

Toward an AI-Native TAC

A proposal for the Michigan OEAA/OEE Technical Advisory Committee

Damian W. Betebenner

Prepared for Andrew Middlestead, Director, OEAA · 10 September
2026

The Michigan TAC meets three times a year. Vendors and state staff present issues; the committee gives advice. The advice is good. What happens to it afterwards is the problem this brief addresses. Part I makes the case that the committee's output should be captured as a durable, queryable corpus rather than as a nine-page document, and sets out what OEAA would have to do and what guardrails the change requires. Part II is a technical appendix for whoever would build it. Both are grounded in a working demonstration built from the April 23–24, 2026 meeting.

Table of contents

Part I — The case	2
What a TAC meeting is for	2
Where the value leaks	2
What “AI-native” actually means here	4
What the cycle would look like	4
What OEAA would have to do	6
Guardrails	6
What it would be worth	7
The recommendation	8
Part II — Technical appendix	9
Design principles	9
The corpus schema	9
Capture and transcription	10
Extraction	11
Agents and their boundaries	11
Retrieval	12
Document generation	12
What this repository demonstrates, and what it does not	13
Sources and further reading	13

Part I – The case

What a TAC meeting is for

Three times a year, Michigan’s Office of Educational Assessment and Accountability and Office of Educator Excellence bring their hardest technical problems to five people who do not work for them.

That is the whole mechanism, and it is a good one. Vendors and state staff present what they are doing or proposing. The committee questions it, and advises. The state is not obliged to take the advice — the committee decides nothing, holds no vote, and signs nothing — but the advice is the product, and its quality is why the arrangement exists.

The April 2026 meeting is a fair sample. Over two days the committee told OEE that its working definition of “meaningfully different” test forms would not survive scrutiny, and supplied a replacement design in some detail. It told OEAA that the fix for high school growth targets was sound but that the adequacy threshold underneath it was set too low. It told a psychometrician that her method was the better of the two she had compared and that the comparison was not the question she should be asking. Each of those is worth considerably more than the meeting cost.

Where the value leaks

Not in the room. In what happens to the room afterwards.

5,613

words of minutes for a two-day meeting

3 weeks

from meeting to circulated record

0

recommendations with a tracked status

There are three distinct losses, and they compound.

The advice evaporates between meetings

The April minutes contain thirty-seven discrete recommendations. They are not labelled as such, not numbered, not gathered anywhere, and not assigned to anyone. They are distributed through nine pages of narrative as bullets beginning “TAC Feedback:”.

By September, answering *“what did we advise about adequate growth targets, and what has happened?”* requires someone to reread the whole document and reconstruct the list from memory. Answering it across three meetings requires rereading three documents. Nobody does this, which means the committee’s advice is effectively re-derived each cycle rather than carried forward.

That is the single largest loss, and it is structural. It has nothing to do with how well the minutes are written.

The record cannot hold the technical content

The most substantial work presented in April was Ji Zeng's comparison of two subscore classification methods — a 52-page deck, two batches of simulation conditions, real-data distribution analysis at three grades.

About thirty of those pages are chart images. Their findings reach the minutes as one sentence: *"The CT-PO method has shown its advantage over the SBAC method."*

What the figures actually show, recovered for this brief by rendering and reading every page:

i What one sentence of minutes was carrying

- Michigan's science subtest score distributions are **visibly non-normal** — multimodal and left-shouldered — which is the empirical warrant for abandoning the multivariate normality assumption that the presenter criticised in the literature. It is the foundation of the whole argument and appears nowhere in the record.
- The unidimensionality check returns a **first eigenvalue near 12.7 against second and third near 1-2.5**. That is the committee's central objection — that subscores are being extracted from something built to measure one thing — stated as a number rather than as an opinion.
- CT-PO beats the Smarter Balanced method in **every one of the simulated conditions**, on both accuracy and consistency. "Has shown its advantage" undersells a clean sweep.
- And the finding that matters most for policy: **CT-PO's advantage collapses as the latent dimensions become less separable**. At realistic separations both methods return classification accuracy too poor to report on an individual student. The better method is still not good enough, in exactly the region Michigan operates in.

The last point is the one a decision-maker needs, and the one most completely lost.

The minutes are not at fault here. A summary document has no mechanism for representing a figure. The information was never available to the format.

The same applies to the growth session. Eight slides were hidden and skipped live; two speaker notes carry material said nowhere aloud — including that **educator evaluations and school grades have been repealed from state law**, which changes the stakes picture for the entire growth-model discussion. None of it is in the record, because none of it was spoken.

The record cannot say who said what

Every question and every recommendation in the April minutes is attributed to "TAC." Five people were in the room with genuinely different expertise. The record renders them as one voice.

There is exactly one sentence in 5,613 words where an individual member is recorded taking

an action.

This matters for a specific reason. Advice from a technical advisory committee carries weight because of who is giving it. “TAC Feedback: Above 65 should be goal of adequate growth” is an instruction from an anonymous body. The same sentence attributed to the member whose career is in growth measurement is an argument a reader can evaluate, and — when OEAA later has to defend the decision to a superintendent or to the federal government — a citation.

None of these three losses is a criticism of the minutes or the person who writes them. They are what happens when a two-day technical conversation must become a nine-page document three weeks later, drafted by one person, with no way to represent a figure and no convention for attribution.

The remedy is not a longer document or a more diligent author. It is a different artifact.

What “AI-native” actually means here

Not a chatbot. Not a summarisation tool bolted onto the existing process. The proposal is a change in what the meeting *produces*.

Today the meeting produces a **document**. A document is a fixed sequence of words optimised for one reading, by one audience, once. Everything not in it is gone.

The proposal is that the meeting instead produce a **corpus**: a structured, cross-linked body of knowledge that persists across meetings, can be queried in natural language by anyone with access, carries every piece of advice with a status attached, and from which documents — minutes included — are *generated* rather than authored.

The distinction is not cosmetic. A document answers the questions its author anticipated. A corpus answers the questions you have later.

💡 This brief is accompanied by a working example

Everything described here has been built for the April 2026 meeting and is published alongside this document. It contains 98 cross-linked pages: one meeting, five sessions, seven source documents, twelve standing topics, thirty-seven recommendations each with a status and an owner, eight open questions, and a glossary.

It was built from the April minutes alone — no transcript, no recording, no new interviews — in a single working session. Its limitations are exactly the limitations of its source, which is the most useful thing about it as evidence.

The [corpus index](#) is the entry point.

What the cycle would look like

Three phases, mapped onto the meeting cycle OEAA already runs.

Before the meeting

The agenda is drafted against the corpus rather than from scratch. Every open recommendation addressed to a presenting office surfaces automatically, so a session on growth targets opens with the six things the committee said last time and what has happened to each.

Presenters' materials are ingested when circulated — which means the committee arrives having been able to ask questions of the deck rather than only of the presenter. A member who wants to know what changed since April gets an answer from the corpus in a sentence rather than by rereading two documents on a plane.

💡 This one is no longer hypothetical

A Copula Approach to the Subtest Indicator Problem is a worked response to Ji Zeng's April session, written from the corpus five months after the meeting and addressed to her by name.

It exists because of the capture this brief argues for. It separates four research questions the session conflated, identifies that the committee's own joint-density recommendation is a copula recommendation without the vocabulary, and proposes a bounded check on data Michigan already holds — and it turns on the Batch 2 gradient, which lives entirely in chart images and reaches no other record of the meeting.

It took minutes to assemble from the corpus. It could not have been assembled from the minutes at all, and re-reading a fifty-two-page PDF would not have surfaced it either.

It is also the artifact the current process has nowhere to put. A TAC member's considered technical response, developed between meetings, currently has no home: it arrives as an email, or it waits five months and arrives as a remark. The corpus gives it a page, a type, and a link from every recommendation it bears on.

During the meeting

The meeting is **recorded and transcribed**. This is the one genuine change in participant experience, and the guardrails around it are covered below.

Transcription alone is not the point. A raw transcript of a two-day technical meeting is less usable than good minutes — it is long, unstructured, and full of false starts. The point is that a transcript is a *substrate*: from it, sessions, recommendations, open questions, and attributions can be extracted with the speaker preserved.

Nothing changes about how the committee works. No one consults an AI during the discussion. The capture is passive.

After the meeting

Draft minutes are generated within a day, not three weeks, and reviewed by the same person who writes them today — a smaller job that starts from a complete draft instead of from

notes, and one that can be done while memory is fresh.

The corpus is updated, and this is the step that matters most: **prior recommendations get their status changed**, with the meeting that changed them named. That is the mechanism by which advice stops evaporating.

And then anything else the state needs — a technical documentation appendix, an ESSA plan amendment narrative, a briefing for a new superintendent, a summary for a legislative committee — is generated from the same corpus rather than reconstructed from the same PDFs.

What OEAA would have to do

Considerably less than the ambition suggests.

The necessary change

Record and transcribe the meeting. Everything else is downstream of having a transcript. Consent from participants, a recording device, and a transcription service. This is the decision that requires OEAA leadership.

The rest

A repository, and a person who owns it. The corpus lives in version control. Maintaining it is a defined task after each meeting, not a new full-time role — the work is reviewing generated content and updating statuses, not authoring from scratch.

What OEAA would **not** have to do: change how the committee works, change who writes the minutes, change vendors, procure a platform, or put student data anywhere near any of it.

Guardrails

These are not afterthoughts. A state education agency has obligations that a technical proposal has to satisfy before it is worth discussing.

! Public records and FOIA

A recording and a transcript of a public body's meeting are records. They are subject to Michigan's Freedom of Information Act on the same terms as the minutes are now, and they are far more detailed.

This is the guardrail that deserves the most thought before adopting the proposal, because it is the one that changes participants' exposure. Two considerations pull in opposite directions:

- A committee that knows it is being transcribed verbatim may hedge. The candour in the April minutes — a vendor reporting its own modelling work as “none of this is great,” the committee telling a state office its working definition would not survive scrutiny — is exactly what makes the advice valuable, and it is exactly what a more exposed record could suppress.
- Conversely, a fuller record is a better defence. When OEAA has to show that a decision was technically advised, a transcript is stronger evidence than a paraphrase.

The workable position is that **the transcript is a working document and the corpus is the record**. OEAA should decide, with its FOIA counsel and before recording anything, what is retained, for how long, and what is published. That decision should be made once, written down, and told to participants in advance.

Consent. Participants are told the meeting is recorded, why, what is retained, and who can access it. Nobody is recorded by surprise. A member who objects is a reason to revisit the policy, not to work around them.

Student data stays out. Nothing in this proposal requires or benefits from student-level data. The corpus holds deliberations — methods, arguments, recommendations — not records. Aggregate figures presented in a deck are fine; microdata has no place in it and the repository should enforce that mechanically rather than by convention.

Vendor confidentiality. DRC, Pearson, and Smarter Balanced present material that is commercially sensitive and sometimes test-secure. Item overlap percentages and form-production strategies are not obviously publishable. The corpus needs a visibility boundary — most plausibly a public layer and a restricted layer — decided with the vendors rather than for them.

AI drafts; humans approve. No generated document becomes official without a named person reviewing and approving it. This is the standing rule, and it is not negotiable. The value of the proposal is that it makes the reviewer’s job start from a complete draft rather than from a blank page — not that it removes the reviewer.

The corpus is auditable. Every claim in it cites the source document it came from. Anyone can check it against the raw material, which is retained unmodified. This is what makes the whole thing trustworthy: not that the extraction is perfect, but that it is checkable.

What it would be worth

Three things Michigan cannot currently do, in rough order of value.

1. Carry advice across meetings. Open a single page, see every recommendation the committee has made, its owner, and its status. Run the meeting against it. This is the difference between a committee that advises and a committee that is *heeded* — and it is entirely a book-keeping problem that no one currently has the tools to solve.

2. Keep the technical substance. The April simulation findings, the hidden-slide context about repealed statutes, the participation table’s actual total — all of it retained rather than compressed away. When someone asks in 2028 why Michigan chose one subscore method

over another, the answer exists with its evidence attached.

3. Answer questions nobody thought to ask. *Which recommendations are blocked on the same ESSA constraint? What has the committee said about parent comprehension across three meetings? Which of last year's advice did we act on?* A document cannot answer these. A corpus can, in seconds, with citations.

The recommendation

Record and transcribe the fall 2026 meeting, and build the corpus from it.

Settle the FOIA and retention policy first, with counsel, and tell participants in advance. Assign one person to own the corpus. Then run the spring 2027 meeting against the register of open recommendations and see whether it changes the conversation.

The cost is a recording device, a transcription service, and a defined post-meeting task. The demonstration accompanying this brief was built from summary minutes alone. With a real transcript, it would be substantially better — and the largest single improvement would be knowing which member said what.

If the fall meeting is too soon, the smaller version is worth doing on its own: **build the recommendation register from the existing minutes and maintain it.** That requires no recording, no policy decision, and no new technology, and it addresses the largest of the three losses. The register in the accompanying corpus is a working starting point.

Part II – Technical appendix

For whoever would build this. The reference implementation is the repository accompanying this brief; paths below refer to it.

Design principles

Four decisions shape everything else.

Plain files, in version control. The corpus is Markdown with YAML frontmatter in a git repository. No database, no platform, no vendor. It is readable by a human in a text editor, by an AI agent, by GitHub’s web interface, and by any Markdown tool. Git supplies history, attribution, review, and rollback for free. The single most important property of an institutional knowledge base is that it outlives the tooling that created it.

Raw sources are immutable. Original documents live in `raw/` with their original filenames, typos and all, and are never edited. Everything derived is regenerable. This is what makes the corpus auditable — any claim can be traced to a source that has not been touched.

One fact, one page. A session page records what happened on a day. A topic page carries the state of an issue across meetings. A recommendation page holds one piece of advice. Nothing is restated; things are linked. This is what keeps the corpus from drifting into contradiction as it grows.

The schema is a file in the repository. `wiki/schema.md` defines the page types, frontmatter fields, link conventions, and log format. Both humans and agents read it before authoring. A convention that lives only in someone's head is not a convention.

The corpus schema

A research knowledge base organises around Source → Method → Finding. A committee's does not. The chain that matters here is:

Artifact → Session → Topic → Recommendation → Follow-up

↓ ↓

People / Organizations

Type	Directory	Role
meeting	meetings/	One per convening — roster, session index, themes
session	sessions/	One per agenda item — presenter, what was asked, what was advised

Type	Directory	Role
artifact	artifacts/	One per source document — provenance, contents, known defects
topic	topics/	A standing thread that outlives any meeting
recommendation	recommendations/	One discrete piece of advice, with status
question	questions/	Posed and unresolved
person / organization	people/, organizations/	Participants and bodies

The two bolded types are what make it a committee corpus rather than a meeting archive.

Topics are shared across meetings. The commonest way this kind of corpus rots is creating growth-targets-fall-2026.md alongside adequate-growth-targets.md. Each topic carries a “state as of” line and a history table with one row per meeting that moved it. Only artifacts and sessions are per-meeting.

Recommendations are the load-bearing type. Each carries:

```

raised: 2026-04-23
addressed-to: OEAA | OEE | DRC | Pearson | Smarter Balanced | TAC
status: open | accepted | in-progress | closed | declined | superseded

```

The rule that makes it work: **a status is never changed silently.** When a later meeting moves a recommendation, the meeting is named in the page body and the change is a git commit with an author and a date. A register that can be edited without trace is not a register.

Capture and transcription

The transcript is a substrate, not an output. Requirements, in priority order:

1. **Speaker attribution.** Diarisation that distinguishes five committee members from a dozen state staff is the whole value proposition. Named speakers beat accurate words — a misheard technical term is fixable from the deck; an unattributed recommendation is not.
2. **Technical vocabulary.** “AGP,” “NTILE,” “CT-PO,” “quadratic weighted kappa” will be mis-transcribed by a general model. Supply a domain glossary — the corpus already has one at wiki/glossary.md — as biasing vocabulary, and expect to correct against the decks.
3. **Timestamps,** so transcript segments align to agenda items.

Room audio for a fifteen-person meeting is the practical constraint. A single table microphone will produce diarisation good enough for the presenter and poor enough for the far

end of the table to be useless. Budget for per-seat or boundary microphones; this is where the proposal succeeds or fails operationally.

The transcript is ingested to `raw/<meeting-date>/transcript/` and never edited. Corrections live in the corpus, with the correction recorded.

Extraction

`scripts/extract-sources.sh` handles the mechanical part:

- PDFs → text via `pdftotext -layout`.
- PowerPoint → slide text, **hidden slides**, and **speaker notes** via direct OOXML extraction. Hidden slides and notes are not incidental: in April they carried the repeal of educator evaluations from state law and the mechanism behind the WIDA projection horizon, neither of which appears anywhere else.
- Word → text.

! The step that is easy to skip and expensive to skip

Image-heavy decks must be rendered and read visually.

The script computes words-per-page for every PDF and flags anything below a threshold, rendering those pages to PNGs for a vision-capable model to read.

For the April simulation deck this was the difference between one sentence and the entire finding. A text-extraction pipeline reports that deck as successfully processed — 680 words extracted, no errors — while silently discarding everything that mattered. The failure is invisible unless you check the ratio.

Agents and their boundaries

Three agents, defined in `.claude/agents/`, with deliberately narrow remits.

meeting-synthesizer turns a transcript into structure. Its core skill is distinguishing what was *presented* from what was *asked* from what was *advised* — the presentation is context, the advice is the product, and most minutes blur them. It splits recommendations aggressively: “move the takeaway to the top and publish the formula” is two recommendations with different owners and different costs.

tac-wiki-maintainer authors and maintains pages. It owns `wiki/`; it never touches `raw/`.

corpus-lint is read-only and audits: broken links, orphan pages, stale statuses, topic drift, contradictions between pages, and — the most important check — **attribution violations**, meaning any page that attributes a recommendation to a named individual where the source attributes it to a body.

That last check deserves emphasis. An AI asked to make a record more useful will readily infer that the growth-measurement specialist made the growth-measurement argument. That

inference is usually right and occasionally wrong, and a fabricated attribution in a technical advisory record is worse than no attribution at all. The rule is absolute: **attribute only what the source attributes**. The April corpus preserves body-level attribution throughout, and documents the loss rather than repairing it.

Retrieval

The corpus is queried with `grep` and file reads. This sounds primitive and is not: 98 densely cross-linked Markdown files with a curated index and a controlled vocabulary are more reliably searchable than the same content in a vector store, because the answer can be *verified* against the cited page.

The discipline, defined in `.claude/skills/tac-query/`:

- Start at the curated index, which is prose grouped by theme, not a directory listing.
- Choose page type by question shape — “what is the state of X” is a topic; “what did the TAC advise” is a recommendation; “what happened on the day” is a session.
- Follow `[[wikilinks]]` outward. The second hop is usually where the answer is.
- **Open every page you cite**. An index line is a pointer, not a source.
- **Cite page paths**. Never assert from memory.
- **Say what is not there**. Naming a gap is a better answer than filling it plausibly.

A session started in the repository runs `.claude/hooks/orient.sh`, which prints the recommendation counts by status, the items explicitly carried to the next meeting, the open questions, and the last few log entries. Context is automatic rather than re-explained each time.

Document generation

Documents are Quarto (`.qmd`) rendering to HTML and PDF from one source, published to GitHub Pages by a CI workflow. The corpus renders as part of the same site, so a generated document can link directly into the evidence behind any claim it makes.

The mechanics worth knowing:

- **Themes are token-based**. `styles/tac-light.scss` and `styles/tac-dark.scss` define only colour custom properties; every rule lives in a shared `_tac-core.scss` that reads them. The PDF preamble `styles/tac-preamble.tex` re-declares the same tokens as LaTeX colours, so the two output channels are one design system rather than two.
- **`[[wikilinks]]` are resolved at render time** by `filters/wikilinks.lua` against a slug map generated from the corpus. The corpus stays readable in a plain text editor and on GitHub while publishing as a proper hyperlinked site.
- **Render PDFs before the website**. A Quarto project render cleans its output directory.

What this repository demonstrates, and what it does not

Demonstrated. That a two-day meeting's material can be turned into a queryable, cross-linked corpus; that recommendations can be tracked with status across meetings; that content locked in figures and speaker notes can be recovered and is substantial; that documents can be generated from the corpus with citations into it; and that the whole thing can be built with plain files and open tooling.

Not demonstrated. Individual attribution — the source cannot support it. Multi-meeting continuity — there is only one meeting; the mechanism exists and is untested. Real-time use during a meeting, which is deliberately out of scope. Vendor-confidential handling, which needs a visibility boundary this public demonstration does not implement.

The honest summary is that this is one meeting's worth of evidence, built from the weakest possible source, and it is already more useful than the source. The interesting question is what it looks like with a transcript.

Sources and further reading

- The corpus: [index](#) · [schema](#) · [register of open recommendations](#)
- The companion document: *A Fuller Record*, a comprehensive summary of the April meeting built from the same material
- The evidence for the losses described in Part I: [what the official minutes lose](#) · [findings recovered from chart images](#) · [hidden slides and speaker notes](#)
- Repository: https://github.com/dbetebenner/Michigan_TAC_042326