

The comparative analysis also reveals two countervailing concerns. A rights model grounded solely in proprietary control may extend too broadly, potentially restricting parody, commentary, and legitimate cultural reuse. Conversely, a narrow focus on misrepresentation may fall short when a technically accurate and clearly labelled output nonetheless appropriates an individual's voice for unauthorized commercial purposes. The Chinese portrait-by-reference structure is instructive because it integrates identification, authorization, reasonable application, and redress within a coherent sequence—while leaving room for each legal system to define its own expressive protections and standards of accountability.

Three foundational limitations are essential. Voice data do not universally qualify as sensitive biometric information; perceptible similarity to an individual does not, by itself, establish legal violation; and failure to meet regulatory expectations does not automatically trigger civil liability. The Yin decision establishes a measured doctrinal foundation, whereas concepts such as personal integrity, technical control, and remedies aimed at restoring control represent normative developments beyond that baseline. These boundaries ensure the approach remains adaptable internationally without presenting domestic legal constructs as prescriptive models.

This delimited comparative perspective further defines the article's intended readership. It does not propose a comprehensive global classification of synthetic-media law nor assert that platform accountability has been fully resolved in any single setting [12]. Instead, it offers international readers a structured method for maintaining clarity across three interrelated questions: under what conditions does an output refer to a specific person; which legal principles or justifications apply to its use; and which parties hold the relevant technical capacity to implement preventive measures or enable redress. This sequencing retains utility even where the substantive rights differ across settings.

8. Conclusion

AI-generated voice should be analysed through a sequence that begins with relational identity and ends with restoration of control. A voice functions as digital identity when recorded or modelled features materially contribute to representing, attributing speech to, or enabling recognition of a natural person in a relevant setting. The capacity to direct how one's vocal characteristics are used explains why authority over a recording does not automatically extend to creating an open-ended generation capability, without conferring ownership of accent, timbre, or style.

Contextual identifiability supplies only the threshold. Courts should define the protected identity connection, select the audience relevant to the challenged use, and triangulate recognition, acoustic, and contextual evidence. They must then decide authorisation, statutory limitation, expressive justification, fault, causation, and remedy separately.

Technical control allocates duties according to the practical capacity to prevent, trace, constrain, or terminate a defined risk at a particular node. Provenance, scoped consent, model documentation, access controls, logs, labels, and contestability make that capacity auditable. Remedies should operate at the node where the defendant can act and should distinguish attributable economic loss from serious impairment of personal agency over vocal representation.

A recent judicial decision grounds this approach but does not resolve all its elements. Its official entry supports contextual recognition, the distinction between recording rights and AI-driven voice generation authority, actor-specific remedies, and rejection of

unproved serious psychological harm. The broader contribution is a bounded human-centred technolaw method that connects social identification, legal authority, and technical architecture while preserving legitimate expression and innovation.

The resulting account resists both technological formalism and personality absolutism. A changed waveform does not erase identity, but an identity connection does not erase lawful expression. Technical architecture can make duties and remedies practicable, but it cannot supply the normative judgment by itself.

Disclosure and Ethics: The author used AI-assisted tools, including ChatGPT and Codex, for Chinese-to-English translation, English-language editing, non-substantive editorial organisation, citation-formatting and document-preparation assistance. The author independently developed the legal analysis, argument and conclusions; reviewed and verified the authorities and all assisted text; and accepts full responsibility for the article. No AI tool is listed as an author. Institutional Review Board Statement: Ethics approval was not required because the article is a doctrinal and normative legal analysis and does not involve human participants or empirical data collection. Funding: The author received no external funding for this research. Conflict of Interest: The author declares no conflict of interest.

References

1. H. D. Krause, "Right to Privacy in Germany — Pointer's for American Legislation?", *Duke Law Journal*, p. 481, 1965.
2. R. Chesney and D. K. Citron, "21st century-style truth decay: Deep fakes and the challenge for privacy, free expression, and national security", *Maryland Law Review*, vol. 78, p. 882, 2018.
3. R. Calo, "Privacy harm exceptionalism", *Colorado Technology Law Journal*, vol. 12, p. 361, 2014.
4. S. Zuboff, "Surveillance capitalism or democracy? The death match of institutional orders and the politics of knowledge in our information civilization", *Organization Theory*, vol. 3, no. 3, p. 26317877221129290, 2022.
5. J. T. Mark, "Constitutional restraints on state right of publicity laws", *Boston College Law Review*, vol. 44, no. 3, pp. 863–900, 2003.
6. J. S. Donath, "Identity and deception in the virtual community", in *Communities in Cyberspace*, Routledge, 2002, pp. 37–68.
7. J. E. Rothman, "The Right of Publicity: Privacy Reimagined for a Public World (Introduction)", 2018.
8. J. Santuber, M. C. Morais, and K. Tica, "Dis/ordering AI–Human Oversight: Materialities, Aesthetics and Contestability", *Law, Technology and Humans*, 2026.
9. A. Hearn, "Meat, Mask, Burden: Probing the contours of the branded self", *Journal of Consumer Culture*, vol. 8, no. 2, pp. 197–217, 2008.
10. H. Nissenbaum, "Privacy as contextual integrity", *Washington Law Review*, vol. 79, p. 119, 2004.
11. D. Jain, "Algorithmic Subversion of Consumer Sovereignty: A Computational and Ethical Analysis of AI-Induced Impulse Buying in the Indian Digital Economy", 2026.
12. J. Chubb, P. Harrison, J. Tompkinson, and E. Thompson, "How close is too close? Digital replicas and protection of the 'unfixed' voice", 2025.
13. H. H. S. Josan, "AI and deepfake voice cloning: Innovation, copyright and artists' rights", *Artificial Intelligence*, 2024.
14. C. McGettigan, S. Bloch, C. Bowles, T. Dinkar, N. Lavan, J. C. Reus, and V. Rosi, "Voice conversion and cloning: psychological and ethical implications of intentionally synthesising familiar voice identities", *Journal of the British Academy*, vol. 13, no. 3, p. a31, 2025.
15. P. Dutt, "Arijit Singh v. Codible Ventures: Bombay High Court Verdict on AI Voice Cloning and Personality Rights in India", *Legal Spectrum Journal*, vol. 4, p. 1, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.