

Theses of the Doctoral (PhD) Dissertation

**Gergely Miklós Nagy
Gödöllő
2026**



HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

A Methodological Framework for the Economic and Operational Evaluation of Enterprise AI

**Gergely Miklós Nagy
Gödöllő
2026**

Doctoral school

Name: Doctoral School of Economic and Regional Sciences

Discipline: Management and Business Administration Studies

Head: **Dr. Sente Viktória PhD**
Professor/PhD
Hungarian University of Agriculture and Life Sciences
Institute for Sustainable Development and Management

Supervisor(s): **Prof. Dr. habil. Zéman Zoltán**
Professor/PhD
John von Neumann University
Doctoral School of Management and Organizational Sciences

Dr. Szalay Zsigmond Gábor
Associate Professor, PhD
Hungarian University of Agriculture and Life Sciences
Institute of Rural Development and Sustainable Economy, Department of Agricultural Digitalization and Extension

.....

Approval of Head of Department

.....

Approval of Supervisor

CONTENTS

1. BACKGROUND AND OBJECTIVES	5
2. MATERIALS AND METHODS	9
3. RESULTS AND DISCUSSION	12
4. CONCLUSIONS AND RECOMMENDATIONS	26
5. NEW SCIENTIFIC RESULTS.....	29
6. RELATED SCIENTIFIC PUBLICATIONS.....	32
7. REFERENCES	34

1. BACKGROUND AND OBJECTIVES

In my research, I examined the problem of evaluating AI-enabled IT developments and AI-supported workflows. The starting premise of the research is that the organizational use of artificial intelligence may simultaneously be understood as a technological, economic, operational, risk, and managerial control issue. The study focuses on the conditions under which the use of artificial intelligence (AI) in a given organizational workflow may be regarded as economically justified, operationally stable, acceptable from a risk perspective, and sufficiently supported for managerial decision-making.

According to Stanford HAI's 2026 Artificial Intelligence Index (*AI Index*) report, a widening gap is emerging between the technological capabilities of AI systems and the preparedness of organizations, regulators, and measurement systems to govern, evaluate, and control these capabilities responsibly (Stanford HAI, 2026). This finding is directly related to the research problem. Beyond technological capability, the organizational value of AI also depends on business realization, integration into workflows, human review, risk control, cost attribution, and decision usability.

European enterprise data likewise indicate that the use of AI is becoming an increasingly important organizational phenomenon, while the maturity of adoption differs substantially by enterprise size and organizational setting. According to Eurostat, in 2025, 19.95% of European Union enterprises employing at least ten people used at least one AI technology; among large enterprises, the corresponding share was 55.03% (Eurostat, 2025). These figures show both the growing organizational significance of AI and the unevenness of its adoption. AI use is spreading, but many

organizations are still developing the measurement, governance, and control practices needed for stable operational integration.

A joint 2025 report by the OECD, BCG, and INSEAD also points out that AI's productivity effects are closely linked to organizational conditions, skills, the data and governance environment, and the institutional conditions of implementation (OECD et al., 2025). This approach is particularly important from the perspectives of management science and information systems research because, in addition to model performance, the value of AI is realized through its integration into organizational workflows, its conditions of use, and its operational consequences.

On the basis of the research objectives, the study seeks to answer the following research questions.

RQ1. What patterns by adoption status, and what preliminary project typology for exploratory evaluation, can be derived from an analysis of the E1-E7 case studies of AI-related IT developments and organizational initiatives?

RQ1 compares the workflow characteristics, adoption statuses, and evaluation differences of the E1-E7 case studies.

RQ2. What conventional and AI-specific cost elements and cost-allocation rules can be identified through a combined interpretation of the literature and the case studies examined, and to what workflow-level unit of evaluation can they be linked?

RQ2 examines a literature-based system of cost elements and allocation rules refined through case-specific problems. Particular attention is paid to one-off implementation costs, fixed recurring costs, and variable costs linked to a processing

event or volume, as well as the resources required for human review, exception handling, data-quality improvement, and control maintenance.

RQ3. How can improvements in operational performance, realized economic value, decision usability, customer value, and process value be distinguished on the basis of concepts from the literature and evaluation problems refined through the case studies?

RQ3 is concerned with a conceptual distinction among outcome dimensions, based on starting points from the literature and refined through case-specific problems. It examines how subprocess acceleration, technological activity, and the production of an output can be distinguished from operational, economic, decision, customer, and process value at the level of the end-to-end workflow, and how the absence of an outcome or the rejection of use can be interpreted.

RQ4. What dimensions of routine error burden, human review, control, and residual risk can be identified in an analysis based on the literature and refined through the case studies, and with what evidentiary status can they be interpreted?

RQ4 examines layers of risk and control dimensions with different evidentiary statuses. It distinguishes routine error burdens and human-review burdens apparent across multiple cases, case-specific control problems, and potential material exposures from observed incidents or losses.

RQ5. What preliminary conventional and AI-specific performance indicators can be associated with the four principal dimensions of cost, outcome, risk, and control, and with the cross-cutting perspective of business

usability, through a combined interpretation of the measurement requirements in the literature and the case studies?

RQ5 concerns the systematization of conventional and AI-specific measurement requirements. The study interprets the performance-measurement literature together with recurring or case-specific measurement problems identified in the case studies.

RQ6. What integrated decision-support framework can be developed for the managerial evaluation of AI-supported IT systems?

RQ6 concerns the design-oriented integration of the findings from the preceding questions. The framework is developed from literature-based starting points, concepts refined through the case studies, and the cross-case synthesis.

2. MATERIALS AND METHODS

The use of a qualitative case-study method is justified because the evaluation problems of AI-supported workflows depend on the organizational context.

The methodology is accordingly delimited: the E1-E7 case studies contribute to the development of the evaluation framework as empirical material for problem exploration, all with the same methodological status. A common case protocol, source identification procedure, and coding hierarchy were applied to all cases; differences in adoption status and the nature of the evidence are themselves subjects of analysis. The cross-case synthesis produces preliminary typological, dimensional, and measurement outputs; statistical validation and a full empirical evaluation are separate research tasks.

The unit of analysis is the bounded organizational workflow into which a specified AI application is integrated.

Participants were selected purposively from managers, project managers, and senior experts with direct knowledge of the relevant organizational, technological, and workflow environment. The primary interviewee and information source for each case was the manager identified by role in the case profile.

Data collection was based on semi-structured interviews. This approach provided common principal themes for comparing the cases while allowing case- and context-specific follow-up questions tailored to participants' responses during the interviews.

In preparing the case descriptions, the researcher compared interview notes, case summaries, on-site observations, available documentary and system sources, and public

organizational information. Comparing these sources supported the internal consistency of the case descriptions and analytical findings. The case profiles are summarized in Table 1.

Table 1. The E1-E7 case studies

Case	Organizational or workflow setting	AI-use status	Primary source; analytical relevance
E1	Insurance claims handling; document preparation	initial, limited use	Operations Manager; efficiency and case handling
E2	Regional service center; invoice processing and exception prescreening	limited pilot use	Controlling Manager; cost impact and control
E3	Banking back-office process; customer-due-diligence prescreening	limited use in simpler cases	Risk and Compliance Manager; risk and control
E4	Healthcare administration; document interpretation and data extraction	limited use for selected documents	Senior IT Expert; integration and system constraints
E5	Telecommunications; classification of customer enquiries and case preparation	partial integration in simpler cases	Process Owner; process efficiency and value realization
E6	Business service center; performance evaluation and workforce management	application-specific rejection	HR Manager; acceptance and rejection rationale
E7	Logistics operations; control and disruption management	application-specific rejection	Logistics Manager; operational fit and rejection rationale

Source: author's own compilation.

The evaluation framework was developed through a mixed deductive-inductive, iterative process. The literature provided the initial concepts, dimensions, and design guidance, while the E1-E7 case studies refined, extended, and connected them. The model-development process used in this dissertation comprises four stages: problem exploration, requirements definition, artifact design, and demonstration.

3. RESULTS AND DISCUSSION

In answer to RQ1, the analysis of the E1-E7 case studies identified two patterns by adoption status: limited, experimental, or partially integrated AI use in E1-E5, and application-specific rejection before production use in E6-E7. From the workflow patterns in E1-E5, the same analysis derived five preliminary workflow types for exploratory evaluation. The typology is exploratory: it does not establish the exclusivity, completeness, or stability of the types, nor their automatic applicability to other cases.

E1: document-preparation and processing workflow. The dominant evaluation question is the relationship between subprocess acceleration and improvement at the level of the end-to-end workflow. The evaluation must therefore examine the extent to which the observed acceleration remains after accounting for the burden of human review and correction.

E2: invoice-processing and exception-prescreening workflow. The evaluation focuses on cost attribution, the cost of exception handling, and realized economic value. Economic value is interpretable at workflow level and requires the joint assessment of cost attribution, the exception-handling burden, and the value actually realized.

E3: customer-due-diligence or decision-preparation prescreening. The evaluation therefore focuses on the minimum control set, control coverage, and the residual risk remaining after controls.

E4: document-interpretation and structured-data-extraction workflow. The principal lesson of the case is the need to distinguish technical processing from an output that can be fitted into the process, integrated, and used in the workflow. Input heterogeneity, the need for repeated processing, manual

data completion, and integration problems emerge as key evaluation dimensions.

E5: workflow supporting the classification of customer enquiries and the preparation of business cases. The dominant evaluation question concerned the usability of recommendations and their impact on the customer process. The value of a recommendation can be captured in more accurate case routing, reduced repeat handling, and improved customer or process value.

E6-E7 refined the boundaries associated with adoption status. The E6 human-resources case emphasized measurement validity and fairness, while the E7 logistics case emphasized network consequences and recoverability.

In answer to RQ2, the cost dimension of AI-supported workflows is multilevel and varies by project type. The economic evaluation used for decision support treats costs as workflow-level unit costs and, in addition to technology fees, accounts for the costs of human review, exception handling, data-quality improvement, integration, control maintenance, and rework.

In answer to RQ3, the outcome dimensions of AI-supported workflows must be treated in layers. Technical or usage activity alone provides information only about the operation of the system. Managerial value becomes interpretable in workflow-level outcomes. Decision-support evaluation therefore measures operational acceleration, capacity effects, realized economic value, technical usability, decision usability, and customer and process value separately.

In answer to RQ4, risk assessment of AI-supported workflows requires a layered risk-control logic. Routine error burden, human-review burden, the minimum control set,

qualitative risk, control, and evidence information, and material residual risk are separate evaluation dimensions. A favorable financial or operational signal cannot compensate for failure to meet applicable legal, security, governance, accountability, or minimum evidence requirements; the local consequence and accountable role must be documented in advance.

In answer to RQ5, the principal conclusion of the KPI catalog is that indicators provide decision-support value only when accompanied by standardized descriptive data. For each KPI, the exact definition, data source, owner, update frequency, intended managerial use, interpretive limitation, relationship to the decision framework, and validation status must be recorded. A KPI provides information for a responsible decision but does not determine its outcome.

In answer to RQ6, the AIOVS decision-support system can be constructed by integrating the answers to the preceding research questions.

The concepts required to interpret the results are presented in Table 2.

Table 2. Key concepts

Concept	Meaning
Workflow unit	A business-meaningful, recurring, completed, or transferable workflow unit to which cost, outcome, review burden, and risk can be assigned.
Technical success	A workflow unit meeting the predefined technical-success criterion.
Business-usable output	A transferable unit or output meeting the predefined business-usability criterion.
Variable cost	The mutually exclusive sum of the technical-use, planned-review, and exception-correction costs per initiated workflow unit.

Routine error burden	Incidence of predefined, mutually distinct routine error classes among completed workflow units.
Expected material residual loss	A non-negative loss amount with a documented local basis for the event, probability, consequence, and timing; no general universal formula applies.
Minimum control set	The minimum set of organizational, technical, and process controls under which an AI-supported workflow can be used responsibly.
Residual risk	The exposure remaining after controls that requires a managerial decision in the specific workflow.
AIOVS	Artificial Intelligence Operational Value Scorecard: a conceptual managerial framework supporting the economic, operational, and risk evaluation of AI's operational value.

Source: author's own compilation.

Comparison with the literature indicates that the contribution of AIOVS lies in integrating established mechanisms of value creation, benefit realization, performance measurement, and control within a workflow unit. It organizes the realized outcome, total cost, output usability, human-oversight burden, qualitative risk, control, and evidence information, and material residual risk into separate but connected categories of managerial information. The framework uses a case-specific governance register and does not prescribe a universal mathematical decision rule.

AIOVS is a decision-support framework for evaluating recurring, measurable organizational workflows into which a specified AI application is integrated. Its scope can be defined in situations where costs, outputs, review effort, errors, controls, and risk events can be linked in a meaningful way. The framework is intended for the coordinated evaluation of financial, operational, technical, and risk considerations and

is distinct from general approaches to AI maturity, algorithm quality, compliance, and strategic portfolio evaluation.

The principal components of AIOVS are a single-period financial calculation; usability ratios based on the same nested set of units; directly reported operational indicators; the costs of human review and exception handling; class-specific routine error burden; a qualitative risk, control, and evidence register; the estimability of residual loss; and a case-specific calibration and managerial governance register. The framework makes divergent signals visible but does not automatically aggregate them into a decision.

AIOVS presents mandatory conditions, informative indicators, contextual considerations, local decision thresholds, and the accountable decision separately. Conditions and thresholds must be recorded before the results are interpreted and may not subsequently be adjusted to produce a desired outcome. The locally authorized decision-maker is responsible for determining the action; the framework does not assign a predetermined action to an indicator value.

Table 3 presents the notation used in AIOVS. N/A means not applicable: under the specified conditions, the value is not interpretable; it is neither zero nor a missing measurement. Ft denotes the Hungarian forint. API is the abbreviation for Application Programming Interface.

Table 3. Complete notation used in AIOVS

Symbol	Precise meaning, unit, denominator, horizon, and condition
T	evaluation horizon: A predefined start and end date to which all compared data and cash flows relate. Unless otherwise indicated, every quantity refers to this single common horizon. Date-to-date; no denominator. The

	horizon must be specified before calculation; no zero-denominator rule applies.
u	usable-state marker: Used exclusively in the subscripts of N_u , N_u^r , and v_u to denote a business-usable unit or output; it is not a general unit symbol. Notational role; no independent unit, denominator, or horizon.
B_0	one-off benefit: A non-negative, one-off benefit amount; a credit or refund is a benefit item, not a negative cost. Ft; no denominator; present value as of the first day of the T analysis horizon. No zero-denominator rule applies.
C_0	one-off implementation cost: A non-negative cost amount deducted once by the RNEV formula; it is not part of the variable cost or fixed horizon cost. Ft; no denominator; present value as of the first day of the T analysis horizon. No zero-denominator rule applies.
N_s	number of initiated workflow units: The total number of workflow units initiated on a consistent basis during the common analysis horizon. Count over the full T horizon; no denominator; non-negative integer. If $N_s = 0$, the separate rules for rates based on it apply.
N_t	number of technically successful units: The number of units in the N_s cohort that meet the predefined technical-success criterion; $0 \leq N_t \leq N_s$. Count over the full T horizon; base N_s ; subscript t denotes technical success.
N_u	number of business-usable units: The number of units among N_t that meet the predefined business-usability criterion; $0 \leq N_u \leq N_t$. Count over the full T horizon; conditional base N_t ; subscript u means usable.
q	technical success rate: $q = N_t / N_s$. Dimensionless; denominator N_s . If $N_s > 0$, calculate; if $N_s = 0$, $q = N/A$.
b	conditional business usability: $b = N_u / N_t$. Dimensionless; denominator N_t . If $N_t > 0$, calculate; if $N_t = 0$, $b = N/A$, not zero.
p	direct usability rate: $p = N_u / N_s$. If both relevant denominators are positive, $p = q \times b$. Dimensionless; denominator N_s . If $N_s = 0$, N/A ; if $N_s > 0$ and $N_t = 0$, $p = 0$, and the product $q \times b$ cannot be formed.
N_t^r	number of reviewed technically successful units: The non-negative number of technically successful units actually reviewed in the usability sample; superscript r denotes the

	reviewed sample. $0 \leq N_t^r \leq N_t$. Count in a sample from the full T horizon; denominator of $\hat{b}$ and numerator of N_t^r / N_t coverage. If $N_t = 0$, coverage is N/A.
N_u^r	number of reviewed and usable units: The number of units in the reviewed technically successful sample that meet the business-usability criterion; $0 \leq N_u^r \leq N_t^r \leq N_t$. Count in a sample from the full T horizon; numerator of $\hat{b}$.
$\hat{b}$	sample-estimated conditional usability rate: The estimate calculated from the reviewed sample, $\hat{b} = N_u^r / N_t^r$. Dimensionless; denominator N_t^r . If $N_t^r > 0$, calculate $\hat{b}$; if $N_t^r = 0$, $\hat{b} = N/A$. If $N_t > 0$, coverage is N_t^r / N_t ; if $N_t = 0$, coverage is N/A.
v_u	unit value per usable output: A non-negative benefit amount per business-usable workflow unit; used only on the benefit side $p \times v_u$. Ft/business-usable output; base N_u ; unit value representative of the full T horizon. No zero-denominator rule applies when this is a documented unit input.
c_{tok}	standard technical-use unit cost: A non-negative cost amount covering routine model, tool, API, or technical use; it excludes planned review and additional runs caused by errors. Ft/initiated workflow unit; denominator N_s ; representative of the full T horizon. An average can be calculated from an aggregate only if $N_s > 0$; if $N_s = 0$, the period average is N/A.
c_{ver}	planned-review unit cost: A non-negative cost amount for the predesignated full, sample-based, risk-based, or approval review; subsequent correction of an identified error is c_{exc} . Ft/initiated workflow unit; denominator N_s ; representative of the full T horizon. An average can be calculated from an aggregate only if $N_s > 0$; if $N_s = 0$, the period average is N/A.
c_{exc}	exception-triggered remediation unit cost: A non-negative cost amount for correction, rerunning, rerouting, restoration, or rework caused by a specific error or exception; it is neither c_{tok} nor c_{ver} . Ft/initiated workflow unit; denominator N_s ; representative of the full T horizon. An average can be calculated from an aggregate only if $N_s > 0$; if $N_s = 0$, the period average is N/A.
c_{var}	total variable unit cost: $c_{var} = c_{tok} + c_{ver} + c_{exc}$. All three are non-negative cost amounts; because the primary resource

	categories are mutually exclusive, the same resource may be counted only once. Ft/initiated workflow unit; denominator N_s ; representative of the full T horizon. An average can be calculated from an aggregate only if $N_s > 0$; if $N_s = 0$, the period average is N/A.
F	volume-independent fixed horizon cost: A non-negative platform, support, continuous-monitoring, maintenance, or control-readiness cost over the full horizon; it cannot be multiplied by volume. Ft over the full T analysis horizon; no unit denominator. No zero-denominator rule applies.
EML	expected material residual loss: A non-negative loss amount with a documented local basis for the event, probability, consequence, and timing, at present value as of the first day of the T analysis horizon; no general universal formula applies. Ft over the full T analysis horizon; no unit denominator. If the basis for quantification is incomplete, record 'not estimable,' not zero; no zero-denominator rule applies.
RNEV	net economic value: If $N_s > 0$, $RNEV = B_0 - C_0 + N_s[(p \times v_u) - c_{var}] - F - EML$. If $N_s = 0$, the volume-dependent term is zero by definition, and $RNEV = B_0 - C_0 - F - EML$. RNEV is a signed result. Ft over the full T analysis horizon, at present value as of the first day of that horizon. If $N_s = 0$, p and the average variable costs per unit are N/A; $0 \times N/A$ cannot be formed.
H	local monetary reference amount: An optional, non-negative, locally approved monetary benchmark for the same horizon and present-value date. When used, $RNEV \geq H$; this is not an automatic managerial decision rule. Ft over the full T analysis horizon, at present value as of the first day of that horizon; no denominator.
t_{base}	baseline total cycle time: At unit level, the full end-to-end cycle time; at cohort level, the equally weighted arithmetic mean of the individual workflow units. Minutes/unit or another documented time unit/unit; $t_{base} > 0$; denominator of CCA. The same inclusion and exclusion rules, composition or a compositionally equivalent baseline, and measurement period are required.
t_{AI}	AI-supported total cycle time: At unit level, the full end-to-end cycle time; at cohort level, the equally weighted

	arithmetic mean of the individual workflow units in the AI-supported state. Same unit and scope as t_{base} ; $t_{\text{AI}} \geq 0$; the same inclusion and exclusion rules, composition or a compositionally equivalent baseline, and measurement period are required.
CCA	cycle-time reduction ratio: $\text{CCA} = (t_{\text{base}} - t_{\text{AI}}) / t_{\text{base}}$. At cohort level, calculate from equally weighted unit-level arithmetic means. Dimensionless; denominator t_{base} ; $t_{\text{base}} > 0$ and $t_{\text{AI}} \geq 0$. N/A with a zero or negative denominator.
h_{base}	baseline total human-hours burden: The sum of all human work hours in the comparable baseline for the same scope, segment, volume, and period. Hours over the full T horizon; $h_{\text{base}} > 0$; denominator of HLR.
h_{AI}	AI-supported total human-hours burden: All human work hours in the AI-supported process for the same scope, with planned review and exception correction reported separately. Hours over the full T horizon; $h_{\text{AI}} \geq 0$; same scope as h_{base} ; does not create an additional cost.
HLR	human workload ratio: $\text{HLR} = h_{\text{AI}} / h_{\text{base}}$. Applies to the total human-hours burden for the same workflow scope, segment, volume, and period. Dimensionless; denominator h_{base} ; $h_{\text{base}} > 0$ and $h_{\text{AI}} \geq 0$. N/A with a zero or negative denominator.
j	routine error-class identifier: A predefined routine error category to be reported separately from other classes. Category symbol; no unit or denominator; must be defined before the T analysis horizon. No zero-denominator rule applies.
n_j	number of completed units affected by a class-j error: The number of completed workflow units in which at least one routine error of class j occurred. Count over the full T horizon; $0 \leq n_j \leq N_{\text{cl}}$; numerator of RDB_j .
N_{cl}	number of completed workflow units: The total number of units completed over the full T horizon and included within the scope of the RDB_j calculation. Count over the full T horizon; denominator of RDB_j . If $N_{\text{cl}} = 0$, $\text{RDB}_j = \text{N/A}$.
RDB_j	class-j routine error-burden rate: $\text{RDB}_j = n_j / N_{\text{cl}}$. Report separately for each class; it is non-financial and may not be aggregated into a score. Dimensionless; denominator N_{cl} ;

	over the full T horizon. If $N_{cl} > 0$, calculate; if $N_{cl} = 0$, $RDB_j = N/A$.
--	---

Source: author's own compilation.

The AIOVS financial calculation relates to a predefined evaluation period T, with a start date, end date, workflow unit, currency, and present-value date set at the first day of the T analysis horizon. B_0 and v_u are non-negative benefit amounts; C_0 , c_{tok} , c_{ver} , c_{exc} , c_{var} , F, and EML are non-negative cost or loss amounts. The formula itself supplies the subtraction signs, while RNEV is a signed result.

Net economic value for the given evaluation period T:

$$RNEV = B_0 - C_0 + N_s[(p \times v_u) - c_{var}] - F - EML$$

If $N_s > 0$, the stated RNEV formula applies. If $N_s = 0$, the volume-dependent term is zero by definition, so $RNEV = B_0 - C_0 - F - EML$; p and the average variable costs per unit are N/A, and $0 \times N/A$ therefore cannot be formed. A credit or refund is recorded as a benefit item, not as a negative cost.

If material residual loss is relevant but the local event, probability, consequence, timing, and present value as of the first day of the T analysis horizon are unavailable, EML is recorded as 'not estimable.' In that case, the formula cannot be reported as a fully residual-risk-adjusted RNEV; only the financial subtotal before deduction of EML may be presented.

Variable cost per initiated workflow unit:

$$c_{var} = c_{tok} + c_{ver} + c_{exc}$$

The formula gives the variable cost per initiated workflow unit as the sum of the mutually exclusive unit costs of the

technical run (c_{tok}), planned review (c_{ver}), and exception-triggered correction (c_{exc}).

Each hour of human work, technical run, and cash flow may be assigned to only one primary category. The review event is c_{ver} ; the specific error correction that follows is c_{exc} ; and volume-independent standby capacity is F .

H is an optional, non-negative, locally approved monetary reference amount that relates to the same T horizon and present-value date as RNEV .

When H is used, the direct comparison is $\text{RNEV} \geq H$. This is not an automatic managerial decision rule, and neither a ratio nor a universal rating can be derived from it.

Technical success and business usability based on a nested set of units:

$$q = N_t / N_s$$

$$b = N_u / N_t$$

$$\text{If } N_s > 0 \text{ and } N_t > 0: p = N_u / N_s = q \times b$$

The three indicators are based on the same nested set of evaluation units, but their denominators are not identical: the denominator of q and p is N_s , whereas the denominator of b is N_t . If $N_s > 0$ and $N_t > 0$, p equals the product of q and b .

Under the predefined technical-success and business-usability criteria, $0 \leq N_u \leq N_t \leq N_s$. In sample-based evaluation, $0 \leq N_u^r \leq N_t^r \leq N_t$; if $N_t^r > 0$, $\hat{b} = N_u^r / N_t^r$; otherwise, $\hat{b} = N/A$. If $N_t > 0$, sample coverage is N_t^r / N_t ; if $N_t = 0$, coverage is N/A . If $N_s > 0$ and $N_t > 0$, $p = N_u / N_s = q \times b$. If $N_s > 0$ and $N_t = 0$, $q = 0$, $b = N/A$, and $p = 0$; the product cannot be formed. If $N_s = 0$, q , b , and p are N/A ; the volume-dependent RNEV term is zero by definition; the average

variable costs per unit are N/A; and $RNEV = B_0 - C_0 - F - EML$.

Unit value per usable output:

$v_u = (\text{total documented monetary benefit attributable to business-usable outputs}) / N_u$, where N_u is the number of business-usable outputs.

If v_u is determined as the ratio of the total monetary benefit attributable to business-usable outputs to N_u , then $N_u > 0$. When a documented unit input is used, the zero-denominator rule is not applicable.

Cycle-time reduction ratio:

$$CCA = (t_{base} - t_{AI}) / t_{base}$$

The same inclusion and exclusion rules, the same measurement period, and the same composition (or an explicitly equivalent composition) must be used in both states; $t_{base} > 0$ and $t_{AI} \geq 0$.

If $CCA > 0$, cycle time decreased; if $CCA = 0$, it remained unchanged; and if $CCA < 0$, it increased. CCA alone does not establish throughput, released capacity, financial value, or managerial action. If throughput is relevant, the directly measured number of completed units must be reported for the specified period.

Human workload ratio:

$$HLR = h_{AI} / h_{base}$$

h_{AI} and h_{base} relate to the same workflow scope, segment, volume, and period; $h_{base} > 0$ and $h_{AI} \geq 0$. All human work

hours must be counted, with hours for planned review and exception correction reported separately.

If $HLR < 1$, the total human-hours burden is lower than the baseline; if $HLR = 1$, it is unchanged; and if $HLR > 1$, it is higher. HLR is a non-financial ratio and not an automatic managerial decision rule; the financial equivalent of planned review work is recorded once as c_{ver} , and that of correction triggered by a specific error is recorded once as c_{exc} .

Class-specific routine error burden:

$$RDB_j = n_j / N_{cl}$$

RDB_j measures the occurrence of the predefined routine error class j : n_j is the number of completed workflow units affected by the error, and N_{cl} is the number of completed workflow units in the period examined.

Each predefined routine error category j must be reported separately. n_j is the number of completed units in which at least one routine error of class j occurred during the period; N_{cl} is the number of completed units. If $N_{cl} = 0$, no rate is available.

Existence, design adequacy, operating effectiveness, and evidentiary support are separate statuses. In the absence of evidence, 'not tested' must be recorded. If a local mandatory condition is not met, a favorable financial or operational indicator cannot override it; the previously documented organizational consequence, accountable role, and reassessment point must be applied and justified.

Expected material residual loss (EML) may be used numerically only when supported by a documented local event, probability basis, material consequence, timing, and present value as of the first day of the T analysis horizon.

EML excludes c_{ver} , c_{exc} , F , and the non-financial RDB; elements. Without adequate support, 'not estimable,' not zero, must be recorded; the subtotal before deduction of EML cannot be described as a fully residual-risk-adjusted RNEV.

The purpose of AIOVS is to present financial, operational, technical, usability, control, and risk information separately and transparently. It distinguishes five governance categories: non-compensable mandatory conditions; indicators that inform but do not determine the decision; contextual managerial considerations; locally and prospectively established decision thresholds; and the accountable managerial decision. The categories do not form an aggregate score or an automatic decision rule.

AIOVS is a conceptual managerial framework based on a synthesis of the literature and qualitative case-study evidence; its numerical illustration is hypothetical. It supports structured managerial judgment but does not determine the decision automatically. The locally authorized managerial role is accountable for the final decision and records, in an auditable manner, the evidence, uncertainties, conditions, and reasons considered.

4. CONCLUSIONS AND RECOMMENDATIONS

The practical conclusions of the research provide organizations with a scientifically grounded theoretical framework for designing AI-supported workflows and making them measurable.

First practical recommendation: before beginning the evaluation, the organization must identify the workflow-level unit of evaluation. It must define precisely what counts as an evaluation unit: an invoice, a document package, a customer-due-diligence case, a data-extraction item, or a customer enquiry. Only then can cost, value, review time, and error burden be linked to a common basis of evaluation.

Second practical recommendation: cost calculation must include the costs of human review, exception handling, repeated processing, data-quality improvement, and control maintenance. A project calculation provides a decision-ready picture when it accounts for operating expenditure in addition to license, implementation, platform, and model-usage costs. Actual operational economics are often determined by the amount of human correction, review, or exception handling required to make the output usable.

Third practical recommendation: the organization must evaluate operational outcomes at the level of the end-to-end workflow. Total lead time, actual capacity release, the number of follow-up queries, the repeat-handling rate, the stability of the customer process, and decision usability together reveal how subprocess acceleration is converted into organizational value.

Fourth practical recommendation: in regulated, customer-impacting, safety-critical, or materially exposed workflows, the applicable mandatory conditions, minimum control set,

operating evidence, and allocation of responsibilities must be documented in advance. The existence, design, operation, timing, capacity, and evidentiary support of controls must be evaluated separately. A favorable financial or operational result cannot compensate for a missing mandatory condition.

Fifth practical recommendation: KPIs must be used with standardized descriptive data in decision support. For each KPI, the definition, data source, owner, update frequency, relationship to the decision, and interpretive limitation must be recorded. Descriptive data ensure that the indicators make value realization, controlled risk, and the workflow usability of the output visible.

Sixth practical recommendation: before interpreting results, use of AIOVS requires documentation of the workflow and its unit; the stage of the initiative; the managerial decision being supported; process criticality; the risk-appetite framework; applicable requirements; minimum requirements for evidence and data quality; minimum control, oversight, and accountability; constraints on operational readiness and capability; evidence supporting each condition; the approving role; the consequence of non-fulfilment; the reassessment point; and an auditable justification.

The limitations of the research arise from its case-study design. The seven qualitative case studies E1-E7 support analytical rather than statistical inference and theoretically justified contrastive comparison. The patterns by adoption status derived from the cross-case analysis are not representative; each type in the preliminary typology developed from the cases involving AI use is represented by a single case.

Reliable application of AIOVS requires suitable data sources, cost-allocation practices, review logs, control testing, and a

clear allocation of accountability. Where measurement conditions are incomplete, the framework primarily performs a diagnostic role.

A further direction for research is an external managerial and empirical evaluation of AIOVS. As part of the research, a structured study involving experts could evaluate the framework's comprehensibility, managerial interpretability, logic of application, and practical clarity.

5. NEW SCIENTIFIC RESULTS

T1. I operationalized the AI-specific extension of the workflow-level evaluation approach.

The scientific contribution of the dissertation is the AI-specific extension and operationalization of the workflow-level evaluation approach: it assigns cost, outcome, technical and business output usability, human-review burden, error burden, evidence supporting the operation of controls, and residual risk to the same business-meaningful process unit. Workflow-level analysis builds on antecedents in the literature; the E1-E7 case studies refined the AI-specific problems of units, usability, oversight, and risk. The contribution therefore lies not in creating workflow analysis as such, but in operationalizing it for an AI-supported environment through common denominators and evidence categories.

T2. Through analysis of the case studies, I identified patterns by adoption status and derived, from the cases involving AI use, a preliminary and exploratory project typology comprising five workflow types.

By analyzing the seven case studies E1-E7, the dissertation identifies patterns of limited, experimental, or partially integrated AI use and of application-specific rejection before production use. The analysis of the workflow patterns in E1-E5 produced a preliminary, exploratory project typology comprising five types; E6-E7 refined the boundaries associated with adoption status. Industry, degree of regulation, dominant role, and adoption status are separate boundary conditions.

T3. I demonstrated the need to treat one-off, variable, and fixed costs; resources for human review and exception handling; routine error burden; the minimum control set and evidence supporting the operation of controls; and material residual risk separately.

My scientific contribution is the workflow-level integration and operationalized separation of cost, risk, and control elements known from the literature and refined through the E1-E7 case studies. The cost and risk categories from the literature were refined by the cases' workflow-specific problems of cost attribution, oversight, fairness, network effects, and control; the normative design of AIOVS links them through common rules for units, periods, and events. Separating routine errors, remediation, control costs, and material consequences also guards against double counting.

T4. I highlighted the need for workflow-level integration and operationalized separation of operational improvement, verified capacity release, realized economic value, technical and decision usability, and customer and process value.

The distinction among outcome layers builds on starting points in the literature, while the E1-E7 case studies refined the differences among subprocess acceleration, usable output, capacity effects, decision uncertainty, and the customer process. This layered interpretation prevents technological activity or subprocess acceleration alone from being presented as full business value. The normative integration in AIOVS assigns the layers to the same workflow unit.

T5. I developed the AI-specific KPI catalog.

My scientific contribution is the development of a preliminary and conceptual catalog of conventional and AI-specific KPIs organized by project type and evaluation dimension. Its proposed managerial use assigns each indicator a definition, denominator, data source, owner, decision purpose, limitation, and validation status. Assessment of content validity, reliability, redundancy, discriminant validity, and priority by project type remains a subject for future research. The performance-measurement literature provided the conventional process, financial, and IT indicators and the standardized descriptive logic; the E1-E7 case studies refined the catalog through measurement requirements relating to output usability, human oversight, repeated processing, and class-specific error burden.

T6. I developed the AIOVS decision-support framework.

My scientific contribution is the development of the AIOVS conceptual managerial framework, which integrates and operationalizes, at workflow level, application-related, typological, and dimensional findings based on the literature synthesis and refined through the E1-E7 case studies, together with financial and operational indicators, usability ratios, qualitative risk, control, and evidence information, and a case-specific governance register.

6. RELATED SCIENTIFIC PUBLICATIONS

Journal articles

Nagy, G. M. (2020): Integration of Business Intelligence Systems. *Controller Info*, 2020(3), pp. 12–14.

Nagy, G. M. és Semjén, B. (2020): Az ERP rendszerek bevezetését segítő vezetői megfontolások. *Controller Info*, 2020(4), pp. 26–29.

Szőke, B., Tóth, E. és **Nagy, G. M.** (2020): Önköltségszámítás folyamata egy adott vállalat példáján keresztül. *Controller Info*, 8(2), pp. 48–52.

Nagy, G. M. (2021): A vezetői számvitel, a költségelemzés és a vállalati információs rendszerek jelene és jövője. *Controller Info*, 9(Special issue), pp. 21–28.

Nagy, G. M. (2022): Az info-kommunikációs technológiák által megalapozott decentralizáció hatása a regionális gazdasági fejlődésre. *Controller Info*, 10(4), pp. 33–38.

Nagy, G. M. (2023a): The Management of Risk and Strategy in the Integration of ERP Systems. *Controller Info*, X(Special issue), pp. 53–56.

Nagy, G. M. (2023b): Transforming Business Operations with Large Language Models: Use Cases and Strategies. *Controller Info*, X(Special issue 2), pp. 56–59.

Nagy, G. M. (2026): Mesterséges intelligenciával támogatott vállalati munkafolyamatok integrált értékelése – szakirodalmi áttekintés az üzleti érték, teljesítmény és kockázat kapcsolatáról. *Magyar Minőség folyóirat*, 35(7), pp.

22–45. ISSN 1416-9576 (nyomtatott), ISSN 1789-5510 (online).

Nagy, G. M. és Szalay, Zs. G. (2026): Nyolcszempontú keretrendszer a vállalati MI-bevezetési döntések komplexitásának kezeléséhez. *Acta Carolus Robertus*. Megjelenés alatt.

Conference publication or conference announcement

Nagy, G. M. (2021): The impact of Big Data on market competitiveness. *Aspects of Financial Literacy: Proceedings of the International Scientific and Practical Conference, March 22–23, 2021*. Sumy State University (SSU), pp. 127–136.

Vajda, G., Thalmeiner, G. és **Nagy, G.M.** (2021): Data asset management and representations in a large enterprise environment. *Aspects of Financial Literacy: Proceedings of the International Scientific and Practical Conference, March 22–23, 2021*. Sumy State University (SSU), pp. 378–388.

Nagy, G. M. (2022): The Development of Strategic and Operational IT-Alignment in the Implementation of ERP Systems for Financial Control. *New Aspects of Financial Literacy: Collection of Studies*. Miskolc, Hungary: FINTELLIGENCE Financial Literacy Center, pp. 133–143.

Szakács, A., **Nagy, G. M.** és Szakács, D. (2026): AI és viselkedési torzítások integrációja a turbulens pénzügyi környezetben: *A XVIII. Pécsi Pénzügyi Napok: IV. Pénzügy és Számvitel Nemzetközi Tudományos Konferencia Tanulmánykötete = XVIII. Finance Days in Pécs: IV. Finance and Accounting International Scientific Conference Book of Studies*. Pécs, Hungary: Pécsi Tudományegyetem Közgazdaságtudományi Kar (PTE KTK), pp. 223–241.

7. REFERENCES

- Eurostat (2025): Use of artificial intelligence in enterprises. Statistics Explained. Data extracted: December 2025. Accessed: 22 May 2026. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises
- OECD (2025): Towards a common reporting framework for AI incidents. OECD Artificial Intelligence Papers, No. 34. Paris: OECD Publishing. <https://doi.org/10.1787/f326d4ac-en>
- Stanford Institute for Human-Centered Artificial Intelligence (2026): The AI Index 2026 Annual Report. Stanford, CA: Stanford Institute for Human-Centered Artificial Intelligence. Accessed: 22 May 2026. Available at: <https://hai.stanford.edu/ai-index/2026-ai-index-report>