

Rethinking the Research Process

How AI Is Changing Research for Students and Professors

Steven Denney · Leiden University

Korea University · AI Literacy for Social Scientists

18 September 2026

Roadmap: AI in research and teaching

- 1 When to automate, when to protect, when to collaborate?

- 2 See how knowledge bases and reusable skills support research.

- 3 Follow a short demo using a recent project of mine.

- 4 Reflect on takeaways and discuss.

From generative AI to agentic AI

Generative AI AI that creates new text, images, audio, or code.

Large language model A model trained on large collections of text to work with and produce language.

Agentic AI AI that can pursue a task through multiple steps, using tools and responding to results.

This talk focuses on agentic AI, which uses generative models in multi-step workflows to use tools, inspect results, and continue work.

From a conversation to carrying out a task

I mainly use Claude Code and Codex in a CLI.



Terminal

A text-based app for running commands.

CLI

A command-line interface used in a terminal.

Repository

A project folder with files and history.

Tool

An allowed action: search, read, write, or run.

What does it mean to use agents for research?

Software development

Code often has fast, repeatable checks: tests, compilers, or visible outputs.

Research

Research relies on evidence, design, and interpretation. Checks are often slower and less mechanical.

Verifiability

Can we inspect what happened?

Validity

Does the evidence support the conclusion?

An example research claim

“Studies of unilateral policymaking similarly show that citizens penalize executive action relative to legislative action”

Cited together: Reeves & Rogowski (2016); Christenson & Kriner (2017).

An excerpt from an earlier draft of my manuscript.

Do these sources support this claim? Are those even *real* sources?

PART 1

Automate, protect, collaborate

Three modes of AI engagement.

Three modes of AI engagement

Mode	Use it when	Examples
Automate	The work is routine and little is gained by doing it yourself.	Schedule management, filing, formatting.
Protect	Doing the work develops or demonstrates understanding.	Reading, creative thinking. Grading and writing.
Collaborate	AI helps with work you guide and judge.	Verifying sources, data analysis, generating feedback, revisions

AUTOMATE

Cognitively Offload

Efficiency is the goal.

Save time and attention

Schedule	Arrange meetings, maintain calendars, and prepare reminders.
----------	--

Track	Update task lists, milestones, deadlines, and routine status reports.
-------	---

File	Name, sort, organize, and convert project materials.
------	--

Format	Prepare references, tables, slides, and documents consistently.
--------	---

PROTECT

Sandbox and protect

Durable learning is the goal.

Research judgment is a practiced capacity

It develops through repeated decisions about evidence.

Interpret

What does this source
actually establish?



Weigh

How much confidence
does the evidence
deserve?



Argue

What claim can the
evidence bear?

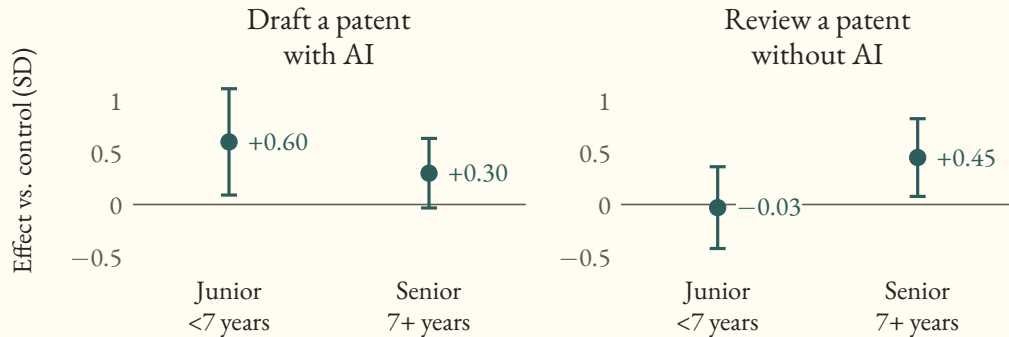
When an agent makes these decisions, it can improve the answer while weakening the capacity to judge it.

Performance and learning need different evidence

Study	What was assessed	Finding
Autor et al. (2026)	Assisted patent drafting and later unaided review.	Senior lawyers improved. The junior effect was unclear.
Bastani et al. (2025)	Assisted practice and a later unaided mathematics exam.	The tutor reduced exam harm. No gain was established.
Bassner et al. (2026)	Programming exercise and learning assessment.	Exercise scores improved. No learning gain was established.

AI assistance and later unaided review

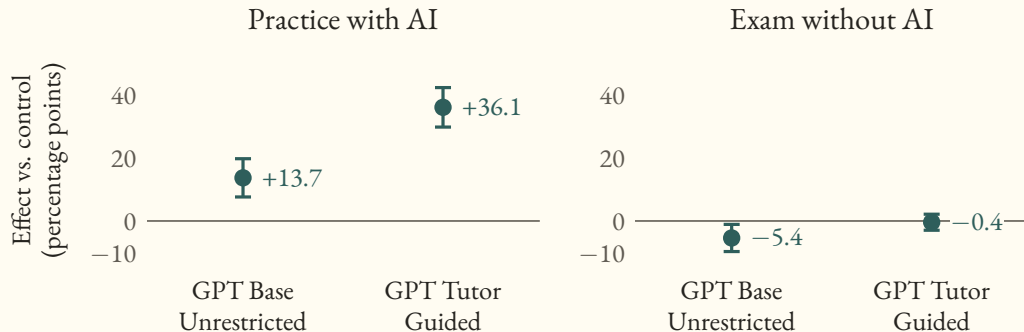
Patent lawyers completed a later patent-review task without AI.



Both groups drafted better with AI, but only senior lawyers improved on unaided review.

Practice with AI, then an exam without it

High-school mathematics. Assisted practice was followed by an unaided exam.



Guided AI raised practice scores, but unaided exam performance did not improve.

COLLABORATE

Combine human judgment with machine
capability

Better, human-machine work is the goal.

Collaboration across research and teaching

Sources

Find literature.
Locate passages.
Compare findings.

Analysis

Organize data.
Write and test code.
Visualize results.

Writing

Structure and revise
drafts. Compare
claims with sources.

Supervision

Helping
augment/structure
feedback and guided
learning.

The machine expands research possibilities. The researcher/teacher remains responsible for what we do and the output.

What might support learning *with* AI?

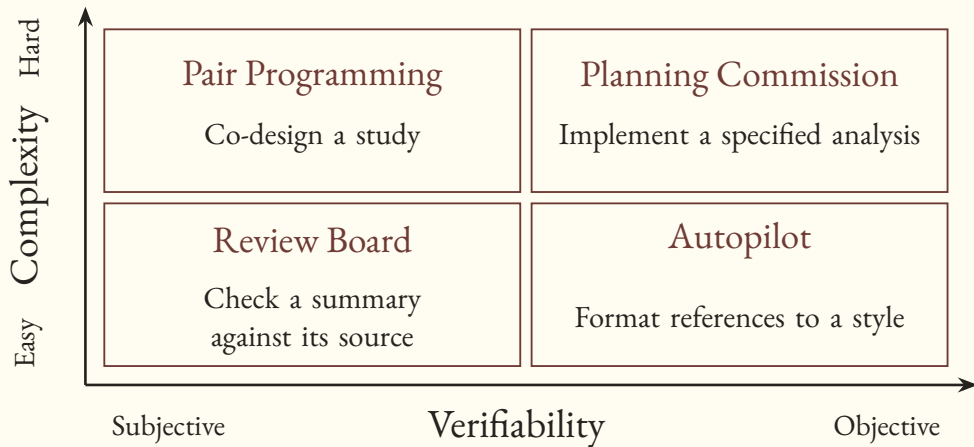
Relevant prior knowledge

Professional seniority is not the same as familiarity with a new task.

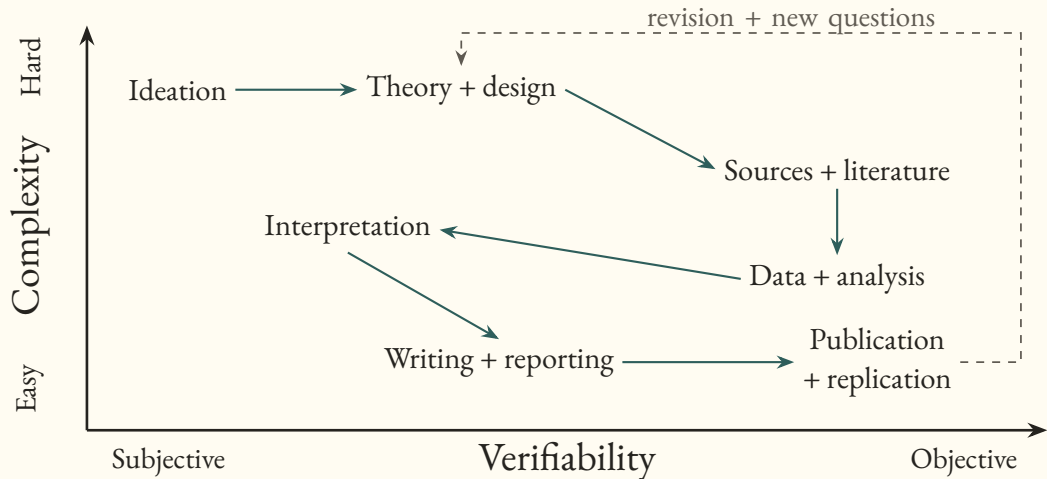
Engagement with the process

Using AI can support learning, but not when it substitutes for engagement with the underlying material.

Match supervision to the research task



Research moves through changing risk terrain



Matching actors to the supervision arrangement

Who plans, who acts, and who reviews?

Arrangement	Plans	Acts	Reviews
Pair programming	Human + agent	Human + agent	Human + agent
Planning commission	Human	Agent	Human
Review board	Agent	Agent	Human
Autopilot	Human instructs	Agent	Agent checks

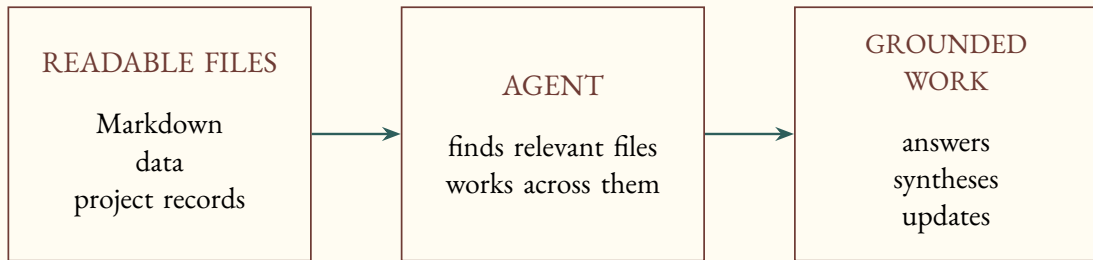
PART 2

Building verifiable AI workflows

From knowledge base to replication.

Building a knowledge base (LLM wiki)

Building a structured file system that an agent can search, read, update, and cite.



The files persist after the LLM session ends.

A folder or repo is not by default a knowledge base

A new project folder

Scattered PDFs

Drafts with unclear provenance

No shared naming rules



A working knowledge base

Readable source files

Stable links to source identity

Instructions for maintenance

The structure must be created and maintained. A great use of agent “skills.”

Repeated and structured task? Make a skill

A skill is a stored procedure for a task you expect to do again.

Instructions What to do and what standard to follow.

Resources Optional scripts, templates, and examples.

Reuse Invoke it when needed; inspect and improve it over time.

What does a skill actually do?

Without

Check whether [claim] is supported.

The model and its agents decide how to do the task.

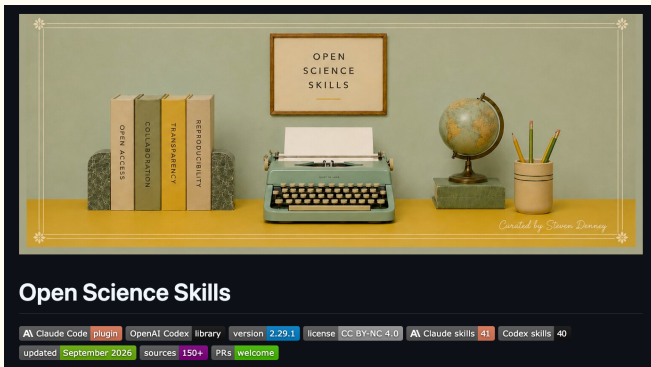
With a skill

`/oss:fact-check manuscript.tex`

The procedure specifies the sources, checks, evidence, and report.

A skill makes the procedure more predictable and reusable.

Open Science Skills



A public library of reusable research procedures for Claude Code and Codex.

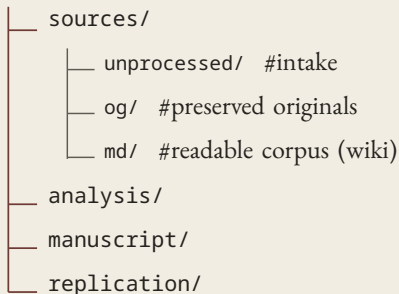
Skills for different parts of the research flow

Knowledge base	research-repo	Preserve and connect the evidence
Source identity	citation-check	Resolve citations and bibliographic records
Claim support	fact-check	Compare manuscript claims with filed sources
Paper and package	paper-review-lite replication-package	Review the paper and execute package checks

research-repo builds the project around your sources

Keep source evidence, analysis, writing, and replication materials in one persistent project.

research-project/



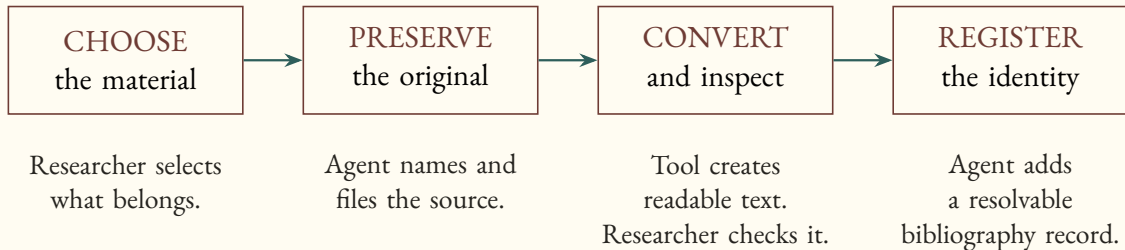
Project repo

Sources, code, drafts, review records, and instructions remain available together.

Project history

Git makes changes recoverable and attributable across sessions. Just use GitHub!

Source collection keeps the researcher “in the loop”



OpenDataLoader PDF creates readable Markdown

Preserved PDF

Abstract: *The formal institutional constraints that Congress and the courts impose on presidential unilateral action are feeble. As a result, recent scholarship suggests that public opinion may be the strongest check against executive overreach. However, little is known about how the public assesses unilateral action.*

Tracked Markdown conversion

Abstract. The formal institutional constraints that Congress and the courts impose on presidential unilateral action are feeble. As a result, recent scholarship suggests that public opinion may be the strongest check against executive overreach.

When a PDF contains figures, extracted image files are relegated to a separate folder with images linked by path.

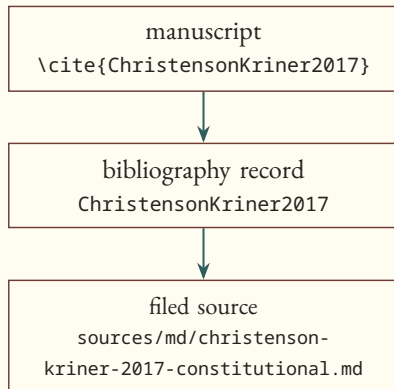
Git connects the knowledge base to writing in Overleaf



The paper and the evidence base can share a versioned project history.

BibTeX keeps source identity connected

```
@article{ChristensonKriner2017,  
  author = {Dino P. Christenson and Douglas L.  
  Kriner},  
  title = {Constitutional Qualms or Politics as  
  Usual?},  
  journal = {American Journal of Political Science},  
  year = {2017},  
  doi = {10.1111/ajps.12262}  
}
```



A citation can be matched to the wrong work

citation-check starts with the keys that actually appear in the manuscript.

CITED KEY	IDENTIFIER	RESOLVED WORK
ChristensonKriner2017	10.1111/ajps.12262	Christenson & Kriner Constitutional Qualms or Politics as Usual?

MATCH The key, bibliographic record, and resolved metadata must match.

fact-check turns prose into a source test

HYPOTHETICAL CASE

MANUSCRIPT CLAIM

“A field experiment increased voter turnout by 12 percentage points” (Author 2025).

FILED SOURCE

“Turnout increased by four percentage points in the treatment group.”

VERDICT

CONTRADICTED

The direction matches.
The magnitude does not.

A valid citation can still support the wrong number.

The reviewers are reviewed too

paper-review-lite checks the manuscript through a verification cascade.

9

specialist reads

argument · numbers · citations
writing · design · figures
preregistration · archive

2

cross-checkers

verify quoted passages
remove false positives

1

pre-submit report

ranked findings
journal readiness

Every critical or recommended finding must point to a quoted manuscript passage.

replication-package prepares a project for deposit

The skill gathers the paper's research materials into a documented replication package.

RESEARCH PROJECT

data/
analysis/
figures/
tables/
manuscript/



1 Find

Materials tied to the paper

2 Build

A separate package structure

3 Document

README, crosswalk,
codebook, and license



replication/

README.md
master.R
data/
code/
docs/
outputs/
LICENSE

Researcher reviews the package, then deposits it
in Dataverse, OSF, Zenodo, or another repository.

PART 3

Follow the evidence

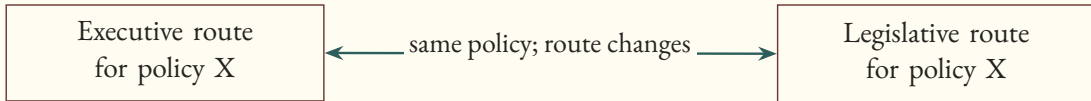
Two mistakes from my own work

What evidence would support this claim?

ARCHIVED DRAFT · Denney, S. (2026), *Governing Immigration by the Rules*

“Studies of unilateral policymaking similarly show that citizens penalize executive action relative to legislative action” (Reeves & Rogowski 2016; Christenson & Kriner 2017).

Sources must compare support for the same policy under executive vs. legislative action.



fact-check uses the knowledge base to check the claim made in the manuscript.

CASE 1

fact-check evaluates each source against the test

The skill retrieves the source, identifies the relevant passage, and tests whether it supports the claim.

Reeves & Rogowski

“Only about a quarter of respondents in any of the surveys supported unilateral policy making”

SUPPORTS Low generalized support

DOES NOT Same-policy route comparison

Christenson & Kriner

The student-loan route “has no significant influence on the probability of a subject backing the president”

SUPPORTS No significant route effect

DOES NOT General executive-route penalty

UNSUPPORTED

Neither source supports the comparative claim as written.

CASE 1

The report recommends. The researcher decides.

fact-check returns a verdict, quoted evidence, and a suggested revision.

SUGGESTED REVISION

Replace the general route-penalty claim with narrower statements that match what each source supports.

RESEARCHER REVISION

BEFORE

Studies of unilateral policymaking similarly show that citizens penalize executive action relative to legislative action \citep{ReevesRogowski2016,ChristensonKriner2017}

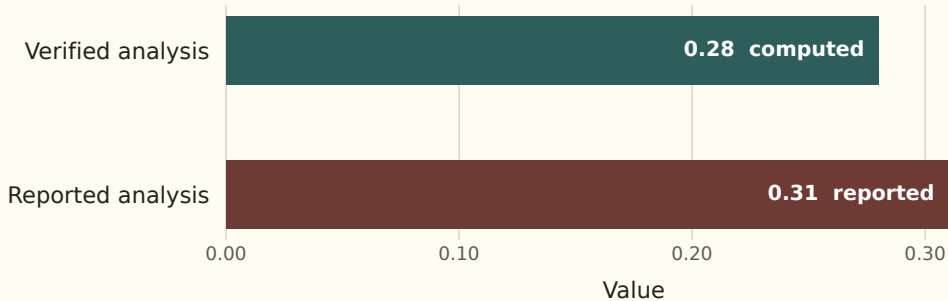
AFTER

Studies of unilateral policymaking find that Americans express low generalized support for unilateral powers \citep{ReevesRogowski2016} while judging particular unilateral acts largely by whether they agree with them \citep{ChristensonKriner2017}

The recommendation is *advisory*. The researcher decides whether and how to revise.

Catching reporting and analysis errors

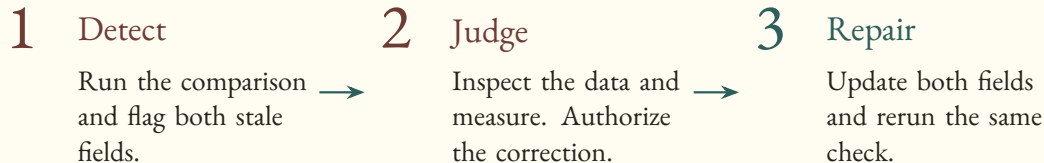
PRE-SUBMISSION REVIEW · paper-review-lite flagged a reported result for verification



FAIL generated 0.28 abstract 0.31 table 0.31

The agent flags the mismatch. The researcher decides.

PRE-SUBMISSION REVIEW



The correction is complete only when the same comparison passes.

PASS generated 0.28 abstract 0.28 table 0.28

Inconsistency detection is highly automatable. Validation remains with the researcher.

Conclusion & Discussion

We're Still Figuring This Out

My takeaways

Fit	Agentic AI is better suited to controlled environments (e.g., software). Research is more brittle and requires repeated, explicit checks.
Verify	Check sources, claims, analysis, reporting, and replication. Passing one check does not validate the next. Verify repeatedly and with agents checking agents.
Preserve	Keep sources, transformations, code, decisions, and revisions connected and recoverable. LLM wikis/knowledge bases are good for research projects.
Disclose	Document who planned, acted, and reviewed; which models, tools, and skills were used; what they accessed; and what was checked.

Use agents to make research verifiable, inspectable, and transparent.

Resources

AI for Research

scdenney.github.io/ai-for-research

Open Science Skills

github.com/scdenney/open-science-skills

Follow-up on Substack

pixelsandpatterns.substack.com

Subscribe for a forthcoming written follow-up and a linked demonstration of setting up a knowledge base.

APPENDIX

References

References 1 of 3

Autor, D. H., Rodchenko, T., Martin, J., Iscenko, Z., Strand, S., Pearl, D., & Ferere, M. (2026). *Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting*. NBER Working Paper 35720. doi:10.3386/w35720.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. doi:10.1073/pnas.2422633122.

Bassner, P., Lenk-Ostendorf, B., Beinstingel, R., Wasner, T., & Krusche, S. (2026). Less stress, better scores, same learning: The dissociation of performance and learning in AI-supported programming education. *Computers and Education: Artificial Intelligence*, 10, 100537. doi:10.1016/j.caeai.2025.100537.

Shen, J. H., & Tamkin, A. (2026). How AI impacts skill formation. *arXiv:2601.20245*. arxiv.org/abs/2601.20245.

Melumad, S., & Yun, J. H. (2025). Experimental evidence of the effects of large language models versus web search on depth of learning. *PNAS Nexus*, 4(10), pgaf316. doi:10.1093/pnasnexus/pgaf316.

References 2 of 3

Reeves, A., & Rogowski, J. C. (2016). Unilateral powers, public opinion, and the presidency. *The Journal of Politics*, 78(1), 137–151. doi:10.1086/683433.

Christenson, D. P., & Kriner, D. L. (2017). Constitutional qualms or politics as usual? The factors shaping public support for unilateral action. *American Journal of Political Science*, 61(2), 335–349. doi:10.1111/ajps.12262.

Denney, S. (2026). *Governing Immigration by the Rules*. Archived manuscript draft.

Kenny, C. T. (2026). *Agentic AI for Political Science Research*. Center for the Study of Democratic Politics, Princeton University, 10 September.

Swaney, C. (2026). *Claude Overload*. Data-Driven Social Science workshop materials, Princeton University.

References 3 of 3

Karpathy, A. (2026). *LLM Wiki*. GitHub Gist, 4 April.

Astro-Han. (2026). *Karpathy LLM Wiki*. Agent Skills-compatible implementation.

MindStudio. (2026). *What Is Andrej Karpathy's LLM Wiki? How to Build a Personal Knowledge Base With Claude Code*. Updated 22 July.

Pocock, M. (2026). *Skills*, including diagnosing-bugs. GitHub repository.

Anthropic. (2026). *Skill Creator Guide*. GitHub repository.