University of Idaho
College of Engineering

FUSING DOMAIN KNOWLEDGE, LLMS, AND GIS: A NEW FRAMEWORK FOR MINERAL PROSPECTING

Weilin Chen
11/19/2025

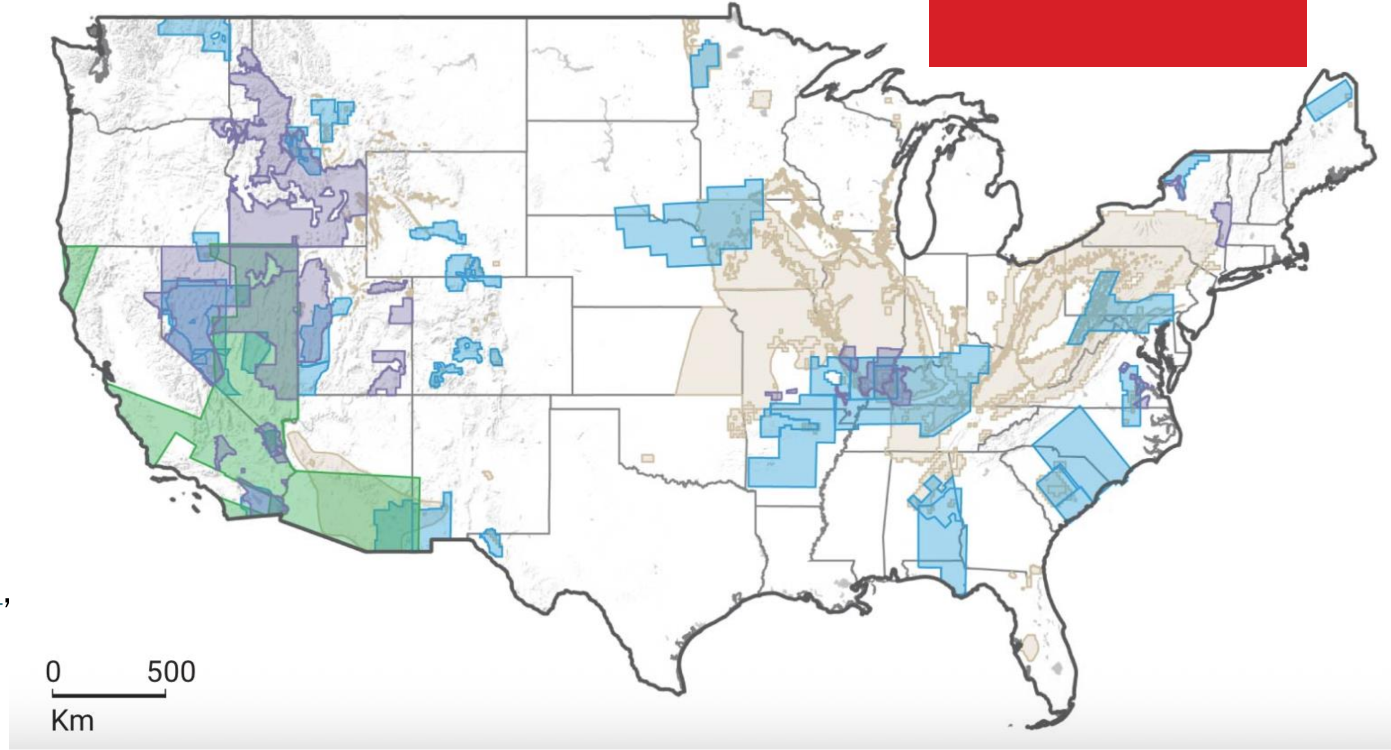
Critical Minerals – A Global Concern

WSJ

By the numbers: Of the 50 elements deemed critical to the American economy and national security by the U.S. Geological Survey,[1] the United States is 100% dependent on foreign suppliers for 12 of them and is more than 50% reliant on non-domestic sources for another 29.[2]

1. USGS, "U.S. Geological Survey releases 2022 list of critical minerals, news release, Feb. 22, 2022.
2. USGS, "Mineral commodity summaries 2025," Jan. 31, 2025.

Science AAAS



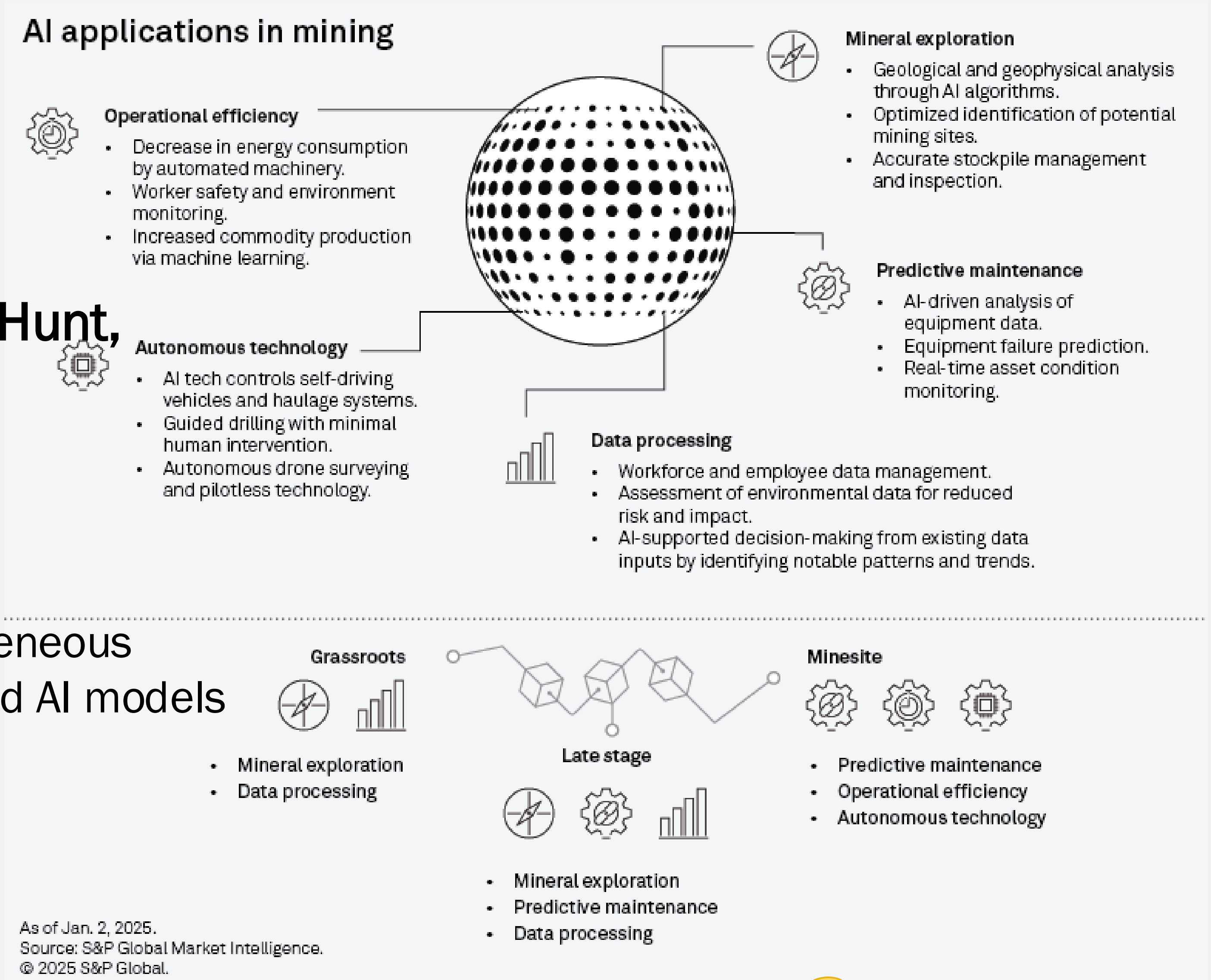
USGS Earth Mapping Resources Initiative

Research Motivation: Minerals & AI

U.S. DEPARTMENT of ENERGY

AI Tool Speeds Up Critical Mineral Hunt, Boosting U.S. Supply

- Critical minerals: Growing demand
- Mineral exploration: Costly, uncertain
- Geoscience data: Complex, heterogeneous
- Needs and trends: Knowledge-infused AI models



AI can do data mining —— but can AI do mining ?

INTRODUCTION
Why Neuro-Symbolic AI?

Relation between AI, ML, DL And NSAI. NSAI is a subfield of AI that combines both symbolic reasoning and DL

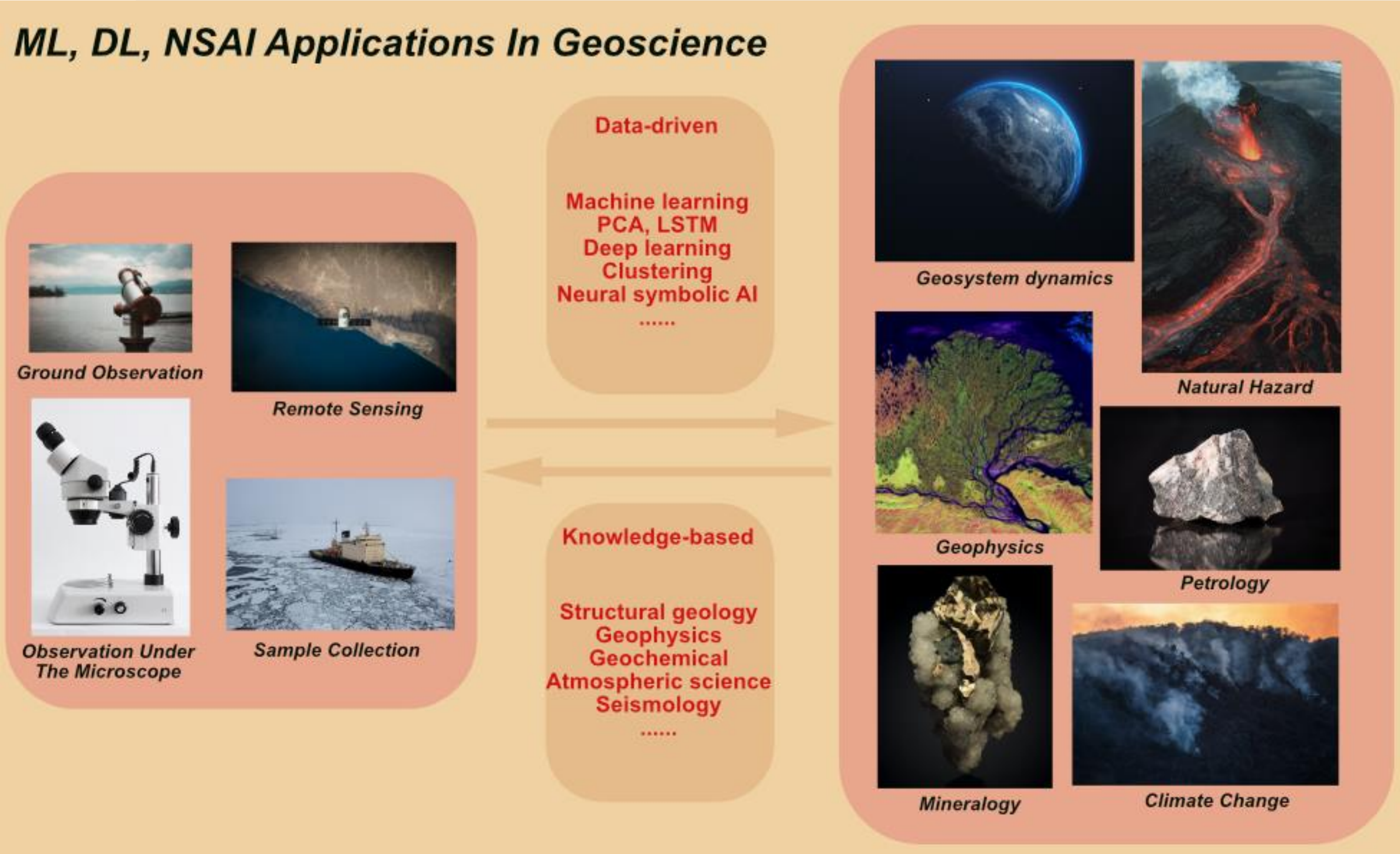
BACKGROUND
Mineral Prediction vs. Exploration

- Prediction: probability-based, model-driven
- Exploration: confirm deposits via fieldwork

Differences Between Mineral Prediction & Mineral Exploration

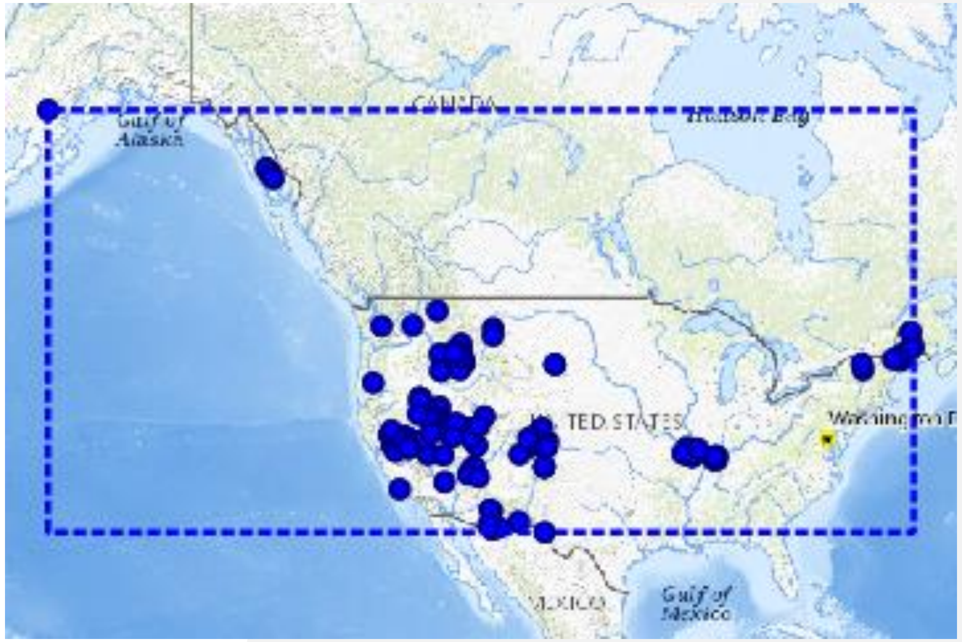
Category	Mineral Prediction	Mineral Exploration
Focus	Assessing likelihood of mineral occurrences	Actively searching for and evaluating deposits
Data Usage	Existing geological and geophysical data	New data collected through fieldwork
Timing	Before/at early stages of exploration	Occurs after mineral prediction
Certainty	Provides likelihood assessment, not definitive	Aims to obtain concrete evidence of mineralization
Scale	Regional or large-scale assessment	Site-specific investigation
Objective	Identify areas with high mineral potential	Discover new deposits and determine their viability
Methods	Geophysical modeling, statistical analysis, geological data	Field surveys, sampling, drilling, geophysical exploration
Results	Predictive maps or models of potential mineral occurrences	Direct evidence of mineral presence and quality
Investment	Lower cost compared to exploration	Higher cost due to fieldwork, drilling, and analysis
Risk	Lower risk as it relies on existing data and modeling	Higher risk as it involves direct exploration and sampling
Decision-making	Guides exploration efforts and target selection	Determines whether to develop or abandon a site

BACKGROUND
Applications of AI in Geoscience



- ML and DL already applied across geoscience domains
- Shows the breadth of AI use in geoscience
- But NSAI applications remain limited → opportunity

RULE-BASED NSAI IMPLEMENTATION
Research Design



Sourced from USGS website.
https://www.usgs.gov/data/national-geochemical-database-ore-deposits-legacy-data

- Data Cleaning:

- Removed duplicates, corrected outliers, and handled missing values.
- Filtered to extract Porphyry Cu-Mo-Au data
- Grouped by DEPOSIT_TYPE to identify unique deposit types, such as:
    - High sulfidation Au-Ag
    - High sulfidation Au-Ag
    - Porphyry Cu or High sulfidation Au-Ag
    - Polymetallic sulfide vein, and more.

- Source: National Geochemical Database: Ore Deposits (USGS).
- Nearly 30,000 samples, 15 mineral system types and 42 mineral deposit types
- Features include geochemical concentrations of Cu, Mo, Fe, and S.
- Task: classify mineral deposit types
- Approach: manually encode symbolic rules + integrate with ML

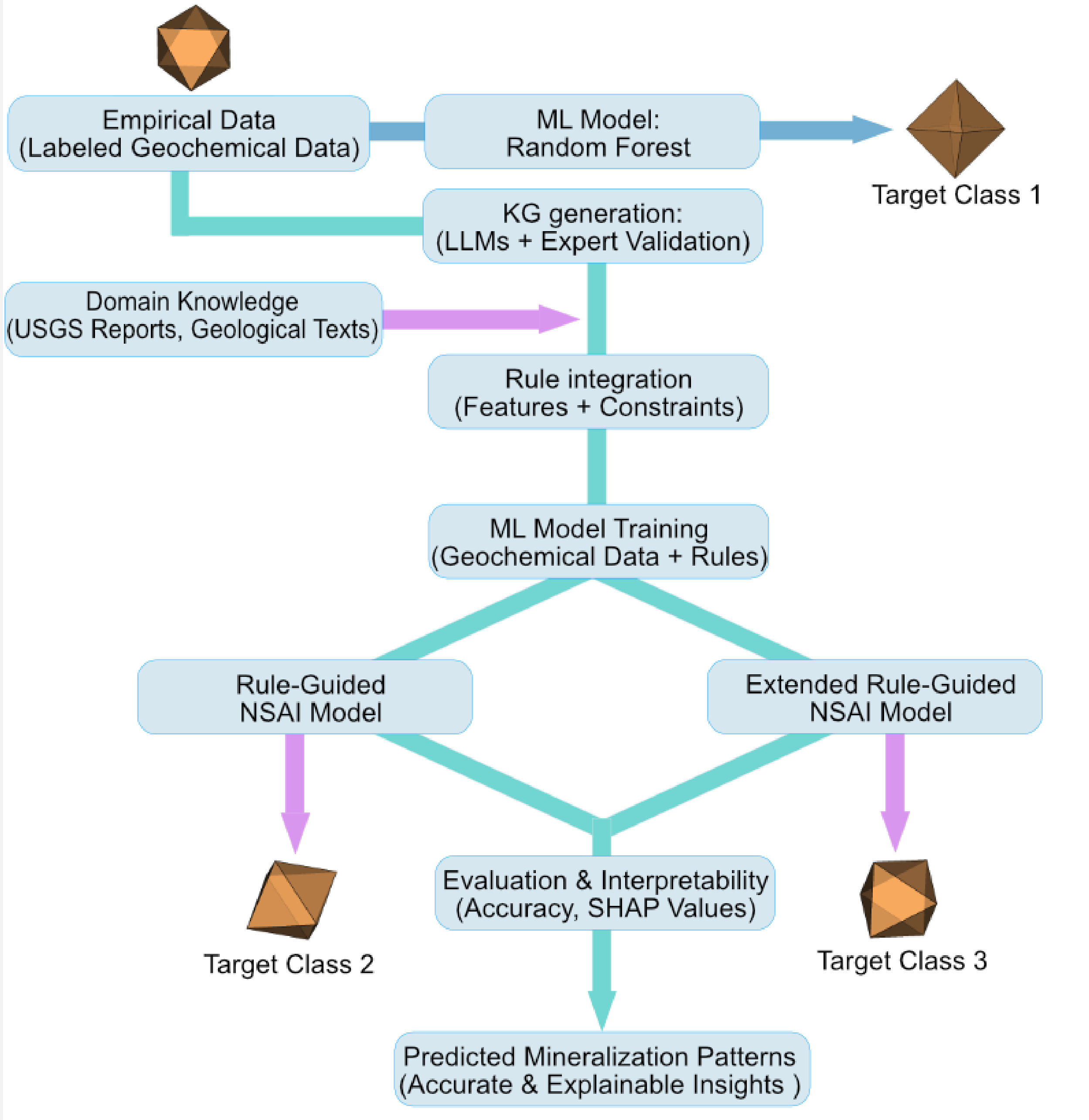
RULE-BASED NSAI IMPLEMENTATION

Framework Workflow

A General Framework Illustrating the Major Parts of the Method Designed in This Study:

(1) Data Collection and Preprocessing
(2) LLM-assisted KG Construction
(3) Rule integration into ML model training

Arrows indicate the flow of data and rule application



RULE-BASED NSAI IMPLEMENTATION
Model Setup

```
Function AddKnowledgeFeatures(dataframe):
    For each row in dataframe:
        If Cu_ppm > 200 AND Mo_ppm > 10 AND
K_pct > 2:
            Set 'PorphyryCuPotential' to 1
        Else:
            Set 'PorphyryCuPotential' to 0
        If W_ppm > 10 OR Fe_pct > 10 OR Ca_pct > 5:
            Set 'SkarnRelatedPotential' to 1
        Else:
            Set 'SkarnRelatedPotential' to 0
        If Cu_ppm > 500 AND Au_ppm > 0.1:
            Set 'HighSulfidationPotential' to 1
        Else:
            Set 'HighSulfidationPotential' to 0
        If Ag_ppm > 50 AND Au_ppm > 0.05:
            Set 'LowSulfidationPotential' to 1
        Else:
            Set 'LowSulfidationPotential' to 0
    Return updated dataframe
```

Rule-guided ML: basic symbolic features

```
Function AddKnowledgeFeatures(dataframe):
    For each row in the dataframe:
        Evaluate thresholds for key geochemical elements to define potential mineralization
types:
            - Porphyry Copper: Cu > 200 ppm, Mo > 10 ppm, K > 2%
            - Porphyry Copper-Molybdenum: Cu > 300 ppm, Mo > 15 ppm
            - Porphyry Copper-Gold: Cu > 300 ppm, Au > 0.15 ppm
            - Skarn-Related: W > 10 ppm OR Fe > 10% OR Ca > 5%
            - Skarn-Tungsten: W > 20 ppm AND Fe > 5%
            - High-Sulfidation Epithermal: Cu > 500 ppm AND Au > 0.1 ppm
            - High-Sulfidation Sulfur: S > 5%
            - Low-Sulfidation Epithermal: Ag > 50 ppm AND Au > 0.05 ppm
            - Low-Sulfidation Antimony: Sb > 10 ppm
            - Intermediate-Sulfidation: Zn > 50 ppm AND Au > 0.05 ppm AND Cu > 100 ppm
            - Polymetallic Sulfide: Pb > 100 ppm AND Zn > 100 ppm AND Ag > 10 ppm
            - Lithocap: Al > 5% AND K > 1%
            - Distal Disseminated: Ag > 30 ppm AND As > 50 ppm
            - High-Alumina Alteration: Al > 8% AND Si > 10%
        Assign binary indicators (1 for potential presence, 0 otherwise) for each potential type.
    Return updated dataframe
```

Extended NSAI: KG-derived rules

RULE-BASED NSAI IMPLEMENTATION
Evaluation Metrics

- Accuracy: Overall prediction correctness.
- Precision: Ability to minimize false positives.
- Recall: Ability to correctly identify actual positive cases.
- F1 Score: Balance between precision and recall.
- Macro Avg F1: Average F1 across all classes equally.
- Weighted F1: Average F1 weighted by class size.
- SHAP values were used to interpret the contribution of features

RULE-BASED NSAI IMPLEMENTATION
Performance Results

Performance Comparison of the Three Models

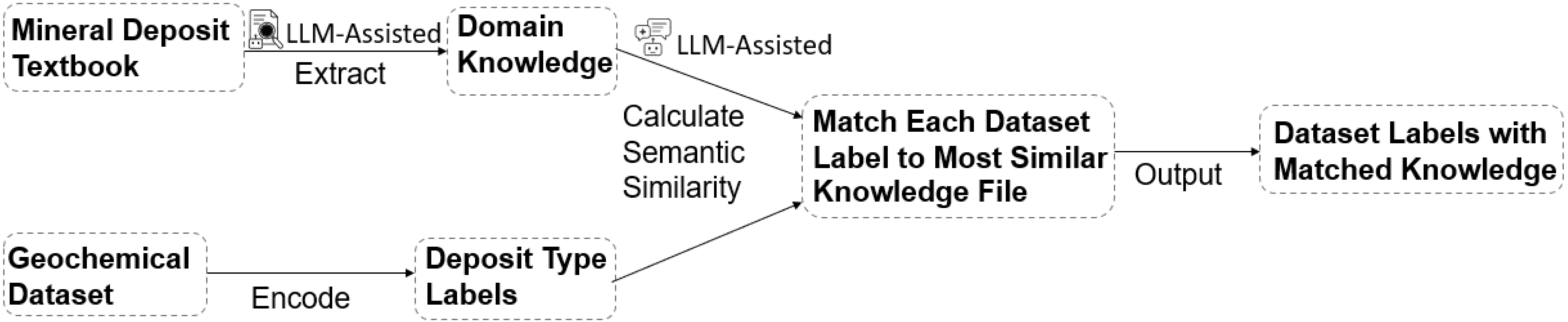
Metric	Model 1	Model 2	Model 3
Accuracy	0.9897 (Test)	0.9897 (Test)	0.9906 (Test)
	0.9874 (Validation)	0.9883 (Validation)	0.9893 (Validation)
Macro Avg F1	0.88 (Test)	0.86 (Test)	0.88 (Test)
	0.75 (Validation)	0.76 (Validation)	0.79 (Validation)
Weighted F1	0.99 (Test)	0.99 (Test)	0.99 (Test)
	0.99 (Validation)	0.99 (Validation)	0.99 (Validation)
Recall	Lower for minority classes	Improved recall for minority classes	Highest recall for minority classes
Recall (High sulfidation Au-Ag (Cu))	0.25 (Validation)	0.50 (Validation)	0.50 (Validation)
Recall (Low sulfidation Au-Ag)	0.78 (Validation)	0.89 (Validation)	0.89 (Validation)
Recall (Polymetallic sulfide skarn/replacement)	0.00 (Validation)	0.00 (Validation)	0.33 (Validation)
Recall (Polymetallic sulfide replacement)	0.75 (Validation)	0.75 (Validation)	1.00 (Validation)

- Accuracy increased with symbolic rules
- Macro F1 increased for minority deposit types

LLM-ASSISTED NSAI FRAMEWORK
Motivation for Automation

- Manual rule extraction is a bottleneck: Time-consuming, bias & inconsistency
- Scalability challenge: cannot keep up with the complexity of large geoscientific datasets
- Opportunity with Large Language Models (LLMs):
  - Automatically extract geological rules from literature
  - Structure rules into machine-readable forms (e.g., JSON, KG)
  - Match rules with dataset labels → enable automated NSAI pipelines
- Goal: make NSAI scalable, adaptable, and less reliant on human intervention

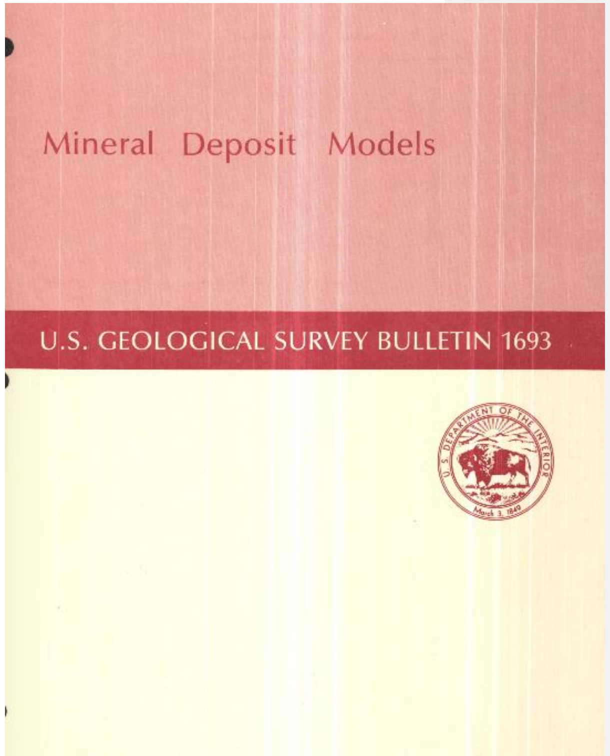
LLM-ASSISTED NSAI FRAMEWORK
LLM-Assisted Workflow



Workflow for LLM-assisted semantic matching between geochemical dataset labels and textbook-derived knowledge items.

LLM-ASSISTED NSAI FRAMEWORK

Descriptive Models



Model 25a

DESCRIPTIVE MODEL OF HOT-SPRING Au-Ag

By Byron R. Berger

DESCRIPTION Fine-grained silica and quartz in silicified breccia with gold, pyrite, and Sb and As sulfides (see fig. 105).

GENERAL REFERENCE Berger (1985).

GEOLOGICAL ENVIRONMENT

Rock Types Rhyolite.

Textures Porphyritic, brecciated.

Age Range Mainly Tertiary and Quaternary.

Depositional Environment Subaerial rhyolitic volcanic centers, rhyolite domes, and shallow parts of related geothermal systems.

Tectonic Setting(s) Through-going fracture systems related to volcanism above subduction zones, rifted continental margins. Leaky transform faults.

Associated Deposit Types Epithermal quartz veins, hot-spring Hg, placer Au.

DEPOSIT DESCRIPTION

Mineralogy Native gold + pyrite + stibnite + realgar; or arsenopyrite ± sphalerite ± chalcopyrite ± fluorite; or native gold + Ag-selenide or tellurides + pyrite.

Texture/Structure Crustified banded veins, stockworks, breccias (cemented with silica or uncemented). Sulfides may be very fine grained and disseminated in silicified rock.

Alteration Top of bottom of system: chalcedonic sinter, massive silicification, stockworks and veins of quartz + adularia and breccia cemented with quartz, quartz + chlorite. Veins generally chalcedonic, some opal. Some deposits have alunite and pyrophyllite. Ammonium feldspar (buddingtonite) may be present.

Ore Controls Through-going fracture systems, brecciated cores of intrusive domes; cemented breccias important carrier of ore.

Weathering Bleached country rock, yellow limonites with jarosite and fine-grained alunite, hematite, goethite.

Geochemical Signature Au + As + Sb + Hg + Tl higher in system, increasing Ag with depth, decreasing As + Sb + Tl + Hg with depth. Locally, NH_4, W.

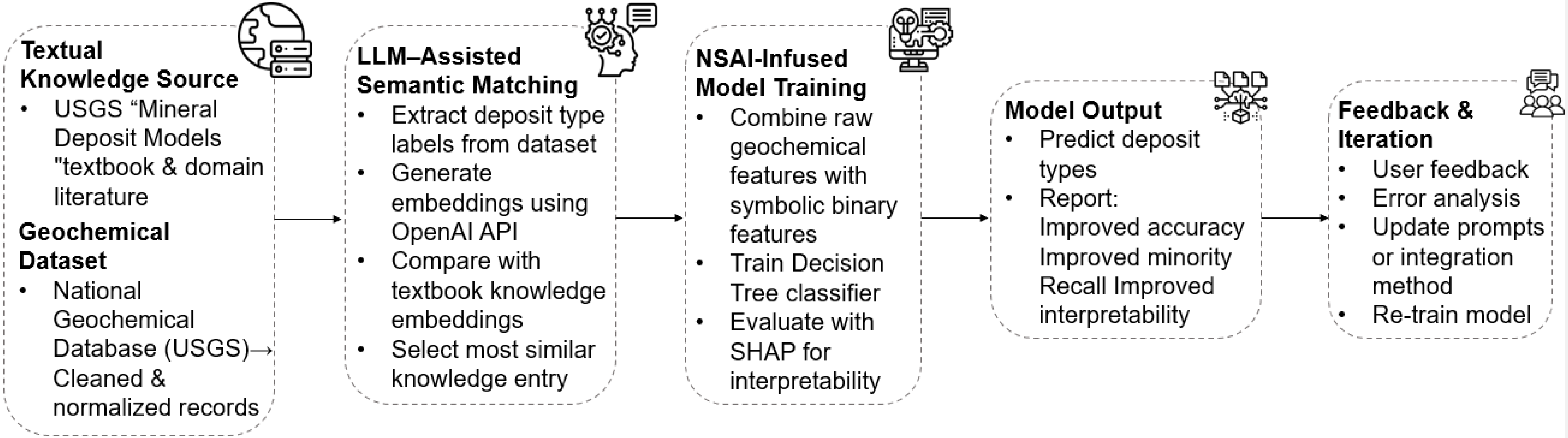
EXAMPLES

McLaughlin, USCA	(Averitt, 1945 and Becker, 1888)
Round Mountain, USNV	(Tingley and Berger, 1985)
Delamar, USID	(Lindgren, 1900)


```
{
  "Model_Name": "DESCRIPTIVE MODEL OF HOT-SPRING Au-Ag",
  "DESCRIPTION": "Fine-grained silica and quartz in silicified breccia with gold,
    pyrite, and Sb and As sulfides.",
  "Rock_Types": "Rhyolite.",
  "Textures": "Porphyritic, brecciated.",
  "Depositional_Environment": "Subaerial rhyolitic volcanic centers, rhyolite domes,
    and shallow parts of related geothermal systems.",
  "Tectonic_Settings": "Through-going fracture systems related to volcanism above
    subduction zones, rifted continental margins. Leaky transform faults.",
  "Associated_Deposit_Types": "Epithermal quartz veins, hot-spring Hg, placer Au.",
  "Mineralogy": "Native gold + pyrite + stibnite + realgar; or arsenopyrite +
    sphalerite + chalcopyrite + fluorite; or native gold + Ag-selenide or tellurides
    + pyrite.",
  "Texture_Structure": "Crustified banded veins, stockworks, breccias (cemented with
    silica or uncemented). Sulfides may be very fine grained and disseminated in
    silicified rock.",
  "Alteration": "Top of bottom of system: chalcedonic sinter, massive
    silicification, stockworks and veins of quartz + adularia and breccia cemented
    with quartz, quartz + chlorite. Veins generally chalcedonic, some opal. Some
    deposits have alunite and pyrophyllite. Ammonium feldspar (buddingtonite) may be
    present.",
  "Ore_Controls": "Through-going fracture systems, brecciated cores of intrusive
    domes; cemented breccias important carrier of ore.",
  "Weathering": "Bleached country rock, yellow limonites with jarosite and
    fine-grained alunite, hematite, goethite.",
  "Geochemical_Signature": "Au + As + Sb + Hg + Tl higher in system, increasing Ag
    with depth, decreasing As + Sb + Tl + Hg with depth. Locally, NH4, W."
}
```


An example of knowledge about the deposit model.

LLM-ASSISTED NSAI FRAMEWORK
Enhanced NSAI Framework



An enhanced NSAI framework that integrates geochemical data with LLM-assisted semantic matching to improve mineral deposit prediction, interpretability, and minority class performance.

LLM-ASSISTED NSAI FRAMEWORK
Performance Results

```
[Baseline ML] 5-fold CV (TRAIN) — mean ± SD
  Accuracy     : 0.843 ± 0.020
  Macro F1     : 0.781 ± 0.044
  Weighted F1  : 0.838 ± 0.022
  Best params  : {'max_depth': None, 'min_samples_split': 2}

[Baseline ML] Final Test (held-out 20%)
  Accuracy     : 0.875
  Macro F1     : 0.850
  Weighted F1  : 0.875
```

📊 Classification Report (Test Set – Baseline ML):

	precision	recall	f1-score	support
Carbonatite	0.79	0.95	0.86	20
High sulfidation Au–Ag	0.71	0.60	0.65	20
High sulfidation Au–Ag (Cu)	0.67	1.00	0.80	4
Lithocap alunite/High sulfidation Au–Ag	0.80	0.80	0.80	20
Lithocap kaolinite/High sulfidation Au–Ag	0.69	0.64	0.67	14
Low sulfidation Au	1.00	0.75	0.86	20
Low sulfidation Au–Ag	0.90	0.95	0.93	20
Low sulfidation epithermal Au–Ag	0.86	0.95	0.90	20
Polymetallic sulfide skarn	0.55	0.67	0.60	9
Polymetallic sulfide skarn/replacement	1.00	0.50	0.67	2
Polymetallic sulfide vein	0.95	0.95	0.95	20
Porphyry Cu	1.00	1.00	1.00	20
Porphyry Cu (Au)	0.90	0.95	0.93	20
Porphyry Cu or High sulfidation Au–Ag	1.00	0.85	0.92	20
Porphyry Cu–Mo	0.91	1.00	0.95	20
Porphyry/skarn Cu	1.00	0.95	0.97	20
Porphyry/skarn Cu and Distal disseminated Au–Au	1.00	1.00	1.00	12
accuracy			0.88	281
macro avg	0.87	0.85	0.85	281
weighted avg	0.88	0.88	0.87	281

```
[Embeddings only] 5-fold CV (TRAIN) — mean ± SD
  Accuracy    : 0.743 ± 0.001
  Macro F1    : 0.593 ± 0.000
  Weighted F1 : 0.677 ± 0.001
  Best params : {'max_depth': None, 'min_samples_split': 2}

[Embeddings only] Final Test (held-out 20%)
  Accuracy    : 0.744
  Macro F1    : 0.593
  Weighted F1 : 0.678
```

📊 Classification Report (Test – Embeddings only):

	precision	recall	f1-score	support
Carbonatite	1.00	1.00	1.00	20
High sulfidation Au–Ag	1.00	1.00	1.00	20
High sulfidation Au–Ag (Cu)	0.00	0.00	0.00	4
Lithocap alunite/High sulfidation Au–Ag	0.29	1.00	0.45	20
Lithocap kaolinite/High sulfidation Au–Ag	0.00	0.00	0.00	14
Low sulfidation Au	0.45	1.00	0.62	20
Low sulfidation Au–Ag	0.00	0.00	0.00	20
Low sulfidation epithermal Au–Ag	1.00	1.00	1.00	20
Polymetallic sulfide skarn	1.00	1.00	1.00	9
Polymetallic sulfide skarn/replacement	0.00	0.00	0.00	2
Polymetallic sulfide vein	1.00	1.00	1.00	20
Porphyry Cu	1.00	1.00	1.00	20
Porphyry Cu (Au)	1.00	1.00	1.00	20
Porphyry Cu or High sulfidation Au–Ag	1.00	1.00	1.00	20
Porphyry Cu–Mo	1.00	1.00	1.00	20
Porphyry/skarn Cu	0.00	0.00	0.00	20
Porphyry/skarn Cu and Distal disseminated Au–Au	0.00	0.00	0.00	12
accuracy			0.74	281
macro avg	0.57	0.65	0.59	281
weighted avg	0.65	0.74	0.68	281

LLM-ASSISTED NSAI FRAMEWORK
Performance Results

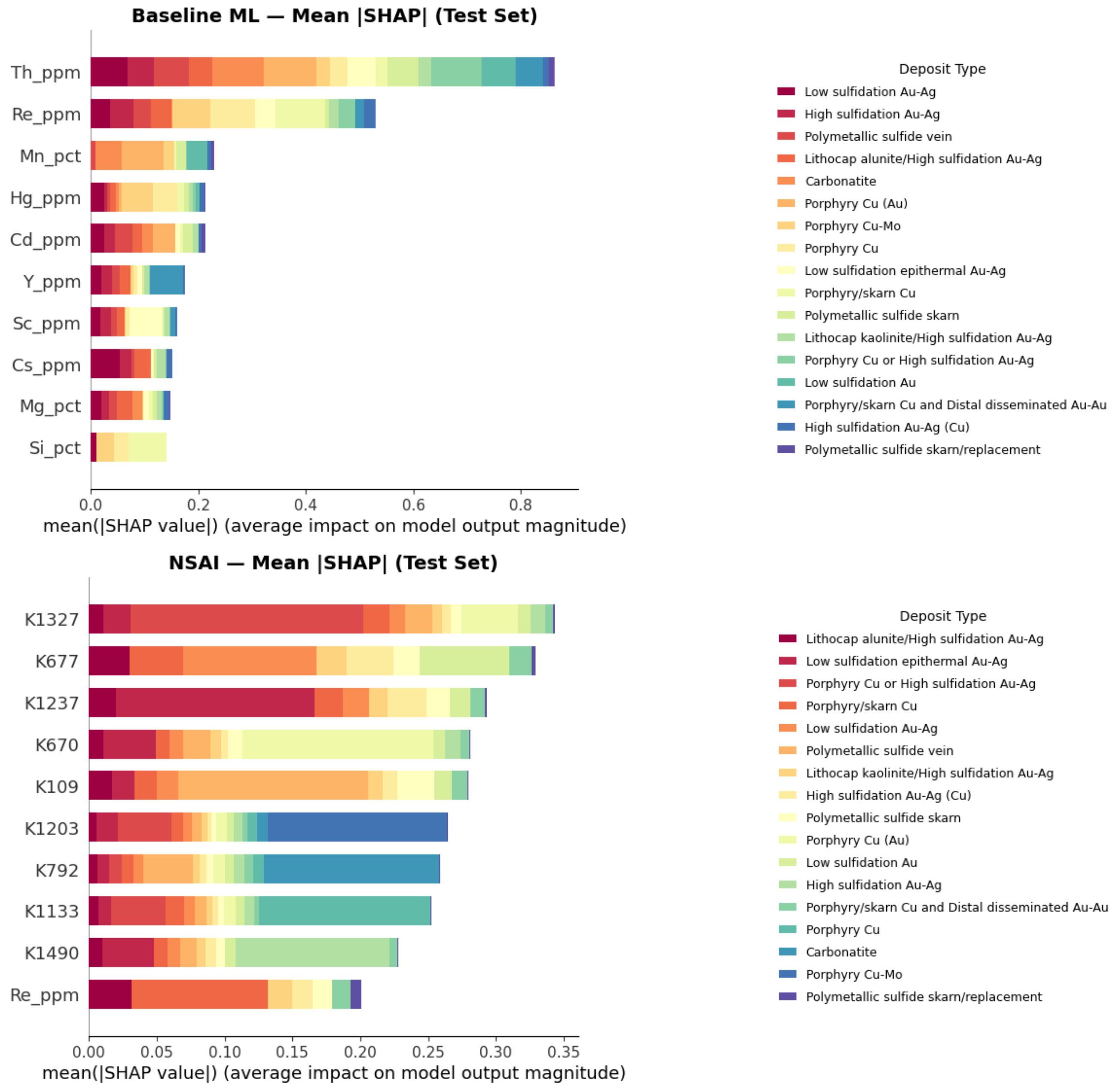
```
[NSAI] 5-fold CV (TRAIN) – mean ± SD
  Accuracy     : 0.979 ± 0.008
  Macro F1     : 0.974 ± 0.013
  Weighted F1  : 0.978 ± 0.008
  Best params  : {'max_depth': None, 'min_samples_split': 5}

[NSAI] Final Test (held-out 20%)
  Accuracy     : 0.979
  Macro F1     : 0.961
  Weighted F1  : 0.978
```

📊 Classification Report (Test Set – NSAI):

	precision	recall	f1-score	support
Carbonatite	1.00	1.00	1.00	20
High sulfidation Au-Ag	1.00	1.00	1.00	20
High sulfidation Au-Ag (Cu)	1.00	1.00	1.00	4
Lithocap alunite/High sulfidation Au-Ag	0.82	0.90	0.86	20
Lithocap kaolinite/High sulfidation Au-Ag	0.85	0.79	0.81	14
Low sulfidation Au	1.00	1.00	1.00	20
Low sulfidation Au-Ag	1.00	1.00	1.00	20
Low sulfidation epithermal Au-Ag	1.00	1.00	1.00	20
Polymetallic sulfide skarn	1.00	1.00	1.00	9
Polymetallic sulfide skarn/replacement	1.00	0.50	0.67	2
Polymetallic sulfide vein	1.00	1.00	1.00	20
Porphyry Cu	1.00	1.00	1.00	20
Porphyry Cu (Au)	1.00	1.00	1.00	20
Porphyry Cu or High sulfidation Au-Ag	1.00	1.00	1.00	20
Porphyry Cu-Mo	1.00	1.00	1.00	20
Porphyry/skarn Cu	1.00	1.00	1.00	20
Porphyry/skarn Cu and Distal disseminated Au-Au	1.00	1.00	1.00	12
accuracy			0.98	281
macro avg	0.98	0.95	0.96	281
weighted avg	0.98	0.98	0.98	281

LLM-ASSISTED NSAI FRAMEWORK
Feature Importance

This SHAP summary plots illustrate the relative importance of geochemical features in the baseline ML model and how the NSAI model integrates symbolic knowledge-derived features (K-features) with geochemical data to drive predictions.

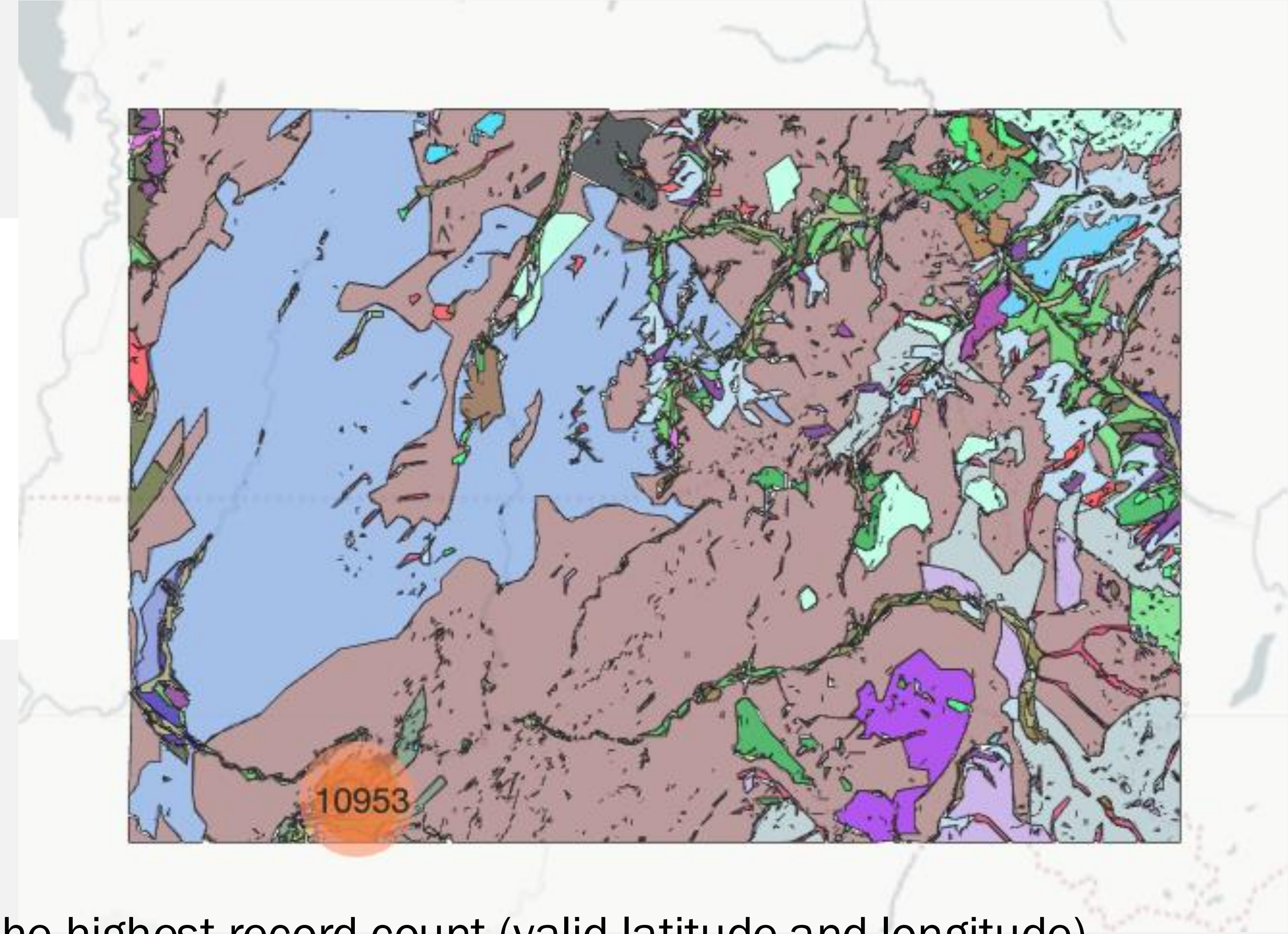


LLM-ASSISTED NSAI FRAMEWORK
Discussion

- NSAI improves generalization, especially for minority deposit types with very limited samples
- LLMs make NSAI scalable, automating rule extraction and reducing reliance on manual encoding
- The hybrid representation (geochemical + symbolic vectors) provides more stable predictions than purely data-driven models
- Confirms the value of embedding expert geological knowledge into ML workflows through automated pipelines

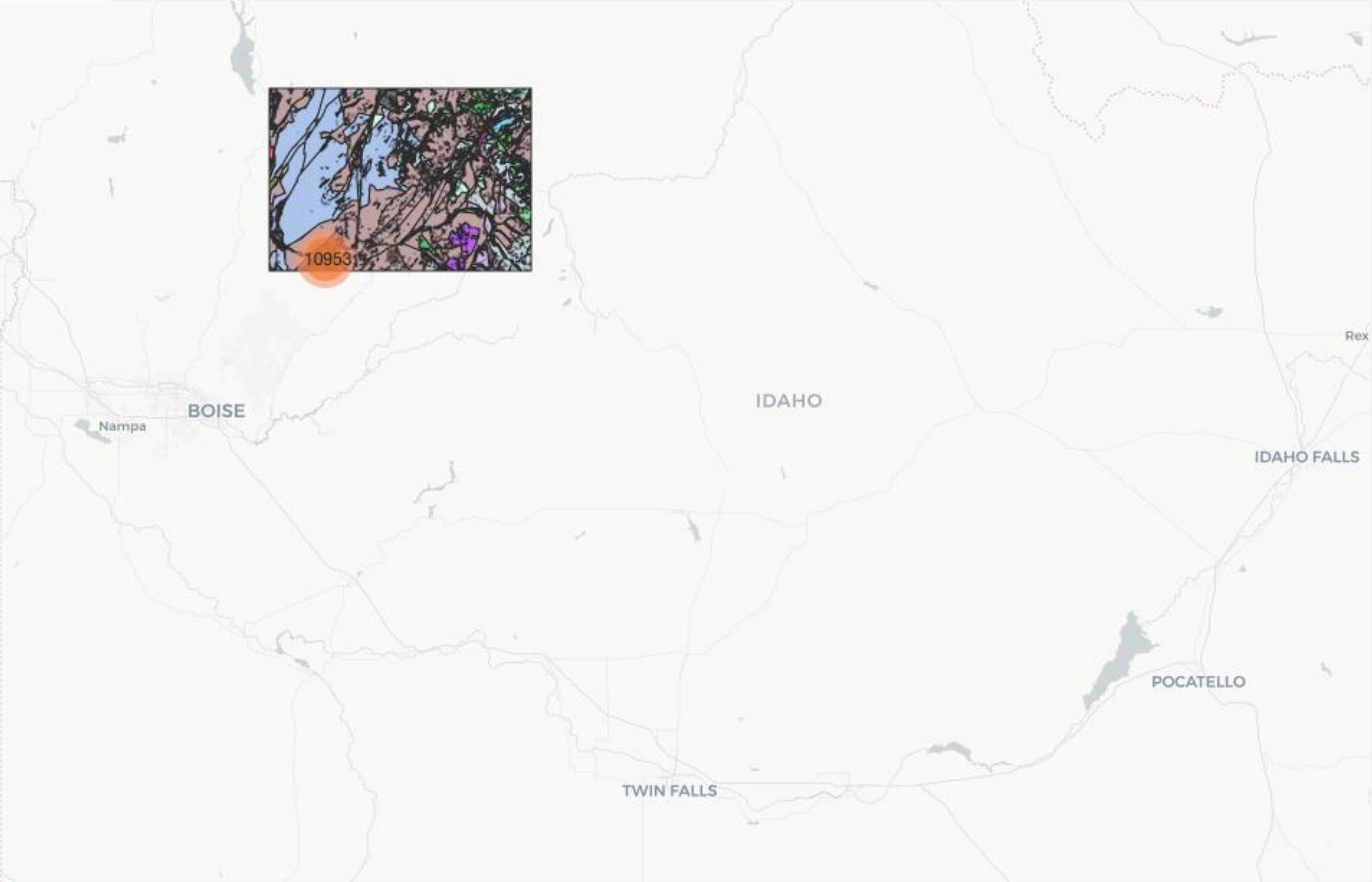
Practical Applications

```
Top 10 DEPOSIT_NAME by record count (valid LAT/LON):
DEPOSIT_NAME
Cumo                          10953
Pebble                         7488
Red Mountain                    944
Ann-Mason                       618
Rain orebody                    563
Betze                           494
Paradise Peak                   453
Cresson (Ruby stockpile)        445
Screamer                        371
Chimney Creek                   326
```

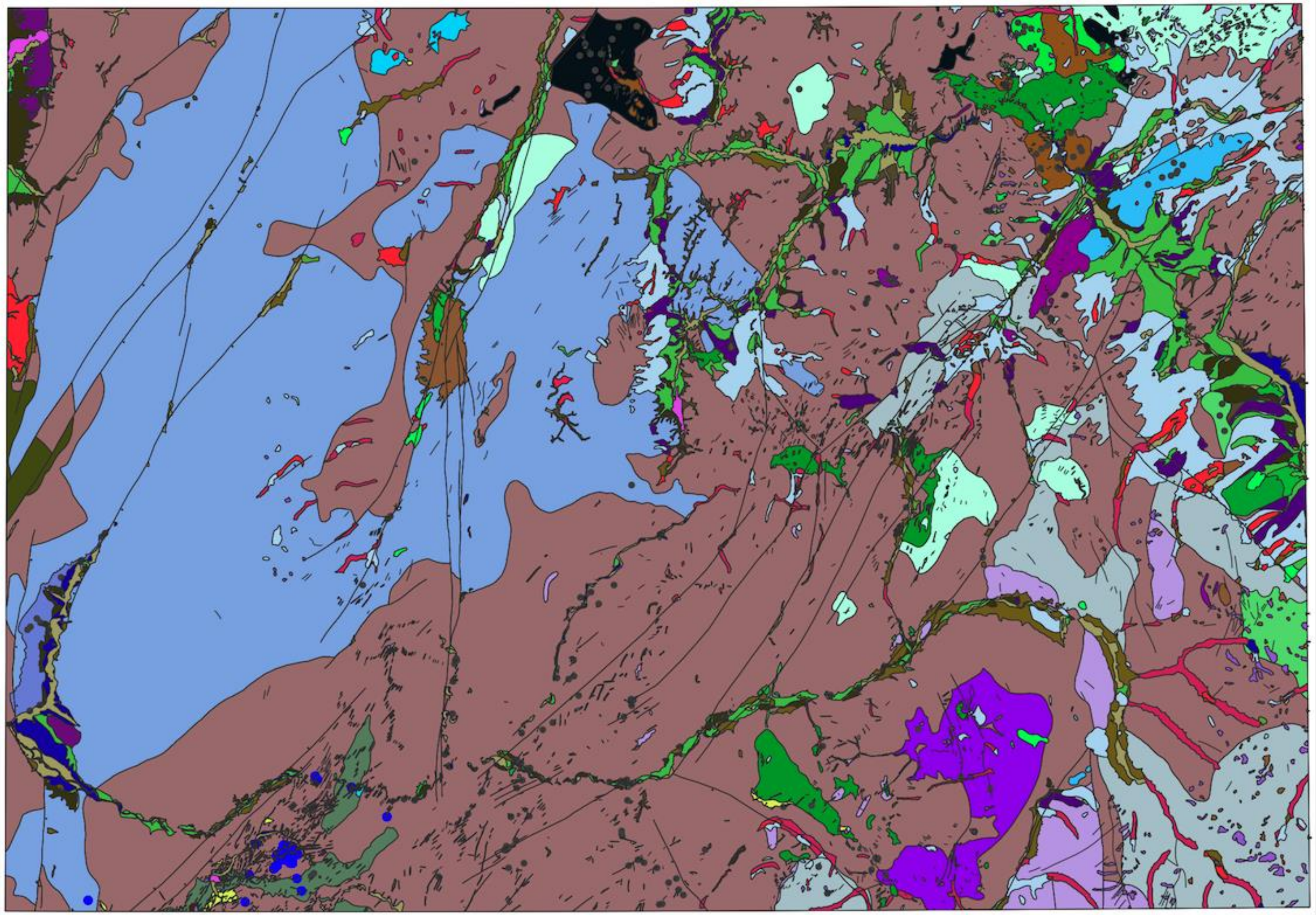


Top 10 deposit names with the highest record count (valid latitude and longitude) from the National Geochemical Database on Ore Deposits: Legacy data

TOWARD PRACTICAL APPLICATIONS
Research Area

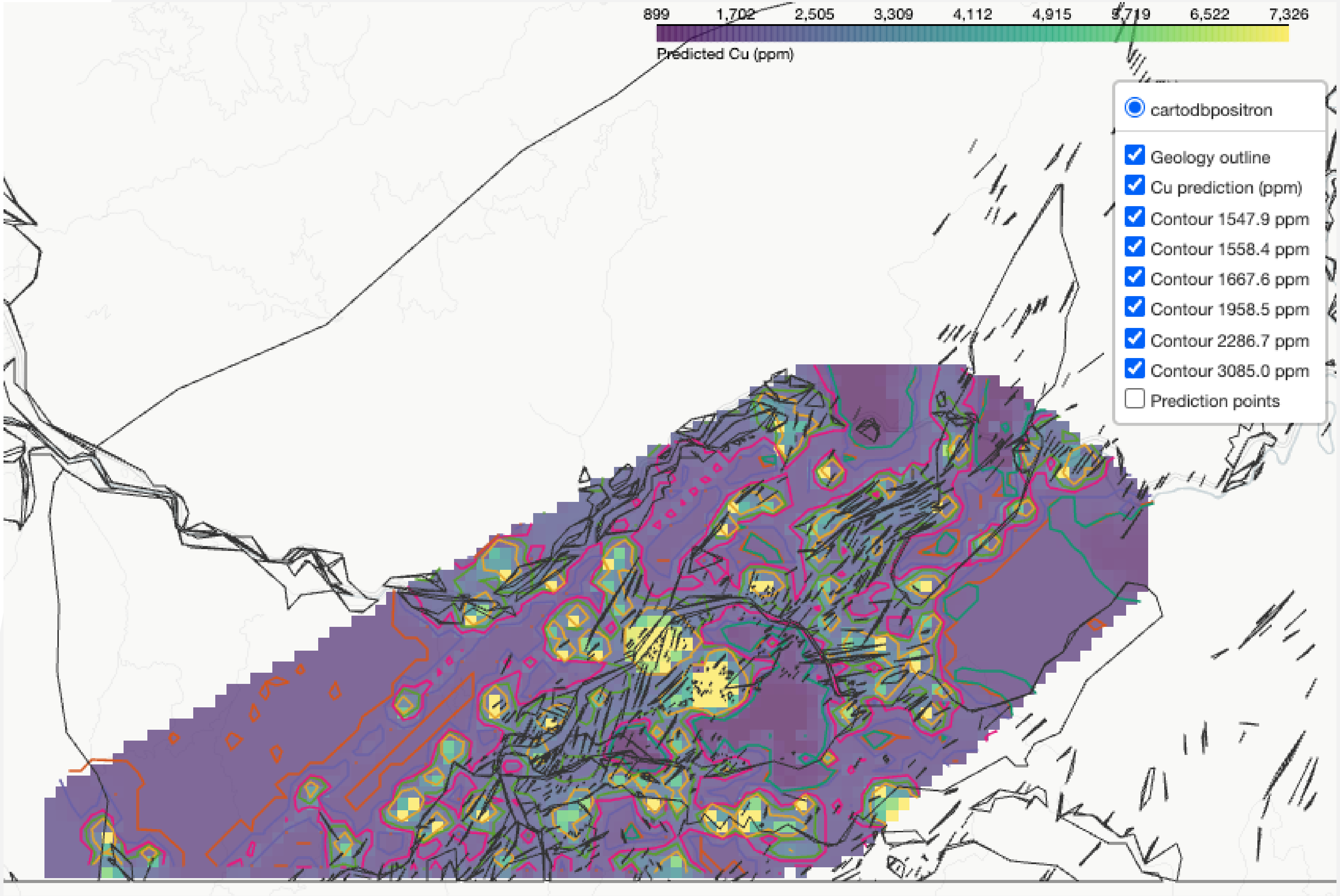


TOWARD PRACTICAL APPLICATIONS
Research Area

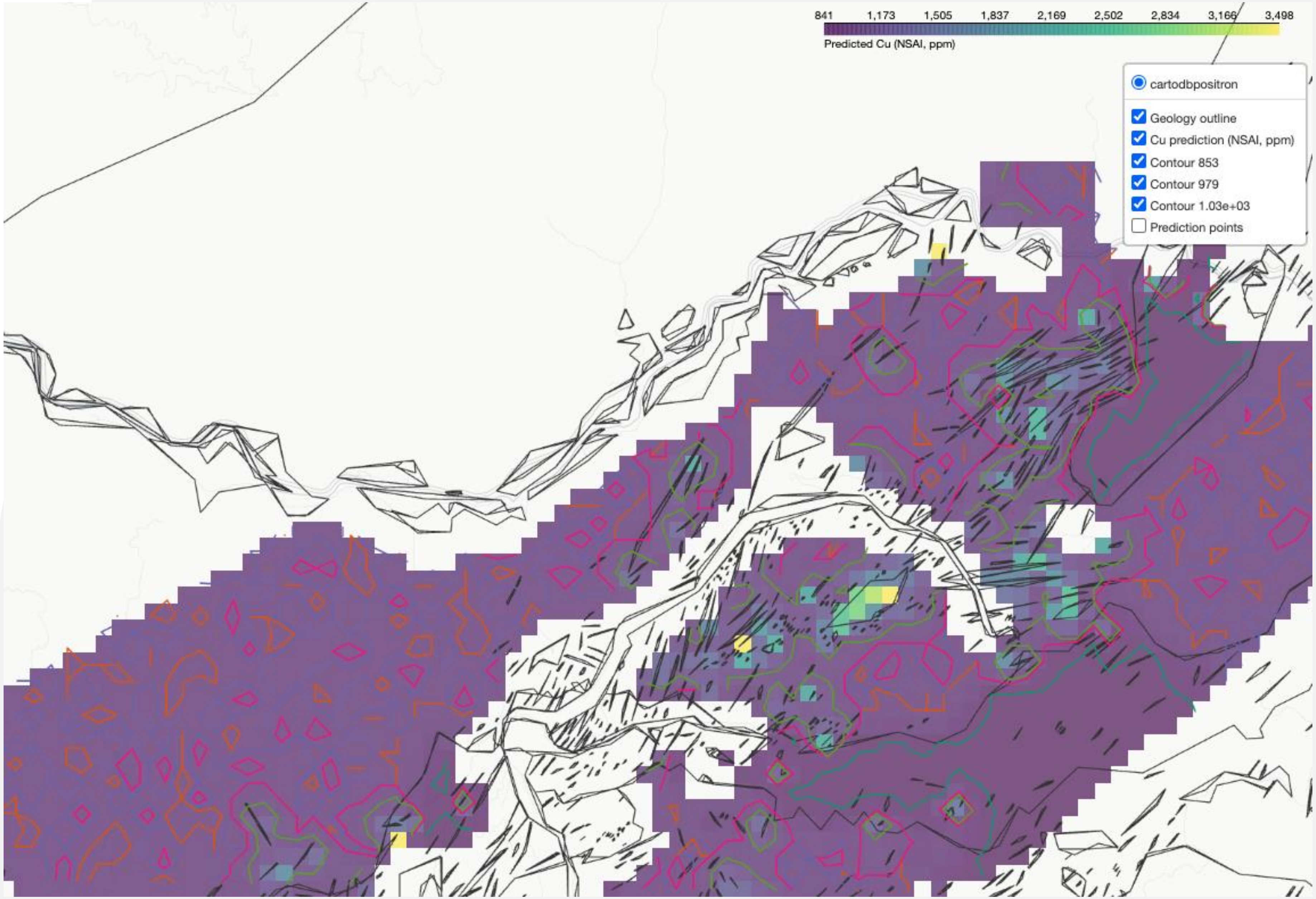


Recorded points on Deadwood River Geologic Map

TOWARD PRACTICAL APPLICATIONS
Results

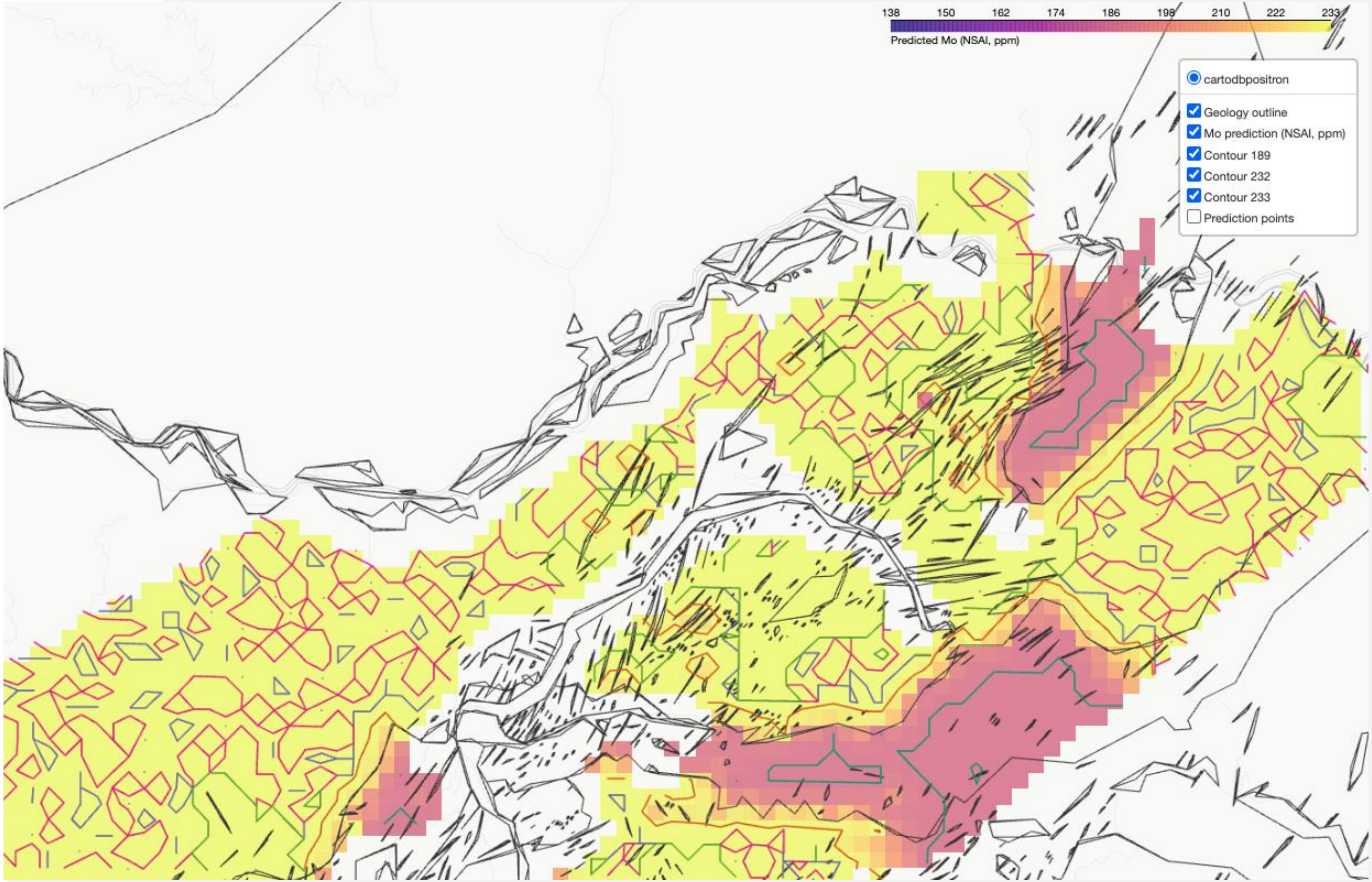


TOWARD PRACTICAL APPLICATIONS
Results



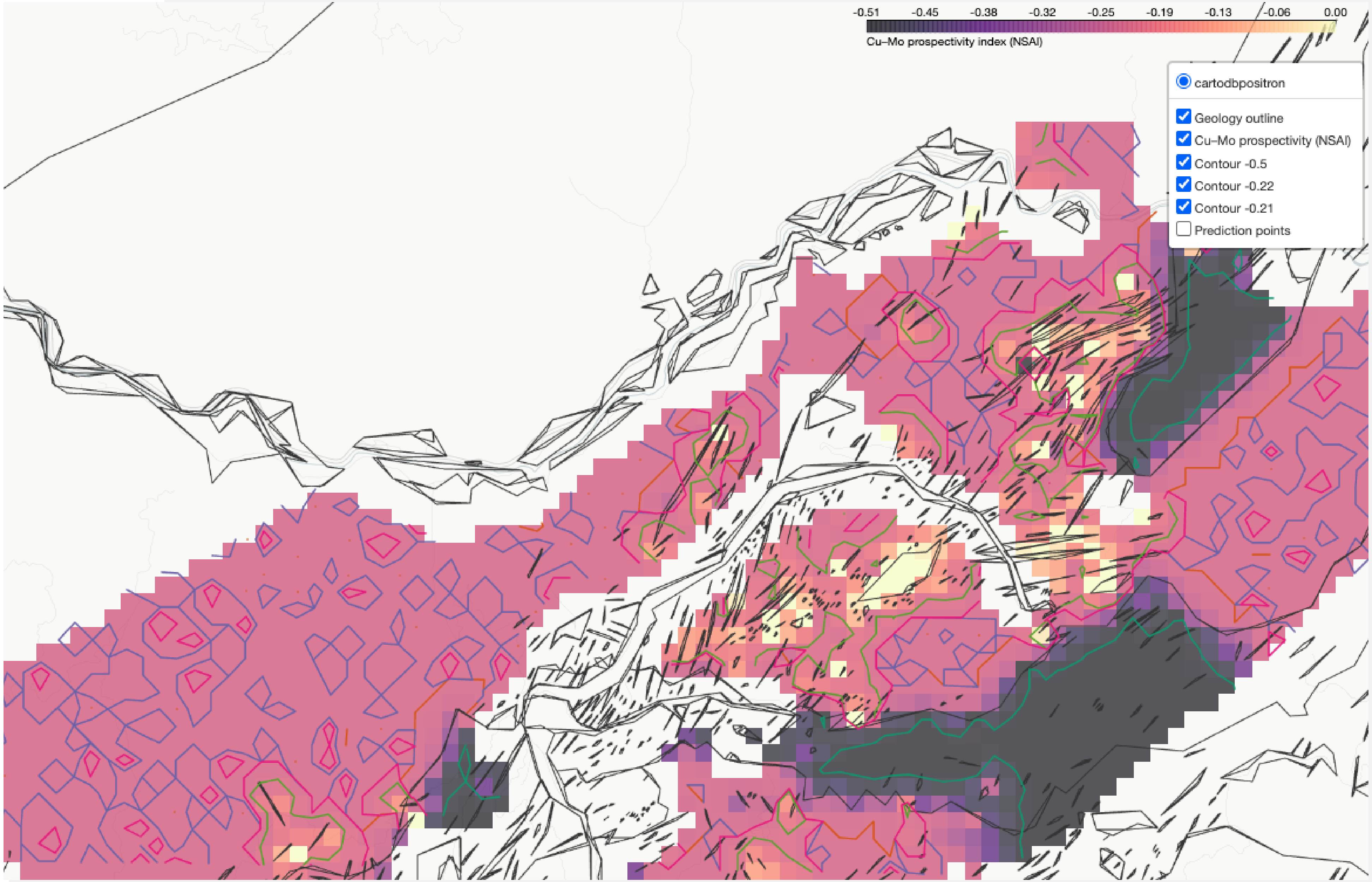
TOWARD PRACTICAL APPLICATIONS

Results



TOWARD PRACTICAL APPLICATIONS

Results



CONCLUSIONS & FUTURE WORK
Conclusion

- NSAI bridges the gap between black-box AI and expert geological reasoning
- Manual NSAI proved feasibility; LLM-assisted NSAI achieved scalability and automation
- Together, they show NSAI can deliver accurate, interpretable, and geologically coherent predictions
- This work lays the foundation for next-generation exploration AI systems trusted by geoscientists

AI can do data mining — and with NSAI, AI can support mineral mining.

University of Idaho
College of Engineering

THANK YOU FOR YOUR ATTENTION!