

Latest Trends in AI Research for EO

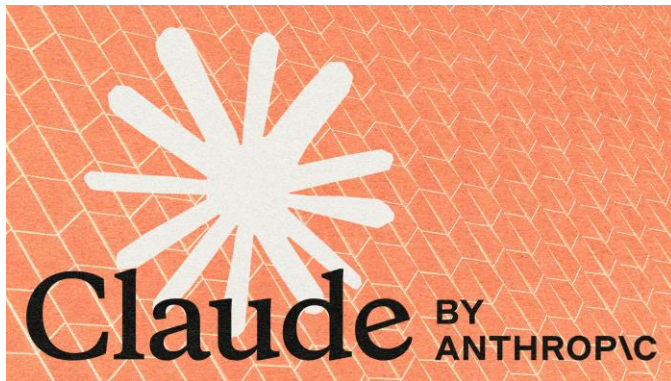
Prof. Dr. Begüm Demir, BIFOLD and TU Berlin



LLM Landscape



ChatGPT



LLM's Performance on Querying Satellite Data



S2 input



Reference map

- Arable land
- Broad-leaved forest
- Coniferous forest
- Mixed forest
- Inland Waters

GPT-5:

The image appears to be a satellite view of a green vegetated area with a river or water body meandering through the center. The bright, light-colored region represents the river or stream, while the surrounding dark to medium green areas indicate dense vegetation or forested land.

LLM's Performance on Querying Satellite Data



S2 input



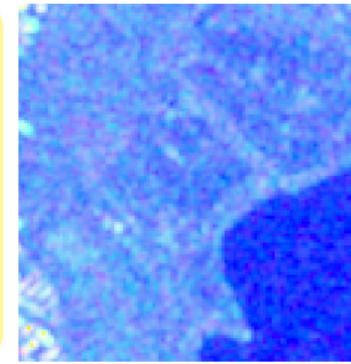
Reference map

- Arable land
- Broad-leaved forest
- Coniferous forest
- Mixed forest
- Inland Waters

Desired Description: *This satellite image, captured during the summer season in Austria, showcases a diverse landscape within the "temperate, no dry season, warm summer" climate zone. The dominant feature is the mixed forest, which covers approximately 905,000 square meters and is distributed across two large areas of 525,000 square meters and 380,000 square meters. The mixed forest shares borders with coniferous forest (298,000 sqm), which is distributed over four individual areas, including one larger area of 236,000 square meters and three marginal areas, inland waters (142,000 sqm), broad-leaved forest (72,000 sqm), and arable land (72,000 sqm). The inland waters are also adjacent to the broad-leaved forest. This complex interplay of forested areas, water bodies, and agricultural spaces highlights the varied ecosystem present in this region of Austria during the summer season.*



Input: Describe the content of the image, including the region, climate zone, and land cover distribution.
Output: This satellite image, captured during the summer season in Switzerland, showcases a diverse landscape within the "temperate, no dry season, warm summer" climate zone. The dominant features are arable land (~526,000 sqm) [...] are adjacent to both inland waters (~305,000 sqm) and urban fabric. Notably, the urban fabric is distributed over three individual marginal areas. The varied landscape presents a mix of agricultural areas, wetlands, water bodies, and artificial surfaces.



Input: How many areas covered by urban fabric can be seen? a) More than five, b) 3, c) 1, d) 0
Output: b

Input: Does any inland water border inland wetlands in this scene?
Output: Yes

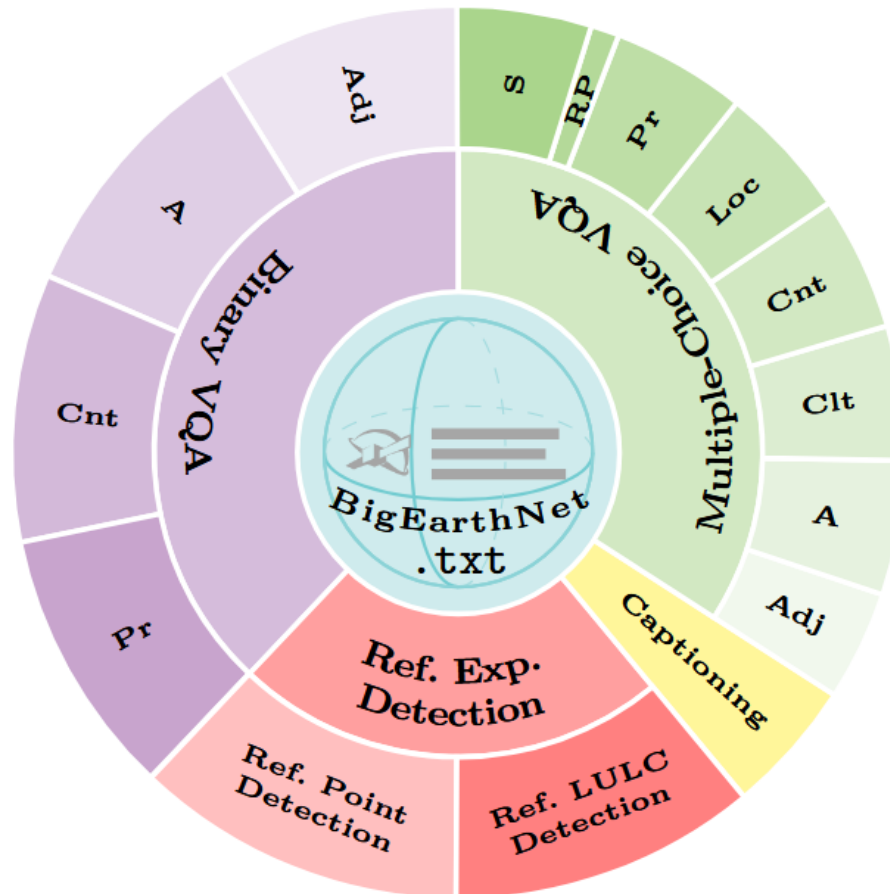
Input: Are there regions of coastal wetlands in the satellite image?
Output: No

Input: Which season is shown in the satellite image?
a) Spring, b) Summer, c) Winter, d) Autumn
Output: b

Input: What is the relative position of the arable land to the inland waters?
a) to the left, b) to the bottom, c) to the top-right, d) to the top
Output: a

Input: Where can the [largest patch of urban fabric](#) be found?
Output: [0.0 0.55, 0.2 1.0]

Input: Output a bounding box enclosing the land cover class instance positioned at [\(0.83, 0.06\)](#).
Output: [0.49 0.0, 1.0 0.2]





BigEarthNet: A Benchmark Archive for EO

✓ To support the studies on search and retrieval, we developed **BigEarthNet** that:

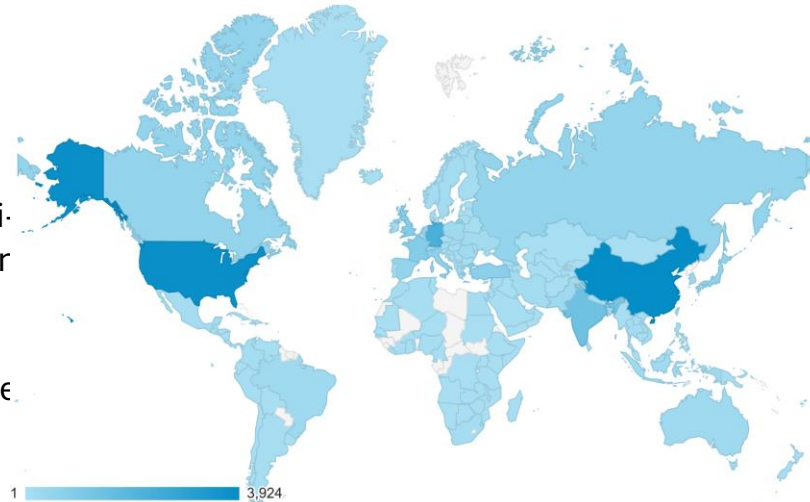
- consists of 549,488 Sentinel-1&2 images.
- advances studies for the analysis of large-scale EO data archives.



 G. Sumbul, et al., BigEarthNet-MM: A Large Scale Multi-Modal Multi-Label Benchmark Archive for Remote Sensing Image Classification and Retrieval. IEEE GRSM, 2021.

 K. Clasen, et al., " reBEN: Refined BigEarthNet Dataset for Remote Sensing Image Analysis", IEEE IGARSS, 2025.

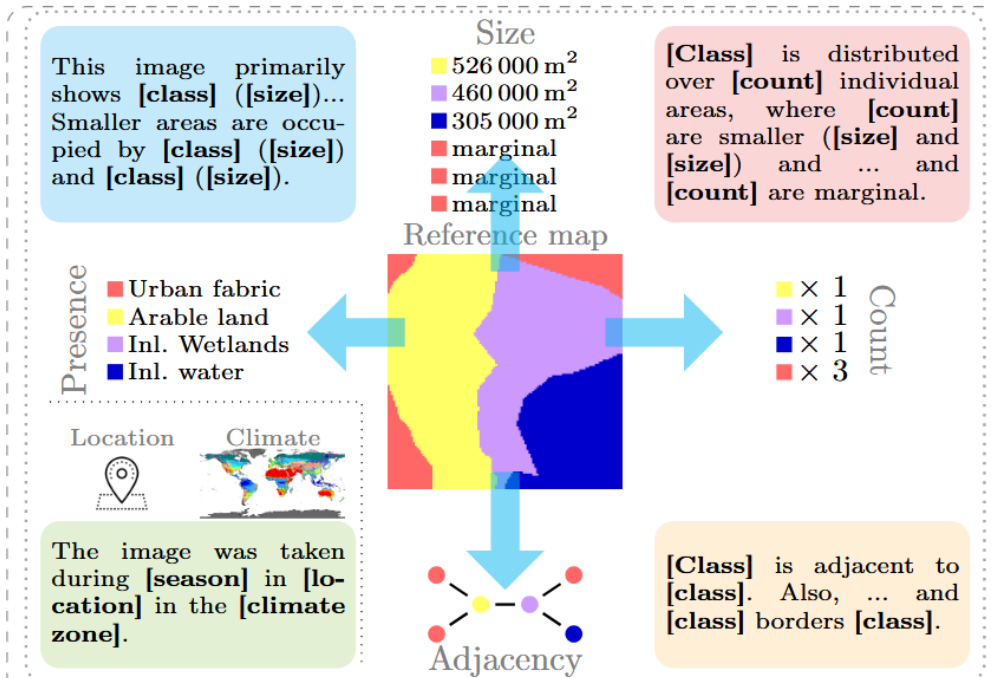
Impact:



Available also at:



Extract Attributes,
Fill & Concatenate Templates



This image primary ... ([size]). [Class] is distributed ... marginal.
 [Class] is adjacent ... [class]. The image was ... [climate zone].

Para-phrase
Refine

You are a Remote Sensing expert. Use the following definitions to increase variance in the output: ... Do not violate these rules: ...

Control your caption. Ensure that you followed the rules. Update the caption by ... Do NOT explain the mistakes.

Final
Caption

This satellite image, captured during the summer season in Switzerland, showcases a diverse landscape within the "temperate, no dry season, warm summer" climate zone. The dominant features are arable land (~526,000 sqm) and inland wetlands (~460,000 sqm), which are adjacent to each other. The arable land borders both inland wetlands and urban fabric (~149,000 sqm). Moreover, the inland wetlands are adjacent to both inland waters (~305,000 sqm) and urban fabric. Notably, the urban fabric is distributed over three individual marginal areas. The varied landscape presents a mix of agricultural areas, wetlands, water bodies, and artificial surfaces.

Results: BigEarthNet.txt

Model	Captioning	Binary VQA	MCQ	Ref. Exp. Detection
SOTA RS	1.66	58.38	35.26	16.18
SOTA CV	0.96	61.96	37.55	31.73
RS-InternVL	34.04	73.29	51.49	65.84

Results on the main tasks of the BigEarthNet.txt benchmark split for the fine-tuned adapted RS-InternVL model from InternVL3-1B as well as best results from VLMs in RS and CV, which are EarthMind and LLaVa for captioning, EarthDial and Qwen for binary VQA, EarthMind and Qwen for MCQ, and EarthMind and GPT for referring expression detection. Reported metric for captioning is BLEU-4, while that for binary VQA, and MCQ is accuracy, and that for referring expression detection is mIoU. All results are in percent (%).

Foundation Models for EO

https://github.com/Jack-bo1220/Awesome-Remote-Sensing-Foundation-Models

FedSense	Towards Privacy-preserved Pre-training of Remote Sensing Foundation Models		ICCV2025
	EarthGPT	EarthGPT: A Universal Multi-modal Large Language Model for Multi-sensor Image	Arxiv2024
RS-vHeat	SkyCLIP	DOFA-CLIP	DOFA-CLIP: Multimodal Vision-Language Foundation Models for Earth Observation Arxiv2025
Copernicus-FM		Falcon	Falcon: A Remote Sensing Vision-Language Foundation Model Arxiv2025
SelectiveMAE	GeoChat	LRS-VQA	When Large Vision-Language Model Meets Large Remote Sensing Imagery: Coarse-to-Fine Text-Guided Token Pruning ICCV2025
SMARTIES	LHRS-Bot	UrbanLLaVA	UrbanLLaVA: A Multi-modal Large Language Model for Urban Intelligence with Spatial Reasoning and Understanding ICCV2025
TerraMind	SkySenseGPT	OmniGeo	OmniGeo: Towards a Multimodal Large Language Models for Geospatial Artificial Intelligence Arxiv2025
SkySense V2		EagleVision	EagleVision: Object-level Attribute Multimodal LLM for Remote Sensing Arxiv2025
AlphaEarth	EarthMarker	SegEarth-R1	SegEarth-R1: Geospatial Pixel Reasoning via Large Language Model Arxiv2025
		RemoteSAM	RemoteSAM: Towards Segment Anything for Earth Observation ACMMM2025
SkySense++	GeoText	DynamicVL	DynamicVL: Benchmarking Multimodal Large Language Models for Dynamic City Understanding Arxiv2025
	Aquila	LISAt	LISAt: Language- Instructed Segmentation Assistant for Satellite Imagery Arxiv2025

The research on EO FMs has been mainly evolved along three directions:

- ✓ **Scaling up** model size and training data to increase representation capacity;
- ✓ Introducing **architectural innovations** that improve the representation accuracy;
- ✓ Integrating **multiple data modalities** to enhance multi-modal learning.

Challenge: Selecting Foundation Models for RS Tasks

https://github.com/Jack-bo1220/Awesome-Remote-Sensing-Foundation-Models

Towards Privacy-preserved Pre-training			
FedSense	EarthGPT	EarthGPT: A Universal Multi-modal Large Language Model for Multi-sensor Image Comprehension in Remote Sensing Domain	Arxiv2024
RS-vHeat	SkyCLIP	DOFA-CLIP	DOFA-CLIP: Multimodal Vision-Language Foundation Models for Earth Observation Arxiv2025
		Falcon	Falcon: A Remote Sensing Vision-Language Foundation Model Arxiv2025
Copernicus-FM	GeoChat	LRS-VQA	When Large Vision-Language Model Meets Large Remote Sensing Imagery: Coarse-to-Fine Text-Guided Token Pruning ICCV2025
SelectiveMAE	LHRS-Bot	UrbanLLaVA	UrbanLLaVA: A Multi-modal Large Language Model for Urban Intelligence with Spatial Reasoning and Understanding ICCV2025
SMARTIES	RS-LLaVA	OmniGeo	OmniGeo: Towards a Multimodal Large Language Models for Geospatial Artificial Intelligence Arxiv2025
TerraMind	SkySenseGPT	EagleVision	EagleVision: Object-level Attribute Multimodal LLM for Remote Sensing Arxiv2025
SkySense V2		EarthMarker	SegEarth-R1
AlphaEarth	GeoText	RemoteSAM	RemoteSAM: Towards Segment Anything for Earth Observation ACMMM2025
SkySense++	Aquila	DynamicVL	DynamicVL: Benchmarking Multimodal Large Language Models for Dynamic City Understanding Arxiv2025
		LISAt	LISAt: Language- Instructed Segmentation Assistant for Satellite Imagery Arxiv2025

- The number of remote sensing foundation models is large and **continuously increasing**.
- Models differ in modalities, applications, resolutions, hardware requirements, etc.
- Model information is scattered (papers, model cards, repositories, etc.) and **inconsistently described**.
- Selection depends on **user-specific needs** (application, region, data, compute) and requires significant manual effort.

Challenge: Selecting Foundation Models for RS Tasks



I want to map different crop types in Africa. I have some Sentinel-2 data and a few labels from last year.



Can you check if this setup actually works?

What's your hardware setup?
Are the labels sparse or well-distributed?



✓ Ran test on Ethiopia + Morocco
Baseline F1: 0.68 → Adapted F1: 0.78
+0.10 gain → Accept adaptation



Running on cloud, no GPU worries. But label data is sparse.



Please deploy the model for my task.

Selected base model: Presto
Issues detected:
label sparsity, task shift, efficiency need
Suggested adapters:
LoRA, Spatial resizing, Pseudo-labeling
Optimized Config:
{ precision: "fp16", batch_size: 8, engine: "ONNXRuntime" }
Performance Estimation:
Expected F1: 0.78, Confidence: Medium

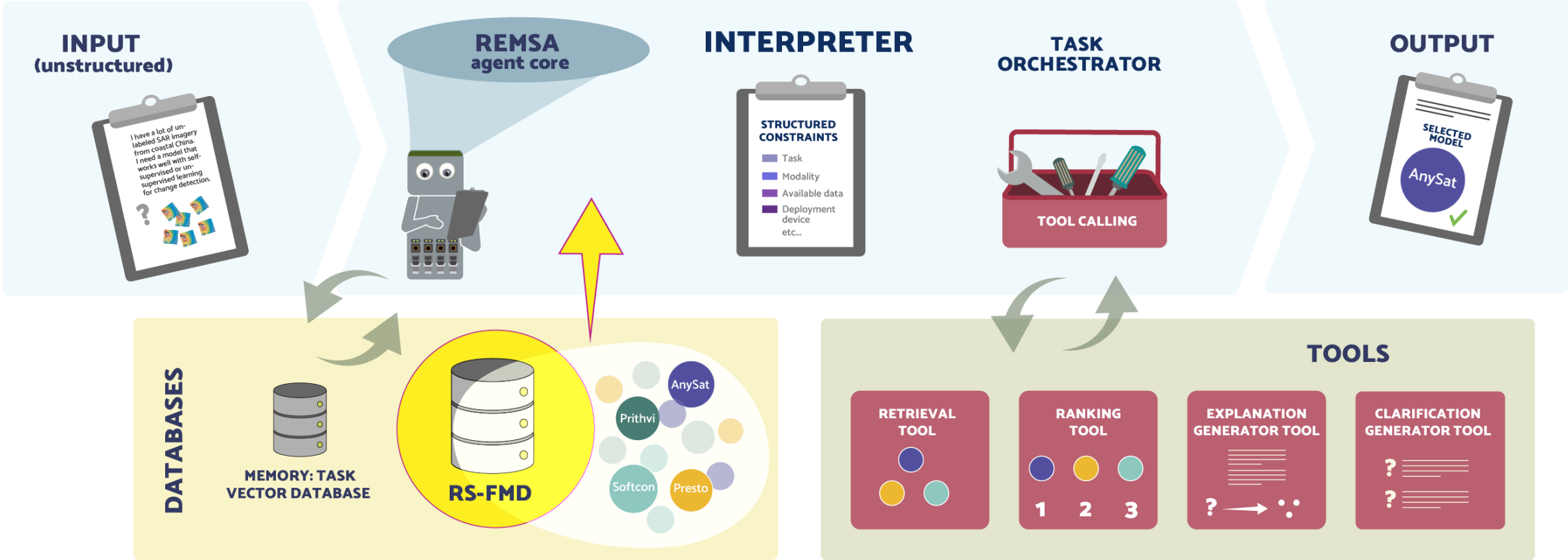


✓ Model deployment.

Final Report:



REMSA (Remote-sensing Model Selection Agent)



B. Chen, et al. "Foundation Model Selection for Remote Sensing via a Constraint-Aware Agent", under review.

FMD (Foundation Model Database)

Field	Type	Description
<i>Main Model Fields</i>		
model_id	string	Unique identifier of the model (free text)
model_name	string	Only the name of the model without extra descriptions (free text)
version	string	Version identifier (free text)
release_date	date	Release date of the model
last_updated	date	Last updated date
short_description	string	Short summary describing the model (free text)
paper_link	URL	URL to the associated publication
citations	integer	Number of citations
repository	URL	URL to the code repository
weights	URL	URL to pretrained model weights
backbone	string	Specific backbone used (free text)
num_layers	integer	Number of layers
num_parameters	float	Model size in millions of parameters
pretext_training_type	string	Type of pretext training strategy (free text)
masking_strategy	string	Masking strategy applied during training (free text)
pretraining	string	Description of pretraining approach (free text)
domain_knowledge	list[string]	Domain-specific knowledge or methods incorporated
backbone_modifications	list[string]	Modifications made to the backbone
supported_sensors	list[string]	Supported satellite sensors
modality_integration_type	string	Integration type (free text)
modalities	list[string]	Input data modalities (free text)
spectral_alignment	{full, partial, none}	Whether the model models spectral continuity
temporal_alignment	{full, partial, none}	Whether the model models temporal sequences
spatial_resolution	string	Spatial resolution of data (free text)
temporal_resolution	string	Temporal resolution of data (free text)
bands	list[string]	Spectral bands used

<i>Nested: PretrainingPhase</i>		
dataset	string	Dataset used for pretraining (free text)
regions_coverage	list[string]	Geographical regions covered
time_range	string	Time range of pretraining data (free text)
num_images	integer	Number of images used
token_size	string	Token size (free text)
image_resolution	string	Input image resolution (free text)
epochs	integer	Number of epochs
batch_size	integer	Batch size
learning_rate	string	Learning rate (free text)
augmentations	list[string]	Augmentations applied
processing	list[string]	Additional preprocessing steps
sampling	string	Sampling strategy (free text)
processing_level	string	Processing level (free text)
cloud_cover	string	Cloud cover filtering (free text)
missing_data	string	Handling of missing data (free text)
masking_ratio	float	Masking ratio

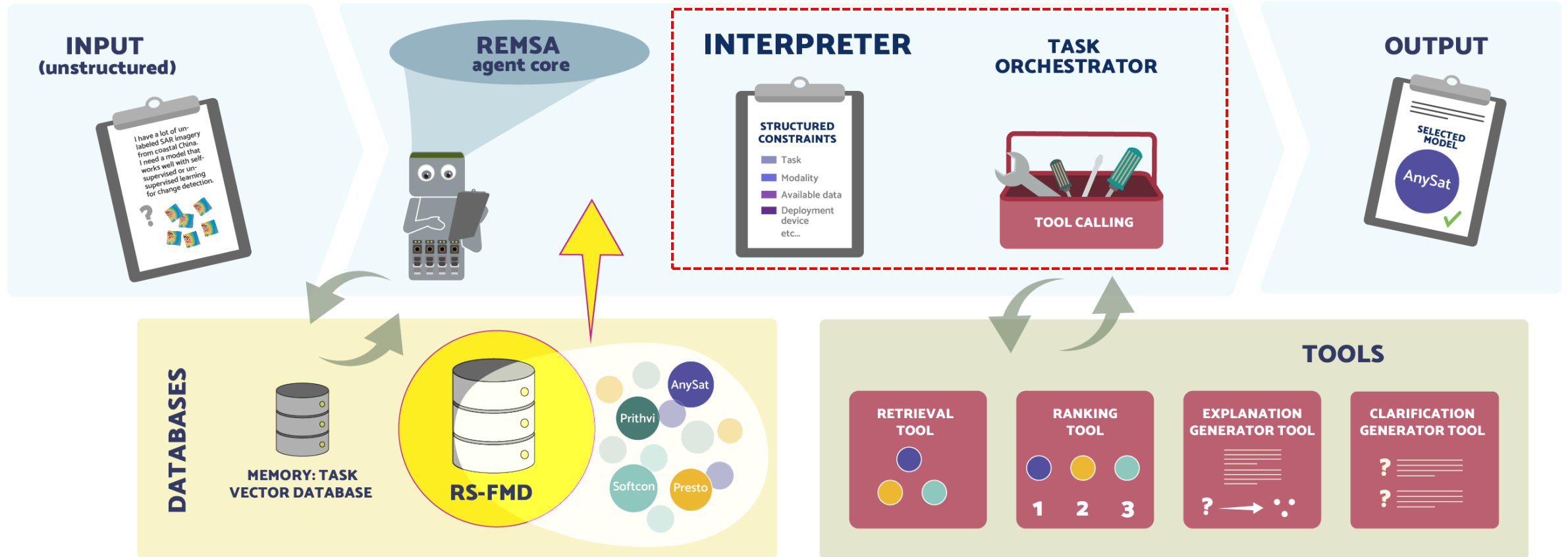
<i>Nested: Benchmark</i>		
task	string	Type of task evaluated (free text)
application	string	Specific application domain (free text)
dataset	string	Benchmark dataset name (free text)
metrics	list[string]	List of evaluation metrics
metrics_value	list[float]	Numeric values for each metric
sensor	list[string]	Sensors used
regions	list[string]	Regions evaluated
original_samples	integer	Total number of samples before sampling
num_samples	integer	Actual number of samples used
sampling_percentage	float	Fraction of dataset retained (0–100)
num_classes	integer	Number of classes
classes	list[string]	Names of each class
image_resolution	string	Input image resolution (free text)
spatial_resolution	string	Spatial resolution (free text)
bands_used	list[string]	Bands used during evaluation
augmentations	list[string]	Data augmentations applied
optimizer	string	Optimizer used (free text)
batch_size	integer	Batch size
learning_rate	float	Learning rate
epochs	integer	Number of epochs
loss_function	string	Loss function (free text)
split_ratio	string	Train/val/test split ratio (free text)



B. Chen, et al. “Foundation Model Selection for Remote Sensing via a Constraint-Aware Agent”, under review.



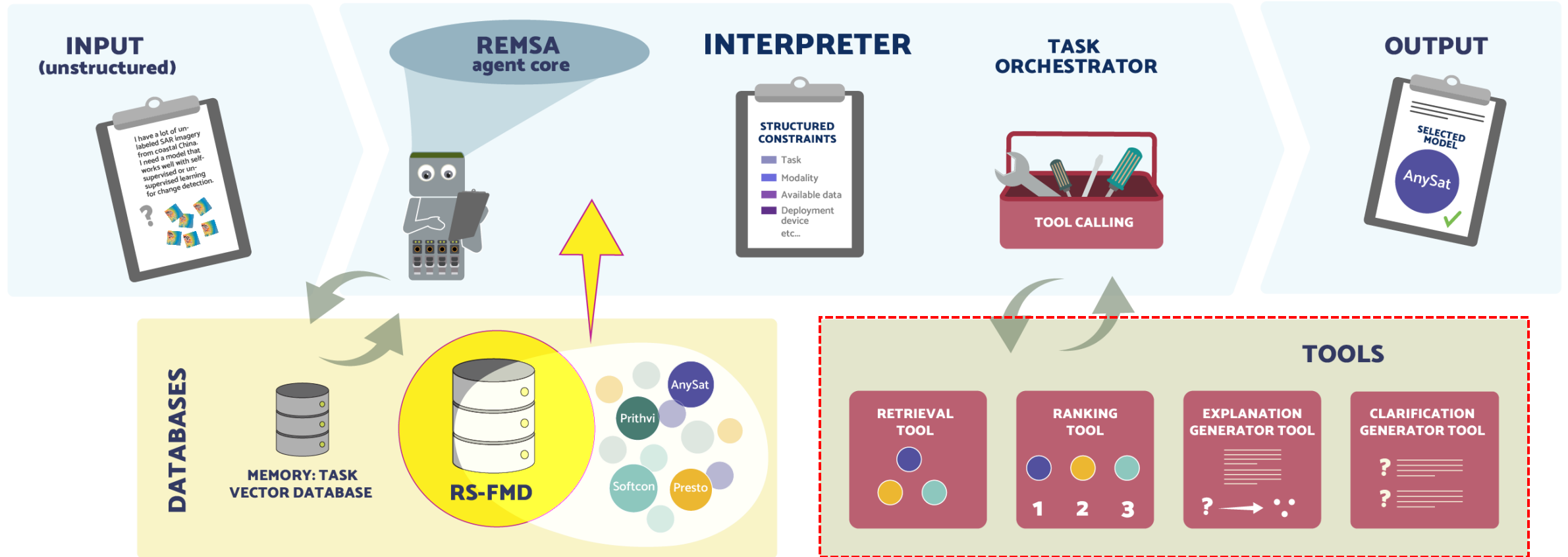
REMSA (Remote-sensing Model Selection Agent)



LLM Agent Core:

- Interpreter: parses user queries into structured constraints.
- Task Orchestrator: manages workflow and tool calls.

REMSA (Remote-sensing Model Selection Agent)



External Tools: • Retrieval Tool: finds candidate FMs from FMD via structured and semantic search • Ranking Tool: applies in-context reasoning to prioritize models • Clarification Generator: asks targeted follow-up questions when needed during reasoning. • Explanation Generator: produces transparent justifications.

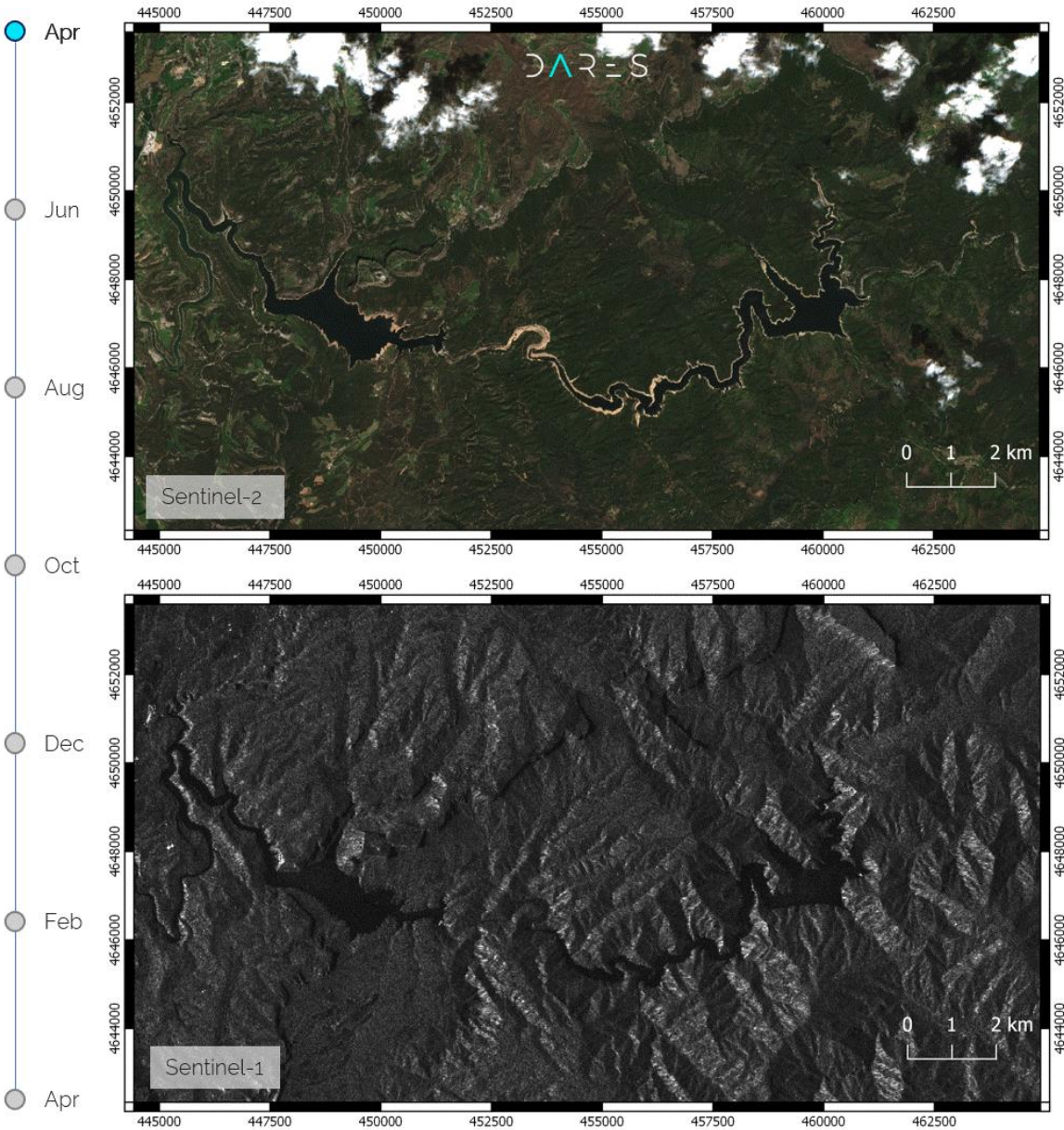


AGENT-BIGEARTH

AI-POWERED ENVIRONMENTAL INTELLIGENCE



Agent-BigEarth



Can you find the most relevant images from the Copernicus data archive to help us monitor drought conditions in Spain?

Yes. I found a time-series of Sentinel-2 images covering the past five years. They include vegetation and soil moisture indicators. Would you like me to prepare a timeline for comparison?



Yes, a timeline would be useful.

Here is the timeline. You can see strong vegetation declines in 2019 and again in 2022.



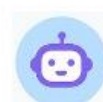
Can you quantify the decline?

Between 2018 and 2023, vegetation density dropped by 28% in the central region. This aligns with rainfall records from Copernicus Climate Data.



Can you generate a short briefing for the Ministry of Agriculture?

Certainly. I prepared a 2-page report with annotated images and charts. It highlights the most affected areas, the timeline of vegetation decline, and potential agricultural risks.



Good. Add a one-page recommendation section focusing on irrigation strategies.

Done. The report now includes suggested irrigation measures based on observed Copernicus satellite patterns and past drought responses.



Agent-BigEarth



Before



After



Journalist

I heard about the flooding in Germany last week. Can you show me the most recent Copernicus' Sentinel images of the affected area?

Yes. I found a Sentinel-2 multispectral image from 3 days ago, just after the peak of the flooding. Would you like me to also show a comparison with how the area looked a month earlier?



Journalist

That would be perfect.

Here's a side-by-side view. On the left, the riverbanks are clearly visible before the flood. On the right, large sections of farmland and roads are underwater.



Journalist

Can you highlight which towns were most affected?

Sure. I've marked three towns where flooding covered more than 40% of the built-up area. Would you like me to export this as an image package for your article?



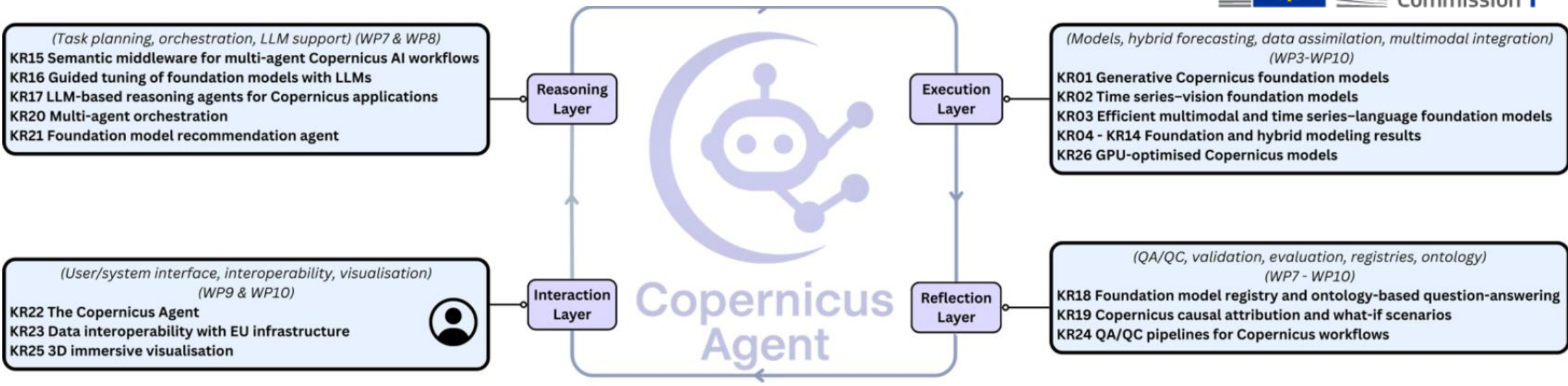
Journalist

Yes, please include captions that explain the scale of the disaster.

Done. I added short captions and percentage estimates of flooded areas, ready for publication.



PROTEUS: The Copernicus Agent for Next-Generation Environmental Intelligence



Agentic AI for Earth Observation

- ✓ Multi-Agents Collaboration for EO
- ✓ LLM-driven planning and workflow automation;
- ✓ Geospatial agent-grounded reasoning;
- ✓ Agent protocols for ground–space coordination;
- ✓ Integration with geospatial Foundation Models;
- ✓ Applications across both upstream mission design and downstream service delivery.



Workshop on Agentic AI for Earth Observation

Date: 19-21 October 2026

Venue: Langenbeck-Virchow-Haus, Berlin

Visit Our Group Webpage
<https://rsim.berlin>

