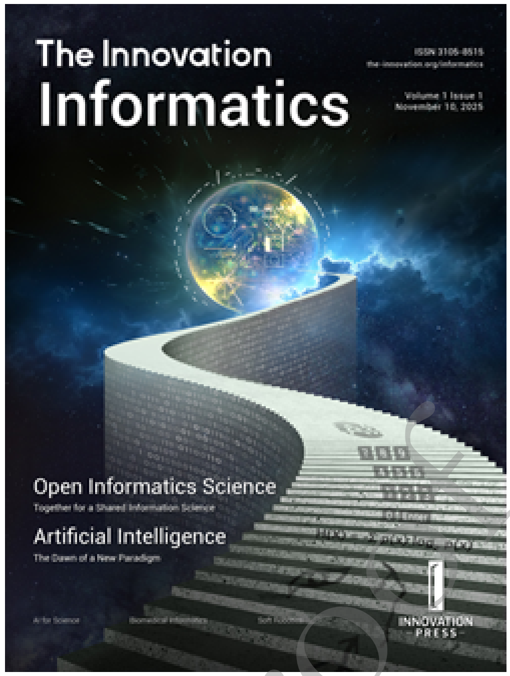


Grounded open-vocabulary 3D LiDAR AI for natural resource surveying

Journal:	*The Innovation Informatics*
Manuscript ID	TII-2026-0037.R2
Manuscript Type:	Commentary
Keywords:	Grounded open-vocabulary AI, 3D LiDAR point-cloud understanding, Survey-grade evidence, Natural resource surveying, Foundation Model
Speciality:	Remote Sensing Image Analysis < Electronic Informatics, Artificial Intelligence, Remote Sensing Sensors < Optics and Optoelectronics

Grounded open-vocabulary 3D LiDAR AI for natural resource surveying

Anquan Xia,[1,2] Yiping Chen,[3] Ting Han,[3, 4,*] Xiang Wei,[1,*] Di Zhao,[1,2] Cheng Ma,[1] Tong Li,[5] Yi Xu,[6] and Meiliu Wu[4]

[1]Development and Research Center, China Geological Survey, Beijing 100037, China

[2]School of Ocean Sciences, China University of Geosciences (Beijing), Beijing 100083, China

[3]School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai 519082, China

[4]School of Geographical and Earth Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom

[5]School of Agriculture and Food Sustainability, The University of Queensland, St Lucia QLD 4072, Australia

[6]Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede 7522NB, Netherlands

*Correspondence: hant23@mail2.sysu.edu.cn (T.H.); weix@mail.cgs.gov.cn (W.X.)

Study Area：Environment & Ecology

Figure 1. Grounded open-vocabulary LiDAR Artificial Intelligence (AI) for natural resource surveying Multi-platform LiDAR point clouds provide complementary 3D structural evidence across diverse surveying tasks, including forest inventory and carbon assessment, terrain and landform characterization, ecological restoration monitoring, and river-lake change analysis. Natural-language prompts enable flexible, long-tail querying, while a human-in-the-loop mechanism supports verification and resolves ambiguous and low-evidence cases. Grounded open-vocabulary LiDAR AI enforces evidence and context-based intelligence, producing consistent semantic understanding and geospatial analysis in natural resource surveying.

STRUCTURAL EVIDENCE GAPS IN NATURAL RESOURCE SURVEYING

Natural resource surveying provides essential evidence for Earth system governance by documenting forests, terrain systems, inland waters, and ecological units whose 3D structures contain information relevant to ecosystem dynamics and environmental change. As extreme events, land-use transitions, and hydrological alterations intensify, surveying systems must move beyond change detection toward defensible and comparable evidence of what has changed, how much it has changed, and whether such changes are consistent across regions, seasons, and acquisition campaigns.

Current observation systems still struggle to accurately acquire and consistently information consistently.[1] Many process-relevant indicators cannot be robustly characterized from 2D observations alone. Even indirect observations are often affected by illumination, viewing geometry, phenology, and sensor-specific bias, causing the same ecological or geomorphic state to appear differently under different observation conditions and weakening long-term comparability.

This limitation reflects a deeper mismatch between the evidence required for natural resource surveys and the representational assumptions of conventional observation pipelines. Natural resource surveying requires evidence that is structurally meaningful, geographically interpretable, and transferable across domains, whereas many current workflows remain optimized for recognition under restricted settings. Therefore, the field needs an observation-and-inference framework that treats 3D structure as a primary source of evidence and makes cross-domain comparability a core requirement. This need motivates a rethinking of 3D LiDAR Artificial Intelligence (AI) for natural resource surveying.[2]

3D LiDAR AI FOR SURVEY-GRADE STRUCTURAL EVIDENCE

3D LiDAR point clouds have become an important source of structural information for natural resource surveying.[3,4] Unlike 2D observations, LiDAR makes 3D structure directly measurable rather than indirectly inferred from spectral signals. Key variables for ecological assessment and resource management can be quantified from point-cloud geometry. Therefore, the value of LiDAR lies not only in producing point clouds, but also in enabling these process-relevant variables to be measured through geometry itself.

The structural value is strengthened by the complementarity of different LiDAR platforms. These platforms form a multiscale evidence chain from regional baselines to local structural precision. Yet LiDAR AI remains constrained by cross-domain heterogeneity. The geometry, density, intensity, and acquisition conditions of LiDAR data vary substantially across platforms, sensors, seasons, and campaigns. Models trained under one condition may therefore not transfer reliably to another. Survey-grade point cloud AI must separate transferable structural information from platform-specific artifacts and campaign-specific shortcuts. Its goal is not only to improve in-domain recognition accuracy, but also to preserve evidentiary validity under realistic deployment.

The challenge is closely related to recent interest in open-vocabulary learning and domain generalization. A Web of Science Core Collection search showed that, since 2025, 7,721 publications have been indexed with topics related to open-vocabulary learning and domain generalization. It indicates that fixed-category models cannot flexibly respond to newly defined survey concepts, while models trained on one data distribution often become unreliable in unseen domains. Cross-dataset studies have shown that differences can cause substantial transfer degradation, with domain-adaptation methods improving target-domain performance by approximately 8-22%. Similar problems have been reported in synthetic-to-real LiDAR transfer and large-scale point-cloud understanding, where domain gaps, non-uniform point distributions, and cross-scene generalization remain major obstacles.

For 3D LiDAR-based natural resource surveying, these limitations are particularly consequential because survey targets are frequently region-specific, and dynamically updated, while observations vary across platforms, sensors, and acquisition campaigns. Therefore, a model that performs well within one campaign cannot be assumed to provide reliable evidence under another observation condition. The challenge is not only to broaden the set of recognizable concepts, but also to ensure that semantic recognition remains grounded in structural evidence that transfers under domain shift.

GROUNDED OPEN-VOCABULARY 3D UNDERSTANDING: AN EVIDENTIARY FRAMEWORK

The central challenge for next-generation natural resource surveying AI is to combine semantic openness with reliable evidence. Survey targets are often diverse and dynamically changing. At the same time, accepted semantic claims must be supported by transferable structural cues rather than prompt sensitivity, dataset bias, or platform-specific correlations. Because domain shift is a defining condition of natural resource surveying rather than an exception, the key problem is not simply to recognize more concepts. It is to make semantic flexibility compatible with survey-grade credibility.

In operational LiDAR deployment, model performance is shaped by interacting factors, including geometry, density variation, sensor effects, and environmental dynamics. These factors can improve inference by revealing fine structural detail, but they can also introduce domain-specific shortcuts. Therefore, better data does not automatically lead to better inference. Reliability depends on whether the model learns transferable structural evidence rather than sensor- or campaign-specific correlations.

Open-vocabulary learning alone cannot solve this problem. It provides a flexible semantic querying, but semantic claim must be constrained by measurable 3D evidence. Natural resource surveying involves both semantic and observational heterogeneity, where survey concepts evolve and the same target may appear differently across acquisition conditions. A grounded open-vocabulary framework translates queries into structural hypotheses and evaluates whether available evidence is sufficient to support a decision, rather than simply matching words to point clouds.

For example, a query such as “storm-damaged crown” should be interpreted through structural cues such as crown discontinuity, height reduction, and boundary asymmetry. When evidence is insufficient due to sparsity or occlusion, the system should express uncertainty and request expert verification. Therefore, we define grounded open-vocabulary LiDAR AI as an evidence-constrained framework rather than a specific architecture. Its implementation can be organized into three coupled mechanisms.

First, domain-aware structural tokenization encodes measurable structural attributes, including height distribution, stratification, curvature, roughness, connectivity, density reliability, and occlusion uncertainty. Unlike generic self-supervised representations, these structural tokens are organized around measurable evidence and explicitly account for domain-sensitive observation factors.

Second, prompt-to-structure formalization converts an open-vocabulary query into inspectable structural predicates before semantic matching. Rather than treating language as a direct label, it translates survey concepts into measurable geometric conditions that can be evaluated against extracted structural tokens. For example, "storm-damaged crown" can be interpreted through evidence such as crown discontinuity, height reduction, and boundary asymmetry.

Third, evidence-aware inference determines whether available evidence is sufficient to support a semantic claim. In survey-grade applications, abstention is preferable to unsupported predictions.[5] Therefore, the system should quantify uncertainty, identify sparse or out-of-distribution observations, and involve human experts when evidence remains ambiguous.

These mechanisms imply a different evaluation paradigm. Grounded open-vocabulary LiDAR AI cannot be assessed solely with conventional in-domain segmentation or classification scores. Evaluation must include cross-platform transfer, cross-season robustness, prompt stability, uncertainty calibration, abstention quality, and the correctness of structural justifications. A strong model should not only predict the correct category, but also remain stable when sensors change, resist spurious prompt wording, know when evidence is insufficient, and point to the structural basis of its decision. This is where the concept becomes operational: grounding is not an abstract philosophical preference, but a set of measurable system properties.

Seen in this light, the broader ambition is not merely to attach language to point clouds, but to build a surveying foundation model in which language serves as an interface to measurement. Such a model would learn not only cross-modal correspondences, but also which correspondences remain trustworthy across changing observational conditions. It would bridge global baselines and local geometric precision through representations that are semantically flexible but evidentially conservative. The promise of open vocabulary, in other words, lies not in maximizing recognizability, but in expanding what can be queried while narrowing what can be accepted without sufficient evidence. That distinction is, in our view, the technical core of survey-grade grounded open-vocabulary 3D LiDAR AI.

DISCUSSION AND OUTLOOK: TOWARD ROBUST SURVEY-GRADE 3D AI

This framing suggests a clear research agenda for 3D LiDAR AI in natural resource surveying. The field should move beyond the assumption that larger datasets, stronger backbones, and broader vocabularies will automatically produce more reliable survey intelligence. The more fundamental task is to design systems that remain trustworthy when the observation process itself changes. This requires a shift from benchmark-centric thinking to deployment-centric thinking.

Several priorities follow. First, datasets and benchmarks should be reorganized around cross-domain evidence consideration rather than only around category coverage. Models should be evaluated across platforms, seasons, ecosystems, and acquisition geometries, with special attention to whether they preserve structural reasoning under shift. Second, future models should report not only outputs but also the evidence used to support them, including structural cues, contextual assumptions, and principal sources of uncertainty. Third, human-in-the-loop design should be treated as part of the model architecture rather than as an external operational patch. In natural resource surveying, credibility depends not only on automation, but on whether the system can identify when expert verification is necessary. If downstream intelligence depends on transferable structural evidence, then future LiDAR acquisition protocols may need to be optimized not only for geometric density, but also for cross-campaign comparability, uncertainty traceability, and multi-platform harmonization. Models should likewise be trained and evaluated within realistic survey workflows, where measurement constraints, annotation uncertainty, expert interpretation, and temporal comparability all matter.

Grounded open-vocabulary 3D LiDAR AI should therefore not be understood as a generic call to combine vision-language models with 3D point clouds. Rather, it defines a technical research agenda centered on a specific question: how can semantic flexibility be made compatible with evidence-based explainability under domain shift? Answering this question would allow natural resource surveying to move from a largely interpretive workflow toward a more scalable, auditable, and scientifically defensible system for structural evidence extraction and change assessment.

REFERENCES

1. Wu K., Zhang Y., Ru L., et al. (2025). A semantic-enhanced multi-modal remote sensing foundation model for Earth observation. *Nat. Mach. Intell.* **7**: 1235–1249. DOI: 10.1038/s42256-025-01078-8

2. Maeda E.E., Brede B., Calders K., et al. (2025). Expanding forest research with terrestrial LiDAR technology. *Nat. Commun.* **16**:8853. DOI: 10.1038/s41467-025-63946-6

3. Prümers H., Betancourt C.J., Iriarte J., et al. (2022). Lidar reveals pre-Hispanic low-density urbanism in the Bolivian Amazon. *Nature* **606**:325–328. DOI: 10.1038/s41586-022-04780-4

4. Frachetti M.D., Berner J., Liu, X., et al. (2024). Large-scale medieval urbanism traced by UAV-lidar in highland Central Asia. *Nature* **634**:1118–1124. DOI: 10.1038/s41586-024-08086-5

5. Chang Y., Zhang J., Ding Y., et al. (2025). Towards intelligent digital ecological civilization: The rise of geospatial foundation models. *Innov. Inform.* **2**, 100023. DOI: 10.59717/j.xinn-inform.2026.100023

FUNDING AND ACKNOWLEDGEMENTS

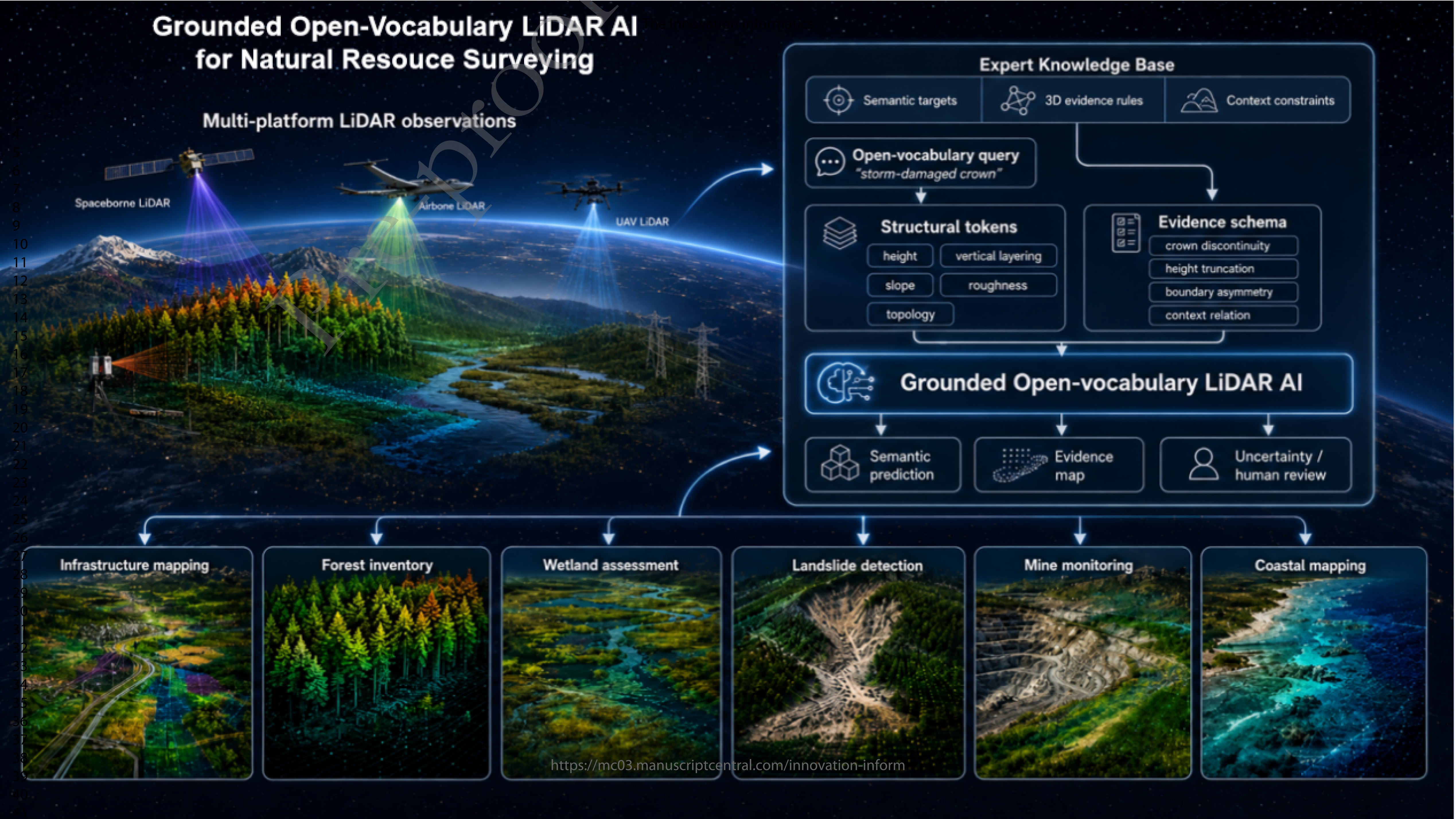
This work was supported by the National Special Project for Scientific and Technological Basic Resources Investigation of China (Grant No. 2022FY1001018), the National Key Research and Development Program of China (Grant No. 2025ZD1007105-05), the Central Fiscal Geological Survey Project of China (Grant No. DD2026101025), the National Natural Science Foundation of China (Grant No. 42371343). The authors gratefully acknowledge the financial support from these projects. We also thank colleagues and collaborators for their valuable discussions and technical assistance throughout the research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, A.X., T.H., and Y.C.; Methodology, X.W., D.Z., C.M., T.L., and Y.X.; Writing—original draft, A.X. and T.H.; Writing—review and editing, A.X., T.H., X.W., and Y.C.; All authors have read and agreed to the published version of the manuscript. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.


Grounded Open-Vocabulary LiDAR AI
for Natural Resouce Surveying
Multi-platform LiDAR observations
Spaceborne LiDAR
Airbone LiDAR
UAV LiDAR
Expert Knowledge Base
Semantic targets
3D evidence rules
Context constraints
Open-vocabulary query
"storm-damaged crown"
Structural tokens
height
vertical layering
slope
roughness
topology
Evidence schema
crown discontinuity
height truncation
boundary asymmetry
context relation
Grounded Open-vocabulary LiDAR AI
Semantic prediction
Evidence map
Uncertainty / human review
Infrastructure mapping
Forest inventory
Wetland assessment
Landslide detection
Mine monitoring
Coastal mapping