

Construcción de un paradigma de diagnóstico inteligente para la evaluación urbana por indicadores mediante la integración de grandes modelos de lenguaje

Autores

WANG Kangxu, ZHANG Shanqi, CHEN Cheng, ZHEN Feng, QIN Xiao, ZHANG Yishu

Resumen

En el contexto de la transformación digital e inteligente de la gobernanza del espacio territorial, el examen y la evaluación de las ciudades están pasando del monitoreo de indicadores al diagnóstico integral. Sin embargo, la evaluación centrada en indicadores sigue tendiendo a describir las anomalías de manera superficial, mientras que la integración entre el análisis de datos y el conocimiento profesional de la planificación continúa siendo débil. Este artículo entiende la evaluación por indicadores como un proceso cognitivo sistemático para interpretar el estado operativo de sistemas urbanos complejos, y propone una ontología diagnóstica S-E-D compuesta por Síntoma, Explicación y Decisión. Sobre esta base, desarrolla un método basado en un gran modelo de lenguaje reforzado por conocimiento, capaz de apoyar el análisis de indicadores, la explicación causal y la generación de estrategias bajo restricciones de conocimiento planificador, mediante el procesamiento estructurado de documentos de planificación y normas técnicas. El estudio de caso de Nanjing muestra que el método propuesto funciona adecuadamente en la integración de múltiples indicadores, la consistencia del razonamiento y la interpretabilidad de las recomendaciones estratégicas. El estudio ofrece una referencia metodológica para transformar la evaluación urbana por indicadores desde una práctica dependiente de la experiencia hacia un diagnóstico inteligente, reproducible y trazable.

Palabras clave

examen y evaluación urbanos; gobernanza del espacio territorial; grandes modelos de lenguaje; diagnóstico inteligente

1 Contexto: transformación digital e inteligente del examen y la evaluación urbanos

1.1 Cambio de paradigma: del monitoreo de indicadores a la cognición inteligente

El examen y la evaluación urbanos constituyen una herramienta importante para supervisar la implementación de la planificación espacial territorial y mejorar la calidad de la gobernanza espacial. En los últimos años, con la consolidación gradual del sistema de seguimiento de implementación en el marco de la "integración de múltiples planes", el examen urbano se ha convertido en un mecanismo regular para comprender el funcionamiento de la ciudad, identificar problemas de ejecución de la planificación y apoyar ajustes adaptativos de política pública [1-3]. Las prácticas locales se han desarrollado en torno al examen de indicadores, el examen de tareas y el examen temático; entre ellos, el examen de indicadores ofrece la vía más directa para percibir el estado operativo del espacio urbano [4].

La evaluación tradicional de indicadores se apoya principalmente en la comparación estadística, la contrastación con objetivos y la elaboración periódica de informes. Es eficaz para registrar si un indicador aislado alcanza su meta, pero tiene menor capacidad para explicar por qué se producen las desviaciones o cómo deben abordarse. Con la expansión de la infraestructura digital, los datos espaciotemporales y las plataformas de operación urbana, el examen de indicadores se desplaza desde

la revisión estadística ex post hacia la percepción dinámica y el monitoreo continuo. La construcción de la red de seguimiento de la implementación de la planificación espacial territorial refuerza aún más la demanda de evaluación dinámica, diagnóstico inteligente y apoyo a la toma de decisiones [5-7]. En este marco, la tarea central del examen de indicadores ya no consiste solo en vigilar valores, sino en construir un proceso cognitivo inteligente que integre identificación de problemas, explicación de mecanismos y generación de estrategias.

1.2 Restricciones prácticas: dependencia de la experiencia e integración insuficiente del conocimiento

Aunque los datos multisource han ampliado la cobertura y la oportunidad del monitoreo, la capacidad diagnóstica profunda de la práctica actual sigue limitada por dos problemas. El primero es la fuerte dependencia de la experiencia profesional. Los sistemas urbanos son complejos, acoplados, multiescalares y no lineales; los cambios en los indicadores pueden involucrar efectos retardados, interacciones intersectoriales y vínculos causales ocultos [8]. Los planificadores suelen tener que interpretar datos fragmentados y de alta frecuencia mediante juicio profesional, pero resulta difícil extraer directamente cadenas causales estables. Por ello, las conclusiones de la evaluación pueden quedarse en una lista de indicadores anómalos y depender excesivamente de la experiencia individual, lo que dificulta identificar causas estructurales y reglas de decisión transferibles [9].

El segundo problema es la integración insuficiente del conocimiento de planificación. Los documentos de política, las directrices técnicas, los planes estatutarios, los informes de evaluación y la experiencia experta constituyen conjuntamente la lógica profesional del diagnóstico urbano. Los métodos digitales existentes son relativamente fuertes en cálculo y visualización, pero débiles para organizar textos de planificación no estructurados e incorporarlos a los procesos de razonamiento. Esto genera una brecha entre los datos de monitoreo y el conocimiento profesional de la planificación [10]. Los métodos basados en datos no representan de forma natural las reglas institucionales, los mecanismos de planificación y la lógica de la gobernanza espacial, mientras que el razonamiento basado en reglas es limitado cuando debe integrar fuentes heterogéneas de conocimiento [11]. Por tanto, se requiere un nuevo marco que conecte las correlaciones de datos con las restricciones del conocimiento.

1.3 Oportunidad técnica: razonamiento inteligente reforzado por conocimiento

Los grandes modelos de lenguaje ofrecen una nueva oportunidad técnica para el diagnóstico de indicadores urbanos gracias a su capacidad de comprensión semántica, asociación de conocimiento y generación de lenguaje natural. Los modelos lingüísticos preentrenados reforzados por conocimiento, la generación aumentada por recuperación, la recuperación densa y el razonamiento basado en agentes ya se han aplicado en campos como las finanzas, la contabilidad, la medicina, la educación y la asistencia a la programación [13-21]. Su valor no reside solo en la generación textual, sino también en organizar conocimiento de dominio, recuperar evidencias pertinentes y construir cadenas de razonamiento estructuradas.

Para el examen urbano, los grandes modelos de lenguaje pueden combinar textos de planificación, estándares de política, datos de indicadores e información contextual, apoyando así la interpretación semántica y el razonamiento en múltiples pasos. Sin embargo, las aplicaciones actuales en planificación se concentran principalmente en la interpretación de políticas, los sistemas de preguntas y respuestas basados en conocimiento o la explicación aislada de indicadores. Todavía no han formado un paradigma completo que conecte comprensión de indicadores, explicación causal y juicio integral de planificación [22-23]. Por ello, este artículo conceptualiza el examen de indicadores como un proceso de diagnóstico

cognitivo de sistemas urbanos complejos, construye una ontología S-E-D y explora un método de análisis inteligente basado en un gran modelo de lenguaje reforzado por conocimiento.

2 Construcción de la ontología cognitiva y mecanismo de refuerzo por conocimiento

2.1 Reconstrucción de la ontología diagnóstica: la ontología cognitiva S-E-D

El examen urbano por indicadores debe entenderse no como un simple monitoreo técnico, sino como un proceso diagnóstico integral orientado a sistemas urbanos adaptativos complejos [24]. Su objeto no es la comparación estática de indicadores aislados, sino las relaciones de acoplamiento, las desviaciones y los riesgos entre subsistemas espaciales, sociales, económicos, ecológicos e institucionales [25-26]. La comparación normativa y la ponderación lineal pueden determinar si un indicador se desvía de una meta, pero no explican suficientemente las causas, las trayectorias ni las implicaciones de gobernanza de esa desviación.

Para responder a esta limitación, el artículo propone una ontología cognitiva para el diagnóstico de indicadores: S-E-D, correspondiente a Síntoma, Explicación y Decisión. La ontología reorganiza el proceso de evaluación en una cadena cognitiva cerrada que va de la percepción del estado a la atribución de mecanismos y, posteriormente, a la respuesta de planificación.

2.1.1 Capa de síntoma: percepción del estado e identificación de desviaciones

La capa de síntoma traduce el estado operativo de la ciudad en señales de problema reconocibles y comparables. Identifica si los indicadores se desvían de los objetivos de planificación, los estándares técnicos, las trayectorias históricas o los rangos esperados, y presta atención a las dinámicas temporales y a las asociaciones entre indicadores. En comparación con la comparación estática de valores, esta capa enfatiza el significado semántico de los indicadores en contextos específicos de planificación. Su resultado es un conjunto de unidades estructuradas de síntoma, como disminución, fluctuación, desajuste, presión de umbral o coordinación insuficiente.

2.1.2 Capa de explicación: comprensión de causas y atribución de mecanismos

La capa de explicación desplaza el análisis desde la pregunta "qué ha cambiado" hacia "por qué ha cambiado". Vincula el desempeño anómalo de los indicadores con la estructura espacial, la dotación de recursos, la etapa de desarrollo, las restricciones institucionales y los procesos socioeconómicos. Mediante restricciones de conocimiento planificador, el modelo puede evitar explicaciones generalizadas y desligadas del contexto local, transformando valores anómalos en interpretaciones causales con evidencia trazable. Esta capa constituye el puente clave entre el monitoreo de datos y el juicio profesional de planificación.

2.1.3 Capa de decisión: generación de estrategias y salida de decisión

La capa de decisión convierte el conocimiento diagnóstico en estrategias de gobernanza. La estrategia debe corresponder a las causas identificadas, los objetivos de planificación, las condiciones institucionales y las rutas de implementación. No debe limitarse a producir recomendaciones genéricas, sino organizar las contramedidas según la cadena de reconocimiento de síntomas, diagnóstico de mecanismos e intervención estratégica. En este sentido, el examen de indicadores se convierte en un proceso de razonamiento planificador basado en evidencias, no en un informe meramente descriptivo.

2.2 Ruta de razonamiento inteligente bajo restricciones de conocimiento

La ruta de razonamiento reforzado por conocimiento contiene tres procesos vinculados. El primero es el modelado y la organización estructurada del conocimiento. Los documentos de planificación, las normas técnicas, los textos de política, los informes de evaluación y la experiencia experta se limpian,

segmentan, normalizan y transforman en unidades semánticas computables. Modelos de incrustación como S-BERT y BGE pueden mapear estas unidades en espacios vectoriales densos y formar una base de conocimiento de planificación específica del dominio [27-30].

El segundo proceso es la recuperación semántica y la programación dinámica del conocimiento. Cuando se detecta una anomalía en un indicador, el sistema recupera fragmentos de conocimiento altamente relevantes según el tipo de indicador, el objeto de planificación, el contexto espacial y la descripción del problema. La recuperación vectorial densa capta similitud semántica, mientras que la recuperación dispersa por palabras clave y las restricciones simbólicas mantienen una correspondencia precisa con términos de política local, estándares y categorías de planificación [31-32].

El tercer proceso es la generación fusionada y el control de la ruta de razonamiento. La generación aumentada por recuperación introduce evidencias de planificación en el contexto del modelo, mientras que las indicaciones estructuradas guían al modelo a seguir la cadena S-E-D [33-36]. Los hiperparámetros de generación restringen adicionalmente el comportamiento de muestreo para que la salida permanezca estable, coherente y anclada en el conocimiento recuperado. En comparación con el ajuste fino por dominio, este enfoque tiene menor costo de entrenamiento y reduce el riesgo de sobreajuste. Cambia el papel del modelo: de almacenar internamente conocimiento planificador a invocarlo, organizarlo y razonar con él.

Tabla 1. Configuración de hiperparámetros de generación del gran modelo de lenguaje y mecanismos de regulación.

Parámetro	Valor	Rango	Función	Mecanismo de regulación
Temperature	0,3	0-2	Controla la aleatoriedad y diversidad de la generación	Un valor bajo mejora la estabilidad del razonamiento y la coherencia lógica.
Top P	0,1	0-1	Trunca el conjunto de tokens candidatos según la probabilidad acumulada	Comprime el espacio de candidatos, reduce expresiones de baja probabilidad y aumenta la determinación de la salida.
Frequency penalty	0	0-1	Reduce la probabilidad de repetir palabras de alta frecuencia	Conserva la recurrencia terminológica para mantener la

				consistencia de la expresión diagnóstica estructurada.
Presence penalty	0	0-1	Favorece la introducción de contenido nuevo	Se ajusta para evitar desviaciones respecto del conocimiento recuperado y mantener el análisis centrado en la información dada.
Max tokens	16384	0-16384	Controla la longitud máxima de una generación	Permite entradas largas y razonamiento en múltiples pasos.

3 Método de razonamiento inteligente basado en la ontología S-E-D

Este artículo desarrolla un método de razonamiento inteligente para la evaluación urbana por indicadores, centrado en grandes modelos de lenguaje. El método integra una base de conocimiento de planificación, información contextual de indicadores e indicaciones de razonamiento restringido, guiando al modelo por la cadena de análisis de síntomas, atribución explicativa y generación de decisiones. El objetivo es hacer que el diagnóstico de indicadores sea más reproducible, trazable e interpretable.

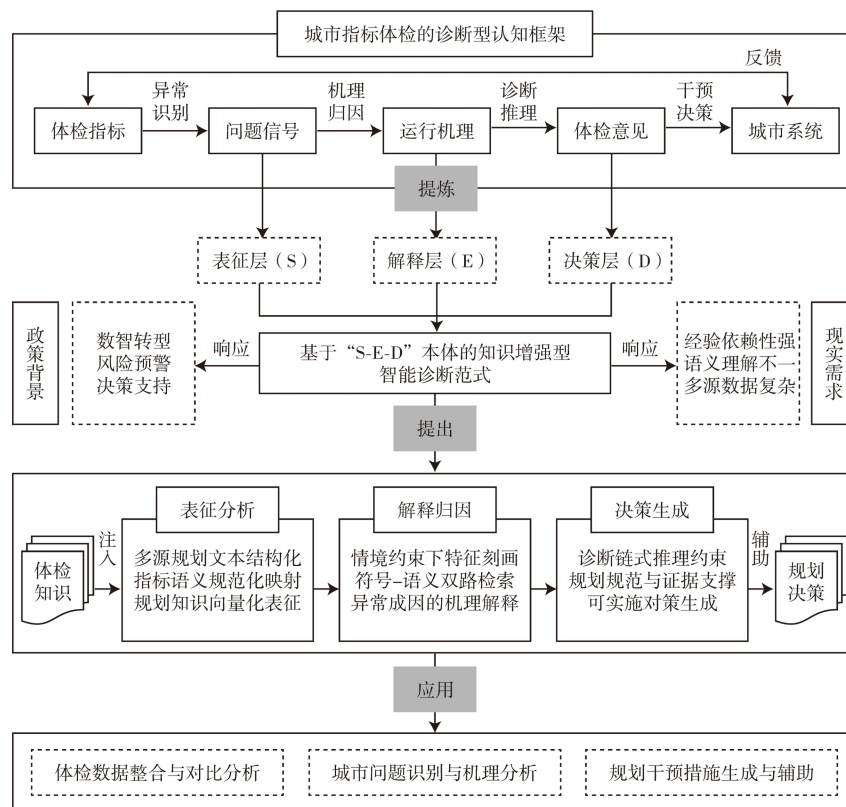


Figura 1. Marco técnico del diagnóstico inteligente para la evaluación urbana por indicadores basado en el ciclo cognitivo cerrado S-E-D.

3.1 Análisis de síntomas: construcción de la base de conocimiento de planificación

El análisis de síntomas comienza con la construcción de una base de conocimiento de planificación. Documentos de política, procedimientos técnicos, textos de planificación estatutaria e informes de examen urbano se transforman en unidades de conocimiento recuperables y computables mediante limpieza textual, análisis estructural y normalización terminológica. Los datos de indicadores se etiquetan por categoría, año, valor, estándar, umbral y objetivo de planificación, mientras que los textos de planificación se descomponen en elementos semánticos como objeto, estado, problema, causa y contramedida.

Para apoyar una recuperación precisa y la reconstrucción contextual, la base de conocimiento adopta una estructura semántica de dos niveles. El nivel superior registra el dominio temático, el objeto de planificación y la fuente institucional de cada unidad de conocimiento; el nivel inferior registra el significado diagnóstico, la lógica causal y la respuesta de política correspondiente. Tras la vectorización, la base puede apoyar la búsqueda por similitud, la recuperación de evidencias y el ensamblaje del contexto para el razonamiento del modelo.

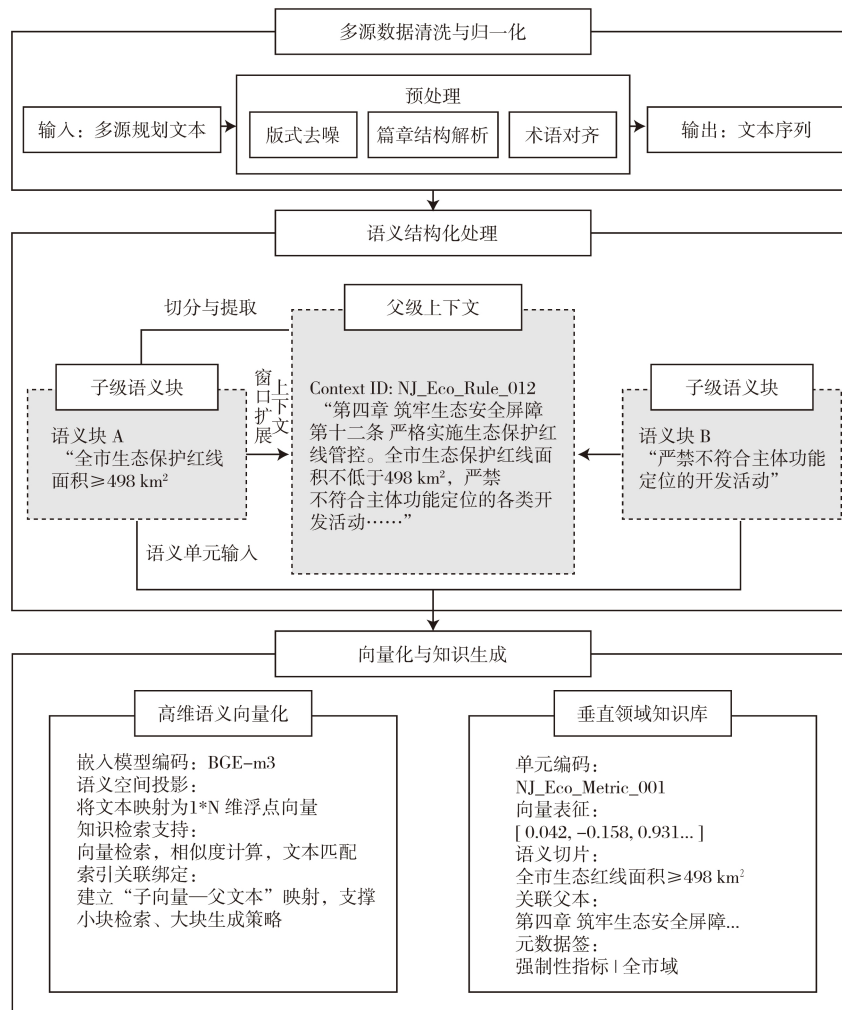


Figura 2. Proceso de deconstrucción semántica y construcción de la base de conocimiento de planificación orientada al diagnóstico por indicadores.

3.2 Atribución explicativa: estrategia de recuperación impulsada por conocimiento

Una vez identificados los indicadores anómalos, el foco analítico se desplaza de si existe una anomalía a por qué se produce. Para evitar que un modelo general produzca explicaciones desvinculadas del contexto espacial, el método propuesto sitúa los cambios de indicadores dentro de condiciones de localización, atributos funcionales, etapas de desarrollo y restricciones institucionales mediante la base de conocimiento de planificación. La recuperación combina restricciones simbólicas y similitud semántica. Las restricciones simbólicas acotan el dominio del objeto y garantizan que las evidencias recuperadas correspondan a la categoría de planificación, la unidad espacial y el contexto de política adecuados. La similitud semántica identifica luego la asociación entre la anomalía del indicador y el conocimiento planificador pertinente. La fusión ponderada produce un conjunto de fragmentos probatorios que respaldan la interpretación causal.

3.3 Generación de decisiones: paradigma de razonamiento diagnóstico basado en evidencias

La generación de decisiones sigue a la identificación de anomalías y a la explicación de sus causas. Convierte el diagnóstico en estrategias de planificación y recomendaciones de gobernanza. La plantilla

de indicaciones estructuradas incorpora reconocimiento de indicadores, análisis causal y generación de contramedidas en el proceso de razonamiento, mientras que el conocimiento recuperado aporta estándares normativos, fundamentos de política y restricciones locales de planificación. Mediante generación aumentada por recuperación y control de hiperparámetros, el modelo reduce saltos no justificados y vacíos de evidencia. La salida resultante es más trazable y más adecuada para la práctica de la planificación.

4 Estudio empírico: exploración experimental en Nanjing

4.1 Panorama del área de estudio y sistema de indicadores

Nanjing, capital de la provincia de Jiangsu, se ubica en el tramo inferior del Yangtsé y en la región de Jianghuai, y constituye una ciudad nodo central del delta del río Yangtsé. A finales de 2024, su población residente permanente alcanzó 9,577 millones de personas y su tasa de urbanización fue de 87,3 %. Desde 2019, Nanjing ha impulsado de forma continua el examen y la evaluación urbanos, formando datos de indicadores relativamente estables e informes anuales de evaluación. Ello ofrece una base adecuada para probar el diagnóstico inteligente.

El experimento seleccionó indicadores básicos estrechamente relacionados con la planificación espacial territorial y con datos continuos y comparables. Tras excluir dos indicadores con datos discontinuos, se conservaron 29 indicadores experimentales. El estudio organizó los informes de examen urbano de Nanjing entre 2019 y 2023 e integró 59 documentos centrales de política y planificación para construir una base estructurada de conocimiento de planificación y una base de normas y políticas.

Tabla 2. Composición del sistema de indicadores de examen y evaluación urbanos de Nanjing en 2023.

Componente	Composición retenida para el experimento	Función diagnóstica
Dimensiones de primer nivel	Seguridad, innovación, coordinación, desarrollo verde, apertura y compartición	Proporcionar la estructura general para evaluar el funcionamiento urbano.
Tipo de indicador	Indicadores básicos e indicadores recomendados	Distinguir los elementos obligatorios de monitoreo de los elementos de evaluación ampliados localmente.
Muestra experimental	29 indicadores básicos tras excluir dos indicadores con datos discontinuos	Apoyar la comparación interanual y el razonamiento basado en el modelo.
Materiales de conocimiento	Informes de evaluación de 2019-2023 y 59 documentos clave de política y planificación	Proporcionar evidencia local, restricciones de política y bases de interpretación planificadora.

4.2 Verificación experimental basada en la ontología S-E-D

El experimento utilizó un gran modelo de lenguaje DeepSeek de 671 mil millones de parámetros reforzado por conocimiento. Se pidió al modelo diagnosticar el desempeño de los indicadores según la

ontología S-E-D, recuperar conocimiento planificador pertinente y generar explicaciones y estrategias con evidencia explícita.

4.2.1 Identificación de indicadores e invocación de información

La superficie de la línea roja de conservación ecológica ofrece un ejemplo representativo. La base de conocimiento define el referente normativo de este indicador como "no disminuir; mantenerse estable o aumentar moderadamente cuando las condiciones lo permitan". Los datos muestran que la superficie de la línea roja de conservación ecológica de Nanjing fue de 505,74 km2 en 2020, disminuyó a 496,64 km2 en 2022 y se recuperó ligeramente hasta 498 km2 en 2024. El modelo reconoció una trayectoria de expansión inicial, posterior descenso y leve recuperación. Al considerar esta tendencia junto con la estabilización o ligera reducción de los terrenos de construcción urbanos y rurales, el modelo juzgó que la presión del desarrollo espacial persistía, pero que la línea de base política no había sido vulnerada de manera fundamental. La salida combinó reconocimiento del estado, juicio sobre la tendencia temporal y evaluación de cumplimiento.

4.2.2 Explicación de problemas y razonamiento causal

En la etapa explicativa, el modelo integró indicadores relativos a seguridad, ecología, uso del suelo, innovación y desarrollo espacial. Situó los valores anómalos dentro de las condiciones de recursos, la etapa de desarrollo y las restricciones institucionales de Nanjing. Por ejemplo, los indicadores sobre tierras agrícolas básicas permanentes, nivel de aguas subterráneas y expansión de suelo de construcción se asociaron con restricciones de recursos naturales y asignación insuficiente de suelo de construcción, revelando presiones sobre la seguridad del suelo y la seguridad ecológica. El crecimiento de los terrenos de construcción urbanos y rurales, la disminución de la proporción de suelo industrial y la baja eficiencia en la disposición de tierras ociosas indicaron que la reestructuración industrial y la renovación del suelo existente iban rezagadas respecto de las necesidades del desarrollo impulsado por la innovación. Indicadores como densidad de población, densidad edificatoria y provisión de servicios públicos mostraron que la expansión de nuevas ciudades avanzaba más rápido que la incorporación de población y funciones, reflejando insuficiente coordinación entre planes y mayor autonomía decisonal a nivel distrital.

A través de este proceso, el modelo transformó anomalías de indicadores en explicaciones causales trazables, en lugar de juicios descriptivos aislados.

Tabla 3. Ejemplos de relaciones de correspondencia multinivel entre indicadores y conocimiento.

Indicador típico	Significado del síntoma	Causa posible y base de conocimiento	Orientación estratégica
Tierras agrícolas básicas permanentes	Reflejan la línea de base de protección de tierras cultivadas y la seguridad del espacio agrícola	Una disminución por etapas puede indicar presión por expansión constructiva o insuficiencia de los mecanismos de compensación de tierras cultivadas; la	Reforzar la protección de las tierras agrícolas básicas permanentes, mejorar la construcción de tierras agrícolas de alto estándar y coordinar la implementación de la compensación de tierras

		interpretación se restringe por la política de protección de tierras cultivadas y los requisitos de la planificación espacial territorial.	cultivadas.
Puntos de anegamiento urbano	Reflejan la resiliencia del drenaje, la capacidad de control de inundaciones y el funcionamiento de la infraestructura	El anegamiento puede estar asociado al envejecimiento de redes de tuberías, insuficiente capacidad de retención de aguas pluviales y débil monitoreo inteligente; la evidencia se vincula con estándares de ciudad esponja y planificación del sistema de drenaje.	Impulsar la construcción de ciudades esponja, intervenir puntos de anegamiento, renovar redes de drenaje y mejorar el monitoreo inteligente y la alerta temprana.

4.2.3 Integración de estrategias y salida de decisión

En la etapa de decisión, el modelo generó recomendaciones estratégicas mediante integración interanual e interdimensional. Para los indicadores de innovación, vinculó la disminución de la proporción de suelo industrial, la ineficiencia del suelo existente y el desajuste espacial entre funciones de innovación y servicios públicos. Propuso mejorar el mecanismo de activación del suelo existente, coordinar la localización de espacios industriales y de innovación, fortalecer la conexión entre distritos de innovación, servicios públicos y redes de transporte, y perfeccionar los mecanismos de apoyo y monitoreo de políticas. En comparación con recomendaciones generales, estas salidas estuvieron más claramente vinculadas con evidencias de indicadores y restricciones de planificación.

4.3 Análisis comparativo de capacidades de los modelos

4.3.1 Diseño experimental y marco de evaluación

Para comparar las capacidades de un gran modelo de lenguaje preentrenado y de un modelo reforzado por conocimiento, el estudio construyó un marco de evaluación jerárquico. La prueba incluyó 50 preguntas objetivas y 5 preguntas de ensayo. Las preguntas objetivas examinaron reconocimiento de indicadores, correspondencia con textos de política, mapeo normativo y razonamiento intertemporal. Las preguntas de ensayo examinaron integración de múltiples indicadores, atribución de problemas y apoyo a la toma de decisiones. La evaluación experta consideró eficacia de la respuesta, claridad lógica, ausencia de errores evidentes, disposición a adoptar la respuesta, confianza en la evaluación y preferencia general.

Tabla 4. Marco de evaluación jerárquico.

Nivel de evaluación	Diseño de preguntas	Capacidad evaluada
---------------------	---------------------	--------------------

Identificación básica	Preguntas objetivas sobre significado de indicadores y correspondencia con textos de política	Reconocimiento de indicadores, recuperación de información y correspondencia con políticas locales
Mapeo normativo y razonamiento intertemporal	Preguntas objetivas sobre estándares, comparación temporal y aplicabilidad institucional	Correspondencia de reglas, lógica temporal y consistencia de evidencias
Diagnóstico integrado	Cinco preguntas de ensayo	Identificación de múltiples indicadores, atribución causal y apoyo integrado a decisiones

4.3.2 Resultados de las preguntas objetivas y análisis

Los resultados de las preguntas objetivas muestran que las diferencias entre los dos modelos fueron más claras en tareas que requerían comprensión de políticas locales y razonamiento intertemporal. En preguntas relacionadas con la correspondencia de textos de política local y el reconocimiento básico de indicadores, ambos modelos se comportaron de manera relativamente estable. El modelo preentrenado cometió un error al asociar incorrectamente un contexto local con otras ciudades, lo que revela un riesgo de deriva semántica cuando se trabaja con múltiples documentos locales.

En el mapeo normativo y el razonamiento intertemporal, el modelo preentrenado cometió cuatro errores. Sus debilidades se concentraron en una aplicabilidad institucional poco clara, uso insuficiente de evidencia clave e inferencia temporal inestable. El modelo reforzado por conocimiento mantuvo mayor corrección en esta parte, lo que sugiere que la base de conocimiento proporcionó anclajes semánticos y restricciones de razonamiento eficaces.

4.3.3 Desempeño en preguntas de ensayo y análisis

La evaluación subjetiva confirma además la ventaja del modelo reforzado por conocimiento. En conjunto, los evaluadores prefirieron este modelo en el 54,5 % de las comparaciones, el modelo preentrenado en el 28,9 %, y juzgaron ambos modelos como similares en el 16,7 % de los casos. La preferencia fuerte por el modelo reforzado por conocimiento también fue superior a la del modelo preentrenado. El modelo reforzado por conocimiento obtuvo mejores puntuaciones en eficacia de respuesta, claridad lógica, ausencia de errores evidentes, disposición a adoptar y confianza de los evaluadores.

La experiencia de los evaluadores influyó en los resultados. Los participantes familiarizados con el examen urbano y el diagnóstico de planificación tendieron más a preferir el modelo reforzado por conocimiento; en el grupo de alta familiaridad, la tasa de preferencia alcanzó el 86,7 %. Los evaluadores profesionales prestaron más atención a la evidencia de indicadores y al soporte de datos, y fueron más capaces de identificar generalizaciones y vacíos probatorios del modelo preentrenado. Los usuarios menos familiarizados se vieron más influidos por la fluidez y la aparente completitud. Esto indica que el refuerzo por conocimiento mejora la densidad informativa, la completitud lógica y el potencial de apoyo a la decisión, pero su valor se vuelve más evidente cuando la tarea exige juicio profesional.

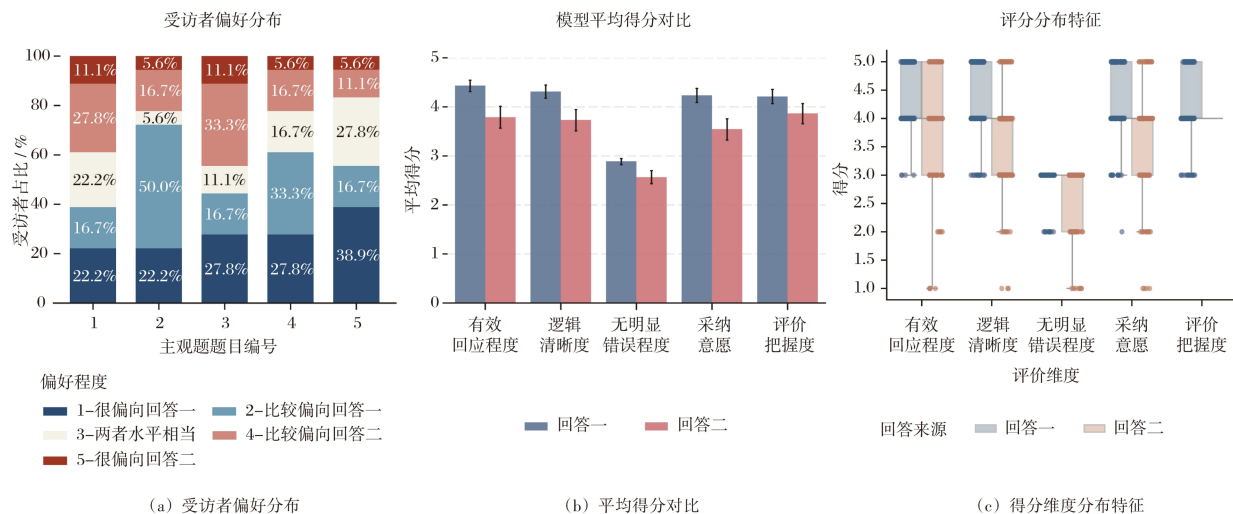


Figura 3. Análisis comparativo del desempeño en preguntas de ensayo.

5 Conclusiones y perspectivas

Este estudio sostiene que el desafío central del examen y la evaluación urbanos no reside solo en la obtención de datos, sino en el vínculo de razonamiento entre monitoreo de indicadores, conocimiento profesional y juicio de planificación. Reconceptualiza el examen de indicadores como un proceso diagnóstico que avanza del reconocimiento de síntomas a la atribución explicativa y luego a la generación de decisiones. Sobre esta base, propone la ontología cognitiva S-E-D y un método basado en un gran modelo de lenguaje reforzado por conocimiento.

El método estructura el conocimiento de planificación y las restricciones contextuales para que la generación del modelo quede delimitada por evidencias de planificación, lógica institucional y contexto profesional local. Apoya la integración de indicadores, la atribución de problemas y el razonamiento estratégico. El experimento de Nanjing muestra que el mecanismo de refuerzo por conocimiento proporciona anclajes semánticos y restricciones de razonamiento, mejorando la estabilidad y la precisión en la comprensión del contexto local, el mapeo de conocimiento normativo y el razonamiento lógico intertemporal. En comparación con el modelo preentrenado, el modelo reforzado por conocimiento produjo salidas con mayor respaldo de datos, análisis más profundo y mejor valor de apoyo a la decisión.

Al mismo tiempo, la mejora no es uniforme en todas las tareas. El método funciona mejor en tareas cuantitativas, locales y ricas en evidencia, pero todavía enfrenta limitaciones en inferencia espacial abstracta, razonamiento causal complejo y optimización expresiva. Los grandes modelos de lenguaje reforzados por conocimiento no pueden sustituir el juicio profesional; más bien proporcionan un mecanismo auxiliar estructurado para el diagnóstico profesional. La construcción y actualización de la base de conocimiento todavía dependen parcialmente de métodos manuales o semiautomáticos, y la adaptación dinámica requiere mejoras adicionales. La prueba empírica se basa en una sola ciudad, por lo que la aplicabilidad interregional necesita verificación posterior. La investigación futura debería fortalecer la fusión de conocimiento multimodal, la validación comparativa entre ciudades y los mecanismos de generación guiados por evidencia, a fin de apoyar un sistema de diagnóstico inteligente más robusto para la gobernanza del espacio territorial.

Referencias

- [1] Wu Z, Guo R, Zhang B, et al. Debate académico sobre la construcción sistemática de la planificación espacial nacional[J]. Urban Planning Forum, 2024(5): 1-11.
- [2] Zhao M, Zhang X. Proceso de desarrollo y operación eficiente del examen y la evaluación urbanos: desde la perspectiva del proceso de política pública[J]. City Planning Review, 2022, 46(8): 65-74.
- [3] Huang Y, Zhang S, Lin Y, et al. Ideas y práctica para construir un sistema municipal de seguimiento de la implementación de la planificación espacial territorial: el caso de Ningbo[J]. Journal of Natural Resources, 2024, 39(4): 823-841.
- [4] Yang M, Wang J, Gu Y. Mecanismo operativo, sistema y métodos del examen y la evaluación urbanos bajo la reforma[J]. Shanghai Urban Planning Review, 2022(1): 16-24.
- [5] Ministerio de Recursos Naturales de la República Popular China. Carta de solicitud de comentarios sobre las Directrices técnicas para la construcción de la red de seguimiento de implementación de la planificación espacial territorial (borrador para comentarios)[EB/OL]. [2025-12-11].
- [6] Cai Y. Ideas básicas y marco general de la construcción de CSPON[J]. China Land, 2024(5): 9-13.
- [7] Wu Z, Shen Q, Wang Y, et al. Nuevo paradigma de planificación: concepto, sistema metodológico y ruta práctica[J]. Urban Planning Forum, 2025(6): 11-17.
- [8] Batty M. The new science of cities[J]. Building Research & Information, 2010, 38: 123-126.
- [9] Liu X, Wu Z. Construcción del paradigma teórico y transformación de los métodos de pensamiento-acción para el diagnóstico del espacio territorial asistido por IA[J]. Urban Planning Forum, 2026(1): 12-20.
- [10] Caldarelli G, Arcaute E, Barthelemy M, et al. The role of complexity for digital twins of cities[J]. Nature Computational Science, 2023, 3(5): 374-381.
- [11] Niu X, Liu S, Sang T, et al. Aprendizaje profesional para grandes modelos: construcción de un modelo de generación de imágenes que integra conocimiento de diseño de morfología espacial urbana[J]. Urban Planning Forum, 2025(1): 55-63.
- [12] Wu Z, Yan J, Xu H, et al. Diez cuestiones anuales clave en el desarrollo de la disciplina de planificación urbana y rural (2024-2025)[J]. Urban Planning Forum, 2024(6): 8-11.
- [13] Hu L, Liu Z, Zhao Z, et al. A survey of knowledge-enhanced pre-trained language models[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(4): 1413-1430.
- [14] Sasaki M, Watanabe N, Komanaka T. Enhancing contextual understanding of Mistral LLM with external knowledge bases[EB/OL]. Research Square, 2024. DOI: 10.21203/rs.3.rs-4215447/v1.
- [15] Han X, Zhang S, Zhu X. Does ChatGPT provide better advice than robo-advisors?[J]. Finance Research Letters, 2024, 58: 104398.
- [16] Bommarito J, Bommarito M, Katz D M, et al. GPT as knowledge worker: a zero-shot evaluation of AI CPA capabilities[EB/OL]. arXiv, 2023.
- [17] Lee J, Stevens N, Han S C. Large language models in finance (FinLLMs)[J]. Neural Computing and Applications, 2025, 37(30): 24853-24867.
- [18] Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine[J]. New England Journal of Medicine, 2023, 388(13): 1233-1239.
- [19] Kung T H, Cheatham M, Medenilla A, et al. Performance of ChatGPT on the United States Medical Licensing Examination: potential for AI-assisted medical education using large language models[J]. PLoS Digital Health, 2023, 2(2): e0000198.

- [20] Gilson A, Safranek C W, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment[J]. JMIR Medical Education, 2023, 9(1): e45312.
- [21] Kasneci E, Sessler K, Kuchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education[J]. Learning and Individual Differences, 2023, 103: 102274.
- [22] Lin Y. Investigación y práctica aplicada de un sistema local inteligente de preguntas y respuestas sobre conocimiento de planificación basado en grandes modelos de lenguaje[J]. Natural Resources Informatization, 2024(6): 63-70.
- [23] Wang T, Song Z, Jiang D, et al. Investigación sobre un método colaborativo para múltiples tipos de examen urbano impulsado por grafos de conocimiento[J]. Shanghai Urban Planning Review, 2022(1): 25-31.
- [24] Ma X, Zheng D, Sun J, et al. Integración de planificación, construcción y gobernanza: práctica innovadora y métodos de la "evaluación planificación-construcción" de Hangzhou[J]. Urban Planning Forum, 2025(2): 81-88.
- [25] Wu J, Wang X, Chen Y, et al. Desafíos del examen urbano en megaciudades y práctica de Shanghai[J]. Urban Planning Forum, 2022(4): 28-34.
- [26] Shi X, Xu Q, Cao Q, et al. Nuevo núcleo de la gobernanza de la planificación: la "primera milla" y la "última milla": el caso de Pekín[J]. Urban Planning Forum, 2023(4): 18-24.
- [27] Wang Q, Mao Z, Wang B, et al. Knowledge graph embedding: a survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [28] Tang J, Qu M, Mei Q. PTE: predictive text embedding through large-scale heterogeneous text networks[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2015: 1165-1174.
- [29] Xiong L, Xiong C, Li Y, et al. Approximate nearest neighbor negative contrastive learning for dense text retrieval[C]//International Conference on Learning Representations. 2021.
- [30] Xiao S, Liu Z, Zhang P, et al. C-Pack: packaged resources to advance general Chinese embedding[EB/OL]. arXiv, 2023.
- [31] Cheng Y, Zhang C, Zhang Z, et al. Exploring large language model based intelligent agents: definitions, methods, and prospects[EB/OL]. arXiv, 2024.
- [32] Karpukhin V, Oguz B, Min S, et al. Dense passage retrieval for open-domain question answering[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020: 6769-6781.
- [33] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[C]//Advances in Neural Information Processing Systems. 2020, 33: 9459-9474.
- [34] Orru G, Piarulli A, Conversano C, et al. Human-like problem-solving abilities in large language models using ChatGPT[J]. Frontiers in Artificial Intelligence, 2023, 6: 1199350.
- [35] Ross S I, Martinez F, Houde S, et al. The programmer's assistant: conversational interaction with a large language model for software development[C]//Proceedings of the 28th International Conference on Intelligent User Interfaces. 2023: 491-514.
- [36] Zhao W X, Liu J, Ren R, et al. Dense text retrieval based on pretrained language models: a survey[J]. ACM Transactions on Information Systems, 2024, 42(4): 1-60.

Traducción asistida por IA