

Construire un paradigme de diagnostic intelligent pour l'évaluation urbaine par indicateurs en intégrant les grands modèles de langage

Auteurs

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

Résumé

Dans le contexte de la transformation numérique et intelligente de la gouvernance de l'espace territorial, l'examen et l'évaluation des villes passent progressivement du suivi des indicateurs à un diagnostic global. Pourtant, l'évaluation centrée sur les indicateurs reste souvent limitée à une description superficielle des anomalies, tandis que l'articulation entre analyse des données et savoir professionnel de la planification demeure insuffisante. Cet article conçoit l'évaluation par indicateurs comme un processus cognitif systématique permettant d'interpréter l'état de fonctionnement des systèmes urbains complexes. Il propose une ontologie diagnostique S-E-D, composée de trois dimensions: symptôme, explication et décision. Sur cette base, il développe une méthode fondée sur un grand modèle de langage renforcé par les connaissances, capable de soutenir l'analyse des indicateurs, l'explication causale et la génération de stratégies dans un contexte de connaissances de planification contraintes, grâce au traitement structuré des documents de planification et des normes techniques. L'étude de cas menée à Nanjing montre que la méthode proposée présente de bonnes performances en matière d'intégration multi-indicateurs, de cohérence du raisonnement et d'interprétabilité des recommandations stratégiques. Elle fournit ainsi une référence méthodologique pour faire évoluer l'évaluation des indicateurs urbains d'une pratique dépendante de l'expérience vers un diagnostic intelligent, reproductible et traçable.

Mots-clés

examen et évaluation urbains; gouvernance de l'espace territorial; grands modèles de langage; diagnostic intelligent

1 Contexte: transformation numérique et intelligente de l'examen et de l'évaluation urbains

1.1 Changement de paradigme: du suivi des indicateurs à la cognition intelligente

L'examen et l'évaluation urbains constituent un outil important pour superviser la mise en oeuvre de la planification spatiale territoriale et améliorer la qualité de la gouvernance spatiale. Ces dernières années, avec la mise en place progressive du système de suivi de la mise en oeuvre dans le cadre de l'« intégration de plusieurs plans », l'examen urbain est devenu un mécanisme régulier pour comprendre le fonctionnement des villes, identifier les problèmes d'exécution des plans et soutenir l'ajustement adaptatif des politiques [1-3]. Les pratiques locales se sont développées autour de l'examen des indicateurs, de l'examen des tâches et de l'examen thématique; parmi eux, l'examen des indicateurs constitue l'entrée la plus directe pour percevoir l'état opérationnel de l'espace urbain [4].

L'évaluation traditionnelle des indicateurs repose principalement sur la comparaison statistique, l'étalonnage par rapport aux objectifs et les rapports périodiques. Elle permet de déterminer si un indicateur donné atteint sa cible, mais explique plus difficilement les causes des écarts et les modalités de réponse. Avec l'expansion des infrastructures numériques, des données spatio-temporelles et des plateformes de fonctionnement urbain, l'examen des indicateurs évolue d'un contrôle statistique a posteriori vers une perception dynamique et un suivi continu. La construction du réseau de suivi de la mise en oeuvre de la planification spatiale territoriale renforce encore la demande d'évaluation

dynamique, de diagnostic intelligent et d'appui à la décision [5-7]. Dans cette perspective, la tâche essentielle de l'examen des indicateurs n'est plus seulement de surveiller des valeurs, mais de construire un processus cognitif intelligent associant identification des problèmes, explication des mécanismes et génération de stratégies.

1.2 Difficultés pratiques: dépendance à l'expérience et intégration insuffisante des connaissances

Bien que les données multisources aient amélioré l'étendue et la rapidité du suivi, la capacité diagnostique approfondie de la pratique actuelle reste limitée par deux problèmes. Le premier est la forte dépendance à l'expérience professionnelle. Les systèmes urbains sont complexes, couplés, multi-échelles et non linéaires; les variations d'indicateurs peuvent impliquer des effets différés, des interactions intersectorielles et des liens causaux cachés [8]. Les planificateurs doivent souvent interpréter des données fragmentées et à haute fréquence à partir de leur jugement professionnel, alors qu'il est difficile d'en extraire directement des chaînes causales stables. Les conclusions d'évaluation peuvent donc s'arrêter à une liste d'indicateurs anormaux et dépendre excessivement de l'expérience individuelle, ce qui rend difficile l'identification des causes structurelles et des règles de décision transférables [9].

Le second problème tient à l'intégration insuffisante des connaissances de planification. Les documents de politique publique, les directives techniques, les plans réglementaires, les rapports d'évaluation et l'expertise constituent ensemble la logique professionnelle du diagnostic urbain. Les méthodes numériques existantes sont relativement fortes en calcul et en visualisation, mais faibles lorsqu'il s'agit d'organiser des textes de planification non structurés et de les incorporer dans des processus de raisonnement. Il en résulte un écart entre les données de suivi et les connaissances professionnelles de la planification [10]. Les méthodes fondées sur les données ne représentent pas naturellement les règles institutionnelles, les mécanismes de planification et la logique de gouvernance spatiale; les méthodes de raisonnement par règles sont, quant à elles, limitées lorsqu'elles doivent intégrer des sources de connaissances hétérogènes [11]. Il faut donc un nouveau cadre reliant les corrélations issues des données aux contraintes du savoir.

1.3 Opportunité technique: raisonnement intelligent renforcé par les connaissances

Les grands modèles de langage offrent une nouvelle possibilité technique pour le diagnostic des indicateurs urbains, grâce à leurs capacités de compréhension sémantique, d'association de connaissances et de génération en langage naturel. Les modèles de langage préentraînés renforcés par les connaissances, la génération augmentée par récupération, la recherche dense et le raisonnement agentique ont déjà été appliqués dans la finance, la comptabilité, la médecine, l'éducation et l'assistance à la programmation [13-21]. Leur valeur ne réside pas seulement dans la génération de texte, mais aussi dans l'organisation des connaissances de domaine, l'extraction de preuves pertinentes et la construction de chaînes de raisonnement structurées.

Pour l'examen urbain, les grands modèles de langage peuvent combiner textes de planification, normes de politique publique, données d'indicateurs et informations contextuelles, afin de soutenir l'interprétation sémantique et le raisonnement en plusieurs étapes. Toutefois, les applications actuelles dans le champ de la planification portent surtout sur l'interprétation des politiques, les systèmes de questions-réponses fondés sur les connaissances ou l'explication isolée d'indicateurs. Elles n'ont pas encore formé un paradigme complet reliant compréhension des indicateurs, explication causale et jugement planificateur global [22-23]. Cet article définit donc l'examen des indicateurs comme un

processus de diagnostic cognitif des systèmes urbains complexes, construit une ontologie S-E-D et explore une méthode d'analyse intelligente fondée sur un grand modèle de langage renforcé par les connaissances.

2 Construction de l'ontologie cognitive et mécanisme de renforcement par les connaissances

2.1 Reconstruction de l'ontologie diagnostique: l'ontologie cognitive S-E-D

L'examen des indicateurs urbains ne doit pas être compris comme un simple suivi technique, mais comme un processus diagnostique global orienté vers des systèmes urbains adaptatifs complexes [24]. Son objet n'est pas la comparaison statique d'indicateurs isolés, mais les relations de couplage, les écarts et les risques entre les sous-systèmes spatiaux, sociaux, économiques, écologiques et institutionnels [25-26]. Les comparaisons normatives et les pondérations linéaires peuvent établir si un indicateur s'écarte d'une cible, mais elles n'expliquent pas suffisamment les causes, les trajectoires et les implications de gouvernance de cet écart.

Pour répondre à cette limite, l'article propose une ontologie cognitive de diagnostic des indicateurs: S-E-D, pour symptôme, explication et décision. Cette ontologie réorganise le processus d'évaluation en une chaîne cognitive fermée allant de la perception de l'état à l'attribution des mécanismes, puis à la réponse de planification.

2.1.1 Niveau du symptôme: perception de l'état et identification des écarts

Le niveau du symptôme transforme l'état de fonctionnement de la ville en signaux de problème reconnaissables et comparables. Il identifie si les indicateurs s'écartent des objectifs de planification, des normes techniques, des trajectoires historiques ou des plages attendues, tout en prêtant attention aux dynamiques temporelles et aux associations entre indicateurs. Par rapport à une comparaison statique de valeurs, ce niveau souligne la signification sémantique des indicateurs dans des contextes de planification précis. Il produit des unités de symptôme structurées, telles que baisse, fluctuation, inadéquation, pression de seuil ou coordination insuffisante.

2.1.2 Niveau de l'explication: compréhension des causes et attribution des mécanismes

Le niveau de l'explication déplace l'analyse de la question « qu'est-ce qui a changé? » vers « pourquoi cela a-t-il changé? ». Il relie les performances anormales des indicateurs à la structure spatiale, aux dotations en ressources, au stade de développement, aux contraintes institutionnelles et aux processus socio-économiques. Grâce aux contraintes des connaissances de planification, le modèle peut éviter les explications générales détachées du contexte local et transformer les valeurs anormales en interprétations causales appuyées par des preuves traçables. Ce niveau constitue le pont essentiel entre le suivi des données et le jugement professionnel de planification.

2.1.3 Niveau de la décision: génération de stratégies et sortie décisionnelle

Le niveau de la décision convertit la connaissance diagnostique en stratégies de gouvernance. La stratégie doit correspondre aux causes identifiées, aux objectifs de planification, aux conditions institutionnelles et aux voies de mise en oeuvre. Elle ne doit pas produire seulement des recommandations génériques, mais organiser les contre-mesures selon la chaîne reconnaissance du symptôme, diagnostic du mécanisme et intervention stratégique. En ce sens, l'examen des indicateurs devient un raisonnement de planification fondé sur les preuves, et non un simple rapport descriptif.

2.2 Chemin de raisonnement intelligent sous contraintes de connaissances

Le raisonnement renforcé par les connaissances comprend trois processus liés. Le premier est la modélisation et l'organisation structurée des connaissances. Les documents de planification, normes

techniques, textes de politique publique, rapports d'évaluation et expériences d'experts sont nettoyés, segmentés, normalisés et transformés en unités sémantiques calculables. Des modèles d'intégration vectorielle tels que S-BERT et BGE peuvent projeter ces unités dans des espaces vectoriels denses et former une base de connaissances de planification propre au domaine [27-30].

Le deuxième processus est la récupération sémantique et l'ordonnement dynamique des connaissances. Lorsqu'une anomalie d'indicateur est détectée, le système récupère des fragments hautement pertinents selon le type d'indicateur, l'objet de planification, le contexte spatial et la description du problème. La recherche vectorielle dense saisit la similarité sémantique, tandis que la recherche clairsemée par mots-clés et les contraintes symboliques maintiennent une correspondance précise avec les termes locaux de politique publique, les normes et les catégories de planification [31-32].

Le troisième processus est la génération fusionnée et le contrôle du chemin de raisonnement. La génération augmentée par récupération introduit des preuves de planification dans le contexte du modèle, tandis que des invites structurées guident le modèle le long de la chaîne S-E-D [33-36]. Les hyperparamètres de génération contraignent en outre le comportement d'échantillonnage, afin que la sortie reste stable, cohérente et ancrée dans les connaissances récupérées. Par rapport à l'ajustement fin par domaine, cette approche réduit les coûts d'entraînement et le risque de surapprentissage. Elle transforme le rôle du modèle: il ne stocke plus seulement des connaissances de planification, mais les invoque, les organise et raisonne avec elles.

Tableau 1. Configuration des hyperparamètres de génération du grand modèle de langage et mécanismes de régulation.

Paramètre	Valeur	Plage	Fonction	Mécanisme de régulation
Temperature	0,3	0-2	Contrôle l'aléa et la diversité de la génération	Une valeur basse améliore la stabilité du raisonnement et la cohérence logique.
Top P	0,1	0-1	Tronque l'ensemble des jetons candidats selon la probabilité cumulée	L'espace des candidats est resserré, les expressions peu probables sont réduites et la détermination de la sortie est renforcée.
Frequency penalty	0	0-1	Réduit la	La récurrence des

			probabilité de répétition des mots à forte fréquence	termes est préservée afin de maintenir la cohérence de l'expression diagnostique structurée.
Presence penalty	0	0-1	Encourage l'introduction de nouveaux contenus	Le réglage évite l'éloignement des connaissances récupérées et maintient l'analyse autour des informations fournies.
Max tokens	16384	0-16384	Contrôle la longueur maximale d'une génération	Le réglage prend en charge les entrées longues et le raisonnement en plusieurs étapes.

3 Méthode de raisonnement intelligent fondée sur l'ontologie S-E-D

L'article développe une méthode de raisonnement intelligent pour l'examen urbain par indicateurs, centrée sur un grand modèle de langage. La méthode intègre une base de connaissances de planification, des informations contextuelles sur les indicateurs et des invites de raisonnement contraintes, guidant le modèle à travers l'analyse des symptômes, l'attribution explicative et la génération de décisions. L'objectif est de rendre le diagnostic des indicateurs plus reproductible, plus traçable et plus interprétable.

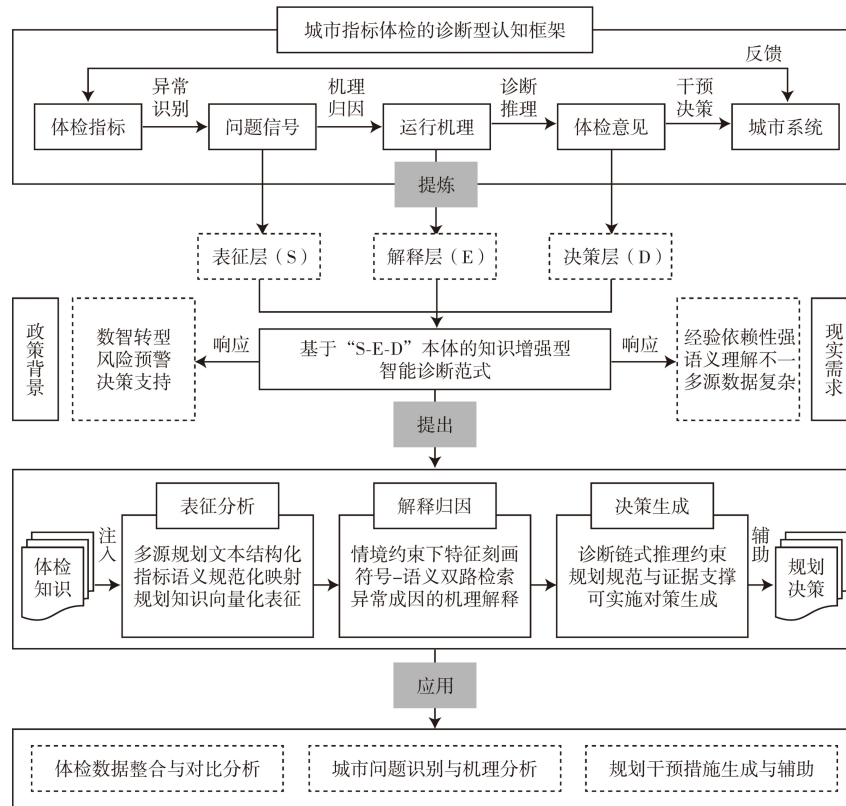


Figure 1. Cadre technique du diagnostic intelligent pour l'évaluation urbaine par indicateurs fondé sur la boucle cognitive fermée S-E-D.

3.1 Analyse des symptômes: construction de la base de connaissances de planification

L'analyse des symptômes commence par la construction d'une base de connaissances de planification. Les documents de politique publique, procédures techniques, textes de planification réglementaire et rapports d'examen urbain sont transformés en unités de connaissance récupérables et calculables par le nettoyage des textes, l'analyse structurelle et la normalisation terminologique. Les données d'indicateurs sont étiquetées par catégorie, année, valeur, norme, seuil et objectif de planification; les textes de planification sont décomposés en éléments sémantiques tels qu'objet, état, problème, cause et contre-mesure.

Pour soutenir une récupération précise et la reconstruction du contexte, la base de connaissances adopte une structure sémantique à deux niveaux. Le niveau supérieur enregistre le domaine thématique, l'objet de planification et la source institutionnelle de chaque unité de connaissance; le niveau inférieur enregistre le sens diagnostique correspondant, la logique causale et la réponse de politique publique. Après vectorisation, la base peut soutenir la recherche de similarité, la récupération de preuves et l'assemblage du contexte nécessaire au raisonnement du modèle.

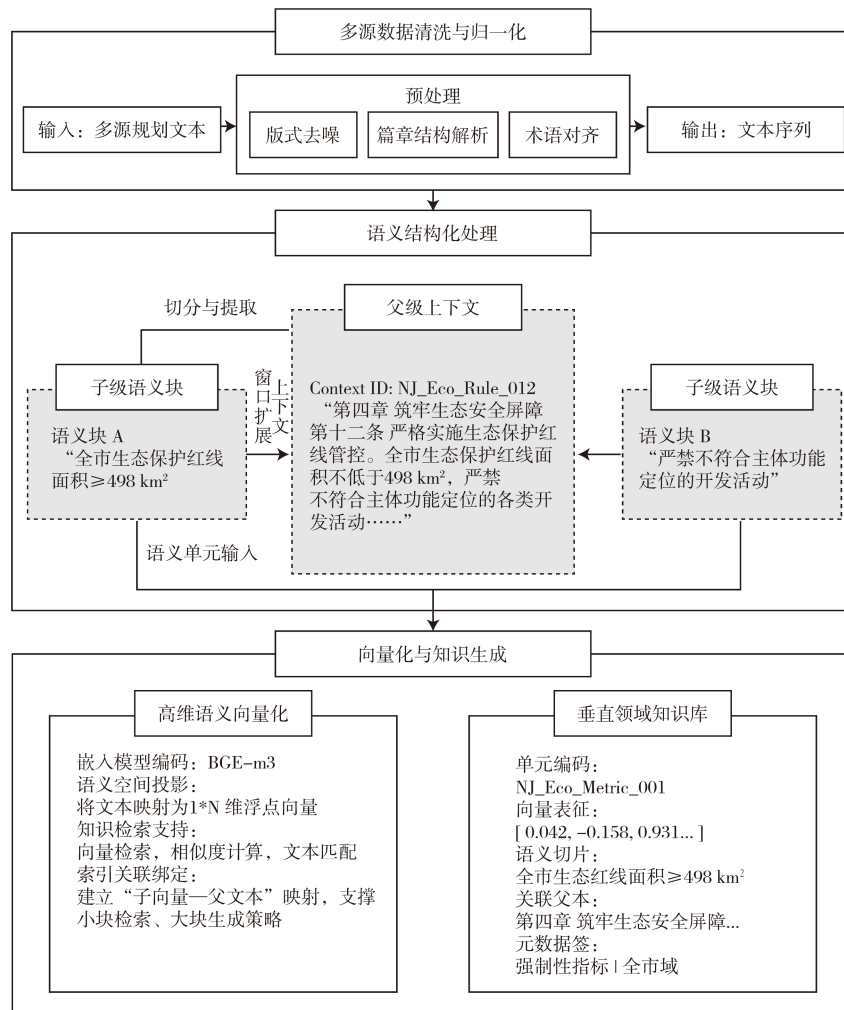


Figure 2. Processus de déconstruction sémantique et de construction de la base de connaissances de planification pour le diagnostic par indicateurs.

3.2 Attribution explicative: stratégie de récupération guidée par les connaissances

Après l'identification des indicateurs anormaux, l'analyse passe de la question de savoir s'il existe une anomalie à celle de comprendre pourquoi elle apparaît. Pour éviter qu'un modèle général produise des explications détachées du contexte spatial, la méthode proposée remplace les variations d'indicateurs dans les conditions de localisation, les attributs fonctionnels, les stades de développement et les contraintes institutionnelles grâce à la base de connaissances de planification. La récupération combine des contraintes symboliques et la similarité sémantique. Les contraintes symboliques réduisent le domaine de l'objet et garantissent que les preuves récupérées correspondent à la bonne catégorie de planification, à l'unité spatiale adéquate et au contexte de politique publique pertinent. La similarité sémantique saisit ensuite l'association entre l'anomalie de l'indicateur et les connaissances de planification concernées. La fusion pondérée produit un ensemble de fragments probants soutenant l'interprétation causale.

3.3 Génération de décisions: paradigme de raisonnement diagnostique fondé sur les preuves

La génération de décisions suit l'identification des anomalies et l'explication des causes. Elle convertit le diagnostic en stratégies de planification et en recommandations de gouvernance. Le modèle d'invite structurée intègre reconnaissance de l'indicateur, analyse causale et génération de contre-mesures dans le processus de raisonnement, tandis que les connaissances récupérées fournissent les normes, les bases de politique publique et les contraintes locales de planification. Grâce à la génération augmentée par récupération et au contrôle des hyperparamètres, le modèle réduit les sauts non étayés et les lacunes de preuve. La sortie devient ainsi plus traçable et plus adaptée à la pratique de la planification.

4 Étude empirique: exploration expérimentale à Nanjing

4.1 Présentation du terrain d'étude et système d'indicateurs

Nanjing, capitale de la province du Jiangsu, est située dans le cours inférieur du Yangtsé et la région du Jianghuai; elle constitue une ville noeud majeure du delta du Yangtsé. À la fin de 2024, sa population résidente permanente atteignait 9,577 millions d'habitants et son taux d'urbanisation était de 87,3 %. Depuis 2019, Nanjing fait progresser de manière continue l'examen et l'évaluation urbains, en constituant des données d'indicateurs relativement stables et des rapports annuels d'évaluation. Ces conditions offrent une base appropriée pour tester le diagnostic intelligent.

L'expérience a retenu des indicateurs de base étroitement liés à la planification spatiale territoriale et disposant de données continues et comparables. Après exclusion de deux indicateurs aux données discontinues, 29 indicateurs expérimentaux ont été conservés. L'étude a organisé les rapports d'examen urbain de Nanjing pour la période 2019-2023 et intégré 59 documents clés de politique publique et de planification afin de construire une base structurée de connaissances de planification et une base de normes et de politiques.

Tableau 2. Composition du système d'indicateurs de l'examen et de l'évaluation urbains de Nanjing en 2023.

Composante	Composition retenue pour l'expérience	Rôle diagnostique
Dimensions de premier niveau	Sécurité, innovation, coordination, développement vert, ouverture et partage	Fournir la structure générale d'évaluation du fonctionnement urbain.
Type d'indicateur	Indicateurs de base et indicateurs recommandés	Distinguer les éléments obligatoires de suivi des éléments d'évaluation étendus localement.
Échantillon expérimental	29 indicateurs de base après exclusion de deux indicateurs aux données discontinues	Soutenir la comparaison interannuelle et le raisonnement fondé sur le modèle.
Matériaux de connaissance	Rapports d'évaluation 2019-2023 et 59 documents clés de	Fournir des preuves locales, des contraintes de politique

	politique publique et de planification	publique et des bases d'interprétation planificatrice.
--	--	--

4.2 Vérification expérimentale fondée sur l'ontologie S-E-D

L'expérience utilise un grand modèle de langage DeepSeek renforcé par les connaissances, comportant 671 milliards de paramètres. Le modèle doit diagnostiquer la performance des indicateurs selon l'ontologie S-E-D, récupérer les connaissances de planification pertinentes et générer des explications et des stratégies avec des preuves explicites.

4.2.1 Identification des indicateurs et mobilisation de l'information

La superficie de la ligne rouge de conservation écologique fournit un exemple représentatif. La base de connaissances définit la référence normative de cet indicateur comme « ne pas diminuer; rester stable ou augmenter modérément lorsque les conditions le permettent ». Les données montrent que la superficie de cette ligne rouge à Nanjing était de 505,74 km2 en 2020, est descendue à 496,64 km2 en 2022, puis a légèrement remonté à 498 km2 en 2024. Le modèle reconnaît une trajectoire d'expansion initiale, de recul ultérieur et de légère reprise. En considérant cette évolution avec la stabilisation ou la légère diminution des terrains de construction urbains et ruraux, il juge que la pression du développement spatial demeure, mais que la ligne de fond politique n'a pas été fondamentalement franchie. La sortie associe reconnaissance de l'état, jugement sur la tendance temporelle et évaluation de conformité.

4.2.2 Explication des problèmes et raisonnement causal

Au stade explicatif, le modèle intègre des indicateurs relatifs à la sécurité, à l'écologie, à l'utilisation du sol, à l'innovation et au développement spatial. Il replace les valeurs anormales dans les conditions de ressources, le stade de développement et les contraintes institutionnelles de Nanjing. Ainsi, les indicateurs concernant les terres agricoles de base permanentes, le niveau des eaux souterraines et l'expansion des terrains de construction sont associés aux contraintes des ressources naturelles et à une allocation insuffisante de terrains de construction, révélant des pressions sur la sécurité foncière et écologique. La croissance des terrains de construction urbains et ruraux, la baisse de la part des terrains industriels et la faible efficacité du traitement des terrains inactifs indiquent un retard de la restructuration industrielle et du renouvellement du stock foncier par rapport aux besoins d'un développement tiré par l'innovation. Les indicateurs de densité de population, de densité bâtie et de services publics montrent que l'expansion des villes nouvelles dépasse l'importation de population et de fonctions, traduisant une coordination insuffisante entre les plans et un renforcement de l'autonomie décisionnelle au niveau des districts.

Par ce processus, le modèle transforme les anomalies d'indicateurs en explications causales traçables, plutôt qu'en jugements descriptifs isolés.

Tableau 3. Exemples de relations de correspondance multi-niveaux entre indicateurs et connaissances.

Indicateur type	Signification du symptôme	Cause possible et base de connaissance	Orientation stratégique
Terres agricoles de base permanentes	Reflètent la ligne de fond de la protection des terres cultivées et la sécurité de	Une baisse par étapes peut signaler la pression de l'expansion de la	Renforcer la protection des terres agricoles de base permanentes,

	l'espace agricole	construction ou l'insuffisance des mécanismes de compensation des terres cultivées; l'interprétation est contrainte par la politique de protection des terres cultivées et les exigences de la planification spatiale territoriale.	améliorer la construction de terres agricoles de haut standard et coordonner la mise en oeuvre de la compensation des terres cultivées.
Points d'accumulation d'eau en ville	Reflètent la résilience du drainage, la capacité de contrôle des crues et le fonctionnement des infrastructures	L'accumulation d'eau peut être liée au vieillissement des réseaux, à une capacité insuffisante de rétention des eaux pluviales et à un faible suivi intelligent; les preuves sont associées aux normes de la ville-éponge et à la planification du système de drainage.	Promouvoir la construction de villes-éponges, traiter les points d'accumulation d'eau, renouveler les réseaux de drainage et améliorer le suivi intelligent ainsi que l'alerte précoce.

4.2.3 Intégration des stratégies et production décisionnelle

Au stade décisionnel, le modèle génère des recommandations stratégiques par intégration interannuelle et interdimensionnelle. Pour les indicateurs d'innovation, il relie la baisse de la part des terrains industriels, l'inefficacité du stock foncier et l'inadéquation spatiale entre fonctions d'innovation et services publics. Il propose d'améliorer le mécanisme d'activation du stock foncier, de coordonner l'implantation des espaces industriels et d'innovation, de renforcer la relation entre quartiers d'innovation, services publics et réseaux de transport, ainsi que d'affiner les mécanismes de soutien et de suivi des politiques. Par rapport à des recommandations générales, ces sorties sont plus clairement liées aux preuves issues des indicateurs et aux contraintes de planification.

4.3 Analyse comparative des capacités des modèles

4.3.1 Conception expérimentale et cadre d'évaluation

Pour comparer les capacités d'un grand modèle de langage préentraîné et d'un modèle renforcé par les connaissances, l'étude construit un cadre d'évaluation hiérarchique. Le test comprend 50 questions objectives et 5 questions de dissertation. Les questions objectives examinent la reconnaissance des indicateurs, la correspondance avec les textes de politique publique, la correspondance normative et le

raisonnement intertemporel. Les dissertations examinent l'intégration multi-indicateurs, l'attribution des problèmes et l'appui à la décision. L'évaluation experte prend en compte l'efficacité de la réponse, la clarté logique, l'absence d'erreurs évidentes, la disposition à adopter la réponse, la confiance dans l'évaluation et la préférence globale.

Tableau 4. Cadre d'évaluation hiérarchique.

Niveau d'évaluation	Conception des questions	Capacité testée
Identification de base	Questions objectives sur la signification des indicateurs et la correspondance avec les textes de politique publique	Reconnaissance des indicateurs, récupération d'information et correspondance avec la politique locale
Correspondance normative et raisonnement intertemporel	Questions objectives sur les normes, la comparaison temporelle et l'applicabilité institutionnelle	Correspondance des règles, logique temporelle et cohérence des preuves
Diagnostic intégré	Cinq questions de dissertation	Identification multi-indicateurs, attribution causale et appui intégré à la décision

4.3.2 Résultats des questions objectives et analyse

Les résultats des questions objectives montrent que les différences entre les deux modèles sont les plus marquées dans les tâches exigeant la compréhension des politiques locales et le raisonnement intertemporel. Pour les questions relatives à la correspondance avec les textes de politique locale et à la reconnaissance de base des indicateurs, les deux modèles se montrent relativement stables. Le modèle préentraîné commet une erreur en associant à tort un contexte local à d'autres villes, ce qui révèle un risque de dérive sémantique lorsque plusieurs documents locaux sont mobilisés.

Dans la correspondance normative et le raisonnement intertemporel, le modèle préentraîné commet quatre erreurs. Ses faiblesses se concentrent dans une applicabilité institutionnelle imprécise, un usage insuffisant des preuves clés et une inférence temporelle instable. Le modèle renforcé par les connaissances conserve une meilleure exactitude dans cette partie, ce qui indique que la base de connaissances fournit des ancrages sémantiques et des contraintes de raisonnement efficaces.

4.3.3 Performance des questions de dissertation et analyse

L'évaluation subjective confirme encore l'avantage du modèle renforcé par les connaissances. Dans l'ensemble, les évaluateurs préfèrent ce modèle dans 54,5 % des comparaisons, le modèle préentraîné dans 28,9 %, et jugent les deux modèles similaires dans 16,7 % des cas. La préférence forte pour le modèle renforcé par les connaissances est également supérieure à celle accordée au modèle préentraîné. Le modèle renforcé par les connaissances obtient de meilleurs scores en efficacité de réponse, clarté logique, absence d'erreurs évidentes, disposition à l'adoption et confiance des évaluateurs.

L'expertise des évaluateurs influence les résultats. Les répondants familiers de l'examen urbain et du diagnostic de planification sont davantage susceptibles de préférer le modèle renforcé par les

connaissances; dans le groupe à forte familiarité, le taux de préférence atteint 86,7 %. Les évaluateurs professionnels accordent plus d'importance aux preuves d'indicateurs et à l'appui des données, et identifient plus facilement les généralisations et lacunes probantes du modèle préentraîné. Les utilisateurs moins familiers sont davantage influencés par la fluidité et la complétude apparente. Cela indique que le renforcement par les connaissances améliore la densité informationnelle, la complétude logique et le potentiel d'appui à la décision, mais que sa valeur apparaît surtout lorsque la tâche exige un jugement professionnel.

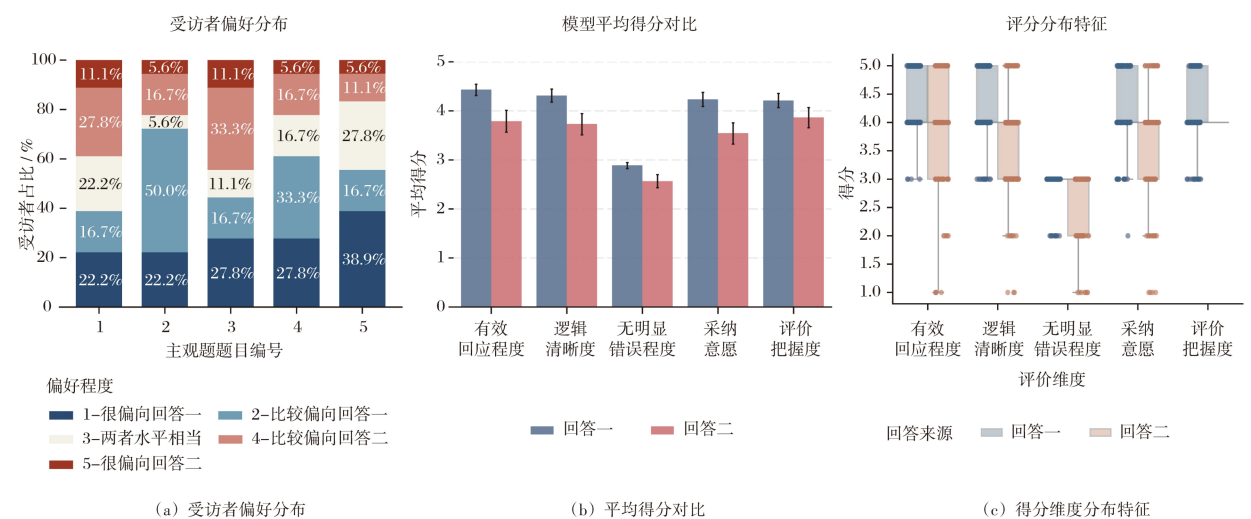


Figure 3. Analyse comparative de la performance aux questions de dissertation.

5 Conclusions et perspectives

Cet article soutient que le défi central de l'examen et de l'évaluation urbains ne réside pas seulement dans l'obtention des données, mais dans la construction d'un lien de raisonnement entre suivi des indicateurs, connaissances professionnelles et jugement de planification. Il reconceptualise l'examen des indicateurs comme un processus diagnostique allant de la reconnaissance des symptômes à l'attribution explicative, puis à la génération de décisions. Sur cette base, il propose l'ontologie cognitive S-E-D et une méthode fondée sur un grand modèle de langage renforcé par les connaissances.

La méthode structure les connaissances de planification et les contraintes contextuelles de sorte que la génération du modèle reste bornée par les preuves de planification, la logique institutionnelle et le contexte professionnel local. Elle soutient l'intégration des indicateurs, l'attribution des problèmes et le raisonnement stratégique. L'expérience menée à Nanjing montre que le mécanisme de renforcement par les connaissances fournit des ancrages sémantiques et des contraintes de raisonnement, améliorant la stabilité et l'exactitude dans la compréhension du contexte local, la correspondance avec les connaissances normatives et le raisonnement logique intertemporel. Par rapport au modèle préentraîné, le modèle renforcé par les connaissances produit des sorties mieux étayées par les données, plus approfondies analytiquement et plus utiles pour l'appui à la décision.

Cependant, l'amélioration n'est pas uniforme dans toutes les tâches. La méthode fonctionne mieux pour les tâches quantitatives, locales et riches en preuves, mais demeure limitée dans l'inférence spatiale abstraite, le raisonnement causal complexe et l'optimisation de l'expression. Les grands modèles de langage renforcés par les connaissances ne peuvent remplacer le jugement professionnel; ils fournissent plutôt un mécanisme auxiliaire structuré pour le diagnostic professionnel. La construction et la mise à

jour de la base de connaissances dépendent encore partiellement de méthodes manuelles ou semi-automatiques, et l'adaptation dynamique doit être améliorée. L'expérimentation empirique repose sur une seule ville; l'applicabilité interrégionale nécessite donc une vérification supplémentaire. Les recherches futures devraient renforcer la fusion des connaissances multimodales, la validation comparative multi-villes et les mécanismes de génération fondée sur les preuves, afin de soutenir un système de diagnostic intelligent plus robuste pour la gouvernance de l'espace territorial.

Références

- [1] Wu Z, Guo R, Zhang B, et al. Discussion académique sur la construction systématique de la planification spatiale nationale[J]. Urban Planning Forum, 2024(5): 1-11.
- [2] Zhao M, Zhang X. Processus de développement et fonctionnement efficace de l'examen et de l'évaluation urbains: perspective du processus de politique publique[J]. City Planning Review, 2022, 46(8): 65-74.
- [3] Huang Y, Zhang S, Lin Y, et al. Idées et pratiques de construction d'un système municipal de suivi de la mise en oeuvre de la planification spatiale territoriale: le cas de Ningbo[J]. Journal of Natural Resources, 2024, 39(4): 823-841.
- [4] Yang M, Wang J, Gu Y. Mécanisme de fonctionnement, système et méthodes de l'examen et de l'évaluation urbains dans le contexte de la réforme[J]. Shanghai Urban Planning Review, 2022(1): 16-24.
- [5] Ministère des Ressources naturelles de la République populaire de Chine. Lettre sollicitant des commentaires sur les Directives techniques pour la construction du réseau de suivi de la mise en oeuvre de la planification spatiale territoriale (projet pour commentaires)[EB/OL]. [2025-12-11].
- [6] Cai Y. Idées de base et cadre général de la construction du CSPON[J]. China Land, 2024(5): 9-13.
- [7] Wu Z, Shen Q, Wang Y, et al. Nouvelle paradigme de planification: concept, système méthodologique et voie pratique[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. Construction du paradigme théorique et transformation des méthodes de pensée-action pour le diagnostic de l'espace territorial assisté par l'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. Apprentissage professionnel pour les grands modèles: construction d'un modèle de génération d'images intégrant les connaissances de conception de la morphologie spatiale urbaine[J]. Urban Planning Forum, 2025(1): 55-63.
- [12] Wu Z, Yan J, Xu H, et al. Dix questions annuelles clés dans le développement de la discipline de l'urbanisme et de l'aménagement 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. Recherche et pratique d'application d'un système local intelligent de questions-réponses sur les connaissances de planification fondé sur les grands modèles de langage[J]. *Natural Resources Informatization*, 2024(6): 63-70.
- [23] Wang T, Song Z, Jiang D, et al. Recherche sur une méthode collaborative de plusieurs types d'examen urbain pilotée par un graphe de connaissances[J]. *Shanghai Urban Planning Review*, 2022(1): 25-31.
- [24] Ma X, Zheng D, Sun J, et al. Intégration de la planification, de la construction et de la gouvernance: pratique innovante et méthodes de l'« évaluation planification-construction » de Hangzhou[J]. *Urban Planning Forum*, 2025(2): 81-88.
- [25] Wu J, Wang X, Chen Y, et al. Défis de l'examen urbain dans les mégapoles et pratique de Shanghai[J]. *Urban Planning Forum*, 2022(4): 28-34.
- [26] Shi X, Xu Q, Cao Q, et al. Nouveau noyau de la gouvernance de la planification: le « premier kilomètre » et le « dernier kilomètre »: le cas de Pékin[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.

Traduction assistée par l'IA