

AMATIS INSIGHT 003

Who Should Do the Analysis Allocating analytical work between people and AI

Evidence, Strategic Insight and Practical Guidance
for Leaders Redesigning Knowledge Work

**AMATIS
TRAINING**

EXECUTIVE SUMMARY

Generative artificial intelligence can produce summaries, draft requirements, compare options and identify patterns at a speed and cost that changes the economics of analytical work. The important leadership question is no longer whether AI can assist analysis. It is which parts of the work should be assigned to AI, which should remain human-led, and what controls make the resulting arrangement trustworthy.

Evidence does not support a universal answer. Controlled studies have found substantial gains on some bounded writing, customer-support and consulting tasks. The same research also shows uneven results. In the studies reviewed, performance deteriorated in some settings when a task fell outside the system's effective capability or experienced workers had to reconcile outputs with complex local knowledge.

This Insight introduces the Amatis Analytical Work Allocation Model. This original Amatis decision aid assesses an analytical task through four factors: ambiguity, context dependence, consequence and verifiability. It has not been empirically validated and is intended to structure judgement rather than produce a mechanical score. Low-ambiguity work with accessible evidence, limited consequences and reliable checks may support greater AI production. As ambiguity, contextual judgement or consequence increases, the allocation should move towards human-led analysis, stronger challenge and independent assurance. When an output cannot be verified at proportionate cost, apparent speed can create hidden risk rather than value.

The model also reframes the value of analytical capability. We argue that, as the cost of producing plausible analysis falls, organisational value moves towards validating evidence, integrating perspectives, making trade-offs explicit and converting analysis into responsible action. These activities require more than reviewing a finished document. They require access to context, sufficient expertise, time to challenge, and authority to reject or reshape an AI-supported conclusion.

Human oversight is therefore not automatically an effective control. A named reviewer can still defer to an apparently confident output, lack the information needed to detect an error, or approve work under unrealistic time pressure. Effective oversight must be designed as work: with clear accountability, independent evidence, explicit decision thresholds and routes to challenge.

The practical implication is to redesign analytical workflows at task level rather than automate whole roles. Leaders should decide where AI may produce, where it may support, where a human must lead, and where independent assurance is required. The objective is not maximum automation. It is better organisational judgement at an acceptable level of cost, speed and risk.

Central Proposition

When producing analysis becomes cheaper, value shifts towards validating it, integrating it and acting on it responsibly.

KEY TAKEAWAYS

- AI performance is task-dependent. Evidence of productivity on one activity does not establish suitability for another.
- Allocate analytical work using four factors: ambiguity, context dependence, consequence and verifiability.
- The correct unit of redesign is the analytical task and decision, not the job title or document type.
- Human oversight is more credible when the reviewer has capability, evidence, time, independence and authority to intervene.
- Cheaper analytical production increases the relative importance of validation, integration, stakeholder judgement and action.

PUBLICATION METADATA

Publication field	Details
Series	Amatis Insight
Edition	003
Publication stage	Published
Publication date	October 2026 version 1.0
Estimated reading time	18-22 minutes
Intended audience	L&D and People leaders; Heads of Business Analysis; transformation, change and digital leaders; public-sector decision-makers; Business Analysts; Product and Programme leaders
Keywords	Generative AI; analytical work; Business Analysis; human oversight; validation; work design; organisational capability; AI assurance

THE STRATEGIC CHALLENGE

Analysis is more than an artefact

Generative AI makes the visible products of analysis easier to produce. A tool can turn notes into requirements, summarise research, draft a business case, propose process changes or create a polished presentation. Because these outputs resemble familiar professional artefacts, organisations may treat production as evidence that analysis has occurred.

That assumption confuses an output with the work required to justify it. Analysis involves deciding what problem is being examined, whose experience matters, what evidence is credible, which assumptions are being made and what consequences follow from action. A fluent document can conceal weak framing, missing stakeholders or unresolved trade-offs.

This distinction matters because generative systems are optimised to produce plausible responses, not to hold organisational responsibility. NIST identifies confabulation, human-AI configuration and information integrity among the risks that require active management in generative AI systems (Autio et al., 2024). The risk is not confined to an invented fact. A response can be factually tidy and still be strategically wrong because it answers the wrong question or omits decisive context.

The task is a better unit than the role

Debate about whether AI will replace Business Analysts, researchers or policy professionals is too coarse for practical work design. Roles contain different tasks, and those tasks contain different combinations of pattern recognition, evidence gathering, interpretation, negotiation and accountability.

The International Labour Organization's task-level analysis reaches a similar conclusion. It finds broad exposure to generative AI but argues that transformation is more likely than wholesale replacement because most occupations contain tasks that still require human input (Gmyrek et al., 2025). Exposure shows that a tool may affect work. It does not prove that the whole task can be delegated safely or that organisational value will follow.

Task automation means that AI performs a task, or a material part of it. Augmentation means that AI supports a person who retains the task and judgement. Occupational change alters the mix or content of tasks within a role. Organisational outcomes are the realised effects on quality, cost, service, risk or benefit. Evidence in one category should not be presented as proof in another.

A requirements workshop illustrates the point. AI may transcribe the discussion, group comments and draft requirement statements. It cannot reliably infer from the transcript alone whether a quiet stakeholder has withheld a concern, whether a proposed rule conflicts with an informal operating practice, or whether agreement reflects genuine alignment. Those judgements depend on context and relationships that may never appear in the prompt.

WHAT THE EVIDENCE TELLS US

Research provides credible evidence that generative AI can improve performance on some knowledge tasks. Noy and Zhang (2023) found that professionals completing mid-level writing tasks with ChatGPT worked faster and produced higher-quality outputs on average. Brynjolfsson, Li and Raymond (2025) found productivity gains among customer-support agents, with larger benefits for less experienced and lower-performing workers.

The evidence is not a general productivity guarantee. Dell'Acqua et al. (2026) studied 758 Boston Consulting Group consultants completing tasks designed to resemble management-consulting work. Participants using AI completed more work, more quickly and at higher quality on tasks inside the technology's effective frontier. For a task outside that frontier, AI users were more likely to produce an incorrect answer. The authors describe a jagged frontier: apparently similar tasks can sit on different sides of the system's capability boundary.

Productivity Evidence Requires Qualification

Other studies reinforce the importance of context. A non-peer-reviewed randomised study of experienced open-source developers working on mature repositories found that access to early-2025 AI tools increased completion time by 19%, despite participants believing that AI had made them faster (Becker et al., 2025). The sample was small and the result should not be generalised to all software work. A February 2026 follow-up could not establish a reliable current effect estimate because participation, task selection and time measurement created serious bias. Its raw results pointed towards improvement with later tools, but the authors described the evidence as weak (Becker et al., 2026).

An NBER working paper reports a randomised field experiment across 66 firms. Assignment to AI access produced a modest reduction in email time, with larger reductions among treated workers who used the tool, but no detectable change in meetings or the overall composition of work (Dillon et al., 2025). Some authors were Microsoft employees when the study examined Microsoft 365 Copilot; NBER records that the authors retained discretion over the results. The ILO's 2026 review similarly concludes that measured gains are uneven and that reported time savings have not consistently translated into organisational output or earnings (Merola et al., 2026). Individual speed and organisational value are related, but they are not the same measure.



The question is not whether AI can produce an answer. It is whether the organisation can know when that answer is good enough to use.



VERIFICATION CAN ABSORB THE SAVING

AI can lower the cost of a first draft while increasing the amount of material that must be checked. This creates verification work: tracing claims to sources, testing assumptions, reconciling contradictions and deciding whether omissions matter. When checks are reliable and inexpensive, this may be a good trade. When they require scarce expertise or extensive stakeholder engagement, the apparent saving can disappear.

Verifiability is therefore different from confidence. A polished explanation may make an output easier to accept without making it easier to prove. A reviewer needs observable evidence, a benchmark or another dependable route to confirmation. Government guidance on AI assurance stresses testing against clear performance benchmarks and building quality assurance throughout the AI lifecycle (Department for Science, Innovation and Technology, 2024).

Knowledge workers surveyed by Lee et al. (2025) described a similar change in practice. Generative AI shifted critical-thinking effort towards information verification, response integration and task stewardship. The study was based on self-reported examples and cannot establish long-term cognitive effects, but it supports a practical point: the analytical work does not disappear. Part of it moves downstream.

HUMAN OVERSIGHT IS NOT A CONTROL BY DEFAULT

Organisations often respond to AI risk by requiring a human to review the output. That safeguard sounds sensible, but its effectiveness depends on how the review is designed. Decades of human-factors research show that people can misuse automation through over-reliance, or reject it after unreliable performance (Parasuraman & Riley, 1997).

More recent experiments show why the phrase human in the loop is insufficient. Sele and Chugunova (2024) found that participants preferred algorithmic recommendations and were less likely to intervene when those recommendations contained larger errors. Bucinca, Malaya and Gajos (2021) found that interventions requiring people to think before seeing or accepting AI advice reduced over-reliance, although participants preferred the less demanding designs.

Interpreting the Evidence ▶

Evidence is not uniform. Alon-Barkat and Busuioc (2023) did not find a simple, general automation-bias effect across their public-sector experiments. One study found stronger adherence when advice aligned with stereotypes, but the effect did not differ significantly between algorithmic and equivalent human-expert advice. Their civil-servant study did not reproduce the selective-adherence finding. Human behaviour around AI varies with task, presentation, prior belief and institutional setting. Oversight should be tested rather than assumed.

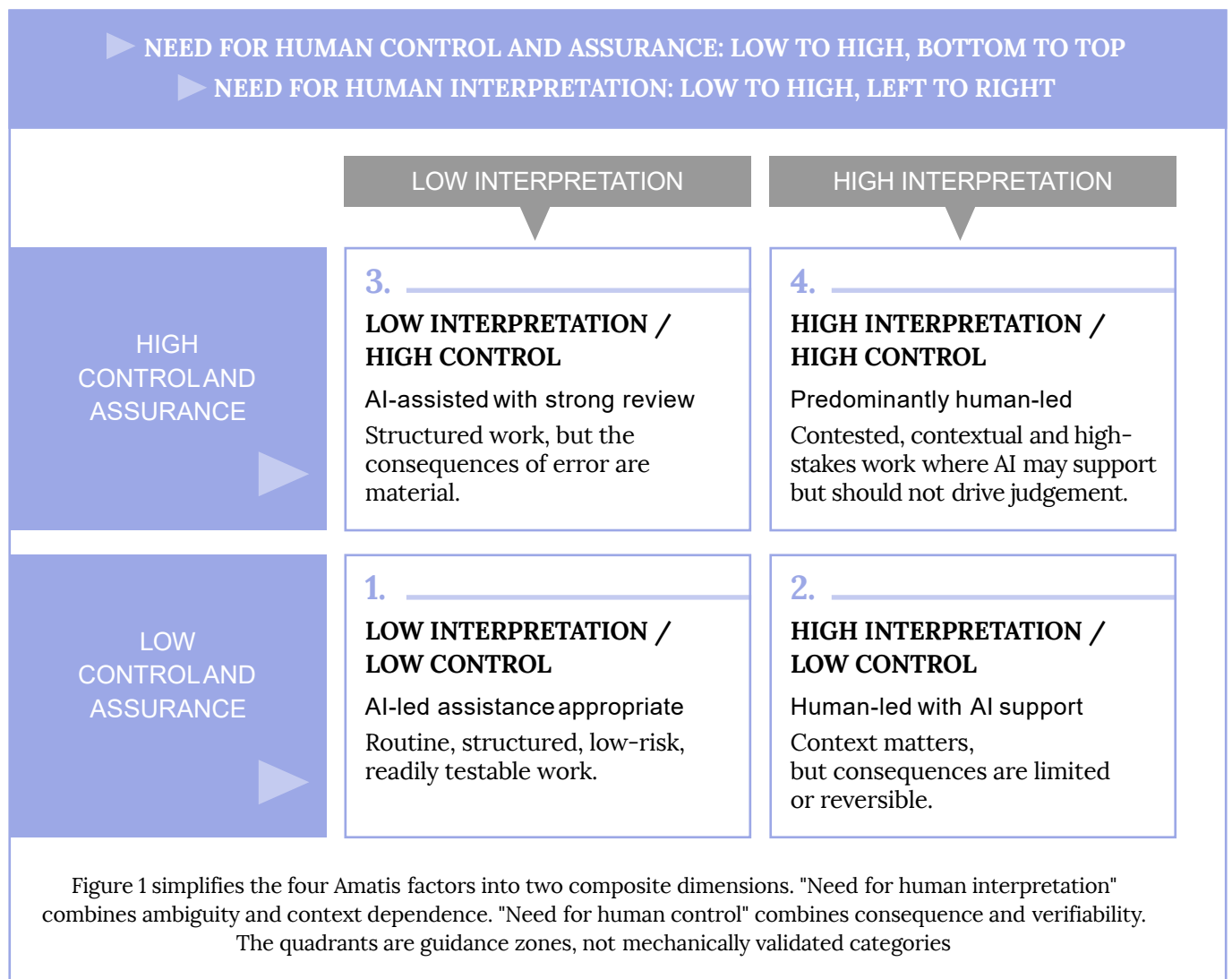
Why Approval Can Fail

A reviewer may fail because they lack domain knowledge, access to source material, time, confidence or authority. They may also be checking hundreds of outputs from a system that is usually right. In that setting, a signature can create the appearance of accountability without a realistic opportunity to exercise it.

THE AMATIS INSIGHT

Analytical work should be allocated according to the characteristics of the task, not enthusiasm for the technology or a blanket rule that every output must be human-approved. The Amatis Analytical Work Allocation Model uses four factors to make that choice visible. It is an original synthesis of research and professional practice, not an empirically validated framework.

Figure 1. The four-factor assessment within the Amatis Analytical Work Allocation Model. Original Amatis conceptual model; not empirically validated (Source: Amatis Insight).



THE FOUR FACTORS

- **Ambiguity:** How settled are the problem, terms, evidence and success criteria? From clear question, stable rules and bounded output to contested problem, incomplete evidence or changing criteria.
- **Context dependence:** How much meaning depends on local history, relationships, tacit knowledge or stakeholder interpretation? From context is explicit, accessible and transferable to essential context is tacit, sensitive, distributed or politically contested.
- **Consequence:** What happens if the analysis is wrong, incomplete or misapplied? From errors are reversible, visible and limited in impact to errors may affect rights, safety, public value, major investment or trust.
- **Verifiability:** Can the output be checked reliably, independently and at proportionate cost? From sources, rules, tests or benchmarks make checking straightforward to there is no dependable ground truth, or checking requires scarce expertise and extensive investigation.

The factors are not a mechanical score. They structure a decision and expose the reason for it. A task may be low in ambiguity but high in consequence. Another may be easy to verify but heavily dependent on confidential context that cannot be shared with the tool. The allocation should respond to the pattern, not an average rating.

FOUR ALLOCATION MODES

Figure 2. Four allocation modes in the Amatis Analytical Work Allocation Model.
Original Amatis conceptual model; not empirically validated. (Source:Amatis Insight)

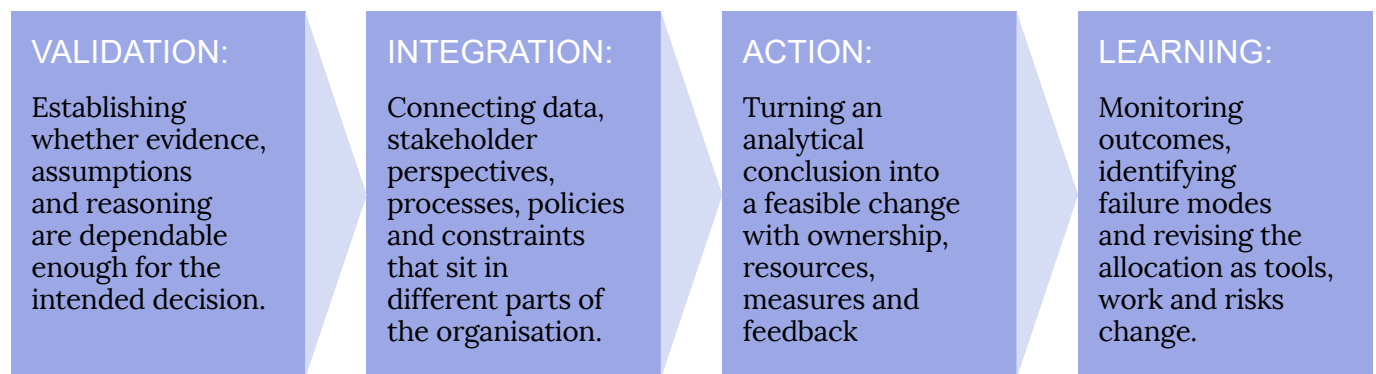
Allocation mode	Typical conditions	AI contribution	Required control
1. AI-produced	Low ambiguity and context dependence; limited consequence; high verifiability	Produce or transform the output at scale	Defined rules, automated tests, sampling and clear exception handling
2. AI-produced, human-validated	Bounded task with moderate consequence and a strong route to verification	Create the first output and surface evidence for review	Reviewer checks sources and criteria, records exceptions and can reject the output
3. Human-led, AI-supported	High ambiguity or context dependence; material trade-offs; consequence is significant	Summarise, generate alternatives, test consistency and challenge assumptions	Human frames the problem, engages stakeholders, integrates evidence and owns the judgement
4. Human-led with independent assurance	High consequence combined with weak verifiability, contested context or irreversible effects	Use only for constrained sub-tasks where benefit is demonstrable	Independent review, documented evidence, escalation, redress and authority to stop or redesign

Mode 1 does not mean unsupervised deployment without accountability. It means that the task's rules and evidence allow assurance to be built into the workflow. Mode 4 does not require excluding AI from every activity. It limits AI's role in the judgement itself while allowing carefully controlled support, such as formatting, retrieval from an approved corpus or comparison against explicit criteria. Independent assurance cannot create dependable ground truth. Where the core judgement cannot be verified at a proportionate level, AI may need to be restricted or excluded from that judgement, with monitoring, escalation and redress retained.

Consequence is the escalation factor. As possible harm increases, organisations should demand stronger evidence before moving towards AI production. Verifiability is the feasibility factor. If an output cannot be checked reliably, assigning it to AI may transfer work from production to uncertain and expensive review.

VALUE MOVES DOWNSTREAM

The Amatis proposition is that, when an organisation can generate more analyses, reports and options, production becomes less scarce. Attention, judgement and organisational capacity do not. The scarce work moves towards four activities.



From a Business Analysis perspective, this shift changes where practitioners may add most value. Producing a requirements document is not the highest-value activity if a tool can create a credible first version. The greater contribution may lie in testing whether the requirement reflects a real need, tracing its effect across the operating model, resolving stakeholder conflict and defining how value will be measured. These disciplines are not limited to people with a Business Analyst job title.



A human signature is not evidence of effective oversight. Oversight must be designed as a capable, resourced and challengeable activity.



PRACTICAL IMPLICATIONS

1. Redesign workflows at task level

Begin with the decision or outcome the analysis must support. Break the workflow into tasks such as evidence collection, classification, interpretation, option development, stakeholder validation and approval. Apply the four factors to each task. A single workflow may legitimately use all four allocation modes.

This avoids two common errors: automating a whole role because some tasks are suitable, and excluding AI because one part of the work is high risk. It also makes dependencies visible. If AI increases the volume of draft outputs, the organisation may need more validation capacity rather than fewer analysts.

2. Define verification before scaling production

For every AI-produced output, specify what acceptable evidence looks like, who can check it and how exceptions will be handled. Suitable mechanisms may include source tracing, comparison with approved data, rules-based testing, stakeholder confirmation, independent sampling and outcome monitoring.

Using the same model to review its own response may identify obvious inconsistencies, but it is not independent verification. High-consequence work needs evidence outside the generation process. The UK Government's Data and AI Ethics Framework similarly links oversight with named responsibility, the ability to intervene and verification of large-language-model outputs (Government Digital Service, 2025).

3. Design meaningful human oversight

As an Amatis design recommendation, meaningful review requires attention to five conditions: relevant capability, access to evidence, protected time, independence from the pressure to approve, and authority to intervene. Weakness in any of these conditions can make review ceremonial.

Workflow design can also prompt active judgement. Requiring an initial assessment before showing the AI recommendation, asking the reviewer to identify disconfirming evidence, or routing uncertain cases to a second reviewer can introduce useful friction. These controls should be tested for effectiveness and workload rather than treated as universal solutions.

4. Protect the development of analytical capability

Some studies have found larger benefits for less-experienced workers on bounded tasks, which may broaden access to useful support. The organisational risk is that junior staff receive fewer opportunities to practise the underlying reasoning while senior staff become responsible for checking a growing volume of outputs.

Learning and development should therefore cover more than tool operation. People need practice in problem framing, evidence quality, source checking, stakeholder analysis, systems thinking, uncertainty and ethical judgement. Organisations should preserve developmental tasks that build the capability needed for future validation and accountability.

5. Measure value beyond production speed

Time saved on drafting is useful but incomplete. Leaders should also measure correction rates, rework, stakeholder acceptance, decision quality, realised benefits, incidents and the effort required for assurance. An output produced in minutes may still be expensive if experts spend hours proving whether it can be trusted.

Measurement should also test whether time is genuinely released. Evidence from workplace experiments suggests that improvements in individual activity do not automatically change coordinated work or organisational outcomes (Dillon et al., 2025). Managers must decide how saved capacity will be used and redesign surrounding processes where necessary.

6. Treat allocation as a governance decision

The chosen allocation mode should be documented with the reasoning behind it, the responsible owner, the permitted data, the verification method and the trigger for reassessment. Changes in model capability, task context, regulation or consequence may alter the appropriate mode.

From a Business Analysis perspective, established disciplines can connect this governance to the operating model. Practitioners in different roles can map affected stakeholders, clarify decision rights, define requirements for evidence and redress, trace benefits and examine how the change affects people, organisation, process, information and technology. The IIBA standard treats business analysis tasks as contextual, iterative and subject to verification and validation (International Institute of Business Analysis, 2022).

A PRACTICAL ALLOCATION CONVERSATION

Table 1. A practical allocation conversation. Source: Amatis Insight.
Original Amatis practical tool.

Conversation step	Leadership question
1. Name the decision	What decision or action will the analysis influence, and who remains accountable?
2. Break down the work	Which tasks collect, transform, interpret, challenge, integrate and approve information?
3. Apply the four factors	Where are ambiguity, context dependence and consequence high, and how strong is verifiability?
4. Select the mode	Should AI produce, produce for validation, support human-led work, or be constrained to subsidiary tasks?
5. Design assurance	What evidence, tests, review capacity, escalation and redress make the arrangement credible?
6. Review in use	Are benefits realised, are errors detected, and has the task or technology changed enough to alter the allocation?

QUESTIONS FOR LEADERS

- Which analytical tasks are we assigning to AI, and have we assessed the task rather than the job title or final artefact?
- Where does essential context sit, and can the people or systems doing the work genuinely access it?
- What is the consequence of a plausible but wrong output, including effects on people absent from the decision?
- Can the output be verified independently at proportionate cost, and have we included that cost in the business case?
- Do reviewers have the expertise, evidence, time and authority required to challenge the system?
- As production becomes cheaper, are we investing in validation, integration, action and future analytical capability?

CONCLUSION

Generative AI changes analytical work because it can produce useful drafts, comparisons and interpretations at low marginal cost. That capability is important, but it does not make every analytical task suitable for delegation.

The evidence points to variation rather than a universal productivity effect. Results depend on the task, the surrounding context, the capability of the user and the organisation's ability to verify what has been produced. The same tool can accelerate one activity and weaken another.

The Amatis Analytical Work Allocation Model gives leaders a practical way to respond. As an original decision aid rather than an empirically validated framework, it structures the judgement without replacing it. Ambiguity and context dependence indicate how much human judgement the task requires. Consequence indicates the strength of evidence and control required. Verifiability determines whether AI-produced work can be checked reliably and economically.

This model does not defend human work from technology or promote automation as an objective. It allocates responsibility according to the characteristics of the work. The Amatis proposition is that, when analytical production becomes cheaper, validation, integration and responsible action become more valuable.

Organisations that respond only by adding a human approval step may create reassurance without control. Those that redesign the whole analytical workflow can use AI where it adds value, preserve human judgement where it matters and build assurance that is capable of finding error rather than merely signing it off.

REFERENCES

- Alon-Barkat, S., & Busuioc, M. (2023). Human-AI interactions in public sector decision making: Automation bias and selective adherence to algorithmic advice. *Journal of Public Administration Research and Theory*, 33(1), 153-169. <https://doi.org/10.1093/jopart/muac007>
- Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.600-1>
- Becker, J., Rush, N., Barnes, B., & Rein, D. (2025). Measuring the impact of early-2025 AI on experienced open-source developer productivity. *Model Evaluation & Threat Research*. https://metr.org/Early_2025_AI_Experienced_OS_Devs_Study-paper.pdf
- Becker, J., Rush, N., Cunningham, T., Rein, D., & Mahamud, K. (2026, February 24). We are changing our developer productivity experiment design. *Model Evaluation & Threat Research*. <https://metr.org/blog/2026-02-24-uplift-update/>
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. *The Quarterly Journal of Economics*, 140(2), 889-942. <https://doi.org/10.1093/qje/qjae044>
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), Article 188. <https://doi.org/10.1145/3449287>
- Dell'Acqua, F., McFowland, E., III, Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Kraymer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science*, 37(2), 403-423. <https://doi.org/10.1287/orsc.2025.21838>
- Department for Science, Innovation and Technology. (2024). Introduction to AI assurance. <https://www.gov.uk/government/publications/introduction-to-ai-assurance/introduction-to-ai-assurance>

- Dillon, E. W., Jaffe, S., Immorlica, N., & Stanton, C. T. (2025). Shifting work patterns with generative AI (NBER Working Paper No. 33795). National Bureau of Economic Research. <https://doi.org/10.3386/w33795>
- Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K., & Troszyński, M. (2025). Generative AI and jobs: A refined global index of occupational exposure (ILO Working Paper 140). International Labour Organization. <https://doi.org/10.54394/HETP0387>
- Government Digital Service. (2025). Data and AI ethics framework. <https://www.gov.uk/government/publications/data-ethics-framework/data-and-ai-ethics-framework>
- International Institute of Business Analysis. (2022). The business analysis standard. <https://www.iiba.org/knowledgehub/the-business-analysis-standard/>
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Article 1121, 1–22. <https://doi.org/10.1145/3706598.3713778>
- Merola, R., Ernst, E., Samaan, D., del Rio-Chanona, M., & Teutloff, O. (2026). The impact of GenAI on jobs, productivity and work organization: A review of the empirical evidence. International Labour Organization. <https://doi.org/10.54394/00034628>
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. <https://doi.org/10.1126/science.adh2586>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
- Sele, D., & Chugunova, M. (2024). Putting a human in the loop: Increasing uptake, but decreasing accuracy of automated decision-making. PLOS ONE, 19(2), e0298037. <https://doi.org/10.1371/journal.pone.0298037>

RELATED INSIGHTS

Amatis Insight 001 – Business Analysis Capability: The Foundation of Successful Public Sector Transformation.

Amatis Insight 002 – Why AI Projects Fail Before They Begin

Insight 002 examined the organisational conditions required before substantial AI investment. Insight 003 continues that argument inside the workflow: once AI is available, organisations must decide how analytical responsibility should be allocated and assured.

ABOUT AMATIS INSIGHT

Amatis Insight is an evidence-led publication series from Amatis Training. It examines important questions at the intersection of Business Analysis, organisational capability, learning, transformation and technology. Each publication translates research and professional practice into practical thinking for leaders and practitioners.

The series is designed to remain useful independently of any training decision. Its purpose is to support clearer questions, stronger organisational judgement and better-informed change.