AN ALGORITHMIC MODEL FOR TECHNICAL, FINANCIAL, AND SCIENTIFIC EXPERTISE IN DIGITAL RESEARCH PROGRAM PLATFORMS

A'zamov Temur Narzullayevich
Advisor to the Director of the Agency for Innovative Development on Digitalization and Artificial Intelligence Implementation, Uzbekistan
https://doi.org/10.5281/zenodo.21903415

Abstract. *Public research-funding competitions combine expert judgment with administrative checks that can be difficult to reconstruct after a decision has been made. This article presents a workflow model that links technical, financial, and scientific expertise within a digital research program platform. Automated checks identify missing documents, rule violations, and scoring conditions, but a named reviewer confirms each decision that changes the application's state. The model specifies workflow states and guarded transitions, criterion-based scoring and aggregation, a trigger for unusually large differences between reviewer scores, schedule alerts, and a composite indicator used to assemble material for council review. The artifact is developed using design-science reasoning. Its main contribution is a traceable process in which state changes are recorded with the responsible actor, time, rule reference, and reason. Operational figures supplied by the platform operator are included as implementation indicators only; they were not produced by a controlled comparison and should be interpreted accordingly.*

Keywords: *research funding, grant review workflow, decision support, score aggregation, process formalization, audit trail*

Introduction

Expert review is central to competitive research funding. In a review of 105 studies, Guthrie et al. (2017) estimated that peer review governs the award of more than 95% of academic medical research funding, and comparable arrangements are used in other fields and jurisdictions. For a funding agency, the quality of the outcome therefore depends not only on reviewers' expertise but also on how the review process is organized.

Research on that process has often concentrated on reliability and bias. Cicchetti (1991) reported low agreement between reviewers across several disciplines; Mutz et al. (2012) found that inter-rater reliability varied widely and was often modest; and Pier et al. (2018) observed little correspondence when experienced reviewers assessed the same set of National Institutes of Health applications. A related literature examines bias and the conditions under which it appears (Lee et al., 2013; Marsh et al., 2008). These findings do not make expert review dispensable, but they do show that reviewer scores can contain substantial variation. For system design, that variation matters because a workflow that simply averages scores can hide disagreement that decision makers may need to see.

Procedural weaknesses create a different problem, especially when part of a competition is handled on paper or through disconnected files. Typical risks include slow hand-offs between units, informal reviewer assignment, inconsistent recording of reasons for returns or rejections, and parallel spreadsheet tracking of deadlines. These are questions about how the process is represented,

controlled, and audited, not about the quality of scientific judgment itself. Workflow engineering provides established methods for addressing problems of this kind (Dumas et al., 2018; van der Aalst et al., 2003).

Digital tools can already support the more mechanical parts of review. Screening systems check formatting, completeness, and originality, while matching algorithms can suggest reviewers for a submission. Checco et al. (2021) survey these developments and also identify an important risk: a model trained on earlier decisions may reproduce the preferences embedded in those decisions. The model developed here therefore treats automated analysis as decision support. A named person remains responsible for confirming and recording the decision.

Two strands of work remain only loosely connected. Studies of review quality usually say little about the way an application moves through an information system, while descriptions of e-government platforms often focus on modules, screens, and services without making the underlying decision rules explicit. When those rules remain only in application code, they are difficult to inspect, challenge, or reimplement.

This study focuses on the expertise stage of a national research program competition. The proposed model links technical, financial, and scientific expertise into one guarded workflow, uses binary decision points so that each branch has a defined outcome, specifies the scoring and reviewer-disagreement rules in implementable form, and reports deployment indicators without treating them as causal evidence.

Method

Research Design

The study uses design-science reasoning as described by Hevner et al. (2004) and Peffers et al. (2007). The research output is a process artifact: a workflow model together with the rules and formal expressions that govern movement through it. Evaluation is conducted in two ways. The analytical part examines the internal consistency of the control logic, including whether defined branches lead to a next state or a terminal outcome. The observational part uses indicators recorded during deployment by the platform operator. Those indicators are treated as evidence of implementation and are not presented as the results of a controlled comparison.

Modeling Notation

The model draws on three established notations. IDEF0 (National Institute of Standards and Technology, 1993) is used to distinguish inputs, controls, mechanisms, and outputs when decomposing functions. IDEF3 (Mayer et al., 1995) informs the description of action sequences and conditional branches. For executable control flow, the article uses binary flowchart decisions consistent with exclusive gateways in BPMN (Object Management Group, 2013) and with the exclusive-choice pattern described by van der Aalst et al. (2003). Keeping these roles separate avoids mixing data, process order, and organizational responsibility in a single representation.

Inputs, Outputs, and Workflow States

The expertise subsystem takes as input the application and its attachments, competition parameters, the round calendar, the classifier of scientific directions, the register of accredited reviewers with their fields and current load, the conflict-of-interest register, the evaluation criteria and their weights, and the scoring scale. Its outputs are a technical conclusion, a financial conclusion, criterion-level scientific scores and comments, an aggregate score, a recorded reason when an application is returned or rejected, and a consolidated package for the scientific and technical council.

Movement of an application is modeled as a finite state machine. Let S be the set of workflow states, E the set of events, and G the set of guard conditions evaluated against application data. The transition function

$$\sigma: S \times E \times G \rightarrow S \qquad (1)$$

A transition is defined only when its guard evaluates to true. In practical terms, this prevents an application from leaving a state through an undefined route or occupying two states at the same time. The model uses the states submitted, technical review, financial review, scientific review, aggregation, additional expertise, returned for revision, forwarded to council, and rejected. Each transition writes the actor, timestamp, module, identifier of the rule that was applied, and resulting state to the record.

Technical Expertise

Technical review begins by checking whether the application is complete and formally admissible. An automated pre-check verifies required documents, the internal consistency of identifiers, signature and format requirements, and compliance with the competition parameters. The result is stored as an analytical attribute and does not change the workflow state by itself. A responsible reviewer is assigned according to field, current load, and conflict-of-interest constraints, then confirms or overrides the pre-check. The state changes only after that confirmation.

Financial Expertise

Financial review uses the same confirmation pattern for the cost plan. The checks cover budget structure, the admissibility of expense items, arithmetic consistency between the budget and calendar plan, and compliance with ceilings for staff costs, equipment, and overheads. Because financial rules are numerous and may change between rounds, they are kept in a versioned rule set outside the application code. A past decision can therefore be reconstructed using the rule version that was in force when it was made.

Scientific Expertise

Scientific review addresses the content of the proposal: the relevance of the problem, novelty of the approach, soundness of the method, credibility of expected results, and fit with the goals of the competition. The platform assembles the reviewer package, applies anonymization when required by the round, sends notifications, tracks deadlines, collects criterion scores and written comments, and aggregates the results. This stage produces graded scores, so it also requires the quantitative rules described below.

Formal Specification of Scoring

Let m be the number of evaluation criteria and n the number of reviewers assigned to an application. Write c_{ij} for the score awarded by reviewer i on criterion j and w_j for the weight of that criterion. The score of a single reviewer is the weighted sum

$$b_i = \Sigma_{j=1}^{m} w_j c_{ij}, \quad \Sigma_{j=1}^{m} w_j = 1 \qquad (2)$$

and the aggregate score of the application is the arithmetic mean across reviewers,

$$R = (1/n) \Sigma_{i=1}^{n} b_i . \qquad (3)$$

Because an average can mask reviewer disagreement, the model evaluates score spread separately. Let $\rho = \max_i b_i - \min_i b_i$, with a tolerance Δ fixed for the round. The rule

$$\rho > \Delta \Rightarrow \text{additional expertise} \qquad (4)$$

When ρ exceeds Δ, one additional reviewer is assigned and R is recomputed over n + 1 scores. In the implementation described here, Δ = 25 on a 100-point scale. The rule can fire only once for an application. If the spread remains above the tolerance after the additional review, the divergence is

noted in the council package instead of starting another review cycle. This restriction ensures that the workflow terminates. When a round contains enough reviewer-level observations, the intraclass correlation coefficient (Shrout & Fleiss, 1979) can also be used as a diagnostic for the criterion set, since persistent disagreement may reflect the measurement instrument as well as individual reviewers.

After expertise, two additional quantities are used. Schedule control compares the actual and planned completion dates for each milestone k,

$$D_k = t_k^f - t_k^p, \quad alert\ if\ D_k > 0 \qquad (5)$$

The council package also includes a composite indicator

$$Q = \alpha R^* + \beta M + \gamma S, \quad \alpha + \beta + \gamma = 1 \qquad (6)$$

Here, R* is the aggregate expertise score normalized to the unit interval, M is a monitoring indicator based on the applicant's previous project record, S is a parameter of scientific effectiveness, and α, β, and γ are weights. Q is used to order and annotate material prepared for the council. It does not determine the council's decision, and the model does not assign an automatic cutoff to Q because the relative importance of its components is a policy choice.

Parameter Setting

Criterion weights w_j are published with the competition documentation before the round opens. The tolerance Δ and the weights α, β, and γ are set by the council and stored as versioned configuration so that a decision can later be reproduced with the parameters used at the time. The article does not claim that these values are optimal; their sensitivity remains to be tested. Figure 1 also shows a score threshold before council transfer. The implementation material available for this study does not specify a single fixed value for that threshold, so no universal threshold is defined or evaluated here.

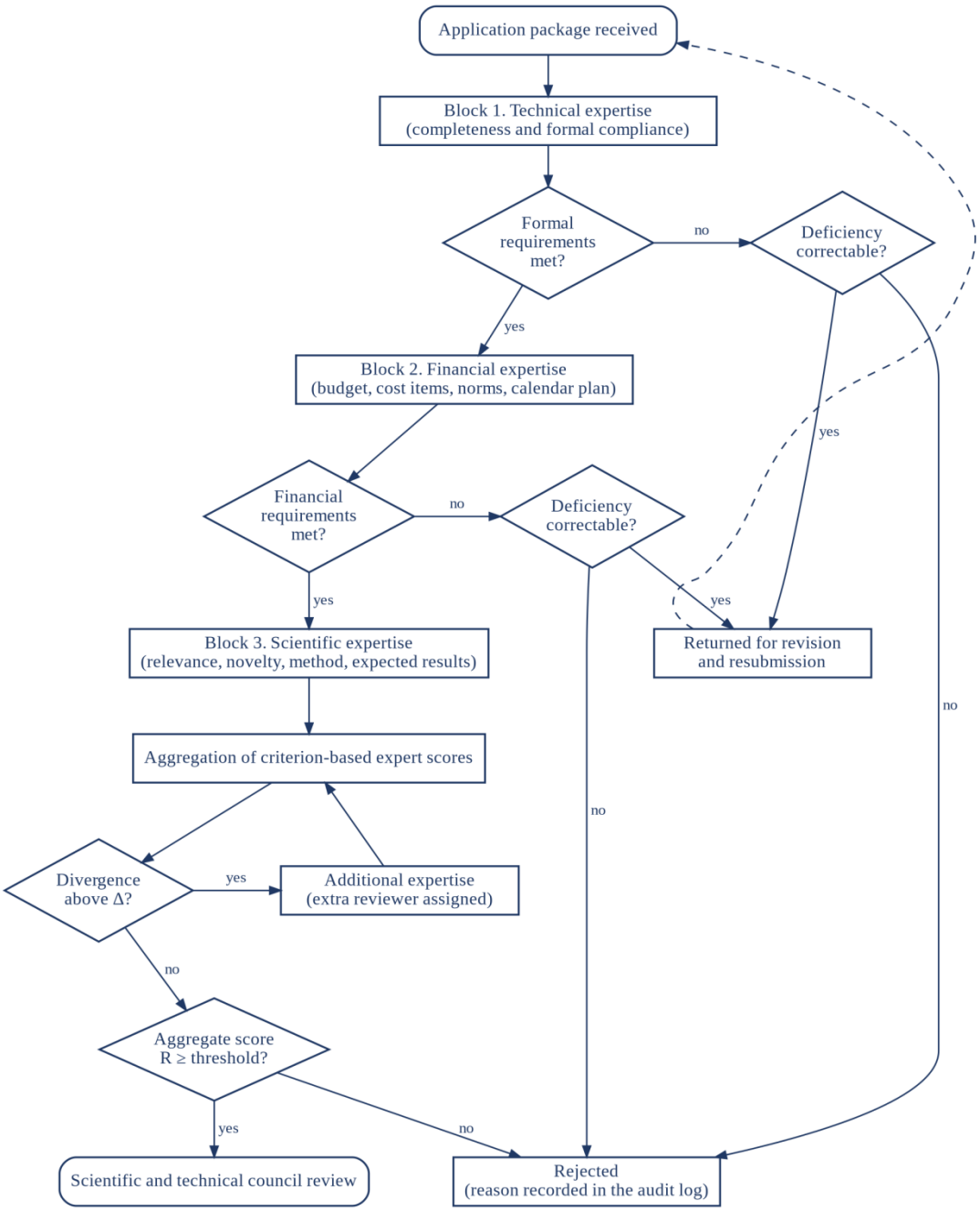
Results

Structure of the Expertise Workflow

Figure 1 presents the technical, financial, and scientific stages as one connected route. The sequence ensures that scientific reviewers receive only applications that have passed the formal and financial checks. Technical and financial review each have three possible outcomes: the file moves forward, is returned for a correctable deficiency with the reason recorded, or is rejected when the deficiency cannot be corrected. Scientific scores are then aggregated and checked for excessive reviewer divergence. After that check, the platform applies the competition's forwarding rule before the application is sent to the council.

Figure 1

The Three-Stage Expertise Workflow From Submission to Council Review



Note. Rectangles denote processing steps, diamonds denote binary decision nodes, and the dashed arrow denotes resubmission after revision. Δ is the tolerated divergence between the highest and lowest reviewer scores. The final score threshold is competition-specific; this article does not assign it a fixed value.

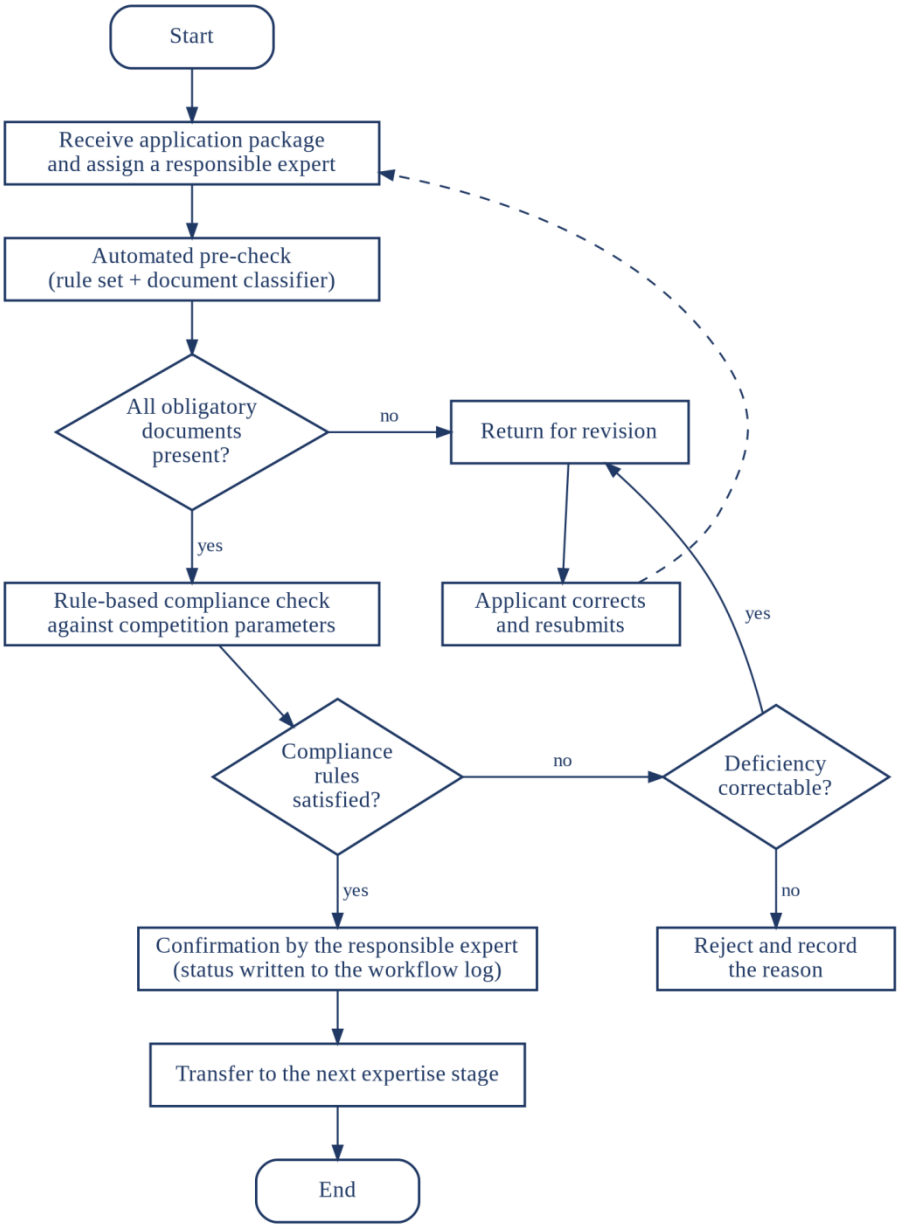
Requirement Checking and Responsible Confirmation

Figure 2 shows the internal logic used in both the technical and financial stages. In each case, the sequence is an automated pre-check, a completeness test, a rule-based compliance test, confirmation by the responsible reviewer, and transfer to the next stage. Every decision node in the specified diagram has two exits, and each exit leads to another defined node, a terminal state, or a return path. This structure reduces ambiguity when the workflow is implemented in software.

Human confirmation is placed after the automated tests and before the workflow state changes. The platform therefore records two separate results: what the automated rules indicated and what the responsible reviewer confirmed. When the two differ, the record preserves that difference instead of allowing the automated result to become an unattributed decision.

Figure 2

Block Scheme of Requirement Checking and Responsible Confirmation Within a Single Expertise Stage



Note. The scheme is instantiated once for technical expertise and once for financial expertise, with different rule sets and different responsible roles.

System Components and Their Methodological Roles

Table 1 summarizes the system components used to implement the model and the methodological role of each component. For example, the audit log does more than display past activity: by recording the actor, time, rule, and outcome, it allows a later reviewer or auditor to reconstruct how a decision was reached.

Table 1

System Components and Their Methodological Role in the Model

Component	Function in the workflow	Methodological role
State register	Holds the current state of every application together with the guards for leaving it	Converts an implicit administrative sequence into a finite set of states with defined transitions
Audit log	Writes actor, time, module, rule identifier, and outcome for each action	Makes a past decision reconstructible, which is the precondition for accountability
Allocation service	Assigns reviewers under field, load, and conflict-of-interest constraints	Replaces informal selection with a stated rule that can be inspected and revised
Scoring service	Collects criterion scores, computes R, evaluates the divergence rule	Separates score calculation from the subsequent decision and keeps reviewer disagreement visible
Integration layer	Exchanges identifiers and documents with registries and financial systems	Preserves the referential continuity of a project across lifecycle stages
Monitoring service	Compares planned and actual dates and raises alerts	Shifts monitoring from retrospective reporting to condition-based signaling

Note. R denotes the aggregate expertise score defined in Equation 3.

Behavior of the Aggregation and Divergence Rules

Consider five reviewers with weighted scores of 82, 78, 75, 71, and 48. Their aggregate score is R = 70.8. The spread is ρ = 34, which exceeds the tolerance of 25, so the application is sent for additional expertise. If a sixth reviewer returns a score of 74, the recomputed aggregate is R = 71.3. The mean changes only slightly, but the council now receives an additional assessment together with a clear indication that the original reviews were unusually dispersed.

The worked case shows why score spread is evaluated separately from the mean. Adding the sixth review changes R from 70.8 to 71.3, yet it gives the council more information about a case in which one assessment differs sharply from the rest. In light of the low agreement reported in controlled studies of grant review (Pier et al., 2018), the divergence rule functions as a safeguard for applications with unusually dispersed evaluations.

Operational Indicators From Deployment

Table 2 reproduces indicators provided by the platform operator for the transition from a paper-based procedure to the digital platform. The operator's implementation documentation also reports a 15.7% acceleration in the introduction of scientific and innovation results into economic sectors and a twofold acceleration in commercialization and sectoral application of scientific developments. These figures are repeated as reported and are not independently validated in this

study. Because the available evidence is not a controlled before-and-after study, the indicators are used to describe implementation and not to estimate a causal effect of the platform.

Table 2

Process Indicators Reported by the Platform Operator Before and After Digitalization

Indicator	Paper-based procedure	Platform	Basis of the figure
Time to submit one application	20–30 min	5–7 min	Operator time records
Duration of the expertise stage	7–10 days	1–2 days	Workflow timestamps
Preparation of a monitoring report	1–2 days, manual	Generated on request	System function
Documents circulating on paper	About 100%	10–15%	Operator estimate
Status visibility for the applicant	On request only	Continuous	System function

Note. Values are as reported by the operator and were not obtained through a controlled before-and-after study.

Discussion

The evidence reported here does not show that digitalization improved the quality of expert judgment. It supports a narrower claim: the platform makes the workflow surrounding that judgment more explicit and easier to reconstruct. A state change is linked to a guard, a versioned rule, confirmation by an identified person, and a log entry. As a result, a later reviewer or auditor can trace how an application moved through the process and which rule was used at each point. The contribution is therefore procedural specification and accountability, not automation of scientific judgment.

This design has implications for disagreement, human oversight, and institutional accountability. First, reviewer disagreement remains visible instead of disappearing inside an average; Equation 4 gives that disagreement a defined procedural consequence. Second, automated pre-checks remain subordinate to confirmation by a named reviewer, which is consistent with the caution raised by Checco et al. (2021) about models that can reproduce patterns in earlier decisions. Recording both outputs also makes systematic differences between automated checks and reviewer confirmations observable. Finally, versioned rules and a complete transition log provide a factual record when a funding decision is challenged or reviewed.

Practical Implications

For agencies considering this design, the largest reported gains in Table 2 occur in clerical activities such as submission, status enquiry, and report preparation. The shorter expertise stage is also notable, although the available evidence does not show how much of that reduction comes from shorter queues, faster hand-offs, or other changes introduced at the same time. The digital record creates an additional analytical opportunity: an agency can measure how often reviewers override

automated pre-checks, in which direction, and under which rules. Those observations would provide a stronger basis for evaluating the rule set itself.

Limitations

The evidence has several limitations. The operational indicators come from one operator and one institutional setting, and they were not collected under a design that isolates the effect of the platform from other changes occurring at the same time. The 15.7% and twofold figures concern outcomes several steps beyond the expertise stage, so the causal link to the workflow is asserted but not demonstrated. Reviewer-level data suitable for inter-rater reliability analysis were not retained for the round from which the model was drawn. The threshold $\Delta = 25$ is an administrative convention, not an estimated optimum, and the same limitation applies to α, β, and γ in Equation 6. Table 2 also does not provide application-level sample sizes or a common observation window, which limits the strength of comparisons based on those figures. Finally, the model assumes that criterion weights are fixed before a round opens and therefore does not address adaptive weighting schemes.

Directions for Further Work

Further work should begin with the data that are currently missing. Retaining reviewer-level scores would allow the intraclass correlation coefficient (Shrout & Fleiss, 1979) to be estimated and would make it possible to set Δ from observed patterns of disagreement within a field. Sensitivity analysis for α, β, and γ could show how strongly the ordering of the council package depends on those weights and whether Q adds useful information beyond R. The platform's event log also provides the input needed for process mining (van der Aalst, 2016). Comparing the process specified in Figure 1 with the process actually executed would reveal where operational practice departs from the model and is a more informative next step than adding further formal detail without new evidence.

Conclusion

Technical, financial, and scientific expertise can be represented as one controlled workflow with guarded state transitions, binary decision points, explicit scoring rules, and confirmation by an identified responsible person. The model does not automate scientific judgment. It formalizes the routing, assignment, timing, and recording that surround that judgment so that the rules can be read, versioned, challenged, and audited. The operator's deployment indicators are consistent with lower clerical and queueing time, but they do not establish a causal effect. The clearest contribution is structural: decisions that were previously dispersed across correspondence and personal knowledge can be recorded in a form that can be reconstructed and evaluated.

REFERENCES

1. Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications, 8*, Article 25. https://doi.org/10.1057/s41599-020-00703-8
2. Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences, 14*(1), 119–135. https://doi.org/10.1017/S0140525X00065675
3. Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). *Fundamentals of business process management* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-56509-4
4. Guthrie, S., Ghiga, I., & Wooding, S. (2017). What do we know about grant peer review in the health sciences? *F1000Research, 6*, Article 1335. https://doi.org/10.12688/f1000research.11917.2

5. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. https://doi.org/10.2307/25148625
6. Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2–17. https://doi.org/10.1002/asi.22784
7. Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist, 63*(3), 160–168. https://doi.org/10.1037/0003-066X.63.3.160
8. Mayer, R. J., Menzel, C. P., Painter, M. K., deWitte, P. S., Blinn, T., & Perakath, B. (1995). *Information integration for concurrent engineering (IICE): IDEF3 process description capture method report*. Knowledge Based Systems.
9. Mutz, R., Bornmann, L., & Daniel, H.-D. (2012). Heterogeneity of inter-rater reliabilities of grant peer reviews and its determinants: A general estimating equations approach. *PLoS ONE, 7*(10), Article e48509. https://doi.org/10.1371/journal.pone.0048509
10. National Institute of Standards and Technology. (1993). *Integration definition for function modeling (IDEF0)* (FIPS PUB 183). U.S. Department of Commerce.
11. Object Management Group. (2013). *Business process model and notation (BPMN) version 2.0.2*. https://www.omg.org/spec/BPMN/2.0.2/
12. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302
13. Pier, E. L., Brauer, M., Filut, A., Kaatz, A., Raclaw, J., Nathan, M. J., Ford, C. E., & Carnes, M. (2018). Low agreement among reviewers evaluating the same NIH grant applications. *Proceedings of the National Academy of Sciences, 115*(12), 2952–2957. https://doi.org/10.1073/pnas.1714379115
14. Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420
15. van der Aalst, W. M. P. (2016). *Process mining: Data science in action* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-49851-4
16. van der Aalst, W. M. P., ter Hofstede, A. H. M., Kiepuszewski, B., & Barros, A. P. (2003). Workflow patterns. *Distributed and Parallel Databases, 14*(1), 5–51. https://doi.org/10.1023/A:1022883727209