

# La estupidez en el uso de los sistemas de información: límites tecnológicos, errores humanos y la vigencia del principio GIGO

## FOCO TEMÁTICO

Actualización académica

## GRUPO

Profesores

## NOMBRE DE LA MATERIA

Gestión de Tecnologías Digitales

## AUTORES

NOMBRE Y APELLIDO	CORREO ELECTRÓNICO
Aníbal Mario Mazza Fraquelli	amf@fraquelli.net

## RESUMEN

El presente artículo analiza críticamente el fenómeno de la “estupidez” en la interacción humano-sistema dentro de los sistemas de información (SI), entendida no como una descalificación del usuario sino, siguiendo a Cipolla (1988), como un patrón de conductas que generan pérdidas para quien las ejecuta y para terceros sin producir beneficios compensatorios. Se revisa la vigencia del principio GIGO (*Garbage In, Garbage Out*) desde su primera aparición documentada en 1957 hasta su reformulación en los entornos de inteligencia artificial generativa y aprendizaje automático. A partir de una revisión teórica se integran cinco marcos de análisis: la taxonomía de errores humanos de Reason (1990), las leyes fundamentales de la estupidez humana de Cipolla (1988), las dimensiones de calidad de datos de Wang y Strong (1996), los principios de diseño centrado en el usuario de Norman (2013) y la taxonomía de alucinaciones en generación de lenguaje natural de Ji et al. (2023). Los hallazgos permiten sostener que el principio GIGO no solo conserva su vigencia, sino que se amplifica en contextos algorítmicos, donde la calidad de los datos de entrenamiento, el diseño de los prompts y el juicio crítico humano determinan la fiabilidad de la salida. Se concluye con implicaciones para el diseño de interfaces, la alfabetización digital y el gobierno de datos en las organizaciones.

## PALABRAS CLAVE

GIGO, sistemas de información, error humano, calidad de datos, inteligencia artificial, *prompt engineering*.

# La estupidez en el uso de los sistemas de información: límites tecnológicos, errores humanos y la vigencia del principio GIGO

## Resumen

El presente artículo analiza críticamente el fenómeno de la “estupidez” en la interacción humano-sistema dentro de los sistemas de información (SI), entendida no como una descalificación del usuario sino, siguiendo a Cipolla (1988), como un patrón de conductas que generan pérdidas para quien las ejecuta y para terceros sin producir beneficios compensatorios. Se revisa la vigencia del principio GIGO (*Garbage In, Garbage Out*) desde su primera aparición documentada en 1957 hasta su reformulación en los entornos de inteligencia artificial generativa y aprendizaje automático. A partir de una revisión teórica se integran cinco marcos de análisis: la taxonomía de errores humanos de Reason (1990), las leyes fundamentales de la estupidez humana de Cipolla (1988), las dimensiones de calidad de datos de Wang y Strong (1996), los principios de diseño centrado en el usuario de Norman (2013) y la taxonomía de alucinaciones en generación de lenguaje natural de Ji et al. (2023). Los hallazgos permiten sostener que el principio GIGO no solo conserva su vigencia, sino que se amplifica en contextos algorítmicos, donde la calidad de los datos de entrenamiento, el diseño de los prompts y el juicio crítico humano determinan la fiabilidad de la salida. Se concluye con implicaciones para el diseño de interfaces, la alfabetización digital y el gobierno de datos en las organizaciones.

**Palabras clave:** GIGO, sistemas de información, interacción humano-computador, error humano, calidad de datos, sesgos cognitivos, inteligencia artificial, alucinaciones, *prompt engineering*.

## 1 Introducción

En la actualidad, los sistemas de información constituyen la infraestructura invisible sobre la cual descansan las operaciones cotidianas de gobiernos, empresas y personas. La transformación digital, la adopción masiva de servicios en la nube y la irrupción de la inteligencia artificial generativa han desplazado la frontera de lo automatizable hacia tareas que, hasta hace pocos años, parecían reservadas al juicio humano. No obstante, en la medida en que estas tecnologías se sofistican, paradójicamente se hace más evidente un factor que parece resistirse a los avances: el usuario humano y sus decisiones.

En la jerga técnica, una frase breve ha acompañado a la computación desde sus orígenes: *garbage in, garbage out*, abreviada como GIGO. Acuñada en los primeros años del procesamiento electrónico de datos, esta máxima recuerda que ningún sistema, por sofisticado que sea, puede compensar la mala calidad de sus entradas. En pleno 2026, con modelos de lenguaje de escala masiva, analítica avanzada y agentes autónomos, la afirmación no solo sigue vigente, sino que adquiere nuevas aristas: los datos de entrenamiento, los prompts, los parámetros de configuración y los supuestos implícitos del usuario son hoy “entradas” tan críticas como lo fueron las tarjetas perforadas en la década de 1960.

La presente ponencia articula un concepto provocador pero académicamente riguroso: el de la *estupidez* aplicada al uso de los sistemas de información. El término, tal como lo desarrolló el

## **Jornada Académica Anual 2026 del Departamento de Sistemas**

historiador económico italiano Carlo Cipolla (1988), no alude a una carencia intelectual sino a un patrón conductual que produce pérdidas simultáneas para el sujeto que lo ejecuta y para quienes le rodean, sin beneficios compensatorios.

Aplicada al ámbito tecnológico, esta categoría permite describir situaciones frecuentes en las organizaciones: decisiones basadas en datos de baja calidad que se asumen confiables, confianza acrítica en salidas algorítmicas, configuración deficiente de controles, uso indebido de hojas de cálculo, y una larga lista de comportamientos que, sin ser deliberadamente malintencionados, generan costos significativos.

El objetivo del artículo es examinar, desde una perspectiva integradora, cómo los límites tecnológicos, los errores humanos y los sesgos cognitivos convergen en la interacción humano-sistema para producir resultados subóptimos, y por qué el principio GIGO sigue siendo una clave interpretativa ineludible. El principio GIGO: del pasado analógico al presente algorítmico

## **2 Origen histórico**

La expresión *garbage in, garbage out* tiene su registro documentado más antiguo en noviembre de 1957, en un artículo del diario estadounidense *The Hammond Times* que describía el trabajo de matemáticos del Ejército con las primeras computadoras electrónicas (Geiger et al., 2019). En dicho texto, el especialista del Ejército William D. Mellin explicaba que las computadoras no pueden pensar por sí mismas y que entradas programadas descuidadamente producen inevitablemente salidas defectuosas. La paternidad exacta de la frase es discutida: a menudo se atribuye al programador e instructor de IBM George Fuechsel, quien la utilizaba durante la década de 1960 para explicar, de forma concisa, que un programa procesa exactamente aquello que recibe, bueno o malo.

Sin embargo, la intuición subyacente al principio es considerablemente más antigua. Ya en el siglo XIX, Charles Babbage, diseñador de la máquina diferencial y considerado un precursor de la computación, reportó que en dos ocasiones se le preguntó si su máquina produciría respuestas correctas cuando se le introdujeran valores incorrectos, a lo que respondió con perplejidad, reconociendo la imposibilidad de tal prodigio.

Esta anécdota, citada con frecuencia en la literatura sobre calidad de datos, anticipa el núcleo conceptual de GIGO: la computación no crea información nueva *ex nihilo* (de la nada), sino que transforma la que se le suministra.

### **2.1 De la regla pragmática al principio conceptual**

Con el paso de las décadas, GIGO dejó de ser solo una advertencia informal para los programadores y se consolidó como un principio teórico del tratamiento de la información. En términos formales, sostiene que la validez lógica de un proceso no garantiza la solidez de sus conclusiones: un argumento puede ser válido y aun así producir conclusiones falsas si parte de supuestos erróneos. Esta distinción, heredada de la lógica aristotélica, tiene enormes consecuencias prácticas en sistemas que automatizan decisiones.

Redman (1998) documentó, en un trabajo ampliamente citado, que los costos asociados a la mala calidad de datos en empresas típicas oscilan entre el 8 % y el 12 % de los ingresos, con impactos que

## **Jornada Académica Anual 2026 del Departamento de Sistemas**

incluyen menor satisfacción del cliente, aumento de costos operativos, peor toma de decisiones, dificultad para ejecutar estrategias, deterioro de la moral de los empleados y erosión de la confianza organizacional. Estas cifras, lejos de ser anecdóticas, cuantifican el precio concreto que pagan las organizaciones cuando ignoran la primera parte del principio GIGO<sup>1</sup>.

### **3 La “estupidez” como categoría analítica**

#### **3.1 Las cinco leyes fundamentales**

En 1976 circuló como panfleto entre colegas, y en 1988 se publicó como libro, el ensayo de Carlo M. Cipolla titulado *The Basic Laws of Human Stupidity*.

El autor, historiador económico de la Universidad de California en Berkeley, propuso cinco leyes fundamentales que, a pesar de su tono humorístico, describen rigurosamente interacciones sociales y económicas con consecuencias medibles.

Dichas leyes pueden sintetizarse así: primero, siempre se subestima el número de personas estúpidas en circulación; segundo, la probabilidad de que una persona sea estúpida es independiente de cualquier otra característica suya; tercero (la ley áurea), una persona estúpida es aquella que causa pérdidas a otra persona o grupo sin obtener beneficio para sí, e incluso con perjuicio propio; cuarto, las personas no estúpidas siempre subestiman el poder dañino de las estúpidas; y quinto, la persona estúpida es el tipo de persona más peligrosa que existe.

Cipolla representó gráficamente estas interacciones en un cuadrante con dos ejes: los beneficios o pérdidas que un agente provoca a sí mismo y los que provoca a otros. En su tipología, el agente inteligente produce ganancias para sí y para los demás; el bandido se beneficia a expensas de otros; el incauto beneficia a otros a costa propia; y el estúpido genera pérdidas para todos, incluido él mismo.

El propio autor observó que el estúpido es más peligroso que el bandido, pues el comportamiento del bandido, al tener lógica desagradable (*nasty rationality*), es predecible; el del estúpido, en cambio, carece de patrón racional y por ello escapa a toda defensa.

#### **3.2 Aplicación al contexto de los sistemas de información**

Trasladar este marco al uso de los sistemas de información resulta productivo. La “estupidez informática”, en sentido cipolliano, no se refiere a un déficit intelectual del usuario, sino a un patrón

---

<sup>1</sup> En agosto de 2012, la firma Knight Capital sufrió uno de los episodios más recordados de falla en el despliegue de software de trading automatizado. Un error en la liberación de código provocó que un sistema automatizado procesara secuencias inválidas de órdenes, generando transacciones erróneas sobre 148 valores y produciendo una pérdida cercana a 440 millones de dólares en aproximadamente 45 minutos. El caso ilustra de forma paradigmática cómo una combinación de entradas defectuosas, procesos automatizados y ausencia de controles humanos efectivos puede traducirse en pérdidas catastróficas. En clave GIGO, no se trató de una falla algorítmica abstracta, sino del resultado predecible de alimentar un sistema con configuraciones incorrectas en ausencia de mecanismos de validación.

## **Jornada Académica Anual 2026 del Departamento de Sistemas**

de interacción humano-sistema que produce pérdidas para el usuario (datos perdidos, decisiones erróneas, sanciones regulatorias) y para terceros (clientes afectados, empleados que heredan información corrupta, organizaciones que toman rumbos equivocados), sin que haya un beneficio compensatorio.

Dentro de esta categoría encajan conductas tan frecuentes como: reutilizar contraseñas triviales en sistemas críticos, deshabilitar controles de validación “para agilizar” una tarea, alimentar un cuadro de mando con datos que se sabe incompletos, copiar y pegar fórmulas sin revisar sus referencias, o aceptar recomendaciones de un algoritmo sin verificar su pertinencia.

La segunda ley de Cipolla adquiere aquí una relevancia particular: la proclividad a este tipo de comportamientos es independiente del nivel educativo, el cargo o la especialidad técnica. Altos directivos, ingenieros con posgrado y analistas con años de experiencia pueden, en el momento menos oportuno, ejecutar acciones que encajan con precisión en el cuadrante del daño mutuo. Esto convierte al fenómeno en un problema sistémico que no se resuelve solo con capacitación individual.

### **4 Taxonomía de errores humanos en la interacción con sistemas**

La investigación sobre errores humanos adquirió carácter sistemático con los trabajos de Donald Norman (1981) y, especialmente, de James Reason (2000). Este último propuso una distinción entre tres tipos de fallos: deslices (*slips*), lapsos (*lapses*) y equivocaciones (*mistakes*).

Los deslices y lapsos corresponden a fallos de ejecución: la intención era correcta, pero la acción se realizó mal (deslices, observables) o se omitió por un fallo de memoria (lapsos).

Las equivocaciones, en cambio, son fallos de planificación: la acción se ejecuta como se pretendía, pero la intención misma era inadecuada para alcanzar el objetivo. Reason apoyó su taxonomía en el marco *skill-rule-knowledge* propuesto por Rasmussen (1983), según el cual el comportamiento humano opera en tres niveles cognitivos.

En el nivel basado en habilidades, las acciones son automáticas y los errores típicos son deslices atencionales.

En el nivel basado en reglas, el usuario aplica procedimientos conocidos a situaciones familiares, y los errores derivan de aplicar una buena regla a una situación inapropiada o una regla defectuosa.

En el nivel basado en conocimiento, el usuario enfrenta situaciones novedosas donde debe razonar desde principios; los errores aquí se relacionan con comprensión incompleta, sesgos de confirmación, sobrecarga cognitiva y exceso de confianza.

Aplicado a los sistemas de información, este modelo explica fenómenos cotidianos: el analista experto que envía un correo al destinatario equivocado por un desliz de autocompletar; el desarrollador que olvida conectar un mecanismo de autenticación tras una actualización (lapso); el administrador que aplica una plantilla de seguridad a un servidor cuyo contexto difiere del habitual (equivocación basada en reglas); o el directivo que, ante un tablero de datos, interpreta erróneamente métricas por falta de experiencia específica (equivocación basada en conocimiento).

## **Jornada Académica Anual 2026 del Departamento de Sistemas**

Resulta decisivo comprender que estos tres tipos de errores exigen intervenciones distintas: los deslices y lapsos rara vez se corrigen con más capacitación, pero sí con mejor diseño y listas de comprobación; las equivocaciones, en cambio, requieren formación, revisión entre pares y validación procedimental.

Reason también distinguió entre fallos activos, cuyos efectos son inmediatos, y fallos latentes, que permanecen ocultos en el sistema hasta combinarse con otros factores para producir un accidente.

Su modelo del “queso suizo<sup>2</sup>”, adoptado por múltiples industrias, sostiene que los grandes incidentes resultan de la alineación fortuita de agujeros en capas sucesivas de defensas. Este marco es particularmente útil para analizar fallas en sistemas complejos como los financieros, sanitarios y de aviación, pero también encuentra aplicación directa en ciberseguridad, donde el incidente suele ser la suma de múltiples debilidades menores.

## **5 Límites tecnológicos y diseño centrado en el usuario**

Cuando un sistema se usa mal, la respuesta reflexiva es culpar al usuario. Sin embargo, Norman (2013), en la edición revisada de *The Design of Everyday Things*, insiste en invertir esa lectura: si un producto se usa mal con frecuencia, el problema está en el diseño. Su propuesta de diseño centrado en el usuario se apoya en principios como visibilidad, retroalimentación, restricciones, mapeo natural, consistencia y *affordances*<sup>3</sup> o potencialidades perceptibles.

La visibilidad implica que las funciones disponibles deben ser discernibles; la retroalimentación, que el sistema debe comunicar el resultado de cada acción; las restricciones, que limitar las acciones posibles evita errores; el mapeo natural, que la relación entre controles y efectos debe ser intuitiva; la consistencia, que operaciones similares deben ejecutarse de la misma forma; y las *affordances*, que los objetos deben sugerir su uso mediante su forma y disposición.

Norman acuñó también las ideas de brecha de ejecución (la distancia entre la intención del usuario y las acciones que el sistema permite) y brecha de evaluación (la distancia entre el estado del sistema y la comprensión que el usuario tiene de él).

Estas brechas, lejos de ser curiosidades académicas, explican un porcentaje significativo de los errores atribuidos a “factores humanos”.

Cuando un sistema presenta opciones visualmente equivalentes pero con efectos radicalmente distintos, la probabilidad de un desliz crece; cuando la retroalimentación ante un comando es tardía o ambigua, el usuario repite la acción o asume un resultado incorrecto. Norman sostiene, con una formulación que merece adopción como norma profesional, que se debe asumir que cualquier error

---

<sup>2</sup> El modelo del queso suizo de James Reason es una teoría de gestión de riesgos que explica cómo ocurren los accidentes en sistemas complejos. Compara las defensas de una organización con rebanadas de queso suizo apiladas. Los “agujeros” en el queso representan fallas o debilidades; los accidentes ocurren solo cuando los agujeros de varias capas se alinean

<sup>3</sup> Posibilidades en inglés.

## **Jornada Académica Anual 2026 del Departamento de Sistemas**

que pueda cometerse será cometido, y planificar en consecuencia. Esta actitud de *diseño para el error* desplaza la responsabilidad desde el individuo hacia el sistema sociotécnico.

Desde una perspectiva de ingeniería de sistemas, esto se traduce en la incorporación de mecanismos como *forcing functions* (funciones forzadoras que impiden acciones inválidas), validaciones en tiempo real, confirmaciones previas a acciones irreversibles, opciones de reversión y mensajes de error informativos que orientan al usuario hacia la corrección. La diferencia entre un sistema resiliente y uno frágil rara vez radica en la complejidad del algoritmo, sino en la calidad de estos mecanismos periféricos.

## **6 La dimensión empírica del GIGO: calidad de datos**

Si el principio GIGO establece que la calidad de la salida no puede superar la de la entrada, corresponde preguntarse qué se entiende por “calidad” en este contexto. Durante décadas la literatura técnica identificó calidad de datos con exactitud, una visión estrecha que Wang & Strong (1996) ampliaron decisivamente. Su marco, publicado en *Journal of Management Information Systems*, partió de un estudio empírico con consumidores de datos y consolidó 15 dimensiones agrupadas en cuatro categorías.

La categoría intrínseca reúne atributos que los datos poseen por sí mismos, como exactitud, objetividad, credibilidad y reputación.

La categoría contextual subraya que los datos deben evaluarse respecto a la tarea en la que se emplean: relevancia, valor añadido, oportunidad, completitud y cantidad apropiada.

La categoría representacional atiende a la forma en que los datos se presentan: interpretabilidad, facilidad de comprensión, consistencia representacional y concisión.

Finalmente, la categoría de accesibilidad incluye el acceso mismo a los datos y la seguridad de ese acceso.

Este marco es relevante para la discusión de la estupidez informática por al menos dos razones.

Primero, muestra que un dato puede ser técnicamente exacto y, sin embargo, de mala calidad si no es oportuno, si no es relevante para la decisión o si se presenta de manera ininteligible. Un reporte contable exacto entregado dos meses tarde no es útil; un indicador calculado correctamente pero sobre una población equivocada tampoco.

Segundo, desplaza el foco desde una visión técnica de la calidad hacia una visión orientada al usuario, lo que conecta directamente con el diseño centrado en la experiencia y con la taxonomía de errores humanos.

Redman (1998) cuantificó los costos organizacionales de los datos defectuosos y subrayó que las causas no son primariamente técnicas, sino organizacionales: ausencia de responsables claros, definiciones inconsistentes de términos de negocio, silos departamentales con criterios divergentes, incentivos desalineados e inversión insuficiente en gobierno de datos.

## **Jornada Académica Anual 2026 del Departamento de Sistemas**

La consecuencia es la coexistencia de múltiples “realidades paralelas” en los sistemas, cada una fielmente registrada pero incompatible con las demás.

### **7 Sesgos cognitivos y la ilusión de competencia digital**

Si se acepta que los errores humanos tienen taxonomía, y que su frecuencia depende tanto del diseño como de la arquitectura organizacional, aún queda un componente que merece análisis específico: los sesgos cognitivos que distorsionan la percepción del propio desempeño.

El efecto Dunning-Kruger, formulado originalmente por Kruger & Dunning (2009), describe la tendencia de las personas con baja competencia en un dominio a sobrestimar sus capacidades, precisamente porque la metacognición necesaria para reconocer la propia incompetencia exige un mínimo de competencia que no se posee.

En el contexto del uso de sistemas de información, este efecto tiene implicaciones serias. Usuarios con conocimientos superficiales pueden asumir que dominan una plataforma analítica, un gestor de bases de datos o una suite de productividad, cuando en realidad ignoran funciones críticas, supuestos subyacentes y riesgos asociados. Una revisión sistemática sobre alfabetización informacional mostró que, en la mayoría de estudios, quienes obtienen peores resultados objetivos sobrestiman significativamente su competencia percibida. Esta brecha entre competencia real y percibida es especialmente peligrosa cuando la decisión que se toma depende de datos que el usuario interpreta sin las salvaguardas adecuadas.

Un hallazgo reciente amplía esta discusión. (Fernandes et al., 2026) reportaron que, en interacciones con sistemas de inteligencia artificial generativa como ChatGPT, el efecto Dunning–Kruger clásico desaparece, y en su lugar aparece una sobreestimación generalizada del desempeño. Más sorprendentemente aún, los usuarios con mayor alfabetización en IA mostraron menor precisión metacognitiva que los novatos: a mayor conocimiento declarativo sobre IA, mayor exceso de confianza en el rendimiento real obtenido con su ayuda. Los autores atribuyen este fenómeno al *cognitive offloading* o descarga cognitiva: la tendencia a delegar el esfuerzo de verificación en el sistema, confiando en sus salidas sin someterlas a contraste crítico.

Este resultado es relevante para la tesis central del artículo. La estupidez cipolliana en el uso de sistemas no deriva necesariamente de la ignorancia, sino también del exceso de confianza mal calibrado. Un experto puede comportarse como usuario “estúpido” si su confianza en el sistema lo lleva a omitir verificaciones que otro, más cauto, realizaría. La calibración adecuada de la confianza, por tanto, se perfila como una competencia crítica en la era de la IA.

### **8 La vigencia del principio GIGO en la era de la inteligencia artificial**

La irrupción de los modelos de lenguaje a gran escala y los sistemas de IA generativa ha traído consigo una reactivación del principio GIGO bajo nuevas formas. En el aprendizaje automático supervisado, la calidad del modelo depende crucialmente de la calidad del conjunto de

## ***Jornada Académica Anual 2026 del Departamento de Sistemas***

entrenamiento: datos sesgados, incompletos o mal etiquetados producen modelos que reproducen y amplifican esos defectos.

Casos documentados de modelos de reconocimiento facial con desempeño desigual entre grupos demográficos, algoritmos de selección de personal que penalizan a mujeres, o sistemas de evaluación crediticia que perpetúan discriminaciones históricas, son manifestaciones del GIGO aplicado a la escala de las redes neuronales profundas.

En los sistemas de generación aumentada con recuperación (RAG) y en los agentes autónomos, el problema se desplaza también hacia la calidad de las fuentes de conocimiento consultadas: si la base documental contiene información obsoleta, incorrecta o inconsistente, la salida generada la heredará. A su vez, el diseño del prompt se convierte en una nueva categoría de entrada crítica: prompts ambiguos, mal estructurados o cargados de presupuestos no explicitados producen respuestas que, aun siendo lingüísticamente correctas, son conceptualmente defectuosas.

Un aspecto particularmente inquietante es la facilidad con la que las salidas de estos sistemas adoptan una apariencia de autoridad. La fluidez lingüística de un modelo generativo puede confundirse con fiabilidad epistémica, llevando al usuario a aceptar afirmaciones inexactas como información verificada.

Aquí confluyen los tres hilos del análisis: 1) el principio GIGO (entradas insuficientes producen salidas defectuosas), 2) la taxonomía de errores humanos (el usuario comete una equivocación basada en conocimiento al no reconocer la situación como novedosa y requerir verificación externa), y 3) la estupidez cipolliana (la acción resultante daña al usuario y a terceros sin beneficio real).

La respuesta a este desafío no es rechazar la tecnología, ni tampoco confiar ciegamente en ella. Pasa por rediseñar la interacción humano-sistema para fomentar la metacognición: interfaces que expliciten incertidumbre, que soliciten verificación en decisiones críticas, que muestren las fuentes consultadas, y que recuerden al usuario que la responsabilidad final del resultado sigue siendo suya. Desde esta perspectiva, la alfabetización en IA no consiste en saber cómo funciona la IA, sino en saber cuándo desconfiar de su salida.

## **9 GIGO en los sistemas de IA: alucinaciones, referencias inventadas y la responsabilidad del prompt**

### **9.1 La queja recurrente: “la IA alucina”**

Resulta difícil exagerar la frecuencia con la que, en foros profesionales, redes sociales y conversaciones académicas, se escucha alguna variante de la siguiente queja: “la inteligencia artificial alucina, inventa datos, fabrica referencias bibliográficas, atribuye citas a autores que nunca las escribieron y, en general, no se puede confiar en ella”. Estas observaciones, en muchos casos, son fácticamente correctas: los modelos de lenguaje a gran escala efectivamente producen contenido fabricado, y la literatura técnica documenta el fenómeno con detalle. Sin embargo, el modo en que esa queja suele formularse contiene un error analítico relevante: presenta el problema como una

## **Jornada Académica Anual 2026 del Departamento de Sistemas**

propiedad exclusiva del sistema, eximiendo al usuario de cualquier responsabilidad sobre la calidad de las entradas que él mismo proporciona.

Ji et al. (2024), en una revisión sistemática publicada en *ACM Computing Surveys*, propusieron una taxonomía hoy ampliamente aceptada en la que distinguen dos clases de alucinaciones: las intrínsecas, en las que la salida contradice la fuente o el contexto proporcionado, y las extrínsecas, en las que la salida introduce información que no puede verificarse contra la fuente.

Esta distinción es crucial para encuadrar correctamente la discusión: las primeras suelen estar asociadas a problemas del modelo y los datos de entrenamiento; las segundas, en cambio, se ven fuertemente moduladas por la calidad del contexto suministrado por el usuario en su prompt (Geiger et al., 2019; Klingbeil et al., 2024; Wand & Wang, 1996).

### **9.2 Cuando el prompt es la “garbage in”**

La aplicación rigurosa del principio GIGO a los sistemas de IA generativa exige reconocer que el prompt es, en sentido técnico estricto, parte de la entrada del sistema. Un modelo de lenguaje no “responde a una pregunta” en el sentido humano del término: predice secuencias de tokens condicionadas por el contexto que recibe. Si ese contexto es vago, ambiguo, contradictorio, carente de marcos de referencia o solicita información sobre la que el modelo no tiene base sólida, la probabilidad de obtener una salida defectuosa aumenta drásticamente.

La frase coloquial “la IA inventó referencias” omite, en la mayoría de los casos, una observación elemental: ¿se le pidió al sistema que verificara fuentes?, ¿se le proporcionó una base documental contra la cual contrastar?, ¿se acotó el dominio de la respuesta?, ¿se especificó qué hacer en caso de no conocer la respuesta?

La investigación reciente ha mostrado, de manera consistente, que la calidad del prompt impacta significativamente la frecuencia de alucinaciones. Estrategias como la asignación de un rol explícito al modelo, la delimitación de fronteras temáticas, la inclusión de contexto verificado mediante generación aumentada con recuperación (RAG), la instrucción explícita de responder “no lo sé” ante incertidumbre, y la solicitud de razonamiento paso a paso, han demostrado reducir mensurablemente la incidencia de errores fabricados. Aunque ninguna técnica de prompting elimina por completo el fenómeno, su omisión sistemática garantiza que el problema se mantenga o agrave.

Esta constatación reactiva el principio GIGO con plena fuerza: el prompt mal diseñado es basura epistemológica que entra al sistema, y la salida fabricada es, en buena medida, la basura que sale.

Cuando un usuario solicita a un modelo de lenguaje “escribe un artículo sobre tal tema con cinco referencias bibliográficas”, sin proporcionar acceso a una base bibliográfica real, sin pedir verificación de existencia y sin instruir al sistema sobre cómo manejar la incertidumbre, está literalmente solicitando un texto plausible, no un texto verdadero. De la misma forma que las personas deben aprender como indica (Argyris, 1991), de idéntica forma deben hacerlo las tecnologías.

Que el resultado contenga referencias inexistentes no es un fallo inesperado del sistema: es la respuesta predecible a una entrada que no requirió fidelidad referencial.

### **9.3 Las causas múltiples de la alucinación**

Sería injusto, sin embargo, atribuir todas las alucinaciones al usuario. La literatura identifica al menos tres niveles causales que conviene mantener analíticamente separados.

En primer lugar, existen causas relativas al modelo y a sus datos de entrenamiento: lagunas de conocimiento, sesgos de cobertura, presencia de información errónea en el corpus, y la naturaleza fundamentalmente probabilística de la generación, que optimiza plausibilidad lingüística antes que veracidad factual.

En segundo lugar, existen causas arquitectónicas y de despliegue: parámetros de muestreo demasiado permisivos, ausencia de mecanismos de verificación contra fuentes externas, falta de calibración de incertidumbre.

En tercer lugar, y este es el nivel sobre el que el usuario tiene control directo, están las causas relativas a la entrada: prompts ambiguos, contexto insuficiente, ausencia de instrucciones de fidelidad, y delegación acrítica de tareas que requieren verificación humana.

La trampa argumentativa en la que con frecuencia caen los críticos de la IA consiste en colapsar los tres niveles bajo la categoría “la IA alucina”, sin preguntarse cuál de ellos es el más relevante en el caso concreto que les afecta.

Esta confusión categorial genera una doble distorsión: por un lado, exime al usuario de mejorar sus prácticas; por el otro, frustra los esfuerzos legítimos de mejora del modelo, pues los problemas se atribuyen al lugar equivocado (Ackerman & Thompson, 2017; Bastani et al., 2024; Ghassemi et al., 2021). Una analogía útil: cuando un buscador devuelve resultados irrelevantes, parte de la responsabilidad puede ser del algoritmo, pero otra parte recae en la formulación de la consulta. Quien escribe una búsqueda de tres letras y luego se queja de que el buscador “no entiende lo que quiere” está omitiendo la primera mitad del intercambio.

### **9.4 La estupidez cipolliana en el uso de la IA**

Aplicando el marco de Cipolla (1988) al uso de los sistemas de IA generativa, emerge un patrón conductual que puede caracterizarse con precisión técnica como estúpido en sentido cipolliano.

Considérese un usuario que: solicita a un modelo la redacción de un informe sin proporcionar contexto, no especifica el formato esperado de las referencias, no instruye al sistema sobre cómo manejar la falta de información, no verifica las citas resultantes contra fuentes reales, presenta el documento ante terceros como propio, y, cuando se descubre que las referencias son inventadas, atribuye la falla exclusivamente a la herramienta.

Las pérdidas resultantes afectan al propio usuario (deterioro de su credibilidad profesional, posibles consecuencias disciplinarias o legales) y a terceros (lectores desinformados, debate académico contaminado, deterioro de la confianza pública en los sistemas de IA). Y, sin embargo, no se obtiene ningún beneficio real: lo que se ahorró en tiempo de redacción se pierde con creces en correcciones, sanciones y reputación (Diario Judicial, 2026; Facultad de Ciencias Jurídicas y Sociales, 2026).

## ***Jornada Académica Anual 2026 del Departamento de Sistemas***

La cuarta ley de Cipolla (las personas no estúpidas siempre subestiman el poder dañino de las estúpidas) cobra aquí especial vigencia. Las organizaciones que despliegan sistemas de IA sin definir políticas claras de uso, sin formar a su personal en *prompt engineering* básico, sin establecer mecanismos de verificación obligatoria y sin asignar responsabilidad sobre las salidas, multiplican exponencialmente la superficie de exposición a este tipo de fallos. La tecnología, por sí sola, no resuelve el problema; lo amplifica.

### **9.5 Hacia un uso epistémicamente responsable de la IA**

De lo expuesto se desprende un conjunto de prácticas que, sin pretender exhaustividad, configuran un uso epistémicamente responsable de los sistemas de IA generativa. P

Primero, formular prompts explícitos que incluyan rol, contexto, restricciones, formato esperado y comportamiento ante incertidumbre.

Segundo, proporcionar al sistema, siempre que sea posible, fuentes verificadas mediante mecanismos de recuperación, en lugar de delegar en su memoria paramétrica.

Tercero, verificar manualmente toda afirmación factual sensible, en particular referencias bibliográficas, datos numéricos, citas literales y atribuciones a personas.

Cuarto, mantener al humano como responsable último de la salida: la firma, la decisión y las consecuencias siguen siendo suyas.

Quinto, formar a los usuarios no solo en el manejo operativo de las herramientas, sino en la metacognición necesaria para identificar cuándo deben desconfiar de ellas.

En suma, el principio GIGO aplicado a la IA no exime al sistema de sus deficiencias intrínsecas, pero tampoco permite al usuario lavarse las manos. Reconocer que el prompt es entrada, y que las entradas determinan las salidas, es el primer paso para superar la dialéctica estéril entre la fascinación acrítica y el rechazo indiscriminado. Lo que las organizaciones necesitan no es una IA que nunca falle —eso es, en el horizonte previsible, técnicamente imposible—, sino usuarios capaces de interactuar con ella sin fallar ellos mismos.

## **10 Conclusiones**

El principio GIGO, lejos de ser una reliquia de los primeros días de la computación, conserva y renueva su vigencia en la era algorítmica. Las entradas relevantes ya no son solo tarjetas perforadas o registros en una base de datos relacional, sino también prompts, datos de entrenamiento, configuraciones de modelos y, de manera decisiva, los supuestos que el usuario introduce de forma implícita en cada interacción. La sofisticación del procesamiento ha aumentado exponencialmente; la calidad de las entradas, en cambio, sigue siendo el cuello de botella.

La interacción humano-sistema, vista a través de los marcos de Cipolla, Reason, Norman, Wang y Strong, Kruger y Dunning, Ji et al. y la investigación reciente sobre IA generativa, revela que los fallos raramente tienen una causa única.

***Jornada Académica Anual 2026 del Departamento de Sistemas***

Los deslices y lapsos del nivel atencional se combinan con equivocaciones de planificación, sesgos metacognitivos, defectos de diseño, carencias en la calidad de los datos, alucinaciones de los modelos generativos y patrones de conducta que Cipolla caracterizaría como estúpidos: acciones que dañan al usuario y a terceros sin beneficio.

El desafío de las organizaciones no es eliminar el error humano ni garantizar la infalibilidad algorítmica, metas utópicas, sino construir sistemas sociotécnicos que resistan errores sin colapsar.

De este análisis se desprenden al menos cinco implicaciones prácticas.

Primero, el diseño de interfaces debe asumir el error como eventualidad y construir defensas en capas, siguiendo el modelo del queso suizo.

Segundo, la alfabetización digital contemporánea debe incluir la calibración metacognitiva: enseñar a los usuarios no solo cómo usar los sistemas, sino cuándo desconfiar de sus salidas.

Tercero, el gobierno de datos debe instituirse como función organizacional con responsables claros, definiciones compartidas y métricas de calidad que abarquen las dimensiones intrínseca, contextual, representacional y de accesibilidad.

Cuarto, el uso de sistemas de IA generativa exige formación específica en diseño de prompts, verificación de salidas y atribución de responsabilidad, dado que el prompt es parte constitutiva de la entrada y, por tanto, parte constitutiva del problema GIGO.

Quinto, la formación de profesionales en sistemas de información debe articular los conocimientos técnicos con una comprensión robusta de los factores humanos, cognitivos y organizacionales que determinan el éxito o el fracaso de cualquier implementación.

En síntesis, la estupidez en el uso de los sistemas de información no es un fenómeno marginal ni exótico, sino un subproducto regular de la interacción entre humanos falibles y tecnologías poderosas, mediada por diseños imperfectos y datos de calidad variable.

Reconocer el principio GIGO como clave interpretativa vigente y situarlo en diálogo con las leyes de Cipolla, las taxonomías de error, los marcos de calidad de datos y la investigación sobre alucinaciones en IA permite abordar el problema con la seriedad analítica que merece y con el pragmatismo que exigen las organizaciones contemporáneas.

Los sistemas no piensan por nosotros; procesan lo que les entregamos. El margen de mejora sigue estando, en gran medida, del lado humano de la interacción.

## **11 Referencias**

- Ackerman, R., & Thompson, V. A. (2017). Meta-Reasoning: Monitoring and Control of Thinking and Reasoning. *Trends in Cognitive Sciences*, 21(8), 607–617. <https://doi.org/10.1016/j.tics.2017.05.004>
- Argyris, C. (1991). Teaching Smart People How to Learn. *Harvard Business Review*, 4(2), 4–15. <https://doi.org/10.1162/152417302762251291>

**Jornada Académica Anual 2026 del Departamento de Sistemas**

- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). *Generative AI Can Harm Learning*. <https://doi.org/10.2139/ssrn.4895486>
- Cipolla, C. M. (1988). *Le leggi fondamentali della stupidità umana*.
- Diario Judicial. (2026). *Abogados: el mal uso de la IA no tiene amparo*. <https://www.diariojudicial.com/news-103168-abogados-el-mal-uso-de-la-ia-no-tiene-amparo>
- Facultad de Ciencias Jurídicas y Sociales. (2026). *Uso de IA en tribunales: sanción de multa y trámite disciplinario*. <https://www.jursoc.unlp.edu.ar/index.php/institucional/politicas-digitales-menu/item/4838-importante-uso-de-ia-en-tribunales-sancion-de-multa-y-tramite-disciplinario.html>
- Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., Kosch, T., Shen, C., & Welsch, R. (2026). AI makes you smarter but none the wiser: The disconnect between performance and metacognition. *Computers in Human Behavior*, 175, 108779. <https://doi.org/10.1016/J.CHB.2025.108779>
- Geiger, R. S., Yu, K., Yang, Y., Dai, M., Qiu, J., Tang, R., & Huang, J. (2019). *Garbage In, Garbage Out? Do Machine Learning Application Papers in Social Computing Report Where Human-Labeled Training Data Comes From?* <https://doi.org/10.1145/3351095.3372862>
- Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, 3(11), e745–e750. [https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9)
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Chan, H. S., Madotto, A., & Fung, P. (2024). *Survey of Hallucination in Natural Language Generation*. <https://doi.org/10.1145/3571730>
- Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, 160. <https://doi.org/10.1016/J.CHB.2024.108352>
- Kruger, J., & Dunning, D. (2009). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. In *Psychology* (Vol. 1). <http://www.scirp.org/journal/psych>
- Norman, D. A. (1981). Categorization of Action Slips. In *Psychological Review VOLUME* (Vol. 88).
- Norman, D. A. . (2013). *The design of everyday things*. Basic Books.
- Rasmussen, J. (1983). *Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models* (Number 3).
- Reason, J. (2000). *Human error: models and management*.

**Jornada Académica Anual 2026 del Departamento de Sistemas**

Redman, T. C. (1998). *The impact of poor data quality on the typical enterprise*.

Wand, Y., & Wang, R. Y. (1996). *Anchoring Data Quality Dimensions in Ontological Foundations*.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–34.  
<https://doi.org/10.1080/07421222.1996.11518099>