

STRATEGIC BLUEPRINT · EMPATHY.AI

A Strategy for AI in Research

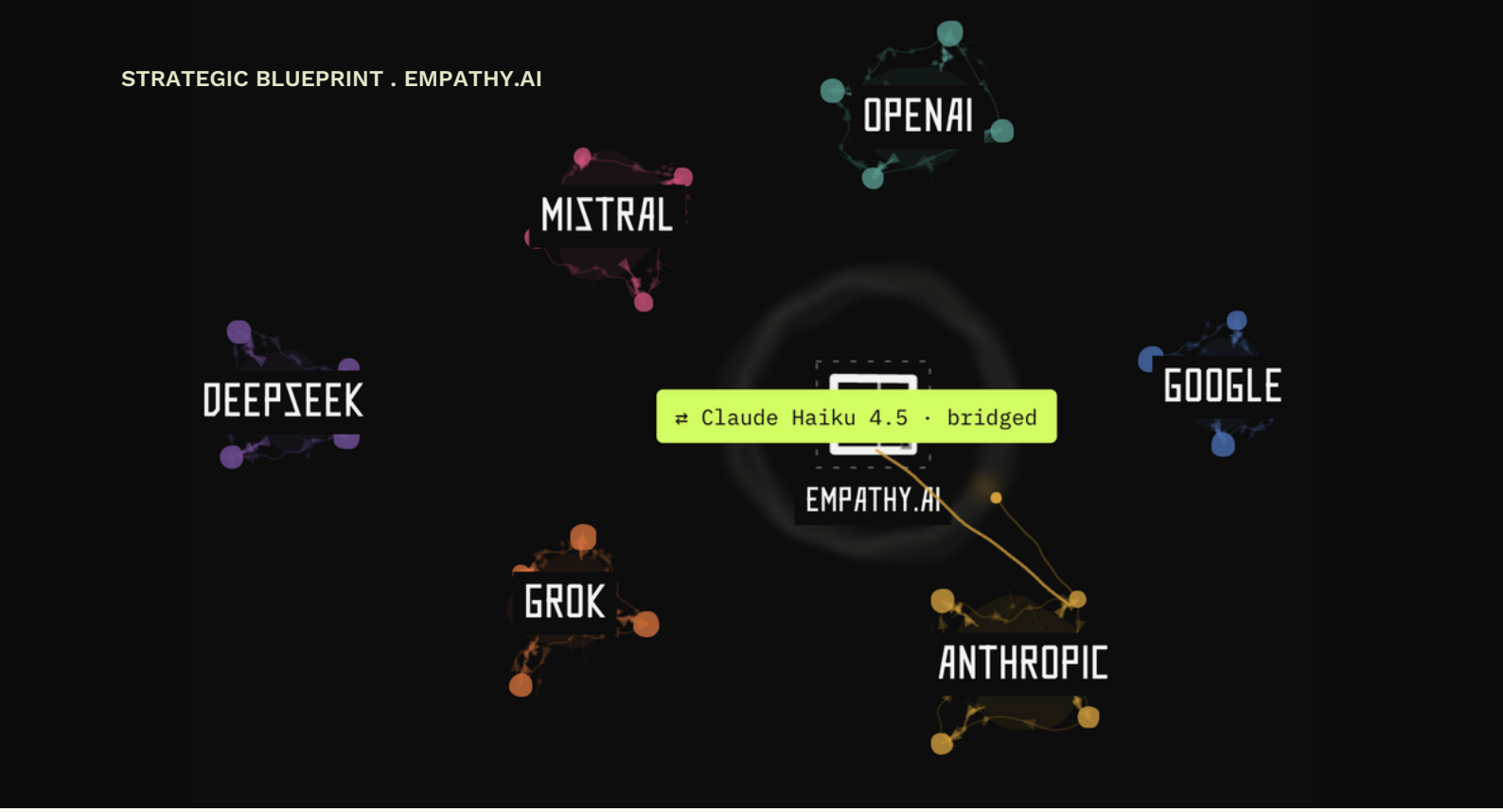
Locally governed intelligence that puts commercial AI in-line to authored knowledge.

A strategic blue print for research institutions, informed by Empathy.AI

Prepared for	Empathy.AI
Intended audience	Research leadership, principal investigators, integrity, libraries, governance and computing teams
Edition	November 2026 · Discussion draft
Nature	Strategic guidance, not a commercial offer or legal opinion

The purpose of AI in research is to equip the researcher to question, investigate and defend what is claimed.

This blueprint proposes a governed approach to AI-assisted research, with local execution and knowledge-centric experiences as practical instruments of institutional autonomy. Institutions can adapt its recommendations to their disciplines, obligations and resources.



Research, Knowledge and LLMs

The organising question for relating with AI when conducting research can be **“What supports this claim?”** A knowledge-centric approach puts authored evidence and accountable interpretation at the centre; the LLM helps researchers find, compare and work with that knowledge.

Adapted from the supplied “Where Answers Come From” illustration, the comparison on the following page contrasts a proposed Local Knowledge-Centric Research workflow with an LLM-Centric Research workflow that has no access to the project archive. The documents and responses are fictional illustrations.

Research, Knowledge and LLMs · Visual guide

SAME QUESTION

Which exclusion rule was approved for Study A?

Local Knowledge-Centric Research

INTERPRET → RETRIEVE → COMPOSE

Approved project archive: Protocol v3, §4; decision log, 12 May; analysis plan v2.

SOURCE-GROUNDED DRAFT

“Exclude samples without a baseline measurement.”

Protocol v3, §4; decision log, 12 May.

LLM-Centric Research

PROMPT → PATTERNS → RESPONSE

Model-led response without access to the approved study records. No project archive is searched in this example.

ILLUSTRATIVE UNSUPPORTED ANSWER

“Exclude incomplete records and statistical outliers.”

No supporting study record.

NEW QUESTION: What did the June amendment change?

REQUIRED MISSING-EVIDENCE RESPONSE

“Not found in this collection.” Records cover approvals through May; inspect or expand the collection.

FAILURE TO TEST: INVENTED DETAIL

“The June amendment relaxed the threshold.” No supporting record; do not accept as evidence.

In both workflows, the researcher must check the source, inference and method before accepting a claim. The left-hand behaviour is an acceptance requirement to test for accuracy and reliability. The right-hand answers illustrate a possible failure mode.

Research, Knowledge and LLMs

Local control and knowledge grounding are different choices

In the proposed Local Knowledge-Centric Research environment, approved protocols, literature, datasets and research records provide the evidence context. The system should retrieve relevant material, expose authors and versions, and distinguish what the sources say from what the model infers. Researchers retain authority over the question, the method and the interpretation.

“Local” adds a separate requirement: documents, prompts, embeddings, retrieved passages, outputs and inference remain within the workstation or institution-controlled service. Any external handoff must be an explicitly approved exception. A local model without governed retrieval can still be LLM-centric; a cloud service with verified retrieval can support a knowledge-centric workflow, although its data boundary is different.

Make uncertainty visible and responsibility inspectable

When evidence runs out, the preferred response is “not found in this collection”, followed by a way to inspect the search or extend its scope. This response describes the limits of the collection and the search. Label conjectures and proposed explanations as exploration, and verify them before accepting them as findings.

Source attribution makes a claim easier to investigate. Researchers retain responsibility for their use of a source, and legal liability depends on the facts and applicable obligations. An error may originate in the source, retrieval, synthesis or subsequent interpretation. Preserve those steps so the team can identify the failure, correct it and trace affected artefacts.

A network of authored research knowledge

The attachment’s network of knowledge owners suggests a further design direction: connect a laboratory archive, partner research group, university library, publisher and research repository through governed access, with each contribution attributed. Each participant would retain stewardship of its collection and determine what may be shared.

For a particular inquiry, the proposed network should exchange only permitted, relevant passages and retain their identifiers, authors, versions and access conditions. The answer record should show which collections were consulted, which passages supported each claim and where evidence remained missing or contradictory. Assess scientific credit through actual contribution and obtain permission separately from attribution.

This proposed research architecture would require validation of its permissions, provenance and cross-institutional data flows before implementation and use.

Local control protects the research boundary. Knowledge grounding makes the evidence inspectable. Human judgement determines whether the claim is justified.

The strategic choice

AI should be adopted for the quality of inquiry it enables. The recommended approach is **local-first where control matters, evidence-first wherever a claim is made, and human-led wherever judgement carries consequence**.

Local execution, source-grounded answers and human oversight address control, evidence and judgement respectively. Scientific validity requires separate evaluation of the method and its results.

The decision for research leadership

Authorise a bounded pilot of a governed research knowledge environment alongside existing approved services. Require evidence of research quality, inspectability, operational security and whole-cost performance before wider adoption.

Select the tool for the task. Prefer conventional software where it is sufficient, and permit approved cloud models where their capabilities justify their risks and the work can be appropriately documented.

Reading guide

Theme	Pages
Research, Knowledge and LLMs	2–4
The research challenge and its ethical response	6–7
Mapping Nature’s local-AI argument into strategy	8
How research artefacts become vulnerable to challenge	9–10
Empathy.AI Spaces and the deployment architecture	11–12
Human-led workflows and responsibilities, with visual guides	13–16
Sustainability, economics and the comparison visual	17–18
A measured pilot	19
Questions for leadership and the strategic commitment	20–21
Annex: provenance record and evidence boundaries	22–23
Practical case study: the frontier bridge	24–27

The primary measure is a defensible research outcome: evidence that can be inspected, a method that can be explained, and a conclusion for which a person accepts responsibility.

The research challenge

Research requires understanding

Research is the work of deciding which question matters, which observations can support a claim, which explanation survives criticism and which uncertainty must remain unresolved.

An AI system can assist each of these activities. The strategic danger is that a fluent result becomes a substitute for the intellectual work needed to justify it: an answer arrives, but the researcher cannot reconstruct the choices through which it became persuasive.

Messeri and Crockett's Nature Perspective identifies this risk: AI can encourage illusions of understanding and scientific monocultures, in which particular methods, questions and viewpoints displace alternatives. Their analysis provides a conceptual framework for examining these risks in specific research settings (Nature, 2024[1]).

Assistance that strengthens authorship

The ethical starting point of the supplied education strategy is worth preserving: the tool should strengthen the person's relationship with knowledge. In research, that means protecting the researcher's authorship throughout the inquiry, including the selection of evidence, methods and interpretations.

Meaningful oversight requires material choices to be visible, contestable and within the researcher's competence to evaluate. This includes the selection of literature, definition of categories and interpretation of results.

Routine conversion, retrieval, formatting and tested computation can release time for harder questions. Researchers should review automated choices that determine what counts as evidence or what a result means.

Protect the capacity to disagree with the system. An assistant that makes its answer easier to question is more useful to science than one that makes its answer harder to resist.

1. Nature, 2024 · https://www.nature.com/articles/s41586-024-07146-0

Three reframes for ethical research AI

Empathy.AI's emphasis on agency, productive effort and active scrutiny translates into three design commitments. These recommended principles can be adapted to each discipline's methods and needs.

Intellectual scaffolding

AI should help researchers formulate alternatives, retrieve evidence and test an argument. A useful system exposes weaknesses in a hypothesis and supports the researcher's judgement about what the inquiry can conclude.

Before requesting an interpretation, the researcher should, where appropriate, state the question, the proposed method and what would count against the expected result. Exploratory AI remains legitimate, but the record should distinguish ideas generated during exploration from hypotheses specified before analysis.

Preserve judgement; remove avoidable friction

Not every difficult task is intellectually valuable. Automating a file conversion or generating a tested plotting script is different from silently deciding which participants to exclude or which contradictory studies to omit.

The system should prepare consequential changes as reviewable proposals. Researchers should approve analytical thresholds, changes to protocols, interpretation of ambiguous observations and the promotion of a draft into the research record.

Active scrutiny

A research interface should distinguish source evidence, generated synthesis, conjecture and unresolved disagreement. It should make it easy to open the passage behind a claim and easy to discover that the passage does not support it.

"Not found in this collection" is an informative result that should make the search limits visible. The corpus may be incomplete, the search may have failed, or relevant work may sit outside the permitted collection.

Human autonomy as an operational requirement

• **Intellectual independence:** Show alternatives and contrary evidence, and support the researcher's ability to question an answer.

• **Authorship and credit:** Keep authors, curators, participants and communities visible behind the material being used.

• **Researcher privacy:** Record the method needed to audit a result, with collection and access limited to that purpose.

• **A recognisable tool:** Present the system as a research instrument with explicit capabilities, limits and human oversight.

• **Successful exit:** Help the researcher leave with a better-supported next step.

A human-led system preserves human authority over the choices that give an operation scientific meaning, while allowing routine mechanical work to be automated.

From Nature’s argument to institutional design

Buhler, Pérez and Boettiger argue that science’s infrastructure needs differ from those of mass-market AI platforms and advocate local, open-weight models as a more accountable alternative. Their Nature Comment presents a strategic argument; cost, environmental impact and capability require comparison on the intended workload (Nature, 2026[1]).

The following mapping translates the argument in the supplied article into proposed institutional choices. The recommendations and acceptance tests are this strategy’s additions.

Reflection in the article	Strategic response	Evidence required before scaling
Scientific workloads have their own scale and infrastructure needs.	Size services around research groups, actual tasks and data boundaries.	Task mix, concurrency, context lengths and acceptable waiting times.
Open-weight models can run on institutional equipment.	Maintain a local execution option and a controlled model catalogue.	Model licences, retained versions, security review and task-specific quality tests.
Decentralisation can increase accountability and control.	Give institutions authority over changes, access, retention and export.	A named operator, change records, access tests and a working recovery plan.
Smaller infrastructure can be a more sustainable path.	Use the smallest adequate tool, share capacity and schedule heavy work.	Energy and whole-life estimates at matched task quality and utilisation.
Reliance on subsidised subscriptions creates dependency.	Preserve exit routes and compare whole costs over a research programme.	Software terms, export tests, renewal scenarios and replacement costs.
Capability increasingly depends on surrounding tools.	Invest in retrieval, tested analytical software and governed workflows alongside the model.	End-to-end evaluation of the complete system, including tool failures.

The boundary of the inference

Treat the article’s user-capacity and model-performance comparisons as illustrations. Establish procurement and capacity commitments through benchmarks on the intended institutional workload.

The practical objective is a proportionate research service with demonstrable control over execution and operation. Its dependencies include upstream training impacts and external model developers.

Local AI is most compelling as an institutional capability: the ability to choose, inspect, preserve and replace the instruments of inquiry.

1. Nature, 2026 · https://www.nature.com/articles/d41586-026-02451-2

How an artefact becomes open to challenge

The integrity of a paper, thesis, dataset or patent application depends on whether its material contributions can be supported, explained and checked. Oxford's policy makes the user responsible for integrity and calls for validation and disclosure (Oxford research policy[1]).

The following fictional scenarios illustrate potential failure mechanisms.

Failure mechanism	Why the artefact can be questioned	Defence to build into the workflow
Unsupported synthesis or fabricated citation	A reviewer cannot find evidence for a central statement.	Check claims against original sources; retain passages and distinguish inference.
Model or service drift	Later users cannot reconstruct how labels, extracts or judgements were produced.	Preserve actual runs; record versions and changes; test output stability.
Opaque analytical decisions	Exclusions, categories or transformations have no defensible rationale.	Require explicit criteria, human approval and a versioned decision record.
Inadequate disclosure	Readers cannot assess the extent to which the result depends on AI.	Describe the tool's substantive role, limitations and validation alongside its name.
Ethics or permission mismatch	Actual data handling differs from participant information or approved methods.	Review data flows and seek an amendment or permission before changing scope.
Contamination of the research record	Generated material is mistaken for observation or primary evidence.	Label synthetic data and drafts; separate them from verified records.

Trace the path from output to consequence

An unsupported model statement may begin as a private note, become a category in a dataset and finally support a published claim. The corrective burden grows as dependent artefacts accumulate; the control point is therefore the moment a suggestion becomes accepted research material.

Responding to a challenged claim

A discovered error may require correction, further validation or withdrawal of a claim. Whether conduct amounts to misconduct, contractual breach or legal liability depends on the facts, applicable standards and the relevant process; an LLM error alone does not settle that question.

Keep generated material provisional until someone can say what supports it, how it was checked and why it belongs in the record.

1. Oxford research policy · https://www.ox.ac.uk/research/support-researchers/research-practice/policy-generative-ai-research

Defensibility across research disciplines

Validation should reflect the discipline and the AI contribution's role in the method. The following fictional Oxford-style examples illustrate that adaptation.

Biomedical research: classification is part of the method

A team asks an LLM to classify narrative clinical records and uses the resulting labels in an association analysis. A journal reviewer could question label accuracy, subgroup errors, consent compatibility and whether a different model version changes the result.

Use an appropriately authorised environment, a clinician-reviewed reference set, pre-specified evaluation criteria and sensitivity analyses. Preserve model outputs and adjudication decisions. Evaluate disclosure controls and label accuracy separately.

Humanities: interpretation must remain attributable

A historian uses a Space to compare letters and draft an account of political influence. A cited letter may mention a meeting without supporting the model's claim that the meeting changed policy.

Retain the passage, archival identifier and interpretive uncertainty, and show competing readings. A source link establishes traceability; the historian must still defend the inference and identify relevant material outside the selected archive.

Social science: categories can silently become findings

A team uses generated themes to code interviews. A funder, ethics committee or methods reviewer could ask whether the categories reflect participants' accounts, the researchers' assumptions or the model's defaults.

Version the codebook, document revisions and inspect disagreements against transcripts. Where appropriate to the methodology, use independent coding or another explicit validation approach to establish corroboration.

Engineering and physical science: validate the analysis

A model generates a data-cleaning script that silently drops extreme measurements. A collaborator or reviewer could challenge the result even if the script executes without error.

Retain the proposed code, review transformations, test against known cases and compare results with an independently checked pipeline. Pinning the model helps reconstruct generation, but the validated code and its dependencies are the primary analytical artefact.

From reproducibility to robustness

Archive what actually happened, test whether a preserved workflow can be rerun, and examine whether plausible variations change the conclusion. These are distinct aims: preserving weights and a random seed does not guarantee bit-for-bit identity across hardware, libraries or inference engines.

A defensible conclusion survives appropriate scrutiny, including scrutiny of the AI's contribution.

Empathy.AI for research

The model reads; the evidence retains its authors

The supplied Empathy.AI materials describe Spaces as a private, knowledge-centric environment for finding, comparing and composing from authored or institution-curated material. This strategy proposes a research adaptation whose suitability would be evaluated through an institutional pilot.

A research Space could contain approved protocols, licensed literature, lab documentation, permitted datasets and prior results. Its purpose would be to make those materials navigable while keeping their authorship, scope and limitations visible.

Evidence scope and open inquiry

“Context-limited” means that asserted factual claims should be supported by the authorised evidence available to the task. The model’s prior training and capacity to introduce unsupported material make claim-level verification essential.

The experience should separate evidence-backed answers from clearly labelled hypotheses and suggestions. When the collection is insufficient, it should offer a deliberate next step: consult a researcher, conduct an approved external search, or expand the collection with a documented decision.

A research-ready Space should provide

• **Knowledge stewardship:** Named curators, scope statements, document versions, rights and inclusion criteria.

• **Claim-level evidence:** Accessible passages whose content supports the specific claim.

• **Honest limits:** Explicit uncertainty, contradictory evidence and “not found in scope” responses.

• **Governed assistance:** Reviewable plans, bounded tools and approvals for consequential changes.

• **Exportable records:** The material needed to examine a run without depending on a live chat service.

Verify these acceptance requirements through a pilot, using independent tests of source grounding, permissions and agentic controls against the relevant research standards.

Preserve the distinction between provenance and truth

A wrong source can be cited accurately, an incomplete collection can be searched correctly, and a plausible inference can exceed its evidence. Retrieval quality, citation support, corpus coverage and scientific validity must therefore be evaluated separately.

These architectural requirements apply equally to Empathy.AI, an institution-built stack and other governed platforms. Assess each option against the same tests.

The authority behind a research claim belongs to evidence and accountable interpretation. The language model is an instrument for working with both.

One strategy, several deployment boundaries

The supplied product materials describe a customer-owned Local AI Workstation and a dedicated hosted Private Space; the proposal also allows Spaces on institution-operated hardware. Each arrangement has its own processing location, data boundary and operating responsibilities.

Deployment	What the institution gains	What still needs to be established
Local AI Workstation or institutional server	Direct control over the execution boundary and retained model versions.	Security operations, patching, access, backups, licences and recovery.
Empathy.AI hosted Private Space	Dedicated managed capacity without owning the hardware.	Exact host and location, processor terms, support access, subprocessors, retention and exit.
Approved cloud service or API	Access to capabilities without operating the inference stack.	Data permissions, feature-specific terms, reproducibility limits, retention and export.

The local boundary must cover the whole workflow

In the proposed local research configuration, documents, prompts, embeddings, retrieved evidence, outputs and inference remain within the workstation or University-controlled service. Demonstrate this acceptance condition across the complete workflow.

Check ingestion, OCR, embeddings, reranking, analytics, crash reports, backups and remote support. Separate signed software updates from research-content access, and verify egress controls through testing.

A bounded external-model exception

Where frontier capabilities add justified value, a proposed bridge can prepare a minimum-context task locally, show the exact outbound content for approval, send only that approved payload and bring the response back for verification. Keep this workflow off by default for restricted projects.

Abstraction reduces the context shared with an external model. A scientific idea, participant or commercial secret may still be inferred from the remaining details. Validate this optional design pattern through a disclosure-risk assessment and separate checks of any anonymisation claim.

Preserve the complete method

Retain weights where licensed, tokenizer, model hash, quantisation, inference engine, prompts, retrieval configuration, corpus snapshot and executed tools. Open weights improve control over execution; they do not necessarily expose training data or make the model’s internal reasoning interpretable.

“Private” describes a controlled relationship. “Local” describes an execution boundary. Neither word, by itself, proves that a scientific method is valid.

A workflow that keeps the researcher present

Governance should be visible at the points where research choices occur. The following workflow is proposed for substantive AI assistance, with controls proportionate to the consequence of an error.

Stage	The system may prepare	The researcher or institution retains
Frame	Candidate questions, related work and alternative explanations.	Research purpose, ethical justification and the distinction between exploratory and confirmatory work.
Retrieve	Evidence with authors, passages, dates and access controls.	Corpus scope, relevance judgements and review of missing or dissenting evidence.
Analyse	Proposed transformations, code, classifications and comparisons.	Method approval, validation criteria and acceptance of consequential changes.
Challenge	Counterexamples, conflicting sources and sensitivity checks.	Interpretation, resolution of disagreements and assessment of uncertainty.
Release	Draft methods, evidence records and declarations.	Publication, external communication, final claims and responsibility.

Four capability lines

• **Evidence discovery:** Navigate approved knowledge and identify where additional searching is needed; do not call retrieval from one Space a complete systematic review.

• **Methods assistance:** Draft protocols, code and analytical alternatives inside a controlled environment; test before use.

• **Research-record stewardship:** Prepare provenance records, identify stale material and trace dependencies when a source changes.

• **Communication support:** Draft summaries and methods descriptions with evidence attached; keep release under named human authority.

Consequential actions need explicit gates

Require approval before altering original data, changing inclusion rules, modifying an approved protocol, committing analytical changes to the accepted record, publishing or communicating externally. Use least-privilege tools and isolated execution for generated code.

Read-only retrieval and reversible draft preparation can operate within pre-approved bounds. Concentrate approval gates on consequential choices so researchers can give those decisions meaningful attention.

Protect the research record and researcher privacy

Retain the prompts, outputs and decisions that materially affect the work under an approved research-record schedule. Separate exploratory scratch work from accepted evidence, restrict access and document deletion rules.

A privacy-first memory design should preserve the method needed for audit under a defined retention schedule. Limit collection, access and retention of private questions to the approved research-record purpose.

The researcher remains present · Visual guide

The system prepares material at each stage; the researcher or institution retains the consequential choices. Downward arrows show the research sequence, while horizontal arrows bring each proposal to human judgement.

STAGE	SYSTEM PREPARES	HUMAN RETAINS
01 Frame	Candidate questions, related work and alternative explanations.	Research purpose, ethical justification and the distinction between exploratory and confirmatory work.
02 Retrieve	Evidence with authors, passages, dates and access controls.	Corpus scope, relevance judgements and review of missing or dissenting evidence.
03 Analyse	Proposed transformations, code, classifications and comparisons.	Method approval, validation criteria and acceptance of consequential changes.
04 Challenge	Counterexamples, conflicting sources and sensitivity checks.	Interpretation, resolution of disagreements and assessment of uncertainty.
05 Release	Draft methods, evidence records and declarations.	Publication, external communication, final claims and responsibility.

Human authority is present throughout the method. Consequential changes and release remain subject to explicit approval.

Responsibility in research

Oxford holds users accountable for AI-generated content and addresses potential breaches through research-integrity procedures. Assess civil, regulatory and contractual liability separately (Oxford research policy[1]).

A proposed responsibility map

Role	Responsibility to assign	Evidence of discharge
Researcher and PI	Validate contributions, supervise use and own scientific claims.	Checked outputs, methods record, decisions and disclosure.
Research leadership	Resource training, review and operation; decide acceptable use by risk.	Approved use cases, owners and escalation routes.
Integrity and ethics	Interpret obligations; review material departures and incidents.	Advice, approvals, amendments and follow-up.
Governance and legal	Assess permissions, contracts, rights, transfers and restrictions.	Assessments, agreements and case-specific decisions.
Computing / operator	Secure the service, preserve configurations and manage change and recovery.	Access tests, releases, incidents and restore tests.
Supplier	Meet agreed security, service, support and contractual obligations.	Contracted commitments and auditable delivery.

Questions for the contract and project review

Assess the allocation of liability, including indemnities, caps, warranties, IP terms, incident duties, support access, termination and record availability.

For sponsored research, patent-sensitive work, clinical information or export-controlled material, obtain case-specific advice before processing. No-training terms alone establish neither a lawful upload nor the absence of disclosure, rights or consent risks.

Local execution reduces some disclosure pathways while preserving the institution’s legal, ethical and contractual obligations. Operating the system adds responsibility for its security and continuity.

When a material error is found

Pause the affected workflow, preserve evidence, identify dependent artefacts and involve the relevant PI and governance teams. Decide on re-analysis, partner notification, correction or other action through the appropriate institutional process.

Accountability needs a person, an assigned duty and an inspectable record. A campus deployment requires explicit responsibility for operation, security and research use.

1. Oxford research policy · https://www.ox.ac.uk/research/support-researchers/research-practice/policy-generative-ai-research

A proposed responsibility map · Visual guide

Each UML-style actor has an assigned expectation and a record that demonstrates it has been met. Associations show the duties and evidence linked to each actor.

Researcher and PI

EXPECTATION

Validate contributions, supervise use and own scientific claims.

EVIDENCE

Checked outputs, methods record, decisions and disclosure.

Research leadership

EXPECTATION

Resource training, review and operation; decide acceptable use by risk.

EVIDENCE

Approved use cases, owners and escalation routes.

Integrity and ethics

EXPECTATION

Interpret obligations; review material departures and incidents.

EVIDENCE

Advice, approvals, amendments and follow-up.

Governance and legal

EXPECTATION

Assess permissions, contracts, rights, transfers and restrictions.

EVIDENCE

Assessments, agreements and case-specific decisions.

Computing / operator

EXPECTATION

Secure the service, preserve configurations and manage change and recovery.

EVIDENCE

Access tests, releases, incidents and restore tests.

Supplier

EXPECTATION

Meet agreed security, service, support and contractual obligations.

EVIDENCE

Contracted commitments and auditable delivery.

Every duty needs an owner and every material decision needs an inspectable record.

A shared research utility

Nature's decentralisation argument invites institutions to shape AI as research infrastructure around their own needs. The article also raises environmental burdens and dependence on subsidised access as reasons to explore alternatives (Nature, 2026[1]).

Start with a research team and a real workload

The relevant local use case is sustained, private knowledge work across a team: frequent evidence retrieval, occasional long analyses and scheduled indexing or batch jobs. A lightly used personal workstation is not automatically a better economic or environmental choice than shared capacity.

Size for concurrent use. Queue heavy jobs and benchmark interactive response times.

Compare equivalent outcomes over a common period

Use a proposed 36-month planning horizon, with sensitivity to utilisation, staffing, renewal prices and equipment life. Compare the full service needed to achieve the same quality and governance requirements for each deployment option.

Cost or constraint	Include in the comparison
Acquisition and continuity	Hardware, installation, replacement, storage, migration and exit.
Ongoing operation	Electricity, cooling, software subscriptions, support, security and staff time.
Research quality	Verification, curation, error remediation and failed-task costs.
Capacity	Actual concurrency, context length, queues, availability and recovery targets.
External use	Approved subscriptions, API use and any hybrid-workflow charges.

Local AI has continuing operating costs. Empathy.AI's supplied proposal includes recurring charges; confirm continuing software and support rights before purchase.

Sustainability must survive measurement

Compare energy per successfully verified task at matched quality, and include idle consumption, utilisation, cooling, hardware life and embodied impacts where data are available. Record the limits of the comparison when cloud-side energy information cannot be obtained.

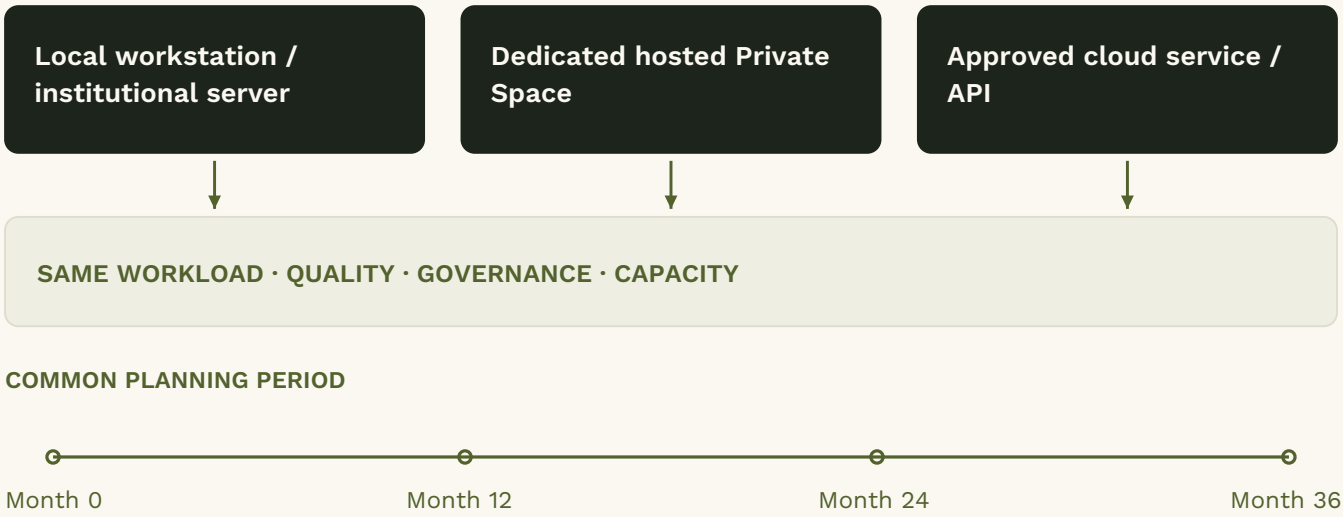
Include the full life cycle in environmental assessments, including solar-powered deployments. Prefer the smallest adequate tool, reuse compute and publish the assumptions behind environmental claims.

Ownership is valuable when it buys control, continuity and useful capacity. Demonstrate those benefits in the institution's workload.

1. Nature, 2026 · https://www.nature.com/articles/d41586-026-02451-2

Compare equivalent outcomes · Visual guide

Compare local, dedicated hosted and approved cloud options over the same proposed 36-month horizon. Match research quality, governance requirements and usable capacity before judging whole-service costs.



COMMON PLANNING PERIOD

FIVE SHARED ASSESSMENT LENSES

Acquisition and continuity	Hardware, installation, replacement, storage, migration and exit.
Ongoing operation	Electricity, cooling, software subscriptions, support, security and staff time.
Research quality	Verification, curation, error remediation and failed-task costs.
Capacity	Actual concurrency, context length, queues, availability and recovery targets.
External use	Approved subscriptions, API use and any hybrid-workflow charges.

Compare whole-service cost at matched outcomes.

Use this framework to compare whole-service costs at matched outcomes. Test utilisation, staffing, renewal-price and equipment-life assumptions before choosing.

A pilot designed to test the strategy

Begin the proposed 12-week evaluation when infrastructure and permissions are ready. Assign a PI, curator, computing owner and governance representatives.

Compare complete research methods

Select two groups with different tasks: for example, evidence synthesis and structured extraction. Use public or appropriately authorised material; sensitive data is not necessary to prove the approach.

Compare a non-generative/manual baseline, the approved cloud workflow and a governed local Space. Use the same tasks and scoring rubric, retain all attempted outputs, and use independent reviewers blinded to the tool where feasible.

Phase	Work	Decision gate
Weeks 1–2	Define scope, permissions, held-out tasks, reference judgements and meaningful failure classes.	Approve the protocol and pre-set success thresholds.
Weeks 3–4	Configure access, corpus, local boundary, versions and evidence export.	Pass access, egress, logging and restore tests.
Weeks 5–8	Run comparisons and test unsupported questions, contradictions and hostile documents.	Stop affected workflows on material leakage or unsafe tool behaviour.
Weeks 9–10	Repeat selected runs, test version changes and complete an independent replay.	Assess stability, provenance completeness and remaining uncertainty.
Weeks 11–12	Review quality, human effort, costs and measured energy.	Adopt, modify, limit or reject each use case.

The scorecard

• **Scientific quality:** Task-specific accuracy and material errors; citation entailment, evidence omissions and appropriate abstention.

• **Researcher agency:** Ability to explain and challenge a sampled output without relying on the chat transcript as authority.

• **Defensibility:** Completeness of the accepted artefact's record and successful reconstruction of sampled workflows.

• **Operational safety:** Permission enforcement, egress controls, incident response and recovery.

• **Efficiency:** Total researcher and reviewer time per accepted result, alongside latency and cost.

Require complete provenance for every accepted pilot artefact, independent human verification of its material claims and no unresolved critical security findings before scaling. Set numerical quality thresholds and sample sizes with the disciplines before testing. These proposed gates are for institutional review and adoption.

A successful pilot identifies the appropriate tool for each task, including conventional software and approved cloud services.

Questions research leaders should ask

Does this imply a ban on ChatGPT Edu?

The strategy supports task-specific selection and controls for approved services. For example, University of Oxford supports ChatGPT Edu use and recognises both its protections and the limitations of cloud-hosted models (University of Oxford research policy[1]).

Is a zero-retention API equivalent to local AI?

No. For example, University of Oxford lists OpenAI API access with Zero Data Retention for highly sensitive research, but the processing remains external and the applicable configuration and terms need checking (University of Oxford developer and research tools[2]).

Assess retention, permission to process the material, model-version control and output validation as separate requirements.

For example, University of Oxford's ChatGPT Edu page currently states that data residency is in the US and advises consultation for sensitive cases. A local deployment changes that data flow, but it does not remove data-protection, confidentiality or research-ethics obligations (University of Oxford ChatGPT Edu[3]).

Can the institution require an exact replay?

Replay requirements depend on what must be reproduced: an exact recorded output, an executable analytical pipeline, a distribution of classifications or a stable scientific conclusion.

Archive actual outputs and relevant dependencies. Local control improves the ability to preserve a setup, but stochastic generation and implementation differences can still affect reruns.

Does a restricted corpus suppress discovery?

It can if its limits are hidden or its curation is one-sided. Make scope visible, preserve dissenting sources and support controlled expansion into new literature.

Use a separate exploration mode for conjecture. Do not let an unsupported suggestion silently become an evidence-backed statement.

Is this only appropriate for experienced researchers?

No, but the design should recognise differing expertise. Doctoral researchers may need explicit mentoring and explanation of methods, while expert users may need efficient inspection of assumptions and exception cases.

Adapt support to the task and the researcher. Provide for accessibility, language needs and legitimate automation while protecting scientific judgement.

What would make the local option unsuitable?

Insufficient task quality, inadequate operating support, unacceptable recovery risk or poor whole-cost performance can all justify another route. Choose a deployment that the institution can maintain and validate throughout its use.

1. University of Oxford research policy · https://www.ox.ac.uk/research/support-researchers/research-practice/policy-generative-ai-research

2. University of Oxford developer and research tools · https://oerc.ox.ac.uk/ai-centre/generative-ai-tools/developer-and-research-tools

3. University of Oxford ChatGPT Edu · https://oerc.ox.ac.uk/ai-centre/generative-ai-tools/chatgpt-edu

A blueprint for institutional action

Institutions should adopt AI tools deliberately, setting the conditions under which they support the production of knowledge.

The education strategy asks how people can relate better with knowledge. For research, the parallel question is how institutions can use AI while preserving the conditions that make knowledge trustworthy: intellectual independence, evidential discipline, methodological openness and accountable interpretation.

Commit to inquiry before output

Judge the system by whether it helps researchers ask better questions, find and challenge evidence, recognise uncertainty and defend a conclusion. Evaluate drafting speed as one contribution to that broader research purpose.

Build the ability to choose

Maintain a local execution option for work that benefits from direct control, alongside approved external capabilities and conventional tools. Avoid making either cloud dependence or local ownership an ideological requirement.

Keep the evidence close to the claim

Use authored knowledge, visible provenance and inspectable decisions to make AI contributions contestable. A researcher should be able to show what the machine produced and why the contribution was accepted.

Preserve responsibility where it belongs

The institution sets the rules, the operator maintains the service, the curator defines the collection and the researcher owns the scientific judgement. The tool should make those responsibilities visible and easier to exercise.

The role for Empathy.AI

Empathy.AI can be positioned as a partner in building that environment: private Spaces, local or dedicated capacity, curated knowledge and governed assistance. Its contribution should be proven through implementation and evaluation, with precise boundaries and no claim that privacy alone guarantees truth.

AI should extend the reach of inquiry while preserving researchers' authorship and their ability to understand and govern the instruments they use.

The immediate recommendation is to approve the bounded pilot, assign its owners and agree its evidence gates. Wider procurement should follow what the pilot demonstrates.

Annex · The AI provenance record

Oxford recommends retaining unedited outputs, dates, versions and uploaded documents for substantive use (Oxford research FAQs[1]). This proposed template extends that record proportionately to the task and data sensitivity.

Record area	Minimum information to preserve
Research identity	Project, artefact identifier, responsible investigator and purpose of the AI contribution.
Authorisation	Applicable ethics decision, data permissions, restrictions and approved processing environment.
System identity	Provider or operator, model identifier and version, run date, deployment boundary; state unavailable details.
Local configuration	Where applicable: weights hash, tokenizer, quantisation, runtime, libraries, hardware and tool versions.
Evidence context	Actual permitted source snapshots, identifiers, authors, versions, corpus scope and access rules.
Retrieval	Query, filters, embedding and reranking versions, selected passages and retrieval configuration.
Run record	Effective prompts where accessible, inputs, parameters, executed tool calls and complete unedited outputs.
Human decisions	Accepted or rejected contributions, consequential changes, decision owner and rationale.
Validation	Reference cases, code tests, source checks, disagreement resolution and sensitivity results.
Lifecycle	Linked downstream artefacts, change history, repository, retention schedule, access and deletion rules.

A declaration that describes the contribution

Suggested wording to adapt to the journal, funder and discipline: “We used [tool, provider, model/version, URL and date] for [specific task] within [processing environment]. Outputs were checked through [validation procedure]; [named role] approved their use. Inputs, outputs and relevant configurations are retained at [controlled repository or access process]. Limitations include [unavailable details and residual uncertainty].”

The auditable record consists of observable evidence, configuration, tool execution and human decisions. Treat generated explanations as outputs requiring verification.

Controlled access to research records

Keep sensitive records in controlled repositories and publish only what is permitted. Explain restrictions and authorised routes for scrutiny.

1. Oxford research FAQs · https://www.ox.ac.uk/research/support-researchers/research-practice/faqs-policy-generative-ai-research

Annex · Evidence boundaries and source notes

This blueprint distinguishes published institutional guidance, external analysis, supplier descriptions and proposed design. That distinction is part of the strategy itself: readers should know which claims they can rely on and which must still be tested.

External foundations

The Nature Comment by Cassidy K. Buhler, Fernando Pérez and Carl Boettiger, published 11 August 2026, provides the local-infrastructure argument mapped on page 8. The supplied PDF ends its substantive excerpt near the beginning of “The agentic future”; this strategy does not reconstruct unseen text, rely on its headline capacity figures or claim a universal environmental advantage (Nature local-AI Comment[1]).

The Nature Perspective by Lisa Messeri and M. J. Crockett, published 6 March 2024, supplies a conceptual framework concerning illusions of understanding and scientific monocultures. Effects in specific research settings require empirical evaluation (Nature research Perspective[2]).

Oxford’s research policy and FAQs are the institutional baseline, checked for this draft on 6 November 2026. The policy expressly addresses responsibility, disclosure, reproducibility and local models; the supplied October data-rules summary is useful context but does not replace the live policy (Oxford research policy[3]).

Supplied Empathy.AI references

• **A Strategy for AI & Education, v12:** Structural and visual reference; the source of the blueprint approach and the ethical emphasis on agency.

• **Artificial Intelligence and Education: Cognitive Debt, Scaffolded Design, and the Case for Locally Governed AI in Schools:** Background for design principles, not evidence of effects on researchers.

• **Human Autonomy vs Machine Automation:** Reference for intermediate choices, meaningful approvals and institution-governed memory.

• **Who We Are · Spaces and Spaces Knowledge-Centric AI Proposal:** Supplier descriptions of the product, deployment choices and operating model.

• **From Feeds to Tools:** Reference for a recognisable machine, authored knowledge, honest limits and successful exit.

• **Knowledge-Centric vs LLM-Centric Experiences:** Illustrative visual reference for the opening comparison, evidence gaps and proposed knowledge network.

Items deliberately not asserted

This document makes no claim of Oxford endorsement, universal cloud illegality, guaranteed absence of hallucinations, zero-carbon inference, automatic research compliance or a proven research outcome from Empathy.AI. It does not transfer school case studies into university research case studies.

Before procurement, verify the current product controls, exact hosting and support arrangements, software and model licences, retention and exit terms, and the full research-specific risk assessment. Policy and service features can change; the institution should retain dated copies of the materials on which its decision relies.

1. Nature local-AI Comment · https://www.nature.com/articles/d41586-026-02451-2

2. Nature research Perspective · https://www.nature.com/articles/s41586-024-07146-0

3. Oxford research policy · https://www.ox.ac.uk/research/support-researchers/research-practice/policy-generative-ai-research

A practical application of Ethical use of AI in Research

The frontier bridge

The frontier bridge combines a knowledge-centric research environment with bounded assistance from a frontier model. The research environment first works with its own evidence, identifies the remaining question and prepares a minimum-context brief for explicit review before an external handoff.

The attached **Frontier Bridge · ia-360** application illustrates this pattern. This case study is based on inspection of its interface, prompts and source code. Live provider testing, a security audit and research-deployment validation remain requirements for adoption. The research example on page 26 is a fictional adaptation of the application's workflow.

What the supplied application implements

Application stage	Mechanism in the supplied code	Research purpose
Work privately in the Space	Ask the selected Space to inspect its documents, name relevant sources and separate private from shareable information.	Establish what the evidence already supports before requesting outside assistance.
Abstract the question	Generate an editable brief, separate the private list of removals, run a browser-side leak check and require approval.	Reduce disclosure and make the outbound question a deliberate research decision.
Ask the frontier model	Show the outbound prompt; provide ChatGPT or Claude handoff and direct provider-API options.	Obtain options, trade-offs and assumptions without automatically sending the original archive.
Bring the answer back	Ask the Space to compare the answer with its documents and propose changes; record accept, partial or reject.	Keep external suggestions provisional until checked against the actual research context.

The local research adaptation

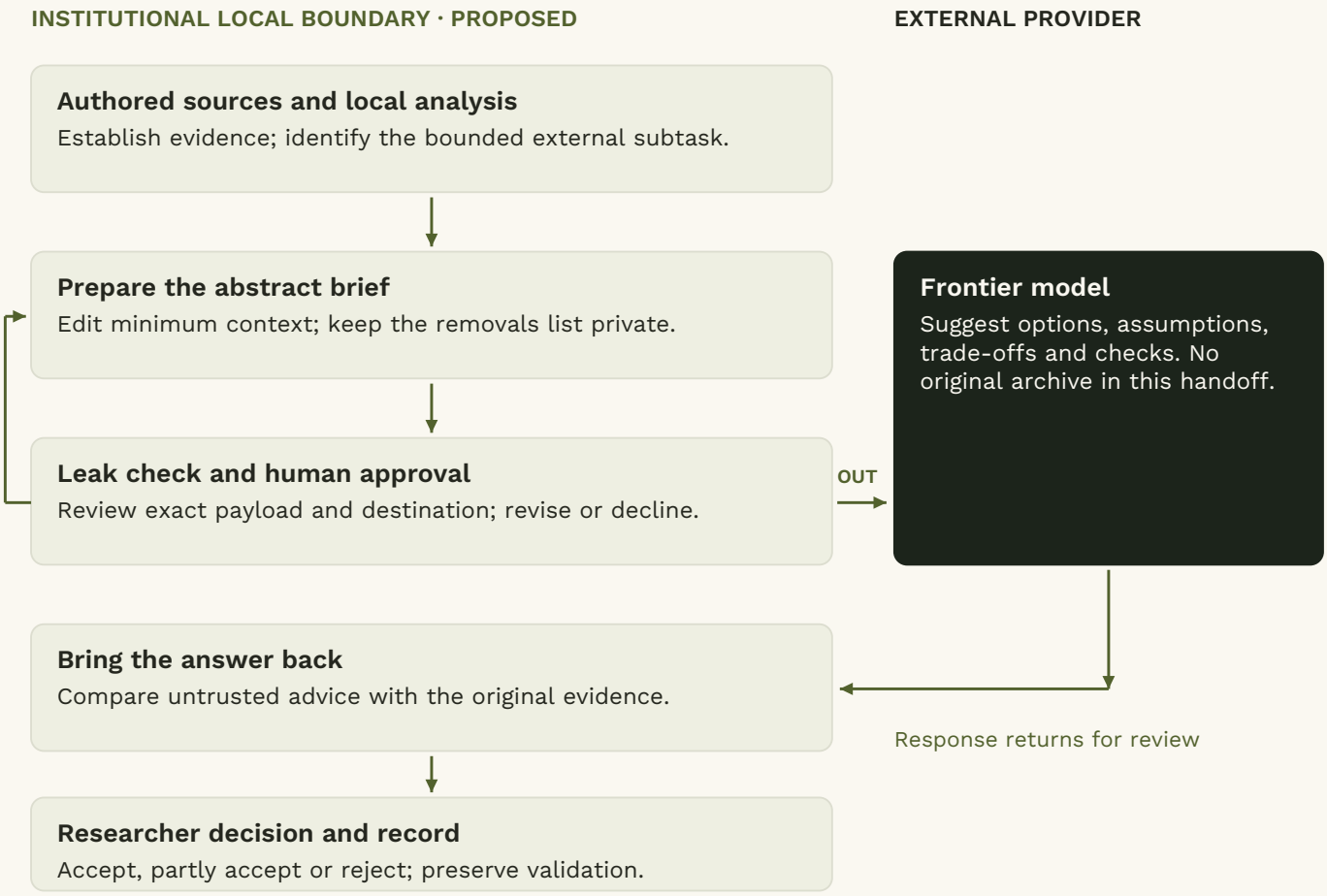
The supplied code sends Space requests to a hosted Empathy.AI Knowledge Engine Agent endpoint. Its processing boundary includes that hosted service. A Local AI Research Experience would need to run or route the knowledge, retrieval and inference services within the institution's approved local boundary and verify all supporting data flows.

The case study illustrates a controlled interaction pattern. A local research deployment would require separate validation of its processing boundary, disclosure controls and research performance.

Prepare a minimum-context question for review. Bring the answer back as a proposal to check against the research evidence.

The frontier bridge · A controlled round trip

The proposed local configuration keeps the research archive and its interpretation inside the institutional boundary. Only a reviewed brief and visible instructions cross to the external model; its response returns for checking before a researcher decides what enters the record.



Approval is a meaningful boundary

The application offers Strict abstraction, replacing names, figures and dates with placeholders, and Balanced abstraction, allowing public entities and published sources while excluding Space-specific text and figures. Both guide the drafting process; anonymisation requires separate assessment and validation.

Its checklist asks the researcher to review identifying details, quotations, synthetic examples, combined identifiability and whether the question stands alone. Approval locks the brief; editing requires reopening approval. High-risk scan findings can be acknowledged and overridden, so the review gate relies on the approver's judgement.

The preview includes a standard instruction wrapper as well as the brief. A research deployment should bind approval to the exact complete payload and destination, and require fresh approval after either changes.

The frontier bridge · A research case study

A methods question without exporting the study

Consider a fictional research team investigating missing measurements in a longitudinal study. Its local Space contains the protocol, analysis plan, codebook and a private summary of missingness. The team wants a critical comparison of analytical options without disclosing participant records, unpublished estimates or the study's identifying details.

The local environment first checks the approved method and identifies a bounded external task: suggest ways to assess sensitivity to missing outcomes. The team retains responsibility for interpreting the actual study data and establishing its findings.

An illustrative brief for researcher approval

Problem: A longitudinal study has repeated outcomes and incomplete follow-up. The team needs to examine how conclusions might depend on assumptions about missing observations.

Constraints: No participant-level data, project identifiers, document extracts, sample sizes or unpublished findings are supplied. The approved research plan remains controlling; any methodological change requires review.

Question: Compare two or three approaches to sensitivity analysis for missing outcomes. Explain their assumptions, trade-offs and diagnostics, identify when each may be unsuitable, and provide dated references that the team can verify.

This brief preserves the type of problem while withholding the particular research instance. The team must still decide whether that combination of structural details is safe to share and whether the remaining context is sufficient to obtain useful advice.

Bring the options back to the evidence

The application's frontier prompt requests options, assumptions, dated citations and checks before application. The researcher must verify the accuracy, currency and validity of any substantive recommendation against original sources and appropriate disciplinary expertise.

The returned answer is brought back into the research environment. Using the actual protocol and data documentation, the local system can identify mismatched assumptions, flag proposed deviations and draft a validation plan. A statistician and responsible investigator then decide whether to accept, partly accept or reject each proposal.

Record the research decision

An accepted proposal might become a separately versioned sensitivity-analysis plan, subject to the project's approvals and independent code checks. Retain rejected options and the reasons for rejection in the decision record. Changes to original data, approved methods or published claims require evidence, validation and approval.

The frontier model supplies candidate approaches. The research team supplies the evidence, validates the method and owns the conclusion.

The frontier bridge · Conditions for responsible use

Scope of the checks

The browser-side scanner flags seven-word overlaps, titles, identifiers, matching figures and some names against retrieved session material. It does not inspect the full archive or establish that paraphrases, distinctive hypotheses or combined details are safe.

Session state lives in browser memory. The visible ledger retains outbound text and send metadata, but the copied record omits exact outbound payloads. Its send entries record actions such as copying a prompt or opening a provider page. Successful provider submission and completion require separate confirmation.

Acceptance requirements for a research deployment

Control	What must be demonstrated before substantive use
Verified local boundary	Local analysis, abstraction, retrieval and return review stay within the approved institutional boundary; hosted services, support and dependencies are explicitly assessed.
Disclosure and permissions	Review the exact payload and destination, source rights, participant commitments and provider terms. Permit declining the bridge and assess residual identifiability after abstraction.
Enforced release gate	Test approval, edits, destination changes, overrides and bypass attempts. Restricted projects may require a second approver or a non-overridable block.
Safe return path	Treat frontier output as untrusted material. Test prompt injection and fabricated references; do not execute code or alter records without review.
Durable provenance	Export complete outbound payloads, instructions, provider/model details, actual outputs, source versions, approval and validation decisions to a controlled repository.
Operational security	Review browser-held API credentials, clipboard and prompt-in-URL handoffs, external scripts and fonts, access enforcement and record retention.

Evaluate the complete bridge workflow

Compare local-only and bridged assistance on the same held-out tasks. Measure validated usefulness, total review time, disclosure failures, appropriate refusal, source support and provenance completeness. Begin with synthetic or public material and include adversarial cases; decline bridging when its benefit does not justify the exposure.

The application illustrates a reviewable external handoff. Ethical compliance requires validation of its controls, with interpretation and acceptance retained within the research environment.

Ethical use is demonstrated through justified disclosure, inspectable evidence and accountable human decisions.