Institutional Guidance on AI Transcription for Research Interviews

A Survey of 130 Universities in the United States, United Kingdom and Canada

Ankur Shrestha

Pepys, United States

Working paper, version 1.0 • October 2026

Abstract

Researchers routinely upload interview recordings to cloud AI transcription tools, but it is unclear what their institutions tell them to do. We surveyed the public guidance of 130 institutions: 100 US R1 universities drawn at random from 187, 15 of the 24 UK Russell Group members, and all 15 Canadian U15 members. Each institution was coded twice, independently, in a language-model-assisted workflow from saved page text, and disagreements were resolved against the page text. Ninety institutions (69.2%, 95% CI 60.8–76.5) published something that touches AI tools and recordings or transcription, but only 73 (56.2%) gave transcription-specific guidance; 40 (30.8%) had none we could find. Sixty-one institutions name at least one tool as approved or prohibited. Zoom's AI features are approved at 41 institutions and Microsoft Teams at 20, while stand-alone meeting bots are barred: Read.ai at 17, Fireflies.ai at 11 and Otter.ai at 20, yet Otter.ai is also approved at 5. Only 25.6% of institutions with guidance say researchers must disclose AI transcription in the ethics protocol, and 36.7% publish no date. Inter-coder agreement was high on whether guidance exists ($\kappa = 0.89$) and moderate on stance ($\kappa = 0.68$). Institutions give researchers different answers about the same products, and much of the guidance is drawn from IT pages on meetings rather than from ethics offices.

Keywords: AI transcription; research ethics; institutional review boards; data governance; qualitative research; document analysis.

1 Introduction

Interviews, focus groups and oral histories are routinely recorded and transcribed. A growing share of that transcription is done by cloud AI services that are cheap, fast and easy to reach from a browser. Researchers who use them are usually bound by an ethics approval, a consent form and an institutional data policy, and cloud transcription touches all three: the audio leaves the institution, may be retained, and may be used to improve the vendor's models. Vendor terms differ on these points, and some are contested in court. In August 2026 a US federal court let privacy claims proceed against Otter.ai in a consolidated class action, an order that recites allegations about retention of recordings for model training [1]; those are allegations, not findings.

What institutions tell researchers is therefore practical. A researcher who asks a research-ethics office whether they may use a given tool will get an answer from a policy page, a library guide or an IT article, and these sources may not agree. Qualitative-methods scholars have flagged the consent side of the problem [2] and compared tools for accuracy, cost and data protection [3, 4], and a scoping review has collected the comparative literature on AI transcription in qualitative research [5]. Locally run alternatives built on open Whisper models exist [6, 7]. General studies of

university generative-AI guidelines exist too [8], but we found no survey of institutional guidance on AI transcription of research data. Three search formulations on 29 September 2026 returned single-institution pages and general AI-policy studies only. That is a limited search, not proof of absence.

We ask four descriptive questions, set out in the study plan before collection: (RQ1) what share of institutions publish guidance on AI or automated transcription of research data; (RQ2) which tools are named as approved or prohibited, and whether institutions contradict each other about the same tool; (RQ3) what they require, including protocol disclosure, consent language, data-classification limits and security review, and whether local processing is preferred or permitted; (RQ4) how current the guidance is and whether it mentions vendor training, retention or litigation.

2 Method

Design and registration. This is a cross-sectional survey of public documents. The design, codebook, sampling frame and analysis plan were written before collection, and their SHA-256 hashes were recorded when collection began on 30 September 2026. The plan files were deposited on OSF after coding and analysis ([**OSF URL**], registered [**OSF date**]), so the registration documents the plan but does not precede the analysis. Section 2.6 lists every deviation.

Sample. The unit is the institution. The US frame is the 187 institutions listed as R1 in the 2025 Carnegie update, as tabulated on Wikipedia's list of research universities in the United States (retrieved 30 September 2026); we drew 100 with Python's `random.sample` after `random.seed(20260930)`. The UK sample is 15 of the 24 Russell Group members, drawn with the same seed. The Canadian sample is all 15 members of the U15 group. Of the US institutions, 75 are public and 25 private; all UK and Canadian institutions are public. No institution was replaced. Because the frame favors large research universities, the results describe those institutions, not higher education in general.

Collection. For each institution a first coder ("Coder A") searched with a fixed protocol: (a) `site:<domain> transcription IRB AI`, (b) `site:<domain> "Otter" OR "Rev" OR "Zoom" transcription research data`, (c) `site:<domain> transcription research interviews`, and up to seven follow-up queries, then read the guidance pages on the institution's own domain. Pages were fetched over HTTP and saved as text with a SHA-256 hash; every quote in the coding had to appear verbatim in a saved page, and we checked this automatically for all 130 institutions. Guidance behind a login, and pages that returned a bot-detection challenge, were treated as not read and recorded as blocked; we did not attempt to bypass challenges. Collection and re-runs took place on 30 September 2026 (UTC). Coders made 859 logged queries in total.

Variables. The codebook (Appendix A) has 16 variables. The main ones are whether guidance exists ("guidance" means a page that says something about AI or automated transcription, or about recordings or transcription in a way that engages AI tools; a general AI policy counts only if it says something about recordings or transcription), the stance (approves named tools, prohibits named tools, both, case by case, general only, none), the named tools, and requirements for protocol disclosure, consent language, data classification, security review and local processing. A tool is "prohibited" only if the text says not to use it, that it is not approved, or that it is blocked; a caution is not a prohibition. Where a page is silent, consent is coded "no" and protocol disclosure and security review "unclear".

Second coding and adjudication. A second coder ("Coder B") re-coded 129 of the 130 institutions from the saved page text alone, without access to Coder A's coding. The two coders disagreed on at least one variable at 110 institutions. For each of those, an adjudicator read both

codings and the saved text and set the final value from the text. Final stance differs from Coder A's at 25 institutions and from Coder B's at 8. Coding and adjudication were language-model-assisted under a written codebook and author supervision (Section 2.6). The remaining 20 institutions, where the coders agreed on everything, keep the agreed coding.

Analysis. Proportions with Wilson 95% intervals; cross-tabulation by country and by public or private control; a tally of vendors by status; the median age of dated guidance; and Cohen's kappa for each categorical variable. No inferential tests were planned.

Deviations from the plan and other limits on the record.

1. *Timing.* The collection window in the first draft of the plan was written as 1 to 3 October; collection started on 30 September. Seventeen minutes after collection began, the plan and codebook were corrected to state the actual window; no other text changed. The hash log records both times. The sampling frame and institution list match their recorded hashes; the corrected plan and codebook are the versions deposited on OSF, and the pre-correction copies were not kept.
2. *Search tool.* The built-in web-search tool reached its session cap partway through collection. Remaining searches used a second search tool (TinyFish) with the same queries. In the first pass, some coders searched by other means (site sitemaps and link crawling) for part of their shard. Those institutions were re-run under the full protocol; ten institutions in the final dataset still have one or more of queries (a) to (c) logged as not executed (ca-119, ca-123, ca-126, uk-102, uk-103, us-029, us-045, us-055, us-057, us-071). Their "no" or "general only" results are weaker.
3. *Overlapping coders and file overwrites.* Several coding passes ran in parallel on the same institutions and overwrote each other's files in some cases. We kept every superseded item (26 first-round items and 19 sweep items are archived), verified the final files against the saved page text, and re-checked every quote.
4. *Request headers.* Some sites returned 403 to a bare user agent but served the page to a browser-style user agent. We allowed retrying with browser-style headers; no CAPTCHA or bot challenge was solved. Three pages at two institutions (two Columbia central IT pages and UNMC's AI guidelines page) were captured by hand from a browser as excerpts, not as machine-saved text; excluding those two institutions, guidance is found at 89 of 128 (69.5%) and transcription-specific guidance at 72 of 128 (56.3%), so no headline figure changes.
5. *Off-domain pages.* The own-domain rule was relaxed for four institutions whose guidance is hosted elsewhere: Boston College and NC State (Google Docs), UCF (a research-administration vendor domain) and Ohio State (a knowledge-base host). Excluding them would change at most 4 of 130 institutions (3.1 percentage points).
6. *Unreadable institutions.* For eight institutions no usable page text could be read: Baylor College of Medicine, UT Health San Antonio, UMass Amherst and Minnesota (bot-detection challenges or unreachable hosts), UCL and Alberta (challenges), Université Laval (the linked guide was behind a sign-in page) and King's College London (the site rate-limited every request during collection). Their codes mean "not found", not "absent". Excluding them, guidance was found at 73.0% (89 of 122).
7. *Language-model-assisted coding.* Every coding pass (first coding, second coding, adjudication) was carried out with a language-model workflow following the written codebook, under the author's supervision, with automated checks that each quote appears verbatim in the saved page. The author reviewed a random sample of 15 institutions [**spot-check result**]. The agreement statistics measure consistency between the two coding passes, not correctness.

Ethics. The study uses public documents only. There are no human participants and no

Table 1: Final stance of 130 institutions.

Stance	Institutions	Share
Approves named tool(s) only	33	25.4%
Names both approved and prohibited tools	26	20.0%
Case by case (no tool named)	12	9.2%
Prohibits named tool(s) only	2	1.5%
General AI or data guidance only	17	13.1%
No guidance found	40	30.8%

Table 2: Tools named by status across institutions (a tool can be counted as both by different institutions). Counts are institutions, not mentions.

Tool family	Institutions approving	Institutions prohibiting
Zoom (AI Companion and built-in transcription)	41	1
Microsoft Teams (including Teams Premium)	20	0
Microsoft 365 (Word Transcribe, Copilot, Stream)	13	1
Whisper-based or local tools (aTrain, noScribe, Whisper apps)	12	1
Rev	6	3
Google (Gemini, Meet, Speech-to-Text)	5	2
Otter.ai	5	20
Read.ai	0	17
Fireflies.ai	1	11
ChatGPT / OpenAI	2	3

personal data, and we did not contact any institution. Institutions are quoted briefly, with their URLs, in the dataset.

3 Results

3.1 How many institutions publish guidance (RQ1)

Ninety of 130 institutions (69.2%, 95% CI 60.8–76.5) published guidance that meets our definition, and 40 (30.8%) had none we could find (Table 1). The 90 are not all about transcription. Seventeen (13.1%) have only general AI or data guidance that mentions recordings or transcription in passing, with no transcription rule. The other 73 (56.2%, 95% CI 47.6–64.4) give transcription-specific guidance: they name approved or prohibited tools, or tell researchers to describe or seek review of the transcription method.

Coverage was 66.0% in the US (66 of 100), 66.7% in the UK (10 of 15) and 93.3% in Canada (14 of 15). With 15 institutions in each non-US group these intervals are wide (Canada 70.2–98.8). Public institutions had guidance more often than private ones (71.4%, 75 of 105, against 60.0%, 15 of 25), but the intervals overlap widely.

3.2 Which tools, and where institutions disagree (RQ2)

Sixty-one institutions (46.9%) name at least one tool as approved or prohibited: 33 approve only, 26 do both, and 2 only prohibit. After merging naming variants, the pattern is that institution-licensed platforms are approved and stand-alone meeting bots are barred (Table 2).

The same product receives opposite treatment at different institutions. Otter.ai is approved at 5 institutions (for example through a university license or on a paid tier) and prohibited at 20; Rev is approved at 6 and prohibited at 3; and two institutions (us-018, us-046) approve and restrict Otter.ai at once, typically by tier or license. Whether the conflict reflects real differences in contracts or only differences in how carefully each page was written cannot be told from the documents. Prohibitions are usually drafted for meeting bots on institutional accounts and often say nothing about researchers who upload a recording. Two institutions prohibit a tool but approve none.

Local processing appears in 25 of the 90 institutions with guidance (permitted at 21, preferred at 4). Northwestern's IT page is an example: it approves Zoom, Teams, Azure speech-to-text and locally run Whisper-based applications, and names Rev and Otter.ai as not to be used with research data.

3.3 What institutions require (RQ3)

Among the 90 institutions with guidance, 23 (25.6%, 17.7–35.4) require researchers to describe or disclose AI transcription in the ethics protocol; 25 (27.8%) require consent language about it; 60 (66.7%) require or point to a security or vendor review; and 75 (83.3%) restrict tools by data classification. The last two figures are inflated by general data-security pages that mention recordings. For protocol disclosure, the remaining 67 pages do not say either way (coded "unclear"), so this figure is a lower bound.

3.4 Currency, vendor training and law (RQ4)

Thirty-three of the 90 guidance institutions (36.7%) show no date. Among the 57 that do, the median age at 30 September 2026 is 175 days, and 22 (38.6%) are older than 12 months. Sixty-two institutions (68.9%) mention vendor training, data retention or recording deletion, a broad definition that includes recording retention and general AI data rules. Twenty-five (27.8%) discuss AI or privacy law beyond naming a statute.

3.5 Hypotheses stated in the plan

3.6 Agreement between coders

The coders agree well on whether guidance exists and moderately on stance. Agreement on protocol disclosure, consent and law is weak, largely because the guidance is general and the coders read "unstated" differently. We therefore treat those variables as indicative and report them with the conventions above.

4 Discussion

Tool-level guidance mostly comes from IT pages, not ethics offices. Two thirds of institutions have something, but the tool-level material mostly comes from IT and security pages: of the 61 institutions that name a tool, 50 (82%) do so on an IT or security page and 23 (38%) on an ethics-office (IRB or HRPP) page, by Coder A's unit coding, which allows several units per institution. Ethics offices more often ask researchers to describe the method than to name products. A researcher who wants a yes or no for a specific product will often not find one from the office that approves their protocol.

Table 3: Hypotheses stated in the study plan before collection.

Hypothesis	Result
H1: fewer than half have transcription-specific guidance	**Not supported.** 73 of 130 (56.2%) do. Fewer than half have guidance that names a tool (46.9%).
H2: at least one vendor is approved by one institution and prohibited by another	**Supported.** Otter.ai (approved 5, prohibited 20), Rev (6, 3), and others.
H3: median guidance page older than 12 months	**Not supported.** Median 175 days among dated pages; 36.7% undated.
H4: fewer than 20% mention vendor training or retention	**Not supported.** 68.9%, under a broad definition that includes recording retention.
H5: most institutions with guidance require protocol disclosure	**Not supported.** 25.6% say so; 74.4% are silent or unclear.

Licensed platforms win; consumer bots lose, and researchers can end up on either side. The tools institutions approve are those they already license (Zoom, Teams, Microsoft 365). The tools they bar are stand-alone meeting bots. Otter.ai is the clearest split: barred at 20 institutions and approved at 5. A researcher moving between institutions, or working with collaborators, meets contradictory rules for the same service.

Requirements are thin and mostly implicit. Only about a quarter of institutions with guidance say that AI transcription must be disclosed in the protocol or covered in consent. That may understate practice, because ethics offices can ask these questions in forms that are not public.

Guidance ages unevenly. About a third of guidance is undated, and the dated guidance is fairly recent, but some pages predate the current litigation and vendor-term changes. The Otter.ai order of August 2026 [1] postdates much of what we read.

4.1 Practical implications

For researchers and graduate students. Treat the absence of a published rule as a question to ask, not a permission. Before uploading participant audio, find your institution's IT or security guidance and its ethics-office guidance, because in our sample they are often different pages with different content. Where the two disagree, ask the ethics office in writing and keep the answer with the protocol. Name the tool and the processing location in the protocol and the consent text; about a quarter of institutions with guidance say disclosure is required, and the rest rarely say either way.

For ethics boards and research offices. Fewer than half of the institutions we reviewed publish guidance that names a tool, and more than a third of what is published carries no date. A short, dated page that says which tools are approved for which classes of data, whether protocol disclosure and consent language are expected, and who to ask, would answer most of what researchers want to know. Reviewing the page when vendor terms change would keep it useful.

Table 4: Agreement between Coder A and Coder B before adjudication (129 institutions). Tool-name sets matched exactly for 69.4% of institution-role pairs, mostly because of naming variants.

Variable	Cohen's κ	Raw agreement
Guidance found	0.89	95.3%
Stance	0.68	74.4%
Local processing	0.71	89.9%
Data-classification rule	0.62	82.2%
Security review required	0.61	78.3%
Vendor training or retention mentioned	0.64	81.4%
Consent language required	0.52	72.9%
Protocol disclosure required	0.41	74.4%
Law or litigation discussed	0.39	76.7%

For transcription-software providers. Institutions in our sample favor tools they already license and tools that run locally. A provider that wants its service to be usable in research can publish its retention, model-training and data-location terms in a form an ethics office can cite, say plainly which plans offer different terms, and support local or self-hosted processing. Nothing in this study tests any provider's product, and no institution's guidance should be read as an endorsement.

4.2 Limitations

The sample covers large research universities in three countries and cannot speak for other institutions. Web pages are a proxy for what ethics offices tell researchers: guidance behind a login, in forms, or given orally is invisible, so "no guidance" means "none found by this protocol". Eight institutions could not be read, ten have a shortened search protocol, and search discovery may have missed pages. All coding was language-model-assisted; agreement between coding passes shows consistency, not correctness, and the human check covered 15 institutions. Adjudicators changed some values on their own judgment; the final dataset records every change. The survey is a snapshot of one week; institutions update pages continually. We did not contact institutions to confirm our readings.

5 Conclusion

Institutional guidance on cloud AI transcription of research data exists at most large research universities, but it is patchy, often drawn from IT pages on meetings, and inconsistent across institutions about the same products. The most actionable gap for researchers is the absence of a clear, dated, research-specific statement of which tools may be used with which kinds of participant data. The dataset lets anyone check our readings and update them.

Statements

Data and materials availability. The coded dataset (final coding, each coding pass with adjudication notes, page URLs, capture times and SHA-256 hashes, and short verbatim quotes), the codebook, the sampling frame and the analysis scripts are released under CC-BY-4.0 (code under MIT) at [**Zenodo DOI**]. Raw page text is not redistributed because it is third-party material; hashes and URLs allow verification, and captures are available on request. The preregistration is at [**OSF URL**].

Ethics. The study uses public documents only. There are no human participants and no personal data, and no institution was contacted. Institutions are quoted briefly, with their URLs, in the dataset.

References

[1] United States District Court, N.D. Cal., "In re Otter.AI Privacy Litigation, no. 5:25-cv-06911," 2026, order granting in part motion to dismiss, Aug. 13, 2026 (Judge Eumi K. Lee); allegations at the pleading stage.

[2] G. Samuel and D. Wassenaar, "Joint editorial: Informed consent and AI transcription of qualitative data," *Journal of Empirical Research on Human Research Ethics*, vol. 20, no. 1-2, pp. 3–5, 2025.

[3] S. Wollin-Giering, M. Hoffmann, J. Höfting, and C. Ventzke, "Automatic transcription of English and German qualitative interviews," *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, vol. 25, no. 1, 2024, art. 8.

[4] K. Kuhn, V. Kersken, B. Reuter, N. Egger, and G. Zimmermann, "Measuring the accuracy of automatic speech recognition solutions," *ACM Transactions on Accessible Computing*, vol. 16, no. 4, pp. 1–23, 2023.

[5] Y. J. Lim and N. C. Tan, "Talk to the bot: A scoping review on using AI-powered transcription tools in qualitative research," *Annals of the Academy of Medicine, Singapore*, vol. 55, no. 4, p. 217, 2026.

[6] J. Sponholz, A. Weilinghoff, and J. Schopf, "Halving transcription time: A fast, user-friendly and GDPR-compliant workflow to create AI-assisted transcripts for content analysis," arXiv:2503.13031, 2025, preprint.

[7] A. Haberl, J. Fleiß, D. Kowald, and S. Thalmann, "Take the aTrain. Introducing an Interface for the Accessible Transcription of Interviews," arXiv:2310.11967, 2023, preprint; published in Journal of Behavioral and Experimental Finance 41 (2024), 100891.

[8] J. Jiao, S. Afroogh, K. Chen, D. Atkinson, and A. Dhurandhar, "The global landscape of academic guidelines for generative AI and large language models," arXiv:2406.18842, 2024, preprint.

A Codebook summary

Institution-level variables: `guidance_found` (yes/no); `stance` (approves_named, prohibits_named, both_named, case_by_case, general_only, none); `approved_tools` and `prohibited_tools` (names, each with a verbatim quote); `requires_protocol_disclosure`, `requires_consent_language`, `security_review_required` (yes/no/unclear); `data_classification_rule` (yes/no); `local_processing` (preferred, permitted, not_mentioned, discouraged); `vendor_training_or_retention_mentioned` and `litigation_or_law_mentioned` (yes/no); `last_updated` (date shown or none shown); `blocked_or_unreadable` (true/false). Full definitions and conventions are in the deposited codebook (version 1.0).

B Final stance by country

Country	Approves	Prohibits	Both	Case by case	General only	None
United States (n = 100)	21	1	23	8	13	34
United Kingdom (n = 15)	3	0	2	2	3	5
Canada (n = 15)	9	1	1	2	1	1