[github.com/google/curie](github.com/google/curie) [arxiv.org/abs/2503.13517](arxiv.org/abs/2503.13517)

Google Research

# CURIE: Evaluating LLMs On Multitask Scientific Long Context Understanding and Reasoning

ICLR

Hao Cui*[1], Zahra Shamsi*[1], Gowoon Cheon*[1], Xuejian Ma[1], Shutong Li[1], Maria Tikhanovskaya[2], Peter Norgaard[1], Nayantara Mudur[2], Martyna Plomecka[3], Paul Raccuglia[1], Yasaman Bahri[1], Victor V. Albert[4,5], Pranesh Srinivasan[1], Haining Pan[6], Philippe Faist[7], Brian Rohr[8], Michael J. Statt[8], Dan Morris[1], Drew Purves[1], Elise Kleeman[1], Ruth Alcantara[1], Matthew Abraham[1], Muqthar Mohammad[1], Ean Phing VanLee[1], Chenfei Jiang[1], Elizabeth Dorfman[1], Eun-Ah Kim[9], Michael Brenner[1,2], Viren Jain[1], Sameera Ponda[1], Subhashini Venugopalan*^[1]

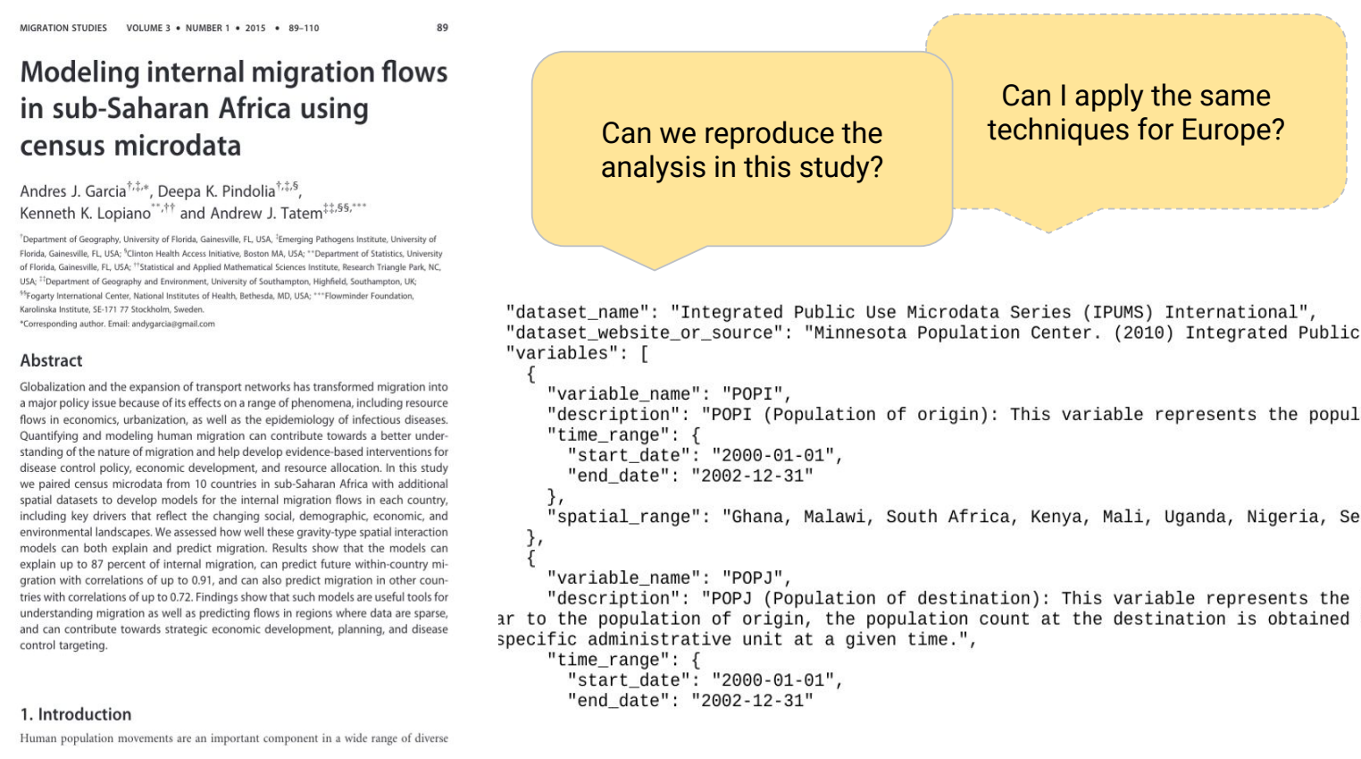
[1]Google, [2]Harvard, [3]University of Zurich, [4]NIST, [5]UMD College Park, [6]Rutgers, [7]FU Berlin, [8]Modelyst, [9]Cornell

{vsubhashini}@google.com

## Can LLMs assist Scientists with some workflows?

### Can we measure problem solving ability?

- Extract details of the data.
- Identify and extract the processes and methodology.
- Write code to solve problem or reproduce study.


MIGRATION STUDIES VOLUME 3 • NUMBER 1 • 2015 • 89–110
Modeling internal migration flows in sub-Saharan Africa using census microdata
Andres J. Garcia, Deepa K. Pindolia, Kenneth K. Lopiano and Andrew J. Tatem
Abstract
1. Introduction
Can we reproduce the analysis in this study?
Can I apply the same techniques for Europe?

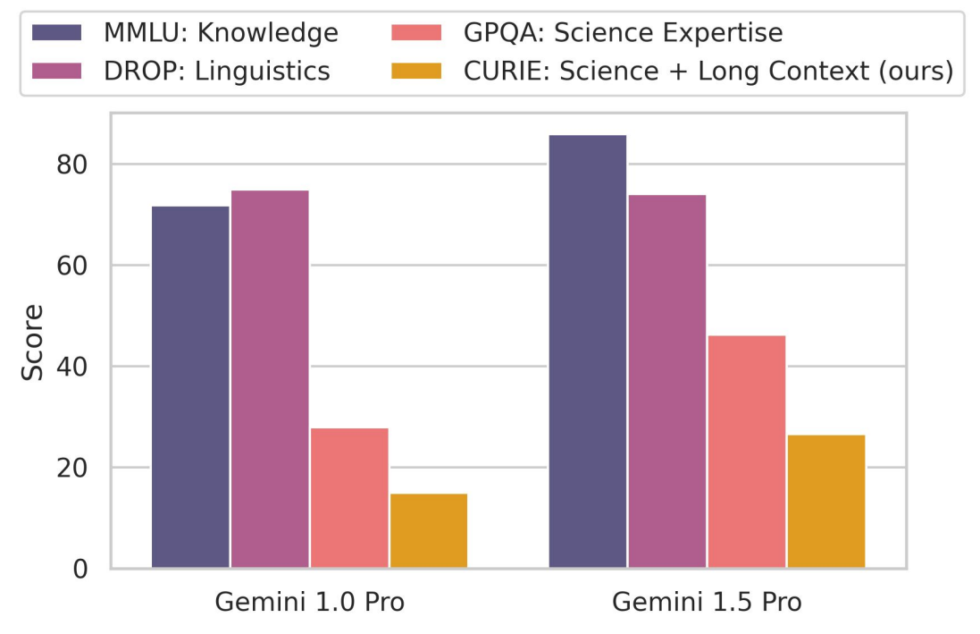
```
"dataset_name": "Integrated Public Use Microdata Series (IPUMS) International",
"dataset_website_or_source": "Minnesota Population Center. (2010) Integrated Public U
"variables": [
  {
    "variable_name": "POPI",
    "description": "POPI (Population of origin): This variable represents the populat
    "time_range": {
      "start_date": "2000-01-01",
      "end_date": "2002-12-31"
    },
    "spatial_range": "Ghana, Malawi, South Africa, Kenya, Mali, Uganda, Nigeria, Sene
  },
  {
    "variable_name": "POPJ",
    "description": "POPJ (Population of destination): This variable represents the po
ar to the population of origin, the population count at the destination is obtained fr
specific administrative unit at a given time.",
    "time_range": {
      "start_date": "2000-01-01",
      "end_date": "2002-12-31"
```

### This requires

- Knowledge of the domain.
- Processing long-context info.
- Reasoning ability to apply knowledge in the context of a problem.

### CURIE: Tests scientific problem solving

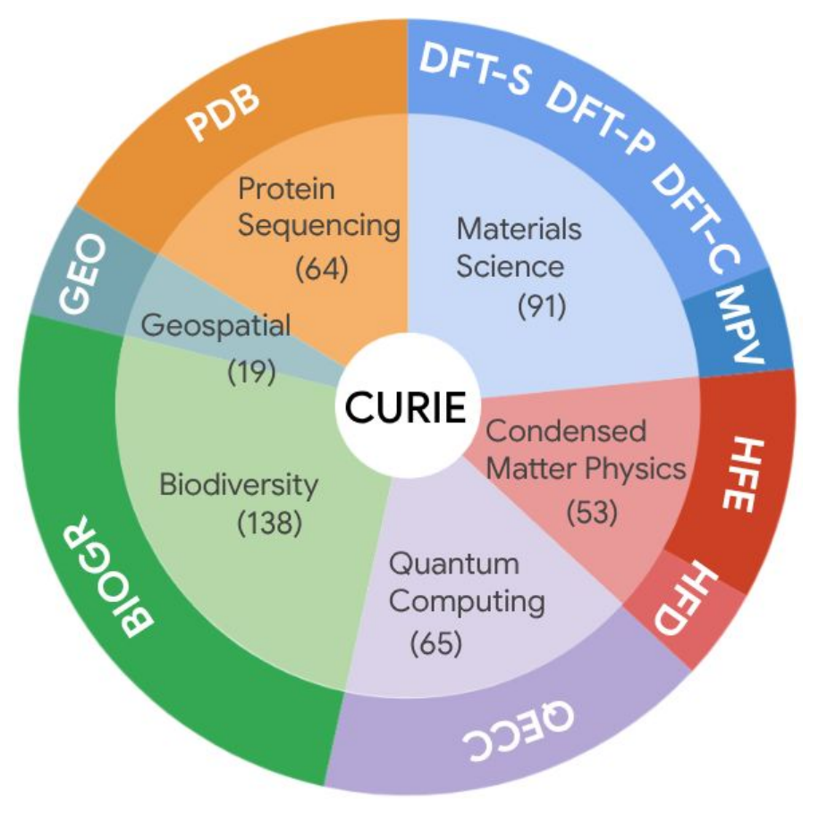
- Existing benchmarks test for knowledge, linguistics.
- CURIE: scientific long-**C**ontext **U**nderstanding **R**easoning and **I**nformation **E**xtraction benchmark

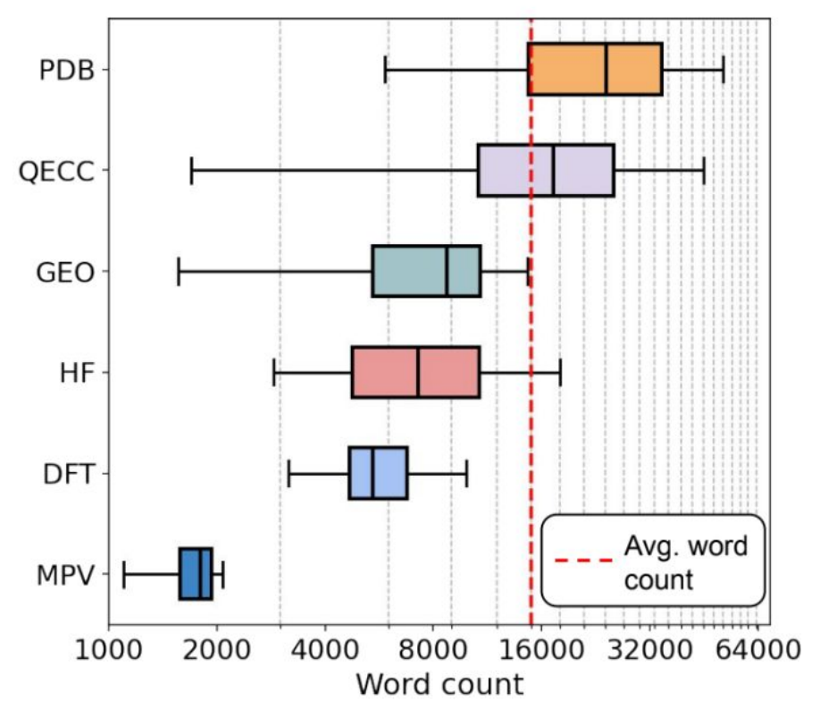


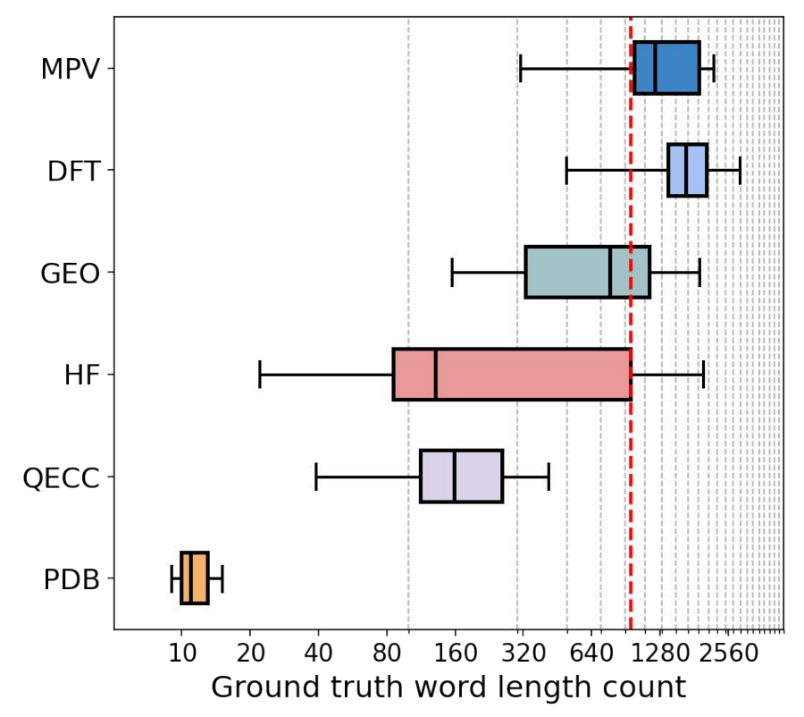
## The CURIE benchmark and dataset

### 580 examples, 429 documents, 10 tasks, 6 domains

- Tasks require expertise + reasoning + long-context understanding
- Avg. input query length ~15k words, gold response length is ~1k



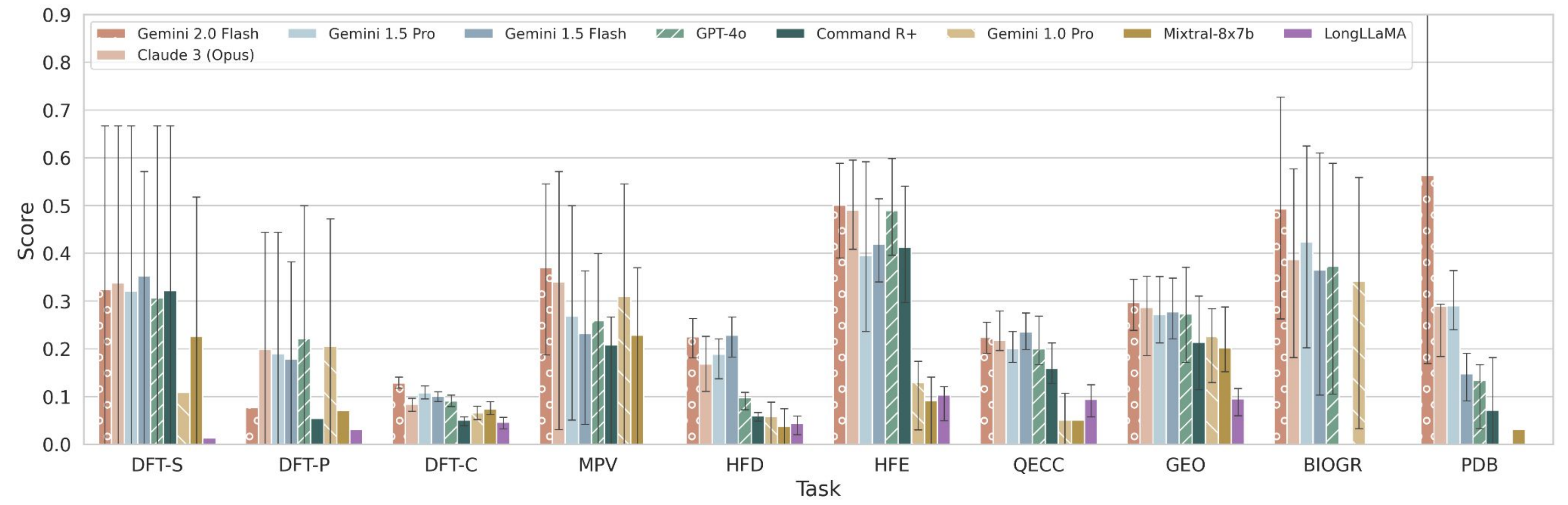




We collaborated closely with domain experts throughout benchmark development. This included:

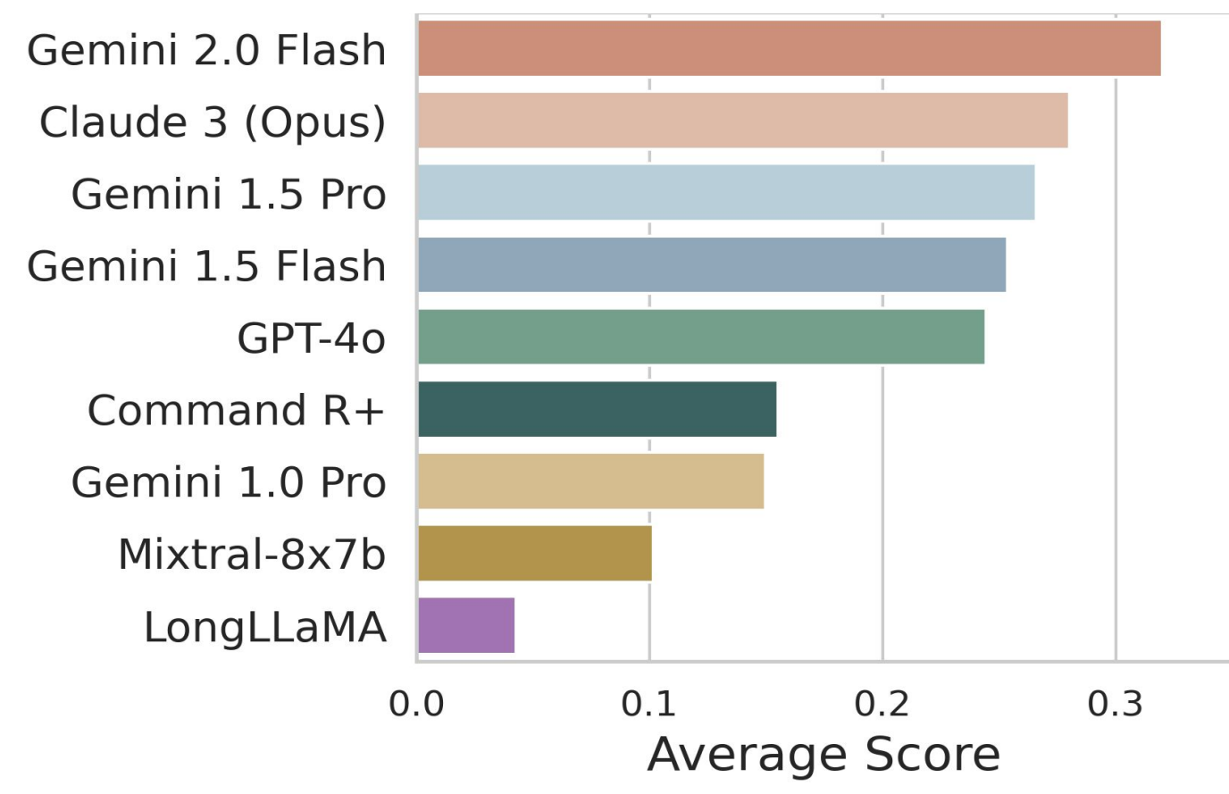
- Defining and identifying tasks reflecting realistic scientific workflows.
- Sourcing relevant papers from the domain.
- Creating accurate, nuanced, and comprehensive ground truth answers.
- Rating task difficulty based on salient features.
- Identifying and verifying evaluation metrics against expert judgments of model responses.

## Results



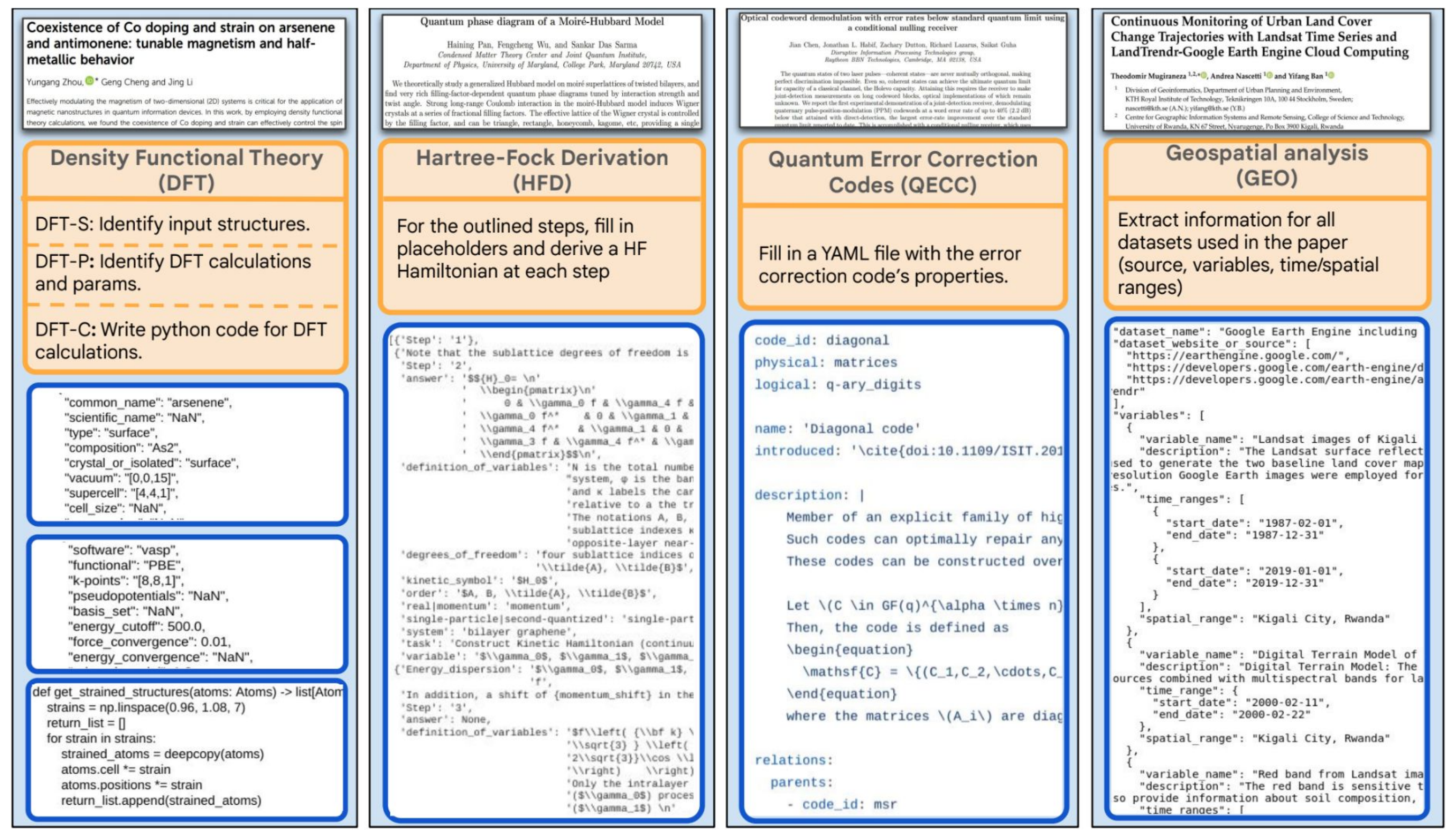
Per task normalized scores of various LLMs on the CURIE benchmark to measure performance on 10 long-context tasks requiring expertise across six scientific disciplines. Higher is better.

- Highest at 32%
- Gemini Flash 2 and Claude-3 do better. They understand the purpose of the extraction tasks and group responses.
- Exhaustive retrieval e.g. DFT-S, MPV, GEO are challenging for all.
- Flash 2 generated code to solve PDB ~50% of the time and was correct. When enumerating sequence it was similar to other models.
- Experts noted that summaries generated by models were succinct while including a multitude of details hard to comb out e.g. in QECC, and easy to remove afterwards.



Average normalized performance of long-context LLMs across the 10 tasks from six scientific domains in CURIE.

## Illustrative examples of tasks in CURIE


Coexistence of Co doping and strain on arsenene and antimonene: tunable magnetism and half-metallic behavior
Quantum phase diagram of a Moiré-Hubbard Model
Optical codeword demodulation with error rates below standard quantum limit using a conditional nulling receiver
Continuous Monitoring of Urban Land Cover Change Trajectories with Landsat Time Series and LandTrendr-Google Earth Engine Cloud Computing

**Density Functional Theory (DFT)**

DFT-S: Identify input structures.

DFT-P: Identify DFT calculations and params.

DFT-C: Write python code for DFT calculations.

```
"common_name": "arsenene",
"scientific_name": "NaN",
"type": "surface",
"composition": "As2",
"crystal_or_isolated": "surface",
"vacuum": "[0,0,15]",
"supercell": "[4,4,1]",
"cell_size": "NaN",
```

```
"software": "vasp",
"functional": "PBE",
"k-points": "[8,8,1]",
"pseudopotentials": "NaN",
"basis_set": "NaN",
"energy_cutoff": 500.0,
"force_convergence": 0.01,
"energy_convergence": "NaN",
```

```
def get_strained_structures(atoms: Atoms) -> list[Atom
  strains = np.linspace(0.96, 1.08, 7)
  return_list = []
  for strain in strains:
    strained_atoms = deepcopy(atoms)
    atoms.cell *= strain
    atoms.positions *= strain
    return_list.append(strained_atoms)
```

**Hartree-Fock Derivation (HFD)**

For the outlined steps, fill in placeholders and derive a HF Hamiltonian at each step

```
[{'Step': '1'},
 {'Note that the sublattice degrees of freedom is
  'Step': '2',
  'answer': '$${H}_0= \n'
       '\\begin{pmatrix}\n'
       '    0 & \\gamma_0 f & \\gamma_4 f &
       '\\gamma_0 f^*    & 0 & \\gamma_1 &
       '\\gamma_4 f^*    & \\gamma_1 & 0 &
       '\\gamma_3 f & \\gamma_4 f^* & \\gam
       '\\end{pmatrix}$$\n',
  'definition_of_variables': 'N is the total numbe
                             'system, φ is the ban
                             'and κ labels the car
                             'relative to a the tr
                             'The notations A, B,
                             'sublattice indexes κ
                             'opposite-layer near-
  'degrees_of_freedom': 'four sublattice indices c
                        '\\tilde{A}, \\tilde{B}$',
  'kinetic_symbol': '$H_0$',
  'order': '$A, B, \\tilde{A}, \\tilde{B}$',
  'real|momentum': 'momentum',
  'single-particle|second-quantized': 'single-part
  'system': 'bilayer graphene',
  'task': 'Construct Kinetic Hamiltonian (continuu
  'variable': '$\\gamma_0$, $\\gamma_1$, $\\gamma_
 {'Energy_dispersion': '$\\gamma_0$, $\\gamma_1$,
                       'f',
  'In addition, a shift of {momentum_shift} in the
  'Step': '3',
  'answer': None,
  'definition_of_variables': '$f\\left( {\\bf k} \
                             '\\sqrt{3} } \\left(
                             '2\\sqrt{3}}\\cos \\l
                             '\\right)    \\right)
                             'Only the intralayer
                             '($\\gamma_0$) proces
                             '($\\gamma_1$) \n'
```

**Quantum Error Correction Codes (QECC)**

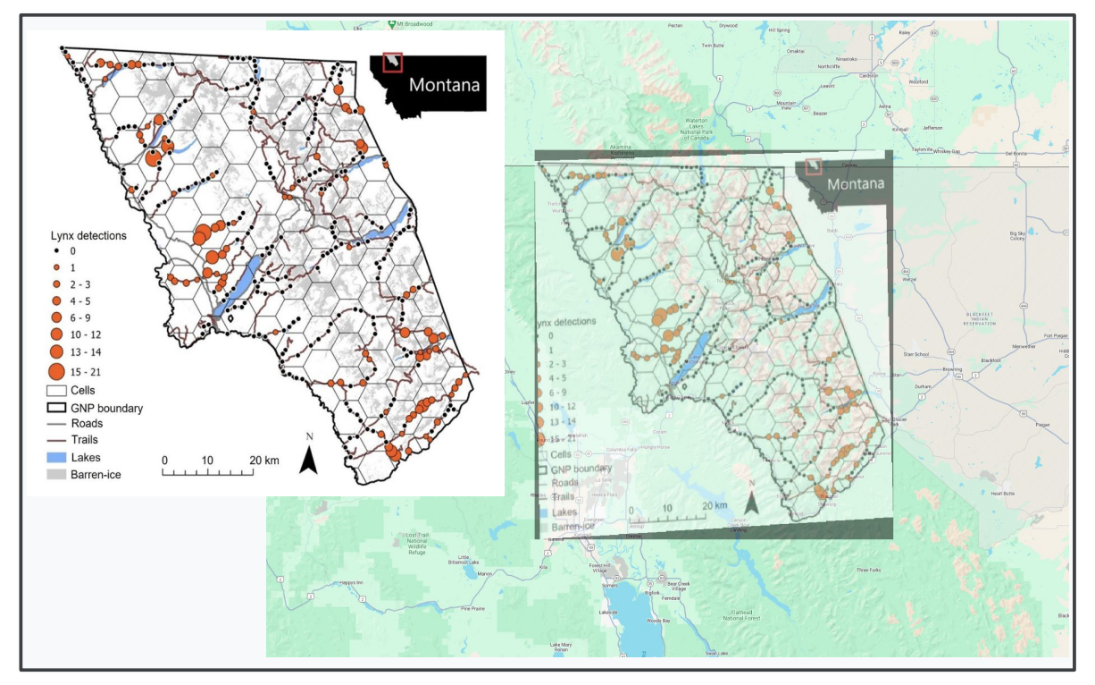
Fill in a YAML file with the error correction code's properties.

```
code_id: diagonal
physical: matrices
logical: q-ary_digits

name: 'Diagonal code'
introduced: '\cite{doi:10.1109/ISIT.201

description: |
  Member of an explicit family of hig
  Such codes can optimally repair any
  These codes can be constructed over

  Let \(C \in GF(q)^{\alpha \times n}
  Then, the code is defined as
  \begin{equation}
    \mathsf{C} = \{(C_1,C_2,\cdots,C_
  \end{equation}
  where the matrices \(A_i\) are diag

relations:
  parents:
    - code_id: msr
```

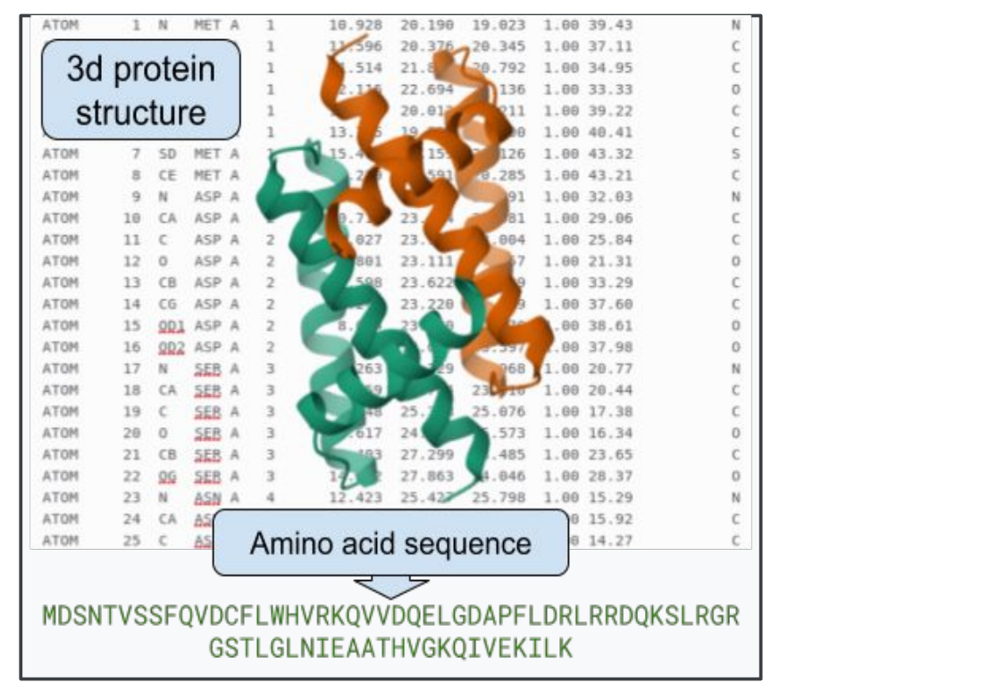
**Geospatial analysis (GEO)**

Extract information for all datasets used in the paper (source, variables, time/spatial ranges)

```
"dataset_name": "Google Earth Engine including
"dataset_website_or_source": [
  "https://earthengine.google.com/",
  "https://developers.google.com/earth-engine/d
  "https://developers.google.com/earth-engine/a
endr"
],
"variables": [
  {
    "variable_name": "Landsat images of Kigali
    "description": "The Landsat surface reflect
sed to generate the two baseline land cover map
esolution Google Earth images were employed for
s.",
    "time_ranges": [
      {
        "start_date": "1987-02-01",
        "end_date": "1987-12-31"
      },
      {
        "start_date": "2019-01-01",
        "end_date": "2019-12-31"
      }
    ],
    "spatial_range": "Kigali City, Rwanda"
  },
  {
    "variable_name": "Digital Terrain Model of
    "description": "Digital Terrain Model: The
ources combined with multispectral bands for la
    "time_range": {
      "start_date": "2000-02-11",
      "end_date": "2000-02-22"
    },
    "spatial_range": "Kigali City, Rwanda"
  },
  {
    "variable_name": "Red band from Landsat ima
    "description": "The red band is sensitive t
so provide information about soil composition,
    "time_ranges": [
```



**BIOGR** task requires identifying latitude and longitude (georeferencing) of a map image.



**PDB** task requires reconstructing a protein's amino acid sequence from the 3D structure.

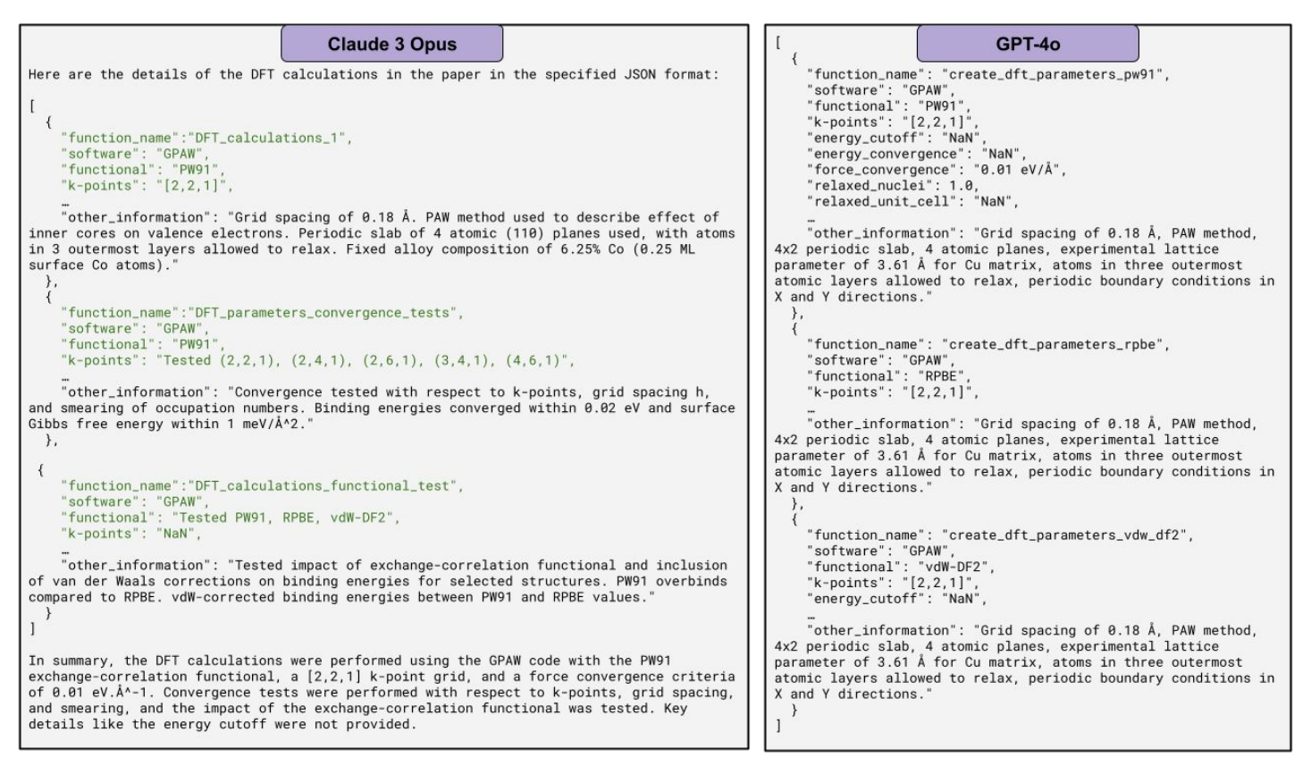
## Brief description of tasks and capabilities measured

| Task | Domain | # Qs | Brief Description | Capability | Output Format | Primary Eval. Metric |
|---|---|---|---|---|---|---|
| DFT-S | Material Science | 74 | Extracts input material structures for DFT calculations. | entity recognition, concept tracking | JSON | LLMSim-F1 |
| DFT-P | Material Science | 74 | Extract parameters for DFT calculations. | concept extraction, tracking, aggregation | JSON | LLMSim-F1 |
| DFT-C | Material Science | 74 | Write functional code for DFT computations. | concept aggregation, coding | TEXT | ROUGE-L |
| MPV | Material Science | 17 | Identify all instances of materials, their properties, and descriptors. | entity recognition, concept extraction, tracking | JSON | LLMSim-F1 |
| HFD | Condensed Matter Physics | 64 | Derive the Hartree-Fock mean-field Hamiltonian for a quantum many-body system. | concept extraction, algebraic manipulation, reasoning | TEXT | ROUGE-L |
| HFE | Condensed Matter Physics | 19 | Extract the most general mean-field Hamiltonian. | concept extraction | TEXT (latex equation) | ROUGE-L |
| QECC | Quantum Computing | 65 | Create a YAML file with the Error Correction Code's properties. | concept aggregation, summarization | YAML | ROUGE-L |
| GEO | Geospecial | 15 | Extract information for all geospatial datasets used along with the spatial and temporal extents. | concept extraction, aggregation | JSON | ROUGE-L |
| BIOGR | Biodiversity | 38 | Determine the latitude, longitude bounding box encompassing the region in the map image. | visual comprehension, reasoning | JSON (lat. lon. co-ordinates) | Intersection-over-Union (IoU) |
| PDB | Protein Sequencing | 138 | Reconstruct a protein's amino acid sequence form the 3D structure. | tracking, aggregation reasoning | Code or TEXT (seq.) | Identity ratio (IDr) |

Tasks in CURIE are varied and have ground truth annotations in mixed and heterogeneous form e.g. as JSONs, latex equations, YAML files, or free-form text. We use programmatic metrics for a predominant number of tasks and propose LLM- based evaluation for others.

## Examples

### DFT-P



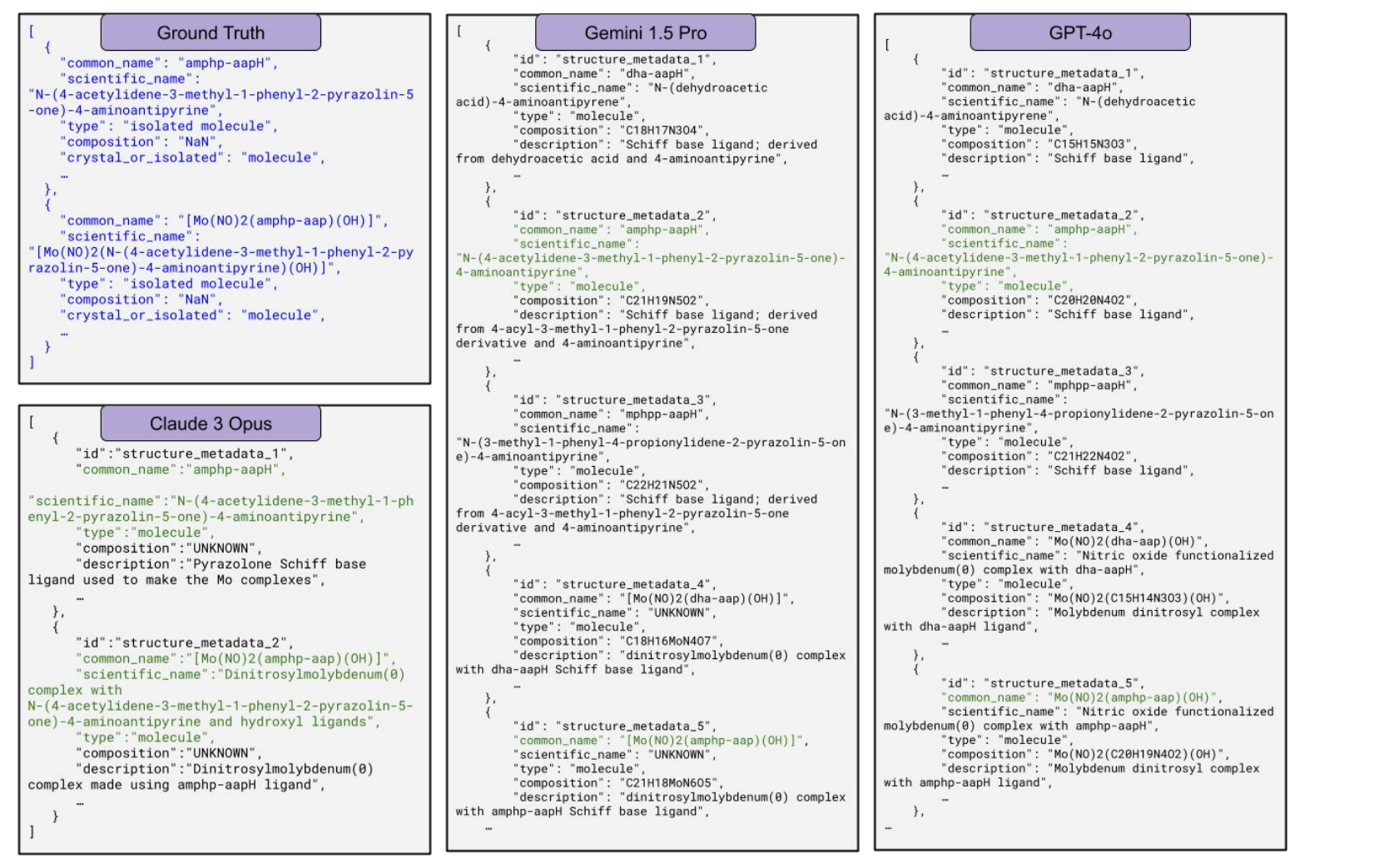
Claude 3 Opus

```
Here are the details of the DFT calculations in the paper in the specified JSON format:

[
  {
    "function_name":"DFT_calculations_1",
    "software": "GPAW",
    "functional": "PW91",
    "k-points": "[2,2,1]",
    …
    "other_information": "Grid spacing of 0.18 Å. PAW method used to describe effect of
inner cores on valence electrons. Periodic slab of 4 atomic (110) planes used, with atoms
in 3 outermost layers allowed to relax. Fixed alloy composition of 6.25% Co (0.25 ML
surface Co atoms)."
  },
  {
    "function_name":"DFT_parameters_convergence_tests",
    "software": "GPAW",
    "functional": "PW91",
    "k-points": "Tested (2,2,1), (2,4,1), (2,6,1), (3,4,1), (4,6,1)",
    …
    "other_information": "Convergence tested with respect to k-points, grid spacing h,
and smearing of occupation numbers. Binding energies converged within 0.02 eV and surface
Gibbs free energy within 1 meV/Å^2."
  },

  {
    "function_name":"DFT_calculations_functional_test",
    "software": "GPAW",
    "functional": "Tested PW91, RPBE, vdW-DF2",
    "k-points": "NaN",
    …
    "other_information": "Tested impact of exchange-correlation functional and inclusion
of van der Waals corrections on binding energies for selected structures. PW91 overbinds
compared to RPBE. vdW-corrected binding energies between PW91 and RPBE values."
  }
]

In summary, the DFT calculations were performed using the GPAW code with the PW91
exchange-correlation functional, a [2,2,1] k-point grid, and a force convergence criteria
of 0.01 eV.Å^-1. Convergence tests were performed with respect to k-points, grid spacing,
and smearing, and the impact of the exchange-correlation functional was tested. Key
details like the energy cutoff were not provided.
```

GPT-4o

```
[
  {
    "function_name": "create_dft_parameters_pw91",
    "software": "GPAW",
    "functional": "PW91",
    "k-points": "[2,2,1]",
    "energy_cutoff": "NaN",
    "energy_convergence": "NaN",
    "force_convergence": "0.01 eV/Å",
    "relaxed_nuclei": 1.0,
    "relaxed_unit_cell": "NaN",
    …
    "other_information": "Grid spacing of 0.18 Å, PAW method,
4x2 periodic slab, 4 atomic planes, experimental lattice
parameter of 3.61 Å for Cu matrix, atoms in three outermost
atomic layers allowed to relax, periodic boundary conditions in
X and Y directions."
  },
  {
    "function_name": "create_dft_parameters_rpbe",
    "software": "GPAW",
    "functional": "RPBE",
    "k-points": "[2,2,1]",
    …
    "other_information": "Grid spacing of 0.18 Å, PAW method,
4x2 periodic slab, 4 atomic planes, experimental lattice
parameter of 3.61 Å for Cu matrix, atoms in three outermost
atomic layers allowed to relax, periodic boundary conditions in
X and Y directions."
  },
  {
    "function_name": "create_dft_parameters_vdw_df2",
    "software": "GPAW",
    "functional": "vdW-DF2",
    "k-points": "[2,2,1]",
    "energy_cutoff": "NaN",
    …
    "other_information": "Grid spacing of 0.18 Å, PAW method,
4x2 periodic slab, 4 atomic planes, experimental lattice
parameter of 3.61 Å for Cu matrix, atoms in three outermost
atomic layers allowed to relax, periodic boundary conditions in
X and Y directions."
  }
]
```

### DFT-S



Ground Truth

```
[
  {
    "common_name": "amphp-aapH",
    "scientific_name":
"N-(4-acetylidene-3-methyl-1-phenyl-2-pyrazolin-5
-one)-4-aminoantipyrine",
    "type": "isolated molecule",
    "composition": "NaN",
    "crystal_or_isolated": "molecule",
    …
  },
  {
    "common_name": "[Mo(NO)2(amphp-aap)(OH)]",
    "scientific_name":
"[Mo(NO)2(N-(4-acetylidene-3-methyl-1-phenyl-2-py
razolin-5-one)-4-aminoantipyrine)(OH)]",
    "type": "isolated molecule",
    "composition": "NaN",
    "crystal_or_isolated": "molecule",
    …
  }
]
```

Claude 3 Opus

```
[
  {
    "id":"structure_metadata_1",
    "common_name":"amphp-aapH",

"scientific_name":"N-(4-acetylidene-3-methyl-1-ph
enyl-2-pyrazolin-5-one)-4-aminoantipyrine",
    "type":"molecule",
    "composition":"UNKNOWN",
    "description":"Pyrazolone Schiff base
ligand used to make the Mo complexes",
    …
  },
  {
    "id":"structure_metadata_2",
    "common_name":"[Mo(NO)2(amphp-aap)(OH)]",
    "scientific_name":"Dinitrosylmolybdenum(0)
complex with
N-(4-acetylidene-3-methyl-1-phenyl-2-pyrazolin-5-
one)-4-aminoantipyrine and hydroxyl ligands",
    "type":"molecule",
    "composition":"UNKNOWN",
    "description":"Dinitrosylmolybdenum(0)
complex made using amphp-aapH ligand",
    …
  }
]
```

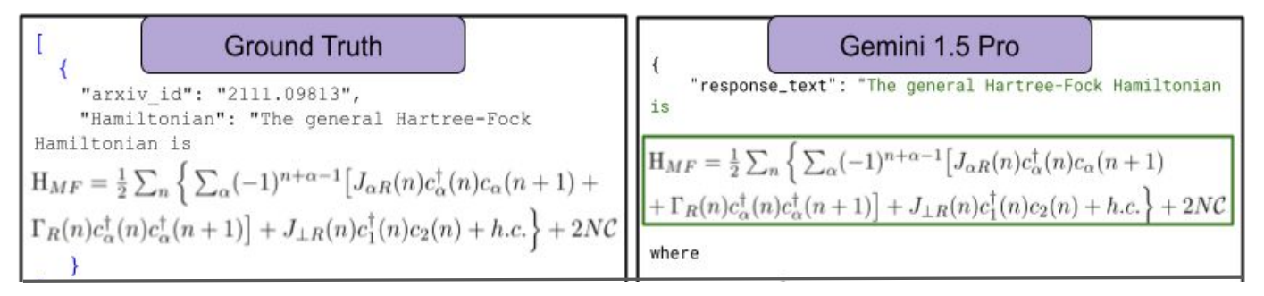
Gemini 1.5 Pro

```
[
  {
    "id": "structure_metadata_1",
    "common_name": "dha-aapH",
    "scientific_name": "N-(dehydroacetic
acid)-4-aminoantipyrene",
    "type": "molecule",
    "composition": "C18H17N3O4",
    "description": "Schiff base ligand; derived
from dehydroacetic acid and 4-aminoantipyrine",
    …
  },
  {
    "id": "structure_metadata_2",
    "common_name": "amphp-aapH",
    "scientific_name":
"N-(4-acetylidene-3-methyl-1-phenyl-2-pyrazolin-5-one)-
4-aminoantipyrine",
    "type": "molecule",
    "composition": "C21H19N5O2",
    "description": "Schiff base ligand; derived
from 4-acyl-3-methyl-1-phenyl-2-pyrazolin-5-one
derivative and 4-aminoantipyrine",
    …
  },
  {
    "id": "structure_metadata_3",
    "common_name": "mphpp-aapH",
    "scientific_name":
"N-(3-methyl-1-phenyl-4-propionylidene-2-pyrazolin-5-on
e)-4-aminoantipyrine",
    "type": "molecule",
    "composition": "C22H21N5O2",
    "description": "Schiff base ligand; derived
from 4-acyl-3-methyl-1-phenyl-2-pyrazolin-5-one
derivative and 4-aminoantipyrine",
    …
  },
  {
    "id": "structure_metadata_4",
    "common_name": "[Mo(NO)2(dha-aap)(OH)]",
    "scientific_name": "UNKNOWN",
    "type": "molecule",
    "composition": "C18H16MoN4O7",
    "description": "dinitrosylmolybdenum(0) complex
with dha-aapH Schiff base ligand",
    …
  },
  {
    "id": "structure_metadata_5",
    "common_name": "[Mo(NO)2(amphp-aap)(OH)]",
    "scientific_name": "UNKNOWN",
    "type": "molecule",
    "composition": "C21H18MoN6O5",
    "description": "dinitrosylmolybdenum(0) complex
with amphp-aapH Schiff base ligand",
    …
```

GPT-4o

```
[
  {
    "id": "structure_metadata_1",
    "common_name": "dha-aapH",
    "scientific_name": "N-(dehydroacetic
acid)-4-aminoantipyrene",
    "type": "molecule",
    "composition": "C15H15N3O3",
    "description": "Schiff base ligand",
    …
  },
  {
    "id": "structure_metadata_2",
    "common_name": "amphp-aapH",
    "scientific_name":
"N-(4-acetylidene-3-methyl-1-phenyl-2-pyrazolin-5-one)-
4-aminoantipyrine",
    "type": "molecule",
    "composition": "C20H20N4O2",
    "description": "Schiff base ligand",
    …
  },
  {
    "id": "structure_metadata_3",
    "common_name": "mphpp-aapH",
    "scientific_name":
"N-(3-methyl-1-phenyl-4-propionylidene-2-pyrazolin-5-on
e)-4-aminoantipyrine",
    "type": "molecule",
    "composition": "C21H22N4O2",
    "description": "Schiff base ligand",
    …
  },
  {
    "id": "structure_metadata_4",
    "common_name": "Mo(NO)2(dha-aap)(OH)",
    "scientific_name": "Nitric oxide functionalized
molybdenum(0) complex with dha-aapH",
    "type": "molecule",
    "composition": "Mo(NO)2(C15H14N3O3)(OH)",
    "description": "Molybdenum dinitrosyl complex
with dha-aapH ligand",
    …
  },
  {
    "id": "structure_metadata_5",
    "common_name": "Mo(NO)2(amphp-aap)(OH)",
    "scientific_name": "Nitric oxide functionalized
molybdenum(0) complex with amphp-aapH",
    "type": "molecule",
    "composition": "Mo(NO)2(C20H19N4O2)(OH)",
    "description": "Molybdenum dinitrosyl complex
with amphp-aapH ligand",
    …
  },
…
```

### HFE



Ground Truth

```
[
  {
    "arxiv_id": "2111.09813",
    "Hamiltonian": "The general Hartree-Fock
Hamiltonian is
```

$$\mathrm{H}_{MF} = \frac{1}{2}\sum_n\left\{\sum_\alpha(-1)^{n+\alpha-1}[J_{\alpha R}(n)c^\dagger_\alpha(n)c_\alpha(n+1) + \Gamma_R(n)c^\dagger_\alpha(n)c^\dagger_\alpha(n+1)] + J_{\perp R}(n)c^\dagger_1(n)c_2(n) + h.c.\right\} + 2NC$$

```
  }
```

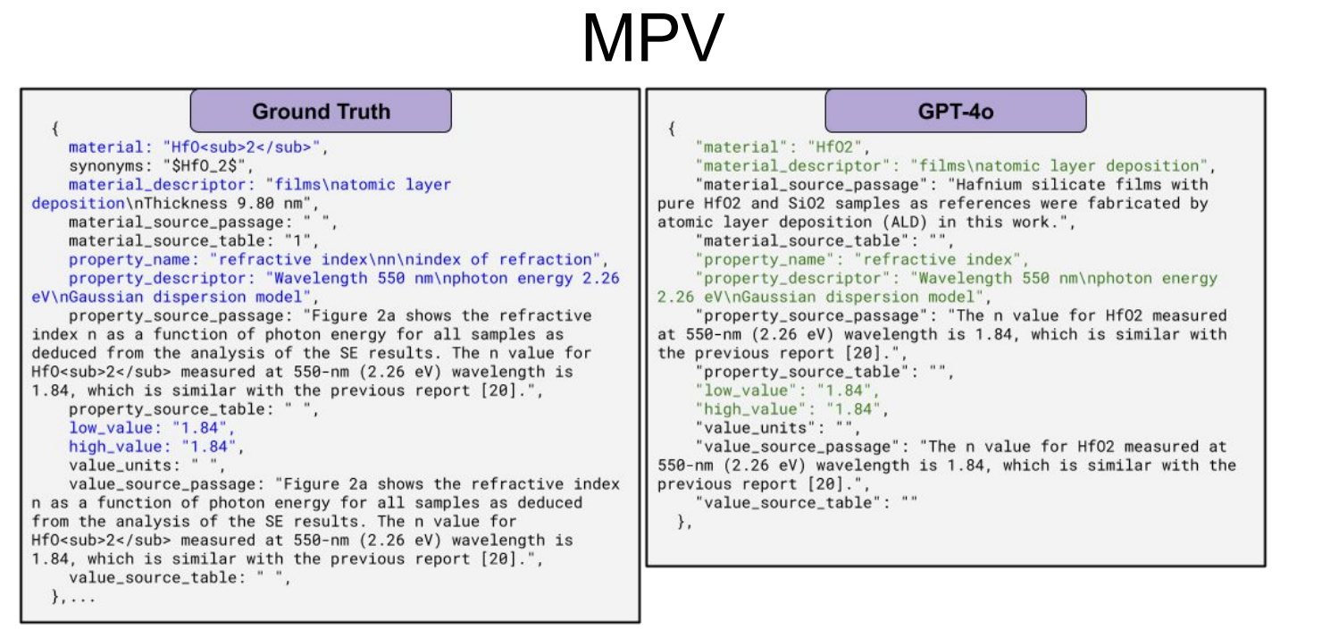
Gemini 1.5 Pro

```
{
  "response_text": "The general Hartree-Fock Hamiltonian
is
```

$$\mathrm{H}_{MF} = \frac{1}{2}\sum_n\left\{\sum_\alpha(-1)^{n+\alpha-1}[J_{\alpha R}(n)c^\dagger_\alpha(n)c_\alpha(n+1) + \Gamma_R(n)c^\dagger_\alpha(n)c^\dagger_\alpha(n+1)] + J_{\perp R}(n)c^\dagger_1(n)c_2(n) + h.c.\right\} + 2NC$$

```
where
```

### MPV



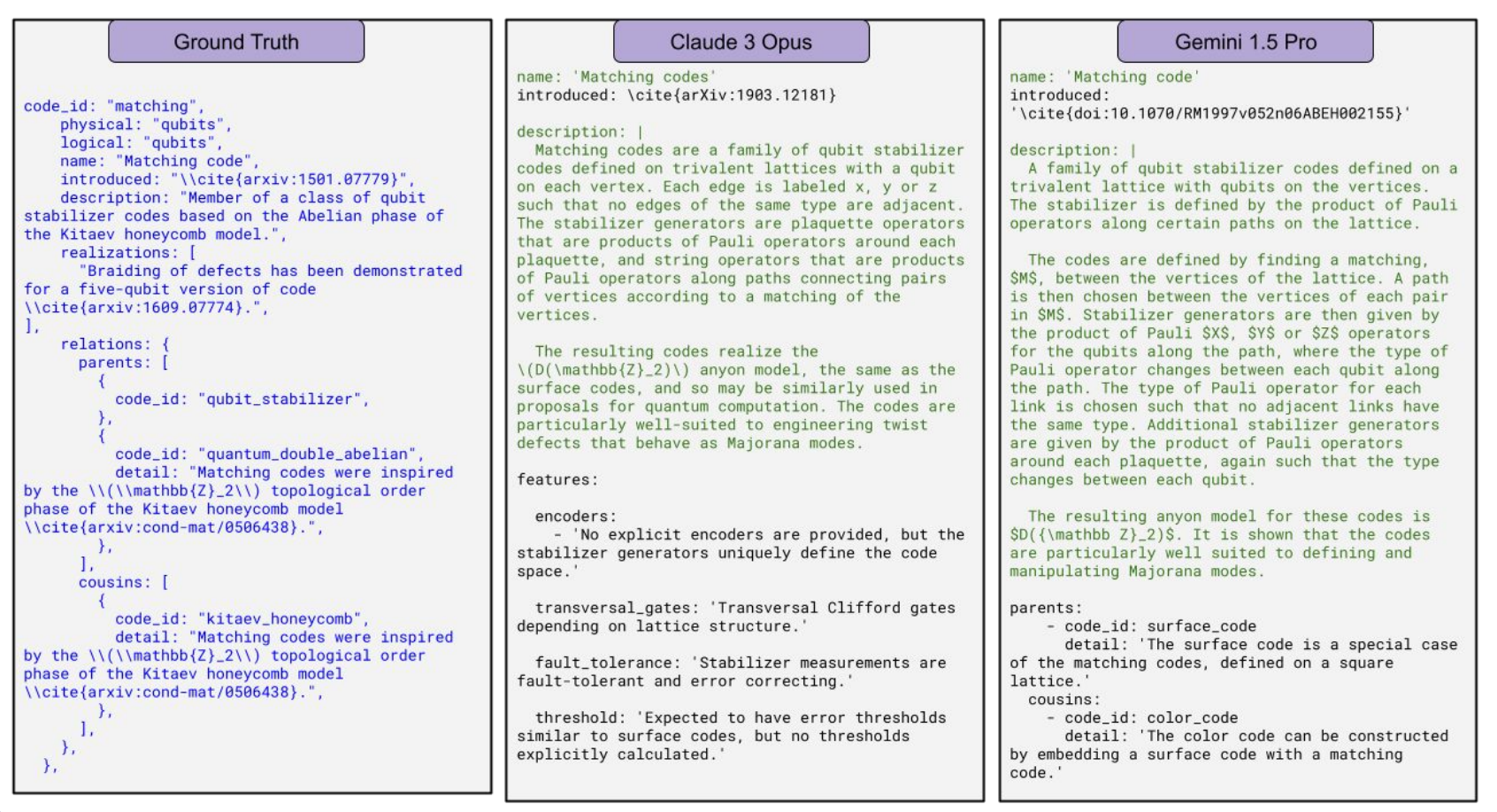
Ground Truth

```
{
  material: "HfO<sub>2</sub>",
  synonyms: "$HfO_2$",
  material_descriptor: "films\natomic layer
deposition\nThickness 9.80 nm",
  material_source_passage: " ",
  material_source_table: "1",
  property_name: "refractive index\nn\nindex of refraction",
  property_descriptor: "Wavelength 550 nm\nphoton energy 2.26
eV\nGaussian dispersion model",
  property_source_passage: "Figure 2a shows the refractive
index n as a function of photon energy for all samples as
deduced from the analysis of the SE results. The n value for
HfO<sub>2</sub> measured at 550-nm (2.26 eV) wavelength is
1.84, which is similar with the previous report [20].",
  property_source_table: " ",
  low_value: "1.84",
  high_value: "1.84",
  value_units: " ",
  value_source_passage: "Figure 2a shows the refractive index
n as a function of photon energy for all samples as deduced
from the analysis of the SE results. The n value for
HfO<sub>2</sub> measured at 550-nm (2.26 eV) wavelength is
1.84, which is similar with the previous report [20].",
  value_source_table: " ",
},...
```

GPT-4o

```
{
  "material": "HfO2",
  "material_descriptor": "films\natomic layer deposition",
  "material_source_passage": "Hafnium silicate films with
pure HfO2 and SiO2 samples as references were fabricated by
atomic layer deposition (ALD) in this work.",
  "material_source_table": "",
  "property_name": "refractive index",
  "property_descriptor": "Wavelength 550 nm\nphoton energy
2.26 eV\nGaussian dispersion model",
  "property_source_passage": "The n value for HfO2 measured
at 550-nm (2.26 eV) wavelength is 1.84, which is similar with
the previous report [20].",
  "property_source_table": "",
  "low_value": "1.84",
  "high_value": "1.84",
  "value_units": "",
  "value_source_passage": "The n value for HfO2 measured at
550-nm (2.26 eV) wavelength is 1.84, which is similar with the
previous report [20].",
  "value_source_table": ""
},
```

### QECC



Ground Truth

```
code_id: "matching",
  physical: "qubits",
  logical: "qubits",
  name: "Matching code",
  introduced: "\\cite{arxiv:1501.07779}",
  description: "Member of a class of qubit
stabilizer codes based on the Abelian phase of
the Kitaev honeycomb model.",
  realizations: [
    "Braiding of defects has been demonstrated
for a five-qubit version of code
\\cite{arxiv:1609.07774}.",
  ],
  relations: {
    parents: [
      {
        code_id: "qubit_stabilizer",
      },
      {
        code_id: "quantum_double_abelian",
        detail: "Matching codes were inspired
by the \\(\\mathbb{Z}_2\\) topological order
phase of the Kitaev honeycomb model
\\cite{arxiv:cond-mat/0506438}.",
      },
    ],
    cousins: [
      {
        code_id: "kitaev_honeycomb",
        detail: "Matching codes were inspired
by the \\(\\mathbb{Z}_2\\) topological order
phase of the Kitaev honeycomb model
\\cite{arxiv:cond-mat/0506438}.",
      },
    ],
  },
},
```

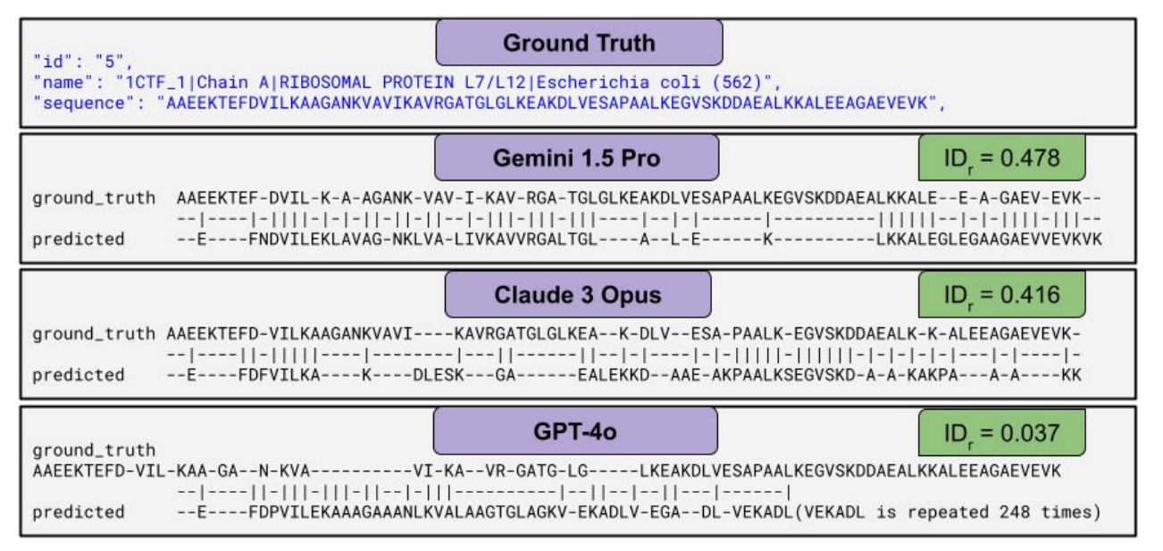
Claude 3 Opus

```
name: 'Matching codes'
introduced: \cite{arXiv:1903.12181}

description: |
  Matching codes are a family of qubit stabilizer
codes defined on trivalent lattices with a qubit
on each vertex. Each edge is labeled x, y or z
such that no edges of the same type are adjacent.
The stabilizer generators are plaquette operators
that are products of Pauli operators around each
plaquette, and string operators that are products
of Pauli operators along paths connecting pairs
of vertices according to a matching of the
vertices.

  The resulting codes realize the
\(D(\mathbb{Z}_2)\) anyon model, the same as the
surface codes, and so may be similarly used in
proposals for quantum computation. The codes are
particularly well-suited to engineering twist
defects that behave as Majorana modes.

features:

  encoders:
    - 'No explicit encoders are provided, but the
stabilizer generators uniquely define the code
space.'

  transversal_gates: 'Transversal Clifford gates
depending on lattice structure.'

  fault_tolerance: 'Stabilizer measurements are
fault-tolerant and error correcting.'

  threshold: 'Expected to have error thresholds
similar to surface codes, but no thresholds
explicitly calculated.'
```

Gemini 1.5 Pro

```
name: 'Matching code'
introduced:
'\cite{doi:10.1070/RM1997v052n06ABEH002155}'

description: |
  A family of qubit stabilizer codes defined on a
trivalent lattice with qubits on the vertices.
The stabilizer is defined by the product of Pauli
operators along certain paths on the lattice.

  The codes are defined by finding a matching,
$M$, between the vertices of the lattice. A path
is then chosen between the vertices of each pair
in $M$. Stabilizer generators are then given by
the product of Pauli $X$, $Y$ or $Z$ operators
for the qubits along the path, where the type of
Pauli operator changes between each qubit along
the path. The type of Pauli operator for each
link is chosen such that no adjacent links have
the same type. Additional stabilizer generators
are given by the product of Pauli operators
around each plaquette, again such that the type
changes between each qubit.

  The resulting anyon model for these codes is
$D({\mathbb Z}_2)$. It is shown that the codes
are particularly well suited to defining and
manipulating Majorana modes.

parents:
  - code_id: surface_code
    detail: 'The surface code is a special case
of the matching codes, defined on a square
lattice.'
  cousins:
  - code_id: color_code
    detail: 'The color code can be constructed
by embedding a surface code with a matching
code.'
```

### PDB



Ground Truth

```
"id": "5",
"name": "1CTF_1|Chain A|RIBOSOMAL PROTEIN L7/L12|Escherichia coli (562)",
"sequence": "AAEEKTEFDVILKAAGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEVK",
```

Gemini 1.5 Pro — $ID_r$ = 0.478

```
ground_truth  AAEEKTEF-DVIL-K-A-AGANK-VAV-I-KAV-RGA-TGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALE--E-A-GAEV-EVK--
              --|----|-||||-|-|-||-||-||--|-|||-|||-|||----|--|-|------|----------||||||--|-|-||||-|||--
predicted     --E----FNDVILEKLAVAG-NKLVA-LIVKAVVRGALTGL----A--L-E------K----------LKKALEGLEGAAGAEVVEVKVK
```

Claude 3 Opus — $ID_r$ = 0.416

```
ground_truth AAEEKTEFD-VILKAAGANKVAVI----KAVRGATGLGLKEA--K-DLV--ESA-PAALK-EGVSKDDAEALK-K-ALEEAGAEVEVK-
             --|----||-|||||----|--------|---||-------||--|-|----|-|-|||||-||||||-|-|-|-|---|-|----|-
predicted    --E----FDFVILKA----K----DLESK---GA------EALEKKD--AAE-AKPAALKSEGVSKD-A-A-KAKPA----A-A----KK
```

GPT-4o — $ID_r$ = 0.037

```
ground_truth
AAEEKTEFD-VIL-KAA-GA--N-KVA----------VI-KA--VR-GATG-LG-----LKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEVK
              --|----||-|||-|||-||--|-|||----------|--||--||-|---||--|------|
predicted     --E----FDPVILEKAAAGAAANLKVALAAGTGLAGKV-EKADLV-EGA--DL-VEKADL(VEKADL is repeated 248 times)
```