36-MONTH RESEARCH PLAN · INFORMATION SCIENCE (CISE)

Testing Informational Conditions for Meaningful AI Contestability

Kerry Thornhill · Graduate Student, Interdisciplinary Studies · contestability.ai · 12 October 2026 (v. 6) · companion to the argument map "From Operative Determination to Contestable Judgment" and the Paper A and Paper B abstracts sheet

Anuradha and Vikas Sinha College of Artificial Intelligence and Advanced Analytics · University of North Texas · adviser Dr. Chakrabarti

RESEARCH IDEA

Institutions use artificial intelligence to inform decisions affecting access to credit and other opportunities. A stated reason may leave a person unable to investigate inaccurate data, an inappropriate proxy, or unsupported reliance on a recommendation. This project belongs to Information Science within Computer and Information Science and Engineering: it studies how people investigate AI-supported decisions and builds the information system that supports that work. Research on algorithm aversion and appreciation examines when people rely on algorithmic rather than human judgment (Dietvorst et al. 2015; Logg et al. 2019). Contestability additionally concerns whether people can question and influence decisions (Kaminski and Urban 2021; Lyons et al. 2021; Alfrink et al. 2023).

Research question. Which informational functions enable people to investigate AI-supported decisions while preserving warranted reliance, and how can those functions become a usable product across disciplines?

I hypothesize that reliance on labeled advice stays contributory rather than preemptive, valid defeaters reducing reliance while warranted reliance is retained; that information matched to an epistemic task improves selective detection of defects, and that omitting the focal operation and its faithful equivalents leaves investigation below a prespecified adequacy floor; that a faithful substitute can perform the same operation adequately; and that a common process can transfer when local standards, evidence, expertise, and procedures are explicitly adapted.

Paper A, *From Notice to Contestable Grounds: Necessary Informational Conditions for Algorithmic Contestability* (preprint, 2026), develops the Necessary Informational Conditions framework, whose six candidate functions are determination and consequence; sociotechnical roles and evidence control; data, source, and quality; governing-standard–operation fit; evidential warrant, uncertainty, and threshold; and contextual applicability. A companion article in preparation, Paper B, *Private Power without Epistemic Preemption: Relational Institutional Answerability in Algorithmic Decision-Making*, develops, from Lange's distinction between contributory and preemptive deference, the argument that the first experiment tests as reliance on labeled advice. This 36-month project will complete the initial two-function study in year one, test the remaining four functions during years two and three, and deliver a working contestability toolkit with evaluated transfer across disciplines. Findings will determine which functions remain distinct and how the product changes.

APPROACH

1 Phase 1 · Months 1–12	2 Phase 2 · Months 13–24	3 Phase 3 · Months 25–36
Function 3 and one Function 4 operation · corpus, pilots, preregistration · two experiments · first toolkit modules · revision point at 12	Functions 1 and 2 · public-benefits cases · belief/action and staged-access tests · prototype · transfer recruitment · replication slot · revision point at 24	Functions 5 and 6, health-care cases (25–30) · field uptake studies where access is secured (25–36) · revision point at 30 · held-out transfer and replication slot (31–36) · toolkit released at 36

PHASE 1 Months 1–12 Complete the Initial Two-Function Study

The initial study tests data, source, and quality and proxy-to-criterion fit (Jacobs and Wallach 2021), a selected operation within governing-standard–operation fit. Months 1–3 cover institutional review board review and a purposive corpus of public credit notices and disputes, then, on approval, cognitive interviews, task pilots, and preregistration. Inductive coding, applied before the register's labels, will identify objections and epistemic tasks and test whether the register covers them; the preregistered rival, a merger of the third, fourth, and fifth functions, is tested incrementally as those functions enter the program. Months 4–9 cover the two experiments below, the first asking whether independent evidence stays in use under labeled advice and the second whether task-matched information enables defect detection; months 10–12 complete analysis, revision decisions, and the first working toolkit modules.

First, adults will make repeated probability judgments using independent evidence and advice labeled as AI or a performance-matched human expert. Source reliability, cue strength, agreement, and valid reasons to discount advice will vary. Three orders separate sequence from commitment: evidence first with an initial judgment, evidence first without one, and adviser first; evidence-only and adviser-only trials set baselines. Primary outcomes are evidence weights, selective response to valid defeaters, and prediction loss against a prespecified integrated-evidence benchmark. Reduced independent-evidence sensitivity with retained adviser influence and movement toward adviser-only judgment is a candidate pattern of preemptive rather than contributory reliance; agreement alone is insufficient.

Second, adults will evaluate synthetic credit cases containing a record defect, a criterion or proxy defect, or no designated defect. Random assignment compares full task-matched information, omission of the focal operation and its faithful equivalents, a faithful substitute, and a conventional reason. Omission replacements are nondiagnostic; comparison content is matched for length, salience, source, and affordances. A verified record extract may substitute for source-record access; an independent proxy-to-criterion comparison may substitute for a complete model. Case materials will not label a defect, and pretesting will check for unintended clues.

Participants will identify a question, select an evidence request, judge warrant, diagnose defects, and state a chosen disposition. Primary outcomes are request fidelity, diagnostic accuracy, defect sensitivity, no-defect specificity, and performance on the targeted versus adjacent task. A bounded necessity claim requires the full route to exceed a prespecified adequacy floor, selective loss under omission, and controlled false challenges; a substitute below the floor is an inadequate delivery, not evidence about necessity, and substitute noninferiority within a prespecified margin tests only the stronger delivery hypothesis. An omission arm that clears the floor defeats the claim for those conditions. These comparisons test the required operation rather than a particular document.

PHASE 2 Months 13–24 Test Two Further Functions

Months 13–24 test determination and consequence through information identifying the operative determination and its effects, and sociotechnical roles and evidence control through records of who generated, supplied, adopted, or changed evidence, automation's actual contribution, and who controls the grounds. Outcomes are decision identification, intervention attribution, and evidence-request routing. Both modules reuse the omission/substitution protocol with supported, no-defect decisions, cases with unavailable evidence or limited reviewer authority, and adjacent-task controls; synthetic public-benefits cases join credit as a second development setting. Cases with limited reviewer authority, in which participants act as reviewers, separately test whether judged warrant and chosen disposition diverge under institutional constraint. A separate comparison, unable by itself to establish necessity, tests whether staged access preserves investigation at lower burden than a complete dossier. Year two delivers both completed tests, the staged-access comparison, and an integrated inquiry and review prototype, and recruits the independent users and domain advisers for the transfer evaluation.

PHASE 3 Months 25–36 Test the Final Two Functions and Transfer

Months 25–30 test evidential warrant, uncertainty, and threshold by varying validation evidence and uncertainty near a cutoff; contextual applicability tests vary the fit between a case and a model's intended population or setting. Both reuse the omission/substitution protocol, with health-care cases entering as a third development setting. Outcomes include warranted reliance, recognition of inadequate support, and detection of contextual mismatch.

During months 31–36, after the month 30 revision, independent users recruited from education and public administration will adapt the tested modules to synthetic scholarship and permit decisions, settings held out from development, scenario authoring, and pretesting; domain advisers will check local criteria and evidence. Using

documentation, users must configure a case, complete inquiry and review, and record a justified disposition. Measures include accuracy, response to grounds, completion, abandonment, assistance, adaptation time, and components reused or changed. Prespecified criteria will assess whether investigative capability is preserved as local standards, evidence, expertise, and procedures replace those of development, and whether the settings' statutory, professional, and contractual standards remain coherent instances of governing-standard–operation fit; failures will guide revision and delimit supported uses.

Where access to institutional records is secured, observational field studies in months 25–36 will trace delays, withheld grounds, evidence entering review, and correction or warranted reaffirmation, measuring institutional uptake separately from informational adequacy without affecting live decisions.

Product and Feasibility

By month 36, I will release a working contestability toolkit for researchers and process designers: connected affected-party and reviewer interfaces, a case generator, a record linking grounds to evidence and disposition, tested modules, scoring tools, and an adaptation guide. Development accompanies each study. Users will configure and evaluate the complete process from notice through inquiry, evidence, reasoned response, and recorded action, including cases with unavailable evidence or limited reviewer authority.

At the University of North Texas, Dr. Chakrabarti provides methods mentorship and the Sinha College of Artificial Intelligence and Advanced Analytics provides secure computing and software development support; recruitment and compensated domain and accessibility advisers are further needs. Shared infrastructure reduces development time. Simulations will set sample sizes; defects and epistemic tasks specified by independent experts, feasible routes, adequacy floors, hypotheses, outcomes, exclusions, thresholds, and analyses will be preregistered. Hierarchical models will account for participants and cases. At months 12, 24, and 30, I will retain, merge, split, add to, or restrict the scope of functions against the evidence, schedule one diagnostic replication, run in the following phase beside its modules, where an adequate test leaves the omission arm's standing against the floor undetermined (two such nulls across prespecified settings downgrade the bounded claim), and preregister changes. Null results will be reported; failed categorical recovery leaves continuous measures primary.

INTELLECTUAL MERIT

The project tests whether investigative capability, rather than explanation, measures informational adequacy; whether reliance on automated advice stays contributory, and whether belief and action diverge under institutional constraint; whether the six informational functions are distinct and conditionally necessary rather than merely helpful; and whether that capability survives transfer. Omission and substitution comparisons test bounded functional requirements. Cross-disciplinary evaluation distinguishes reusable structure from local criteria and institutional dependencies. Conclusions remain limited to tested tasks and settings.

BROADER IMPACTS

The open toolkit, cases, analysis code, and teaching materials will support research and training across disciplines. Consumer advocates and accessibility testers will shape screen-reader access, keyboard navigation, language burden, and supported use. Reporting will include harmful and unequal effects. The product will expose evidence gaps and authority limits; it cannot confer remedial power. Studies will not affect live decisions.

REFERENCES

Alfrink K et al. 2023. Contestable AI by Design. Minds Mach 33.

Dietvorst B J et al. 2015. Algorithm Aversion. J Exp Psychol Gen 144.

Jacobs A Z and Wallach H. 2021. Measurement and Fairness. Proc ACM FAccT.

Kaminski M E and Urban J M. 2021. The Right to Contest AI. Colum Law Rev 121.

Logg J M et al. 2019. Algorithm Appreciation. Organ Behav Hum Decis Process 151.

Lyons H et al. 2021. Conceptualising Contestability. Proc ACM Hum-Comput Interact 5.

One argument map, three uses: Paper A · Paper B · research plan. Kerry Thornhill · Sinha College of AI and Advanced Analytics, University of North Texas · 12 October 2026 (v. 6)