ISBN: 978-631-6557-84-1

Recibido:13 de julio de 2026 / Aceptado: 10 de septiembre de 2026 / Publicado: 28 de septiembre de 2026

Área: Ciencias Sociales - Administración

CAPÍTULO 10

INTELIGENCIA ARTIFICIAL GENERATIVA Y RESPONSABILIDAD GERENCIAL: UN MODELO DE SUPERVISIÓN HUMANA BASADA EN RIESGO

Generative artificial intelligence and managerial accountability: a risk-based human oversight model

José Luis Villegas Chumpitaz, Nataly Cárdenas Carrión, Wilfredo Pablo Quiroz Cristobal, Silvia Liliana Salazar Llerena, Erick Eduardo Aguilar Álvarez
DOI: 10.55204/pmea.134.c219

José Luis Villegas Chumpitaz[1] *0000-0002-9773-470X*
jvillegas@untels.edu.pe

Nataly Cárdenas Carrión[2] *0000-0003-2632-7100*
ncardenas@untels.edu.pe

Wilfredo Pablo Quiroz Cristobal[3] *0000-0003-2137-4688*
wquirozc@unmsm.edu.pe

Silvia Liliana Salazar Llerena[4] *0000-0002-1481-0368*
silvia.salazar@autonomadeica.edu.pe

Erick Eduardo Aguilar Álvarez[5] *0009-0003-8627-1951*
erick.aguilar@pucp.edu.pe

[1]Universidad Nacional Tecnológica de Lima Sur, Facultad de Ingeniería y Gestión. Villa el Salvador, Lima, Perú.
[2]Universidad Nacional Tecnológica de Lima Sur, Facultad de Administración de Empresas. Villa el Salvador, Lima, Perú.
[3]Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Administrativas. Lima, Perú.
[4]Universidad Autónoma de Ica, Ica, Perú.
[5]Universidad Nacional de Cajamarca, Facultad de Ciencias Económicas Contables y Administrativas. Cajamarca, Perú.

Resumen: Introducción: La integración acelerada de la inteligencia artificial generativa (IAG) en las organizaciones ha reconfigurado los procesos decisionales directivos. No obstante, la literatura científica revela un vacío conceptual relativo a cómo graduar la intensidad del control humano según el riesgo y la criticidad de la decisión gerencial. Objetivo: Analizar los factores determinantes del nivel de supervisión humana requerido en decisiones gerenciales asistidas por IAG, considerando su criticidad, riesgo y el grado de autonomía tecnológica, con el propósito de formular un modelo conceptual de supervisión gerencial basada en riesgo. Métodos: Enfoque cualitativo de síntesis documental y análisis analítico-sintético sobre la literatura científica reciente (2024–2026) y marcos institucionales de gobernanza (NIST AI 600-1). Resultados: Se propone un modelo estructurado en cuatro niveles jerárquicos (Bajo, Moderado, Alto y Crítico) que ope-

ISBN: 978-631-6557-84-1

racionalizan la transición desde la autonomía operativa supervisada por muestreo hasta el control humano reforzado con prohibición de prescripción autónoma. Conclusiones: La responsabilidad gerencial permanece inalienable en la figura humana, exigiendo esquemas formales de gobernanza, trazabilidad y auditoría proporcional al riesgo de la decisión.

Palabras clave: *Inteligencia artificial generativa, responsabilidad gerencial, supervisión humana, riesgo decisional, gobernanza organizacional.*

Abstract: Introduction: The rapid integration of generative artificial intelligence (GAI) into organizations has reshaped managerial decision-making processes. However, the scientific literature reveals a conceptual gap regarding how to adjust the level of human oversight based on the risk and criticality of the managerial decision. Objective: To analyze the factors determining the level of human oversight required for GAI-assisted managerial decisions, considering their criticality, risk, and the degree of technological autonomy, with the aim of formulating a conceptual model of risk-based managerial oversight. Methods: A qualitative approach involving literature synthesis and analytical-synthetic analysis of recent scientific literature (2024–2026) and institutional governance frameworks (NIST AI 600-1). Results: A model structured into four hierarchical levels (Low, Moderate, High, and Critical) is proposed, which operationalizes the transition from operational autonomy supervised by sampling to reinforced human control with a prohibition on autonomous decision-making. Conclusions: Managerial accountability remains inalienably vested in human agents, requiring formal governance, traceability, and audit frameworks proportional to the risk of the decision.

Keywords: *Generative artificial intelligence, managerial accountability, human oversight, decision risk, organizational governance.*

1. INTRODUCCIÓN:

La adopción de la inteligencia artificial generativa (IAG) en el ámbito organizacional ha transitado velozmente de la experimentación operativa hacia el soporte directo de la alta dirección. De acuerdo con la literatura científica reciente, la IAG se incorpora sistemáticamente en el análisis de información de entornos complejos, la simulación de escenarios corporativos, la síntesis de corpus normativos y la formulación de alternativas estratégicas (Schulte & Kanbach, 2026; Li & Tian, 2026). Sin embargo, este despliegue de capacidades analíticas avanzadas plantea interrogantes teóricos y prácticos sobre la preservación del control directivo y la asignación formal de la responsabilidad por las consecuencias de las decisiones adoptadas (García López & Nájera Ochoa, 2026).

Un primer aspecto que requiere mayor precisión es la naturaleza particular de la IAG frente a otras formas de analítica automatizada. Mientras los sistemas predictivos tradicionales suelen operar sobre variables y resultados previamente delimitados, los modelos generativos producen respuestas abiertas, probabilísticas y dependientes del contexto de interacción. Esta característica amplía la capacidad de exploración, síntesis y formulación de alternativas, pero introduce simultáneamente una incertidumbre epistemológica que el directivo debe reconocer: un resultado lingüísticamente convincente puede contener inferencias débiles, omisiones o asociaciones no verificadas. Feuerriegel et al. (2024) señalan que el valor de la IAG se extiende desde la creación de contenido hasta el apoyo decisional, aunque su aprovechamiento exige comprender tanto sus capacidades como sus límites. En la misma línea, Mariani y Dwivedi (2024) advierten que la expansión de la IAG hacia actividades de innovación y diseño organizacional abre oportunidades relevantes, pero también problemas de agencia, sesgo y uso no

ISBN: 978-631-6557-84-1

ético. Por ello, trasladar los resultados generados por un modelo hacia una decisión gerencial no constituye una operación neutra: supone convertir una salida probabilística en una acción organizacional con consecuencias concretas.

La incorporación de IAG también debe analizarse como un proceso organizacional y no únicamente como una decisión de adquisición tecnológica. La evidencia reciente muestra que los beneficios esperados dependen de capacidades internas, condiciones de adopción, habilidades del personal y mecanismos que permitan integrar los resultados de la IA con el conocimiento acumulado de la organización. Leoni et al. (2024), a partir de evidencia obtenida en organizaciones de diferentes países, identifican que la IA puede fortalecer los procesos de gestión del conocimiento y la toma de decisiones, pero su valor está condicionado por factores que facilitan o dificultan su adopción. De forma complementaria, Kumar et al. (2025) muestran que la intención de adopción de IAG en contextos empresariales se relaciona con beneficios percibidos y con reservas vinculadas a la calidad de la información y a su posible carácter engañoso, mientras que Hughes et al. (2026) destacan el peso de la complejidad, las competencias del personal y la capacidad organizacional para gestionar el cambio. En consecuencia, una política de supervisión no puede diseñarse separada de la madurez organizacional: la misma herramienta puede representar niveles de exposición distintos según la calidad de los datos, la experiencia de quienes la utilizan y la existencia de controles institucionalizados.

Esta perspectiva conduce a replantear la relación entre autonomía tecnológica y autoridad gerencial. La literatura contemporánea ha dejado de tratar la colaboración humano-IA como una simple suma de capacidades y la concibe, cada vez más, como una arquitectura de distribución de tareas, derechos de decisión y responsabilidades. Hillebrand et al. (2025) integran la investigación sobre colaboración humano-IA y gestión algorítmica para mostrar que la interacción con sistemas inteligentes transforma simultáneamente la decisión y el control organizacional. Raisch y Fomina (2025) añaden que, en problemas exploratorios y poco estructurados, los resultados varían según la forma en que se combine la búsqueda humana con la algorítmica, lo que impide asumir que una mayor autonomía tecnológica producirá siempre mejores soluciones. Esta idea encuentra respaldo empírico en Krakowski et al. (2025), quienes muestran que el desempeño puede mejorar cuando la interacción humano-IA se adapta a las características cognitivas y al contexto de trabajo, pero puede deteriorarse cuando dicha interacción se implanta sin ese ajuste. Así, la supervisión humana debe entenderse como una propiedad del diseño organizacional.

A ello se suma el problema de la confianza. Para que la IAG contribuya a una decisión de mayor calidad no basta con que el sistema sea útil; el usuario debe calibrar adecuadamente cuánto peso otorgar a sus recomendaciones. Wen et al. (2025) encontraron que la confianza en la IA incrementa el peso decisional asignado al sistema en contextos de gestión, lo que resulta funcional cuando la confianza corresponde con la calidad real de la herramienta, pero riesgoso cuando se transforma en deferencia automática. La evidencia experimental de Zercher et al. (2025) muestra, además, que la incorporación de IA generativa a equipos puede reducir ciertas asimetrías informativas, aunque también genera nuevas dificultades vinculadas con la desconfianza, el procesamiento inadecuado de los aportes algorítmicos y la forma en que el equipo integra esa información. Esta tensión se intensifica ante los sesgos presentes en los modelos generativos, cuya identificación y mitigación continúan siendo desafíos relevantes para

ISBN: 978-631-6557-84-1

la gestión de información empresarial (Wei et al., 2025). Por tanto, la supervisión eficaz no implica desconfiar sistemáticamente de la IAG, sino establecer condiciones para una confianza calibrada, verificable y proporcional al impacto de la decisión.

Desde la gobernanza, el desafío consiste en traducir principios generales la transparencia, la responsabilidad y la supervisión humana en procedimientos que funcionen dentro de la operación cotidiana. Papagiannidis et al. (2025) sostienen que la gobernanza responsable requiere prácticas estructurales, procedimentales y relacionales a lo largo del ciclo de vida de las aplicaciones de IA; es decir, no se agota en una declaración ética. La evidencia de Hadley et al. (2025) sobre juntas internas de revisión algorítmica refuerza este argumento al mostrar que la gobernanza efectiva depende de su integración con procesos existentes, del respaldo del liderazgo y de la capacidad de ajustar la intensidad de revisión al nivel de riesgo. Del mismo modo, Salgado-Criado (2025) propone entender la supervisión desde una perspectiva operativa, articulando exigencias técnicas, de negocio, seguridad y cumplimiento. En el sector público, de Almeida y dos Santos Júnior (2025) encontraron que la formación de decisores, desarrolladores y usuarios se asocia con estadios más avanzados de gobernanza de IA. En conjunto, estas contribuciones sugieren que la responsabilidad gerencial solo se vuelve operativa cuando existen roles, reglas de escalamiento, registros y competencias que permiten ejercerla de manera efectiva.

El vacío que emerge de esta discusión no se limita, entonces, a la existencia o ausencia de control humano, sino a la falta de criterios suficientemente claros para determinar cuánto control resulta razonable en cada situación. La colaboración humano-IA puede aumentar la inteligencia organizacional cuando se aprovechan capacidades complementarias y se conserva la posibilidad de explicación y revisión (Kolbjørnsrud, 2024). Asimismo, la evidencia reciente indica que el desempeño derivado de la IAG está mediado por la calidad y eficiencia de las decisiones, y que sus resultados cambian según las condiciones de gobernanza institucional (Liu et al., 2026). Más recientemente, Leon (2026) muestra que los mecanismos sociotécnicos de control pueden incorporar políticas, restricciones técnicas y flujos de trabajo capaces de clarificar responsabilidades y orientar el uso de IAG. Estas aproximaciones respaldan la necesidad de pasar de un principio genérico de "humano en el circuito" a un esquema graduado que relacione consecuencias, autonomía y trazabilidad. Precisamente en ese punto se sitúa la contribución del presente capítulo: ofrecer una estructura que permita diferenciar cuándo basta una revisión muestral, cuándo se requiere validación previa y cuándo la decisión exige control humano reforzado y colegiado.

A pesar del vertiginoso incremento de investigaciones en torno al binomio IAG y gestión organizacional, la problemática central no reside en determinar si la IAG posee la capacidad técnica para asistir el proceso decisional, sino en definir en qué condiciones institucionales y con qué gradiente de intensidad debe ejercerse la supervisión humana. Una supervisión humana uniforme resulta ineficiente e imprudente para las organizaciones: la aplicación de controles rígidos y continuos sobre tareas rutinarias de bajo impacto genera ineficiencias y cuellos de botella operativos; por el contrario, la supervisión tenue o ausente sobre decisiones estratégicas de alta complejidad expone a la organización a severos riesgos legales, financieros, éticos y reputacionales (Mendoza-Rejas et al., 2026; Srivastava, 2026).

Frente a este escenario, la pregunta de investigación que orienta este estudio es: ¿qué

ISBN: 978-631-6557-84-1

factores determinan el nivel de supervisión humana requerido en las decisiones gerenciales asistidas por inteligencia artificial generativa, considerando su criticidad, riesgo y el grado de autonomía tecnológica? Por consiguiente, el objetivo general de este capítulo es analizar los factores asociados al nivel de supervisión humana en la toma de decisiones gerenciales asistidas por IAG, articulando su criticidad y riesgo implícito con la autonomía de la tecnología, a fin de proponer un modelo conceptual de supervisión gerencial basada en riesgo.

2. METODOLOGÍA:

Para la consecución del objetivo trazado, el presente trabajo adoptó un enfoque cualitativo de carácter descriptivo y analítico-sintético, fundamentado en una revisión documental e integración de la literatura científica producida entre los años 2024 y 2026. El proceso metodológico se estructuró en tres fases continuas:

En la primera fase, se realizó la identificación y selección del corpus bibliográfico mediante búsquedas en bases de datos científicas internacionales (Scopus, Web of Science, IEEE Xplore y motores académicos), priorizando artículos revisados por pares y marcos de referencia estandarizados. Se delimitaron como ejes temáticos centrales: la colaboración humano-IA en decisiones complejas (Li & Tian, 2026), los roles decisionales de la IAG (Schulte & Kanbach, 2026), los protocolos de supervisión episódica (Taş et al., 2026), la delegación y control de la autonomía (Srivastava, 2026), la opacidad administrativa y riesgos éticos (Mendoza-Rejas et al., 2026), y el marco de gestión de riesgos de IAG del NIST (Autio et al., 2024).

En la segunda fase, se procedió a la extracción de datos y clasificación analítica mediante una matriz comparativa de antecedentes directos. Se sistematizaron los enfoques principales, aportes disciplinares, limitaciones declaradas y vacíos teóricos identificados en cada investigación seleccionada.

En la tercera fase, mediante un proceso de triangulación conceptual y abstracción inductiva, se articularon los hallazgos de la literatura para construir la propuesta conceptual. Las dimensiones de criticidad de la decisión, nivel de autonomía tecnológica y mecanismos de control humano se integraron dentro de una matriz jerárquica de riesgo-supervisión que dota de validez teórica y operabilidad a la propuesta final.

3. RESULTADOS Y DISCUSIÓN:

La comprensión científica del impacto de la IAG en la toma de decisiones gerenciales exige desglosar la literatura en tres dominios interrelacionados: el proceso decisional reconfigurado por algoritmos generativos, los modos de colaboración e interacción humano-tecnología, y los esquemas de gobernanza organizacional y rendición de cuentas.

Respecto al proceso decisional gerencial, Schulte y Kanbach (2026) sostienen que la IAG trasciende el rol tradicional de herramienta analítica pasiva para constituirse en un agente generativo que interviene activamente en fases como la identificación del problema, la síntesis de evidencia y la recomendación de alternativas de acción. No obstante, García López y Nájera Ochoa (2026) advierten que la toma de decisiones .ªumentada"por IA altera la estructura de autoridad organizacional, requiriendo redefinir las competencias y la firma directiva responsable.

En el ámbito de la colaboración humano-IA, Li y Tian (2026) destacan que los modelos de interacción deben graduarse en función de la complejidad cognitiva del problema. Asimismo, Srivastava (2026) profundiza en el dilema de la delegación, subrayando que otorgar excesiva

ISBN: 978-631-6557-84-1

autonomía a sistemas algorítmicos sin un marco explícito de control erosiona el juicio reflexivo gerencial. Para evitar esta pérdida de agencia humana, Taş et al. (2026) proponen un protocolo de supervisión episódica en flujos de trabajo generativos, demostrando la necesidad de instaurar hitos de control directivo en puntos críticos del proceso.

Por otra parte, la gobernanza y los riesgos asociados a la opacidad algorítmica constituyen una preocupación preponderante en la literatura reciente. Mendoza-Rejas et al. (2026) identifican que el uso indisciplinado de la IAG genera riesgos severos de opacidad, sesgos y alucinaciones en la fundamentación de actos administrativos y gerenciales. Ante esto, la literatura sobre gobernanza de IA (Batool et al., 2025) y las directrices normativas del NIST (Autio et al., 2024) concuerdan en que la gestión del riesgo debe sustentarse en principios de transparencia, trazabilidad y supervisión humana diferenciada.

Tabla 1

Matriz de síntesis y sistematización de antecedentes principales (2024–2026)

Autor(es) y Año	Enfoque Principal	Aporte	Espacio que aborda este capítulo
García López & Nájera Ochoa (2026)	IA y toma de decisiones gerenciales.	Integra decisiones aumentadas, gobernanza y responsabilidad gerencial.	Precisar cómo el riesgo de una decisión se traduce en un nivel concreto de supervisión humana.
Schulte & Kanbach (2026)	GenAI y procesos de decisión organizacional.	Revisa tareas y roles de GenAI a través de componentes del proceso decisional.	Conectar las etapas decisionales con criterios de autonomía y supervisión basada en riesgo.
Li & Tian (2026)	Colaboración humano-IA.	Sistematiza paradigmas de colaboración y toma de decisiones.	Traducir los modos de colaboración en niveles de control gerencial según criticidad.
Taş, Memmert & Bittner (2026)	Supervisión humana en workflows de GenAI.	Propone un protocolo de supervisión en distintas etapas del trabajo.	Adaptar la supervisión al contexto de decisiones gerenciales y sus consecuencias.
Srivastava (2026)	Delegación, autonomía y control.	Examina decisiones sobre cuánto delegar a sistemas de IA.	Integrar autonomía tecnológica, riesgo decisional y responsabilidad gerencial.
Batool, Zowghi & Bano (2025)	Gobernanza de IA.	Revisa marcos y mecanismos de gobernanza.	Vincular los principios generales de gobernanza con reglas de supervisión aplicables a decisiones.
Mendoza-Rejas et al. (2026)	GenAI y opacidad administrativa municipal.	Aborda transparencia, trazabilidad y supervisión en gobiernos municipales.	Ampliar el análisis a decisiones gerenciales organizacionales y a una matriz de riesgo-supervisión.
Urbanovič & Holubčík (2026)	IA en decisiones gerenciales y sostenibilidad.	Revisa la IA en decisiones relacionadas con modelos de negocio sostenibles.	Diferenciar el foco en supervisión y responsabilidad del foco en sostenibilidad.
Autio et al. (2024)	Gestión de riesgos en IA generativa (NIST AI 600-1).	Establece perfiles y acciones de mitigación de riesgo para IAG.	Sustentar institucionalmente la graduación de controles y la trazabilidad de la IAG.

3.1. Desarrollo del modelo conceptual de supervisión gerencial basada en riesgo:

El modelo conceptual propuesto postula que el nivel de supervisión humana no es una variable fija ni discrecional, sino un requerimiento que debe calibrarse en función de la interacción entre tres dimensiones esenciales: a) la criticidad y riesgo de la decisión gerencial, b) el grado de autonomía permitido a la IAG, y c) el marco transversal de gobernanza organizacional.

3.1.1. Dimensiones e interacción del modelo:

1. **Criticidad y Riesgo de la Decisión:** Representa la magnitud de las consecuencias potenciales de la decisión para la organización y sus partes interesadas. Se evalúa considerando la reversibilidad del impacto, los compromisos financieros, las responsabilidades legales, las consecuencias sobre las personas y la exposición ético-reputacional.

ISBN: 978-631-6557-84-1

2. **Grado de Autonomía de la IAG:** Define el rango de libertad operativa otorgado al sistema algorítmico. Varía desde la ejecución autónoma restringida a tareas operativas hasta la mera asistencia analítica sin autoridad prescrita para tomar acciones finales.

3. **Nivel de Supervisión Humana Requerido:** Representa el mecanismo de control activo, validación o veto ejercido por el responsable gerencial antes, durante o después del uso de los insumos generados por la IAG.

4. **Responsabilidad Gerencial (*Accountability*):** Principio deontológico y legal inalienable. La incorporación de IAG en cualquier etapa de la decisión no exime ni diluye la responsabilidad del gerente, quien responde formalmente por los resultados finales ante la organización y terceros.

Tabla 2

Matriz de niveles de riesgo, autonomía de la IAG y supervisión humana requerida

Nivel de riesgo	Ejemplos ilustrativos	Papel y autonomía de la IA	Supervisión humana requerida	Mecanismos de gobernanza y trazabilidad
Bajo	Resúmenes de información, clasificación de documentos, borradores de comunicaciones administrativas internas.	Alta autonomía operativa dentro de límites y plantillas predefinidos.	Revisión básica, pasiva o por muestreo a posteriori.	Registro automatizado de interacción (prompt/output) y auditoría muestral periódica.
Moderado	Priorización de proyectos, análisis preliminar de proveedores, contraste de escenarios operativos.	Genera análisis comparativos, resúmenes de alternativas y recomendaciones estructuradas.	Validación humana directa (Human-in-the-loop) antes de su adopción o ejecución.	Verificación de alucinaciones y sesgos en variables clave; archivo de fuentes consultadas.
Alto	Decisiones estratégicas de RR. HH., asignación presupuestaria significativa, reestructuración operativa.	Soporte analítico profundo; estrictamente sin autoridad ni facultad decisoria final.	Revisión sustantiva, crítica y aprobación explícita e indelegable por parte del gerente.	Documentación formal de la lógica de decisión humana que valida o rectifica el insumo de la IAG.
Crítico	Decisiones con efectos legales, financieros, humanos o reputacionales graves y de difícil reversión.	Procesamiento restringido de datos; prohibición explícita de prescripción autónoma.	Control reforzado (Human-on-the-loop estricto), revisión colegiada o comités de supervisión.	Trazabilidad integral, auditoría técnica previa, expediente completo de responsabilidad directiva.

3.2. Discusión e implicaciones para la administración:

El Modelo de Supervisión Gerencial basada en Riesgo ofrece implicaciones teóricas y prácticas sustanciales para la ciencia de la administración y la práctica corporativa.

En primer lugar, el modelo contribuye a visualizar desde otra perspectiva la tensión teórica entre la eficiencia operativa prometida por la IAG y la imperiosa necesidad de control institucional. Al establecer una gradación explícita (Tabla 2), se evita el sobrecontrol de tareas insignificantes, que neutraliza la ventaja competitiva de la tecnología, y el subcontrol de decisiones críticas, previniendo el sesgo de automatización (*automation bias*) documentado en la literatura (Srivastava, 2026).

En segundo lugar, se reafirma el carácter inalienable de la responsabilidad gerencial. En concordancia con los hallazgos de García López y Nájera Ochoa (2026), la asunción de que un modelo generativo sofisticado recomendóüna acción no constituye un eximente ético ni legal para la alta dirección. La matriz de supervisión actúa como un escudo de responsabilidad, asegurando que a mayor criticidad exista una firma humana consciente y documentada.

ISBN: 978-631-6557-84-1

En tercer lugar, para la gestión administrativa pública y privada, la implementación del modelo demanda el fortalecimiento de la gobernanza organizacional de IA (Batool et al., 2025; Mendoza-Rejas et al., 2026). Esto implica adecuar los reglamentos internos, crear registros de trazabilidad de prompts/outputs para decisiones de Niveles 3 y 4, e instituir comités de supervisión ética encargados de auditar periódicamente los sesgos y la veracidad de las herramientas generativas empleadas.

Una lectura más profunda del modelo permite señalar que su principal aporte no consiste únicamente en clasificar decisiones en cuatro niveles, sino en convertir el riesgo en un criterio de diseño de los derechos de decisión. Esta distinción es relevante porque la IAG puede intervenir en varias etapas, búsqueda de información, estructuración del problema, generación de alternativas, comparación de escenarios o recomendación, sin que todas impliquen la misma cesión de autoridad. Hillebrand et al. (2025) muestran que la interacción con IA modifica tanto la colaboración como los mecanismos de control, mientras que Raisch y Fomina (2025) evidencian que distintas configuraciones de interacción humano-IA producen diferentes patrones de solución. Desde esta perspectiva, la matriz propuesta evita una equivalencia problemática entre "usar IAG" y "delegar la decisión": la tecnología puede tener una participación cognitiva intensa sin adquirir, por ello, autoridad decisoria final. La variable crítica pasa a ser qué parte del proceso se delega, con qué posibilidad de reversión y bajo qué obligación de justificación humana.

Esta interpretación resulta consistente con la evidencia empírica que cuestiona los enfoques binarios de automatización. Krakowski et al. (2025) muestran que una misma tecnología puede producir efectos opuestos cuando cambia el modo en que se organiza la interacción entre personas y sistemas de IA. Del mismo modo, Zercher et al. (2025) encuentran que incorporar IA a equipos no garantiza mejores decisiones: el resultado depende de la distribución del conocimiento y de la manera en que los miembros humanos procesan los aportes del sistema. Luong et al. (2026) añaden que las capacidades de IA y los incentivos organizacionales condicionan la colaboración humano-máquina en decisiones algorítmicas. Estos hallazgos fortalecen el supuesto central del modelo: la supervisión no debería definirse solo a partir de la potencia técnica de la herramienta, sino considerando la estructura de la tarea, los incentivos, la competencia de los usuarios y las consecuencias de un eventual error. Una organización que mantenga el mismo mecanismo de control para todas las decisiones corre el riesgo de vigilar excesivamente lo trivial y, al mismo tiempo, normalizar delegaciones inadecuadas en asuntos de alta criticidad.

También es necesario precisar que el aumento de supervisión no debe confundirse con la búsqueda de una intervención humana meramente formal. El valor del control humano depende de que el decisor conserve capacidad real para cuestionar el resultado, solicitar evidencia adicional, modificar el curso de acción o rechazar la recomendación. Wen et al. (2025) muestran que una mayor confianza puede traducirse en un mayor peso otorgado a la IA; por ello, el problema no es únicamente generar confianza, sino calibrarla. La advertencia es especialmente pertinente cuando los modelos presentan sesgos sistemáticos o producen respuestas plausibles que ocultan incertidumbre, cuestión subrayada por Wei et al. (2025). En contextos públicos, Wang et al. (2026) encuentran además que la aceptación de configuraciones colaborativas depende de elementos como la explicación de resultados y la atribución de responsabilidad. Para la administración, esto implica que la supervisión de niveles Alto y Crítico debe evaluar tanto

ISBN: 978-631-6557-84-1

el contenido de la recomendación como las condiciones bajo las cuales fue producida, evitando que la revisión humana se reduzca a una firma posterior que legitime automáticamente la salida del sistema.

Las implicaciones de gobernanza son igualmente importantes, por lo que se detalla un modelo basado en riesgo que vincule cada nivel con procedimientos, responsables y evidencias verificables. Papagiannidis et al. (2025) plantean que la gobernanza responsable se concreta mediante prácticas estructurales, procedimentales y relacionales, lo que coincide con la necesidad de combinar políticas formales, circuitos de aprobación y coordinación entre áreas. La experiencia documentada por Hadley et al. (2025) indica que las juntas de revisión algorítmica son más útiles cuando se integran con mecanismos ya existentes de riesgo, privacidad, auditoría o cumplimiento, y cuando poseen respaldo directivo suficiente para incidir en decisiones reales. En esa misma lógica, Leon (2026) encuentra que los guardarraíles sociotécnicos pueden mejorar la claridad de la responsabilidad cuando articulan reglas, restricciones técnicas y flujos de trabajo. En términos administrativos, esto cuestiona la creación de comités puramente nominales: si un órgano de supervisión carece de información, competencias o capacidad de detener una implementación, su existencia puede incrementar la apariencia de control sin reducir sustantivamente el riesgo.

La aplicación práctica del modelo exige, por tanto, una secuencia de gestión previa a la utilización de la herramienta. Primero, la organización debería inventariar los procesos en los que la IAG interviene y distinguir entre generación de información, recomendación y ejecución. Segundo, debe valorar la severidad y reversibilidad de las consecuencias, incluyendo efectos financieros, regulatorios, laborales, reputacionales y sobre terceros. Tercero, corresponde asignar un responsable humano con autoridad suficiente para validar o escalar el caso. Cuarto, deben definirse evidencias mínimas de trazabilidad, como fuentes consultadas, versiones relevantes, instrucciones dadas al sistema y fundamento de la decisión final. Esta lógica es coherente con la perspectiva operativa de Salgado-Criado (2025), que concibe la supervisión como una función de control organizacional adaptable a la aplicación concreta, y con los hallazgos registrados por de Almeida y dos Santos Júnior (2025), quienes resaltan la importancia de procesos combinados y de la formación de los actores. Asimismo, la propuesta de Kolbjørnsrud (2024) sobre colaboración humano-IA recuerda que la explicación y la complementariedad deben diseñarse de forma deliberada, no asumirse como propiedades automáticas de la tecnología.

El modelo también permite matizar la relación entre control y desempeño, debido a que resultaría simple asumir que una mayor supervisión reduce necesariamente la eficiencia o que una mayor autonomía produce, por definición, una ventaja competitiva. Liu et al. (2026) muestran que la capacidad organizacional habilitada por IAG puede mejorar tanto la calidad como la eficiencia decisional, pero que las condiciones de gobernanza modifican la forma en que ese valor se materializa. Hughes et al. (2026) llegan a una conclusión compatible al identificar que la adopción organizacional depende de la capacidad de cambio y de las competencias disponibles, más que de la disponibilidad aislada de la herramienta. En paralelo, Leoni et al. (2024) muestran que la IA puede potenciar la gestión del conocimiento y apoyar decisiones más informadas cuando se integra a procesos organizacionales estructurados. En consecuencia, la matriz riesgo-supervisión no debe interpretarse como un freno a la innovación, sino como un mecanismo de asignación eficiente de la atención directiva: concentra controles intensivos donde el

ISBN: 978-631-6557-84-1

costo esperado del error es mayor y libera tareas de bajo riesgo de revisiones innecesariamente onerosas.

Existe, sin embargo, un riesgo adicional que merece atención crítica: convertir la IAG en una respuesta predeterminada para problemas cuya naturaleza todavía no ha sido comprendida. Adobor (2026) denomina a esta tendencia una trampa de adopción basada en el "solucionismo", en la que problemas organizacionales complejos se reformulan prematuramente como problemas susceptibles de resolverse mediante algoritmos y datos. Esta advertencia complementa a Mariani y Dwivedi (2024), quienes sitúan la agencia, el uso indebido y el diseño organizacional entre los temas centrales de la investigación sobre IAG. Feuerriegel et al. (2024), por su parte, resaltan que las posibilidades generativas deben analizarse junto con sus riesgos y limitaciones. El modelo propuesto ofrece una defensa parcial frente a ese sesgo porque obliga a preguntar, antes de delegar, cuál es la naturaleza de la decisión y cuáles serían las consecuencias de una salida errónea. No obstante, el propio proceso de clasificación debe evitar automatizarse completamente: determinar criticidad es un juicio contextual y normativo que requiere comprender a las partes afectadas, el marco regulatorio y la estrategia de la organización.

Finalmente, las implicaciones para la responsabilidad gerencial exceden el acto de aprobación final. La *accountability* supone capacidad de reconstruir por qué se utilizó la IAG, qué información fue considerada, quién tenía autoridad para intervenir y qué razones explican la aceptación o rechazo de una recomendación. La evidencia de Wang et al. (2026) sobre atribución de responsabilidad en decisiones colaborativas y los resultados de Leon (2026) sobre claridad de responsabilidades mediante guardarraíles refuerzan la necesidad de una trazabilidad comprensible, no meramente técnica. A su vez, Salgado-Criado (2025) recuerda que la supervisión debe integrar objetivos empresariales, requisitos técnicos y exigencias de cumplimiento. Por ello, en los niveles Alto y Crítico del modelo la responsabilidad no debería descansar en una sola revisión documental, sino en una cadena de control que permita detectar errores, escalar discrepancias y documentar el juicio humano. Este enfoque fortalece la defensa organizacional frente a fallos sin trasladar ficticiamente la responsabilidad hacia el sistema generativo.

La propuesta debe interpretarse como un marco conceptual sujeto a validación y ajuste contextual. La clasificación de riesgo puede variar entre sectores, jurisdicciones y tipos de organización; una recomendación que es reversible en una actividad comercial puede ser materialmente irreversible en un proceso laboral, financiero o público. Asimismo, la velocidad con la que evolucionan los modelos generativos puede modificar el balance entre capacidad técnica y posibilidad de supervisión. Por ello, la utilidad administrativa de la matriz dependerá de que sus niveles sean revisados periódicamente a partir de incidentes, auditorías y evidencia de desempeño. Más que fijar una taxonomía permanente, el modelo ofrece una lógica de decisión: a mayor severidad potencial, menor tolerancia a la autonomía no controlada y mayor necesidad de justificación, trazabilidad y revisión independiente. Esta lógica conserva la flexibilidad requerida para innovar, pero establece un límite institucional claro frente a la delegación acrítica de autoridad gerencial (Papagiannidis et al., 2025; Hadley et al., 2025).

4. CONCLUSIONES:

El desarrollo del presente capítulo permite formular las siguientes conclusiones principales:

ISBN: 978-631-6557-84-1

En primer lugar, la incorporación de la inteligencia artificial generativa en los procesos decisionales exige abandonar los enfoques de control uniforme o discrecional, reemplazándolos por un modelo dinámico de supervisión humana calibrado en función del riesgo y la criticidad de la decisión.

Por otro lado, el Modelo de Supervisión Gerencial basada en Riesgo articulado en este trabajo operacionaliza la interacción entre riesgo, autonomía de la IAG y control humano a través de cuatro niveles rigurosos (Bajo, Moderado, Alto y Crítico). Este esquema contribuye una intervención eficiente de la IAG en actividades rutinarias y potencia la protección ética y legal ante situaciones de alto impacto.

Además, la responsabilidad gerencial (*accountability*) conserva una conexión a las atribuciones y decisiones fundamentadas por el responsable a cargo. La IAG debe constituirse exclusivamente en un instrumento de ampliación cognitiva, permaneciendo la autoridad, la convalidación crítica y la rendición de cuentas bajo el dominio del directivo.

En cuarto lugar, se recomienda para investigaciones futuras la validación empírica cuantitativa y cualitativa de la matriz de riesgo propuesta mediante estudios de caso en sectores altamente regulados (banca, salud y sector público), evaluando el impacto de los protocolos de supervisión en la reducción de errores decisionales y en el desempeño organizacional.

5. FINANCIACIÓN:

Los autores no recibieron financiación para el desarrollo de la presente investigación.

6. CONFLICTO DE INTERESES:

Los Autores declaran que no existe conflicto de intereses.

7. CONTRIBUCIÓN DE AUTORÍA:

		Villegas J.	Cárdenas N.	Quiroz W.	Vento J.	Aguilar E.
Participar activamente en:						
Conceptualización	X	X	X	X	X	
Análisis formal	X	X	X	X	X	
Adquisición de fondos						
Investigación	X	X	X	X	X	
Metodología	X	X	X	X	X	
Administración del proyecto	X	X	X	X	X	
Recursos	X	X	X	X	X	
Redacción –borrador original	X	X	X	X	X	
Redacción –revisión y edición	X	X	X	X	X	
La discusión de los resultados	X	X	X	X	X	
Revisión y aprobación de la versión final del trabajo.	X	X	X	X	X	

8. REFERENCIAS:

Adobor, H. (2026). Navigating the AI adoption trap: A framework for organizational practice. *Organizational Dynamics*, *55*(2), 101232. https://doi.org/10.1016/j.orgdyn.2026.101232

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

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics*, *5*, 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

de Almeida, P. G. R., & dos Santos Júnior, C. D. (2025). Artificial intelligence governance: Unders-

tanding how public organizations implement it. *Government Information Quarterly*, *42*(1), 102003. https://doi.org/10.1016/j.giq.2024.102003

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*, 111–126. https://doi.org/10.1007/s12599-023-00834-7

García López, I. M., & Nájera Ochoa, J. (2026). Inteligencia artificial y toma de decisiones gerenciales: Implicaciones para la gestión estratégica en las organizaciones. *Universita Ciencia*, *14*(40), 68–92. https://doi.org/10.5281/zenodo.21463159

Hadley, E., Blatecky, A., & Comfort, M. (2025). Investigating algorithm review boards for organizational responsible artificial intelligence governance. *AI and Ethics*, *5*, 2485–2495. https://doi.org/10.1007/s43681-024-00574-8

Hillebrand, L., Raisch, S., & Schad, J. (2025). Managing with artificial intelligence: An integrative framework. *Academy of Management Annals*, *19*(1), 343–375. https://doi.org/10.5465/annals.2022.0072

Hughes, L., Davies, F., Li, K., Gunaratnege, S. M., Malik, T., & Dwivedi, Y. K. (2026). Beyond the hype: Organisational adoption of generative AI through the lens of the TOE framework–A mixed methods perspective. *International Journal of Information Management*, *86*, 102982. https://doi.org/10.1016/j.ijinfomgt.2025.102982

Kolbjørnsrud, V. (2024). Designing the intelligent organization: Six principles for human-AI collaboration. *California Management Review*, *66*(2), 44–64. https://doi.org/10.1177/00081256231211020

Krakowski, S., Haftor, D., Luger, J., Pashkevich, N., & Raisch, S. (2025). Human-centered artificial intelligence: A field experiment. *Management Science*, *72*(1), 57–72. https://doi.org/10.1287/mnsc.2022.03849

Kumar, A., Shankar, A., Hollebeek, L. D., Behl, A., & Lim, W. M. (2025). Generative artificial intelligence (GenAI) revolution: A deep dive into GenAI adoption. *Journal of Business Research*, *189*, 115160. https://doi.org/10.1016/j.jbusres.2024.115160

Leon, M. (2026). Governing generative AI in organizations: A design theory and quasi-experimental field study of sociotechnical guardrails. *The Journal of Supercomputing*, *82*, Article 641. https://doi.org/10.1007/s11227-026-08760-7

Leoni, L., Gueli, G., Ardolino, M., Panizzon, M., & Gupta, S. (2024). AI-empowered KM processes for decision-making: Empirical evidence from worldwide organisations. *Journal of Knowledge Management*, *28*(11), 320–347. https://doi.org/10.1108/JKM-03-2024-0262

Li, H., & Tian, F. (2026). Advancing decision-making through AI-human collaboration: A systematic review and conceptual framework. *Group Decision and Negotiation*, *35*, Article 26. https://doi.org/10.1007/s10726-026-09980-1

Liu, D. Y., Kamal, M. M., Dou, J., & Zhang, J. Z. (2026). Generative AI, digital dexterity, and organizational future performance: The roles of decision-making and institutional governance for sustainable supply chains. *Technological Forecasting and Social Change*, *227*, 124656. https://doi.org/10.1016/j.techfore.2026.124656

Luong, A., Kumar, N., & Lang, K. R. (2026). Human–machine collaboration and algorithmic decision-making: The role of organizational incentives and AI capabilities. *Information & Management*, *63*(6), 104384. https://doi.org/10.1016/j.im.2026.104384

Mariani, M., & Dwivedi, Y. K. (2024). Generative artificial intelligence in innovation management: A preview of future research developments. *Journal of Business Research*, *175*, 114542. https://doi.org/10.1016/j.jbusres.2024.114542

Mendoza-Rejas, J. N., Mendoza-Rejas, A. R., & Vasquez-Hernandez, B. R. (2026). Inteligencia artificial generativa en el riesgo de opacidad administrativa en las municipalidades: Una revisión de literatura. *Espacios*, *47*(2), 296–308. https://doi.org/10.48082/espacios-a26v47n02r08

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, *34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Raisch, S., & Fomina, K. (2025). Combining human and artificial intelligence: Hybrid problem-solving in organizations. *Academy of Management Review*, *50*(2), 441–464. https://doi.org/10.5465/amr.2021.0421

Salgado-Criado, J. (2025). Human oversight of artificial intelligence: An operations management perspective. *Journal of Industrial Engineering and Management*, *18*(2), 285–304. https://doi.org/10.3926/jiem.8567

Schulte, N., & Kanbach, D. K. (2026). Rethinking organizational decision-making: The emerging roles and tasks of generative artificial intelligence. *Management Review Quarterly*. Advance online publication. https://doi.org/10.1007/s11301-026-00611-2

Srivastava, A. (2026). Governing AI-enabled decision making: Delegation, autonomy, and control at the operations–marketing interface. *Production and Operations Management*. Advance online publication. https://doi.org/10.1177/10591478261473004

Taş, E., Memmert, L., & Bittner, E. (2026). Episodic oversight in generative AI workflows: A nine-step protocol for preserving human agency (OP-9). *Electronic Markets*, *36*, Article 58. https://doi.org/10.1007/s12525-026-00915-x

Urbanovič, M., & Holubčík, M. (2026). Artificial intelligence in managerial decision-making for sustainable business models: A systematic literature review. *Systems*, *14*(3), 245. https://doi.org/10.3390/systems14030245

Wang, S., Zhang, Y., Xiao, Y., & Liang, Z. (2026). From comparison to collaboration: Experimental evidence on human and artificial intelligence in public decision making. *Information & Management*, *63*(5), 104364. https://doi.org/10.1016/j.im.2026.104364

Wei, X., Kumar, N., & Zhang, H. (2025). Addressing bias in generative AI: Challenges and research opportunities in information management. *Information & Management*, *62*(2), 104103. https://doi.org/10.1016/j.im.2025.104103

Wen, Y., Wang, J., & Chen, X. (2025). Trust and AI weight: Human-AI collaboration in organizational management decision-making. *Frontiers in Organizational Psychology*, *3*, 1419403. https://doi.org/10.3389/forgp.2025.1419403

Zercher, D., Jussupow, E., Benke, I., & Heinzl, A. (2025). How can teams benefit from AI team members? Exploring the effect of generative AI on decision-making processes and decision quality in team–AI collaboration. *Journal of Organizational Behavior*. Advance online publication. https://doi.org/10.1002/job.2898