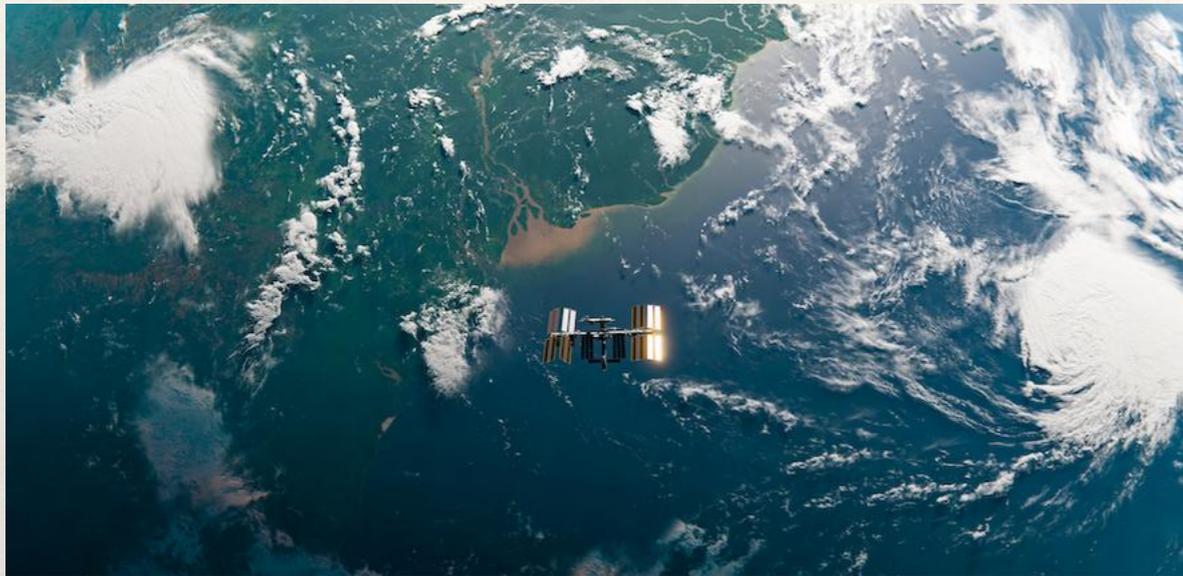


# Finding all life on earth (with OxCaml!)

WG2.8, Portugal  
March 2026



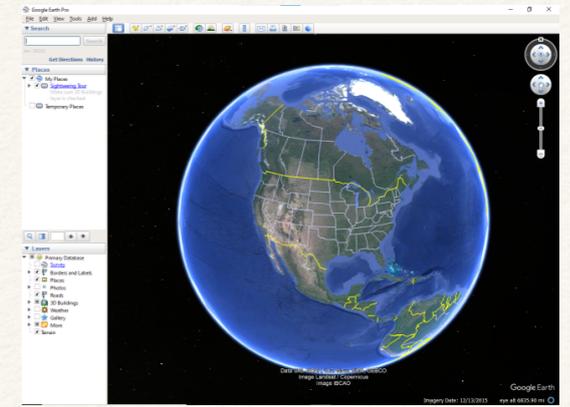
**Anil Madhavapeddy**, Professor of Planetary Computing, [anil.recoil.org](http://anil.recoil.org)

Department of Computer Science & Technology,  
University of Cambridge ([cst.cam.ac.uk](http://cst.cam.ac.uk))  
Cambridge Conservation Initiative ([conservation.cam.ac.uk](http://conservation.cam.ac.uk))

With much help from a diverse and interdisciplinary team from Computer Science, Plant Science, Ecology, Conservation and Policy, including Sadiq Jaffer, Srinivasan Keshav, Cyrus Omar, Clement Atzberger, Andrew Balmford, Thomas Ball, Andrew Blake, Neil Burgess, Alec Christie, David Coomes, Michael Dales, Lynn Dicks, Alison Eyres, Frank Feng, Patrick Ferris, Ryan Gibb, Radhika Iyer, Jovana Knezevic, Madeline Lisaius, Ian Ondo, Sam Reynolds, William Sutherland, Thomas Swinfield, Shane Weisz, Robin Young, and more!

**2004**

OpenStreetMap  
Google Maps  
Google Earth



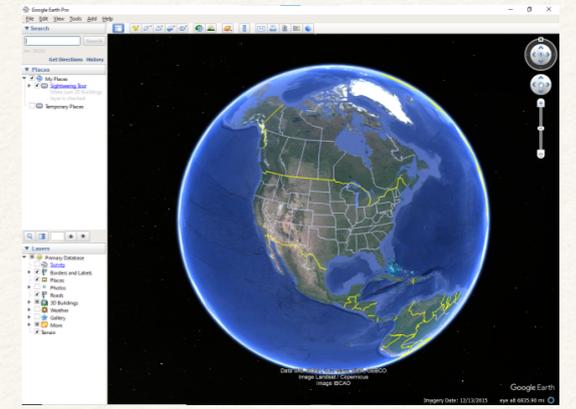
???

**2026**

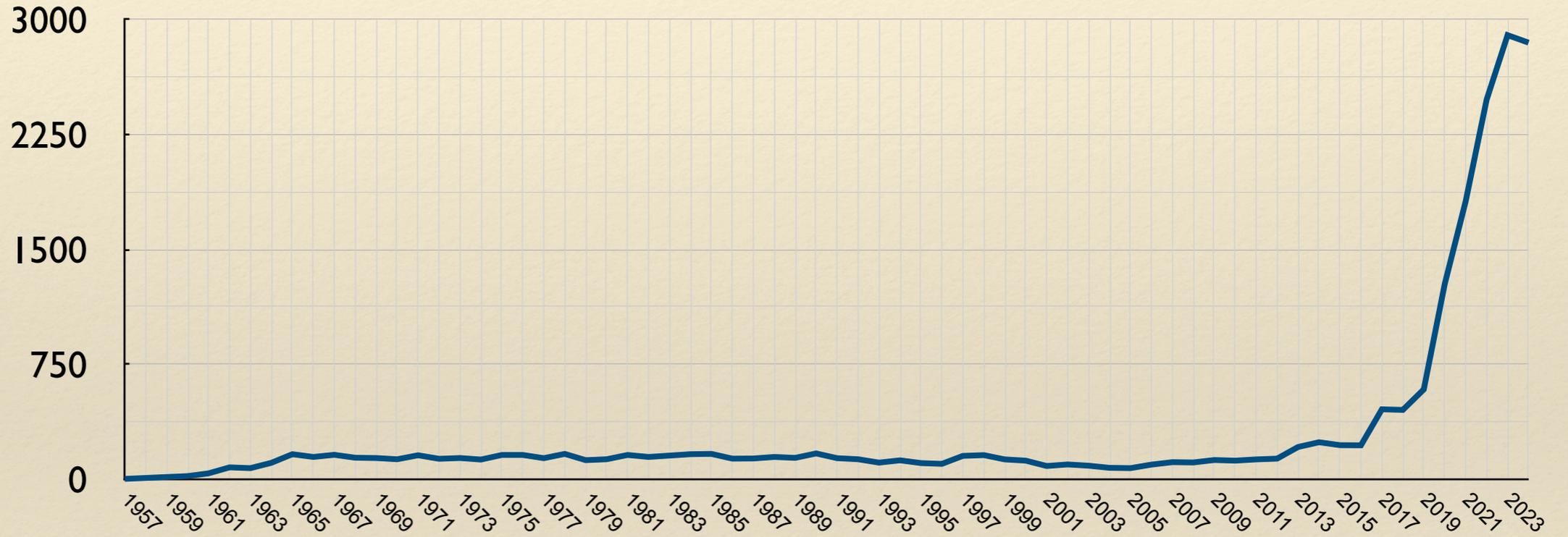


2004

OpenStreetMap  
Google Maps  
Google Earth



Number of satellite launches per year



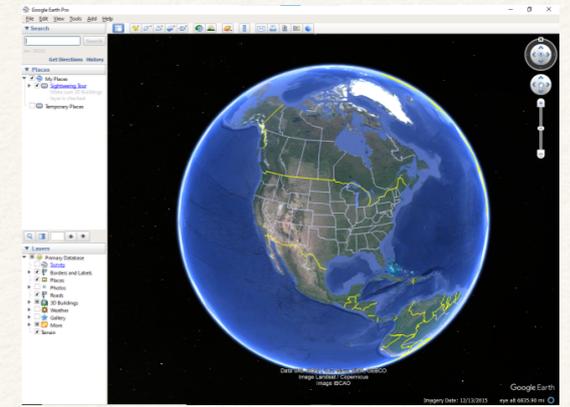
???

2026

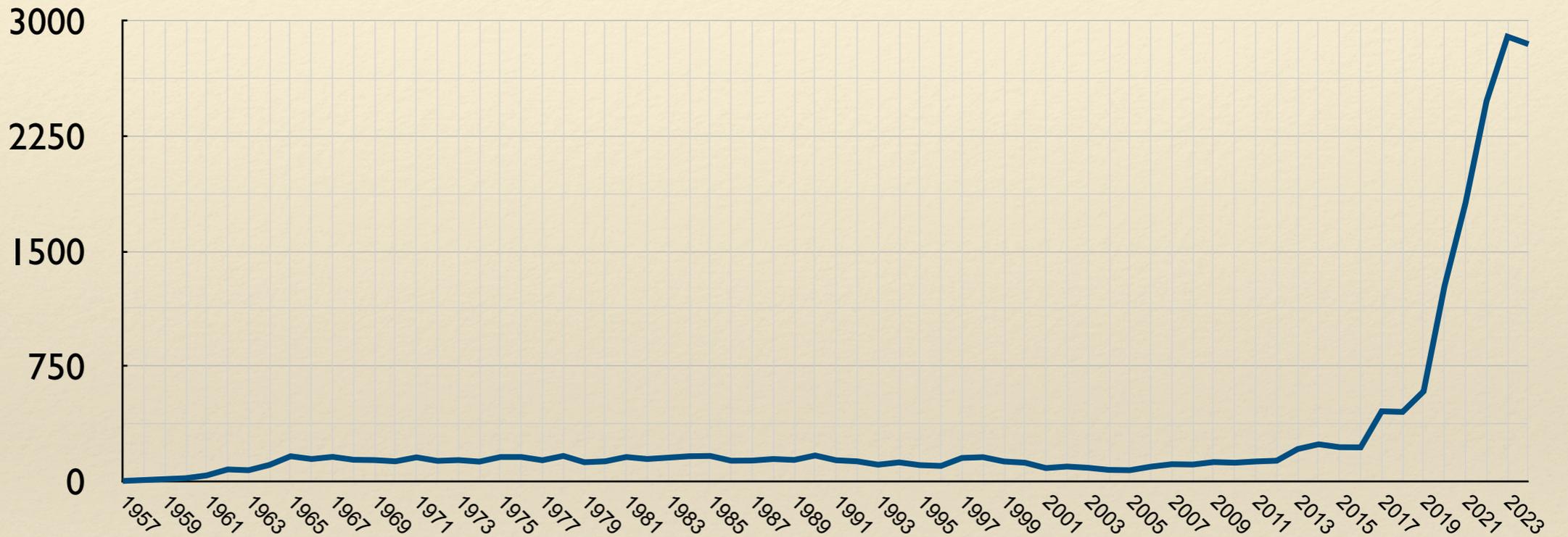


2004

OpenStreetMap  
Google Maps  
Google Earth



Number of satellite launches per year



???

2026



# The Bitter Lesson

The most significant advancements in AI have come from focusing on general methods that leverage computation rather than human-designed representations and knowledge

*Richard Sutton, 2019, <http://www.incompleteideas.net/Incldeas/BitterLesson.html>*



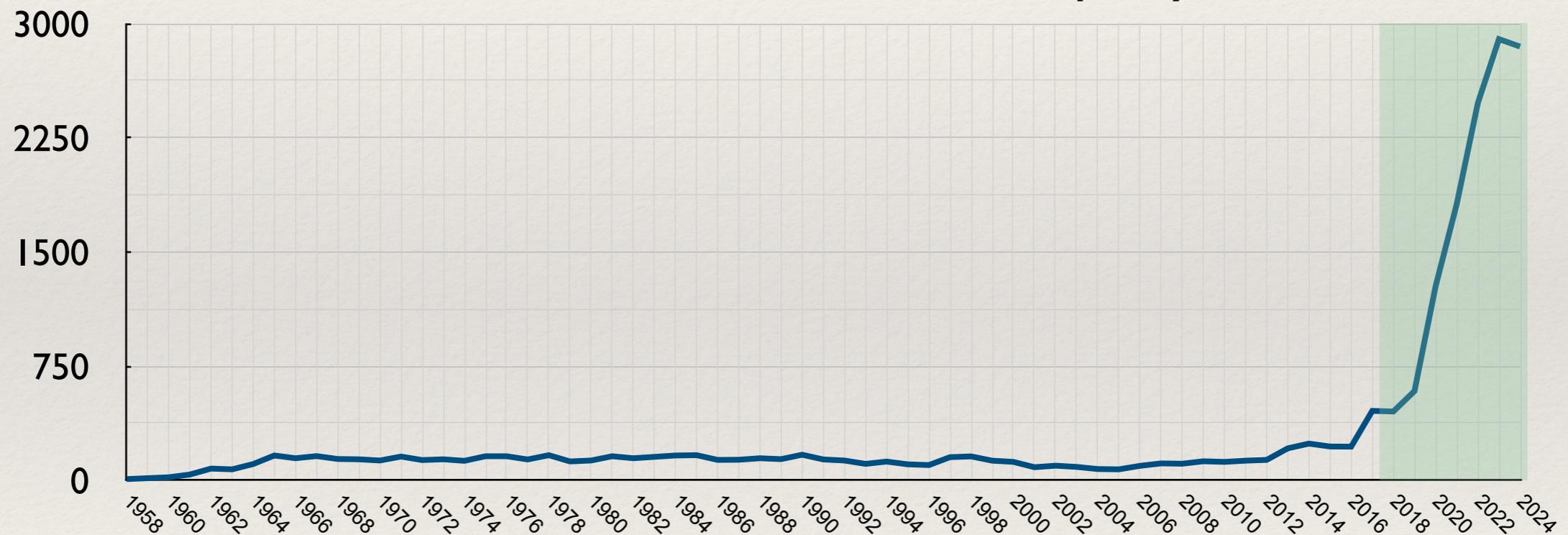
# The Bitter Lesson

The most significant advancements in AI have come from focusing on general methods that leverage computation rather than human-designed representations and knowledge

*Richard Sutton, 2019, <http://www.incompleteideas.net/Incldeas/BitterLesson.html>*



## Number of satellite launches per year



**5TB/day/satellite = exabytes of observations**



**v1 Can we identify activity anywhere?**

**Interactively!**

**v2 Can we spot changes in realtime?**

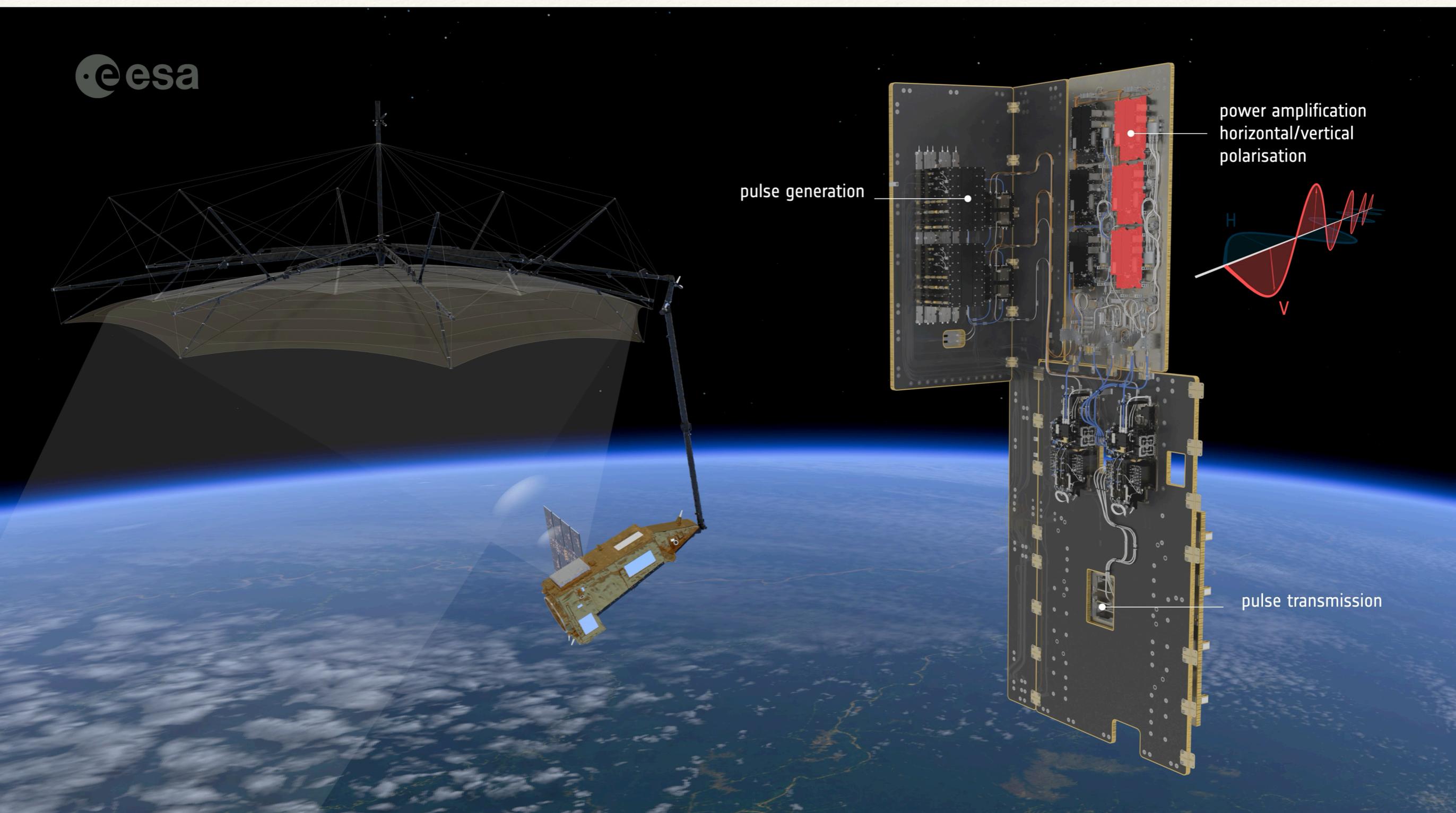
**Mega scaling!**

**v3 Can we predictively model the planet?**

**Emergent reasoning!**



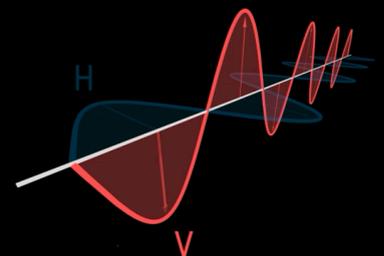
# The Bitter Lesson for Satellite Data: the signal is noisy



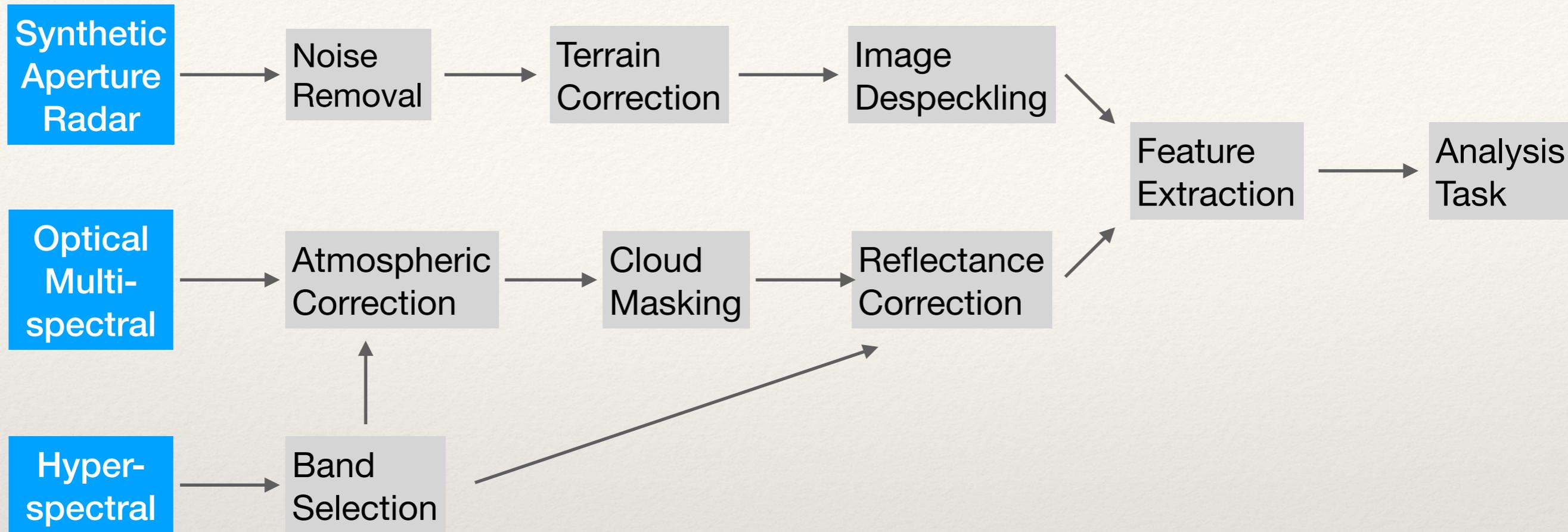
pulse generation

power amplification  
horizontal/vertical  
polarisation

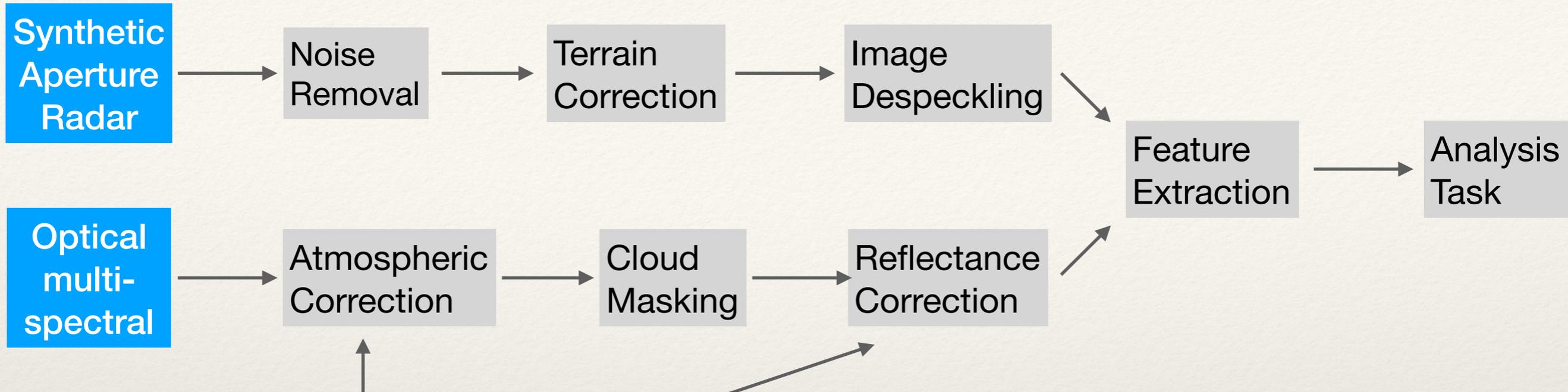
pulse transmission



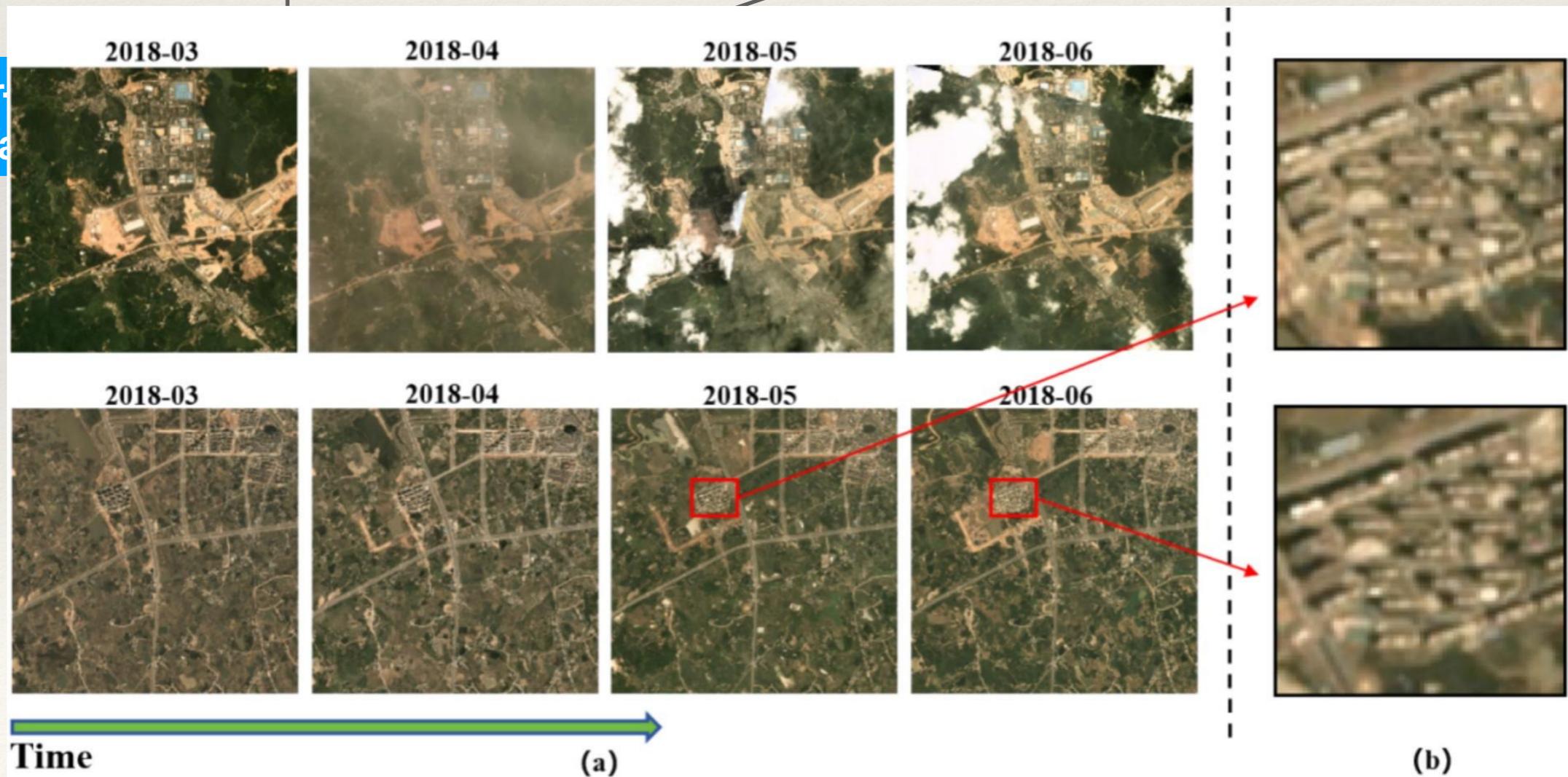
# The Bitter Lesson for Satellite Data: every GIS task is manual



# The Bitter Lesson for Satellite Data: signals are temporal



Hyper  
spectra

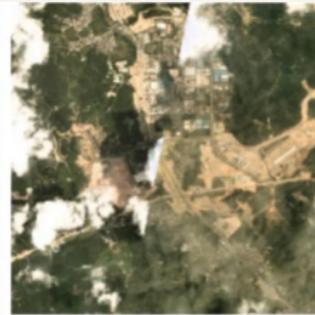


# The Bitter Lesson for Satellite Data: life's too short *(to do anything other than self supervised learning)*

← Unlabelled Self-Supervised Learning

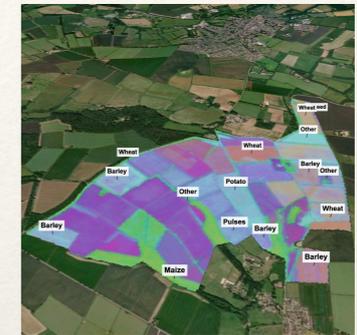
→ Few-shot labels

*Petabytes*



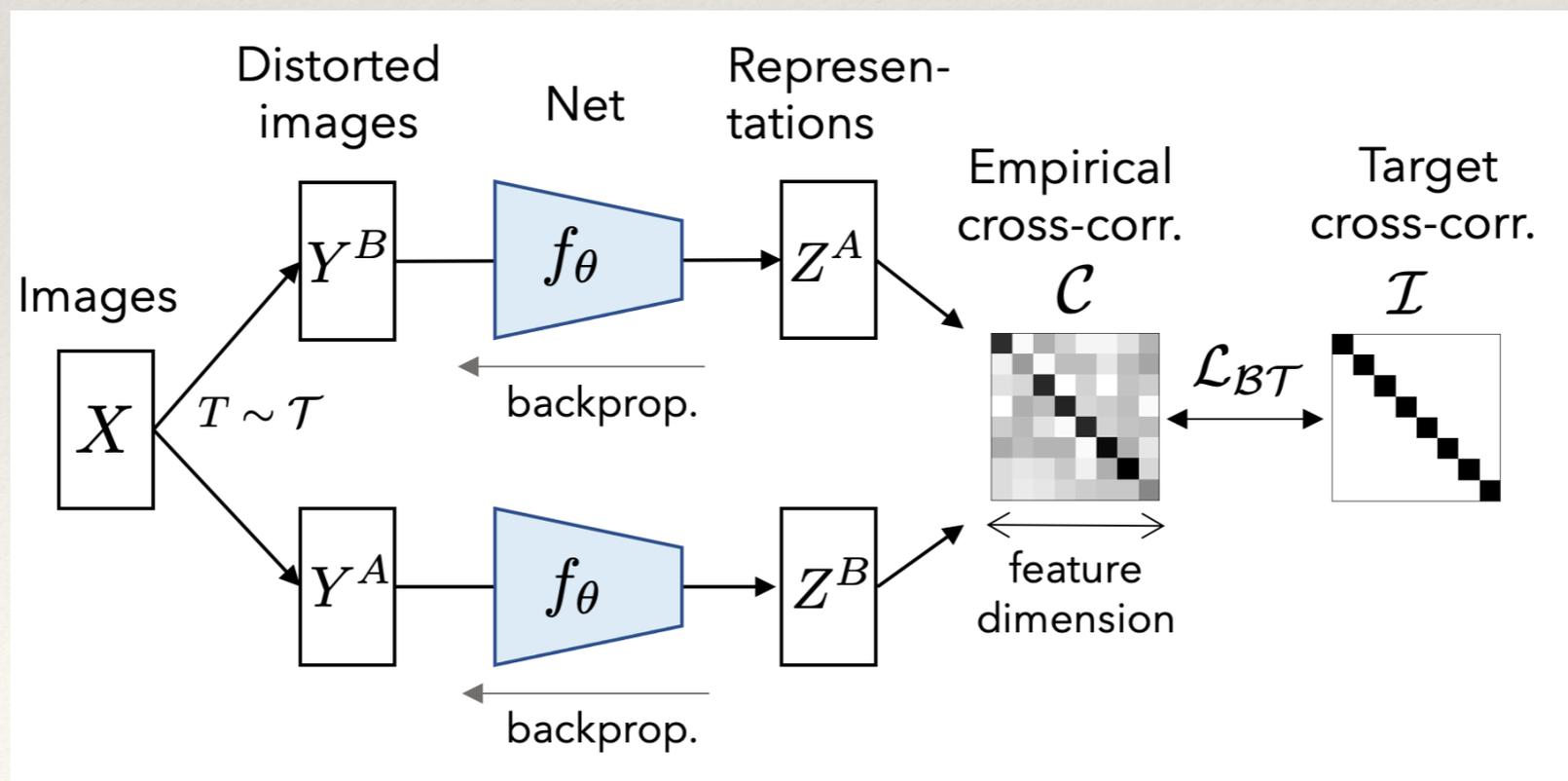
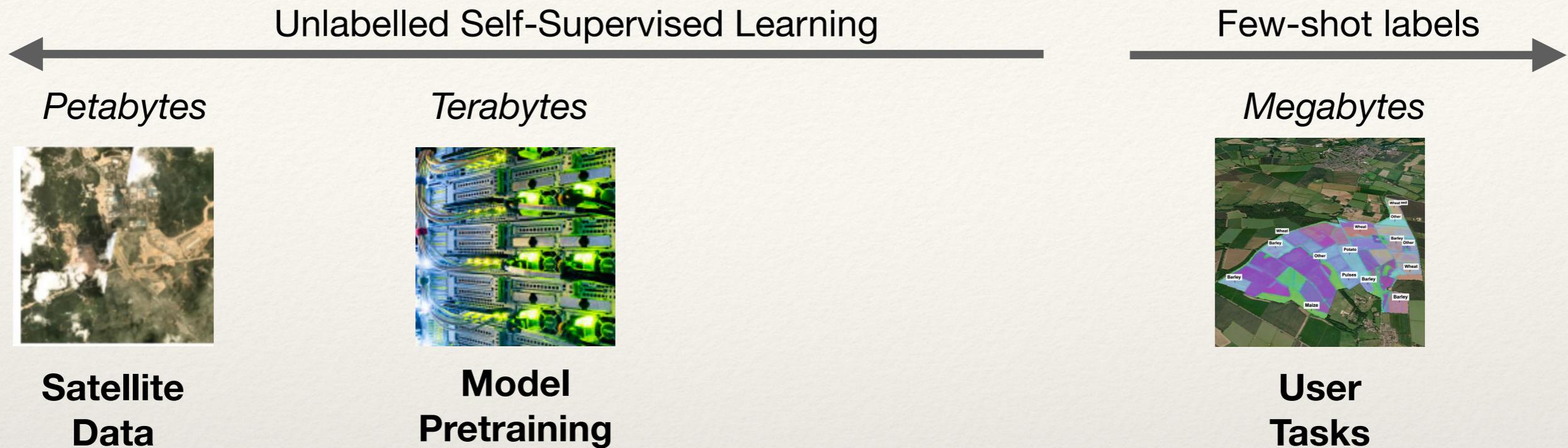
**Satellite  
Data**

*Megabytes*



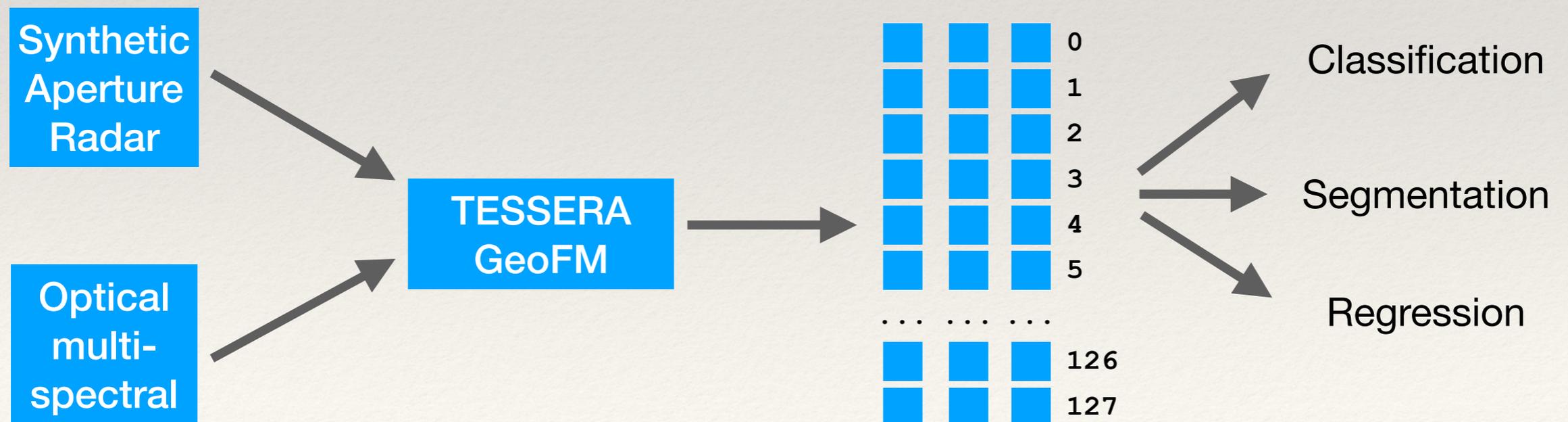
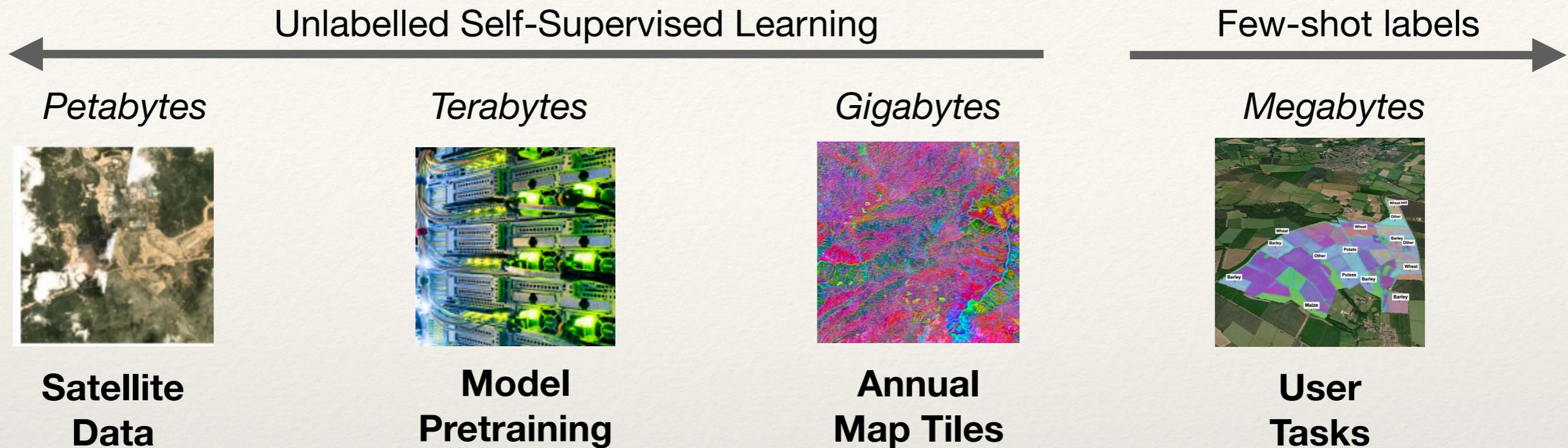
**User  
Tasks**

# The Bitter Lesson for Satellite Data: life's too short (to do anything other than self supervised learning)



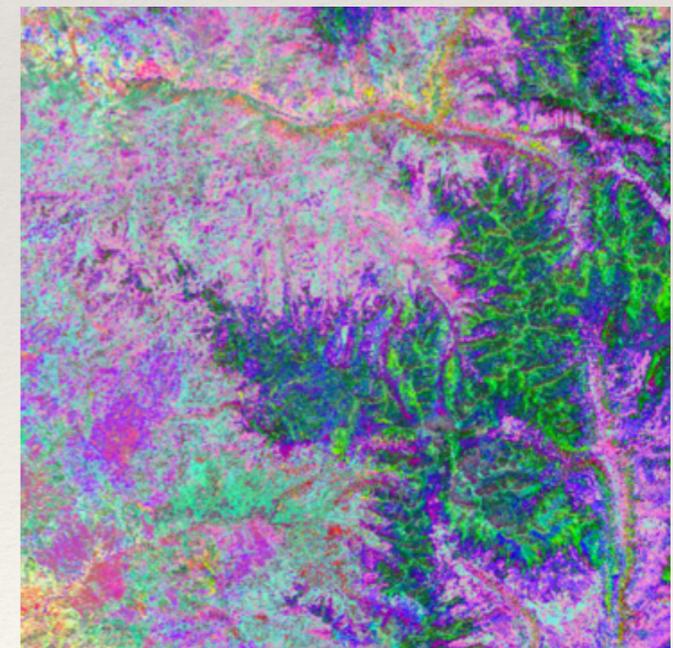
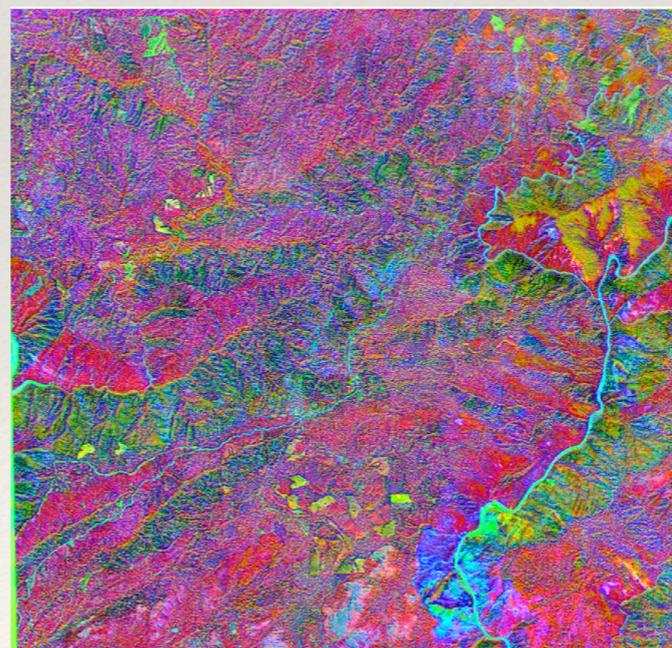
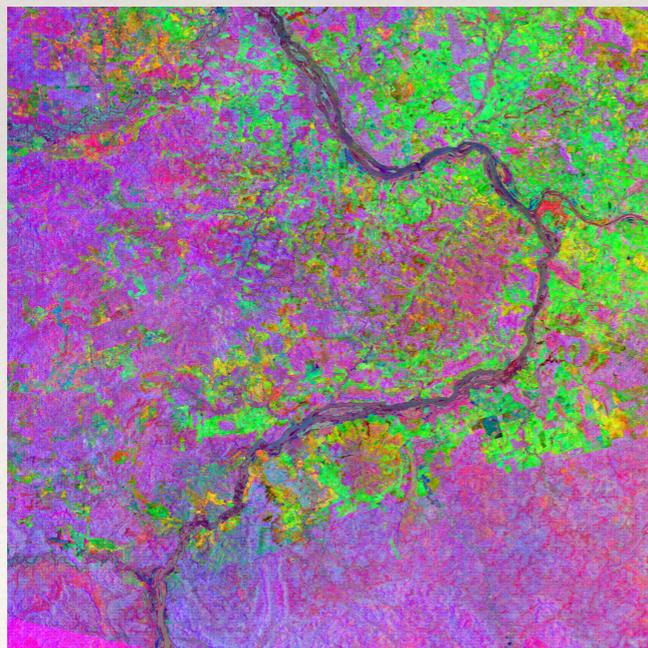
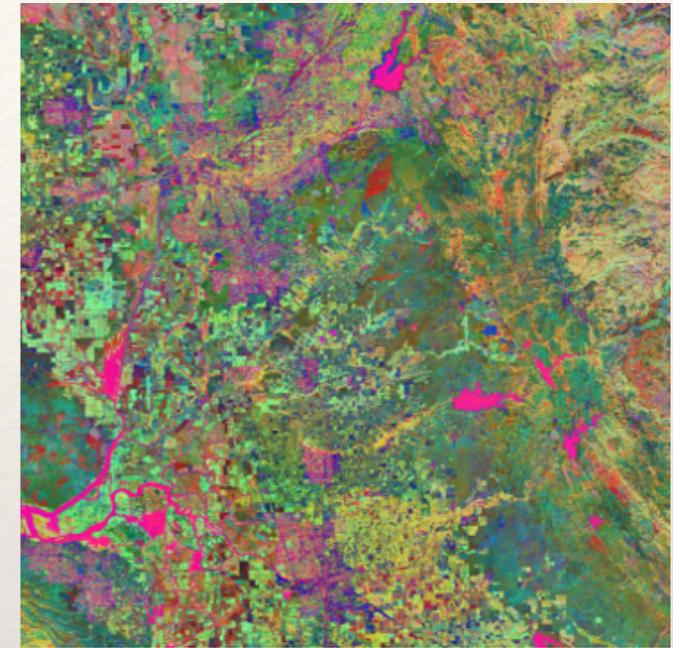
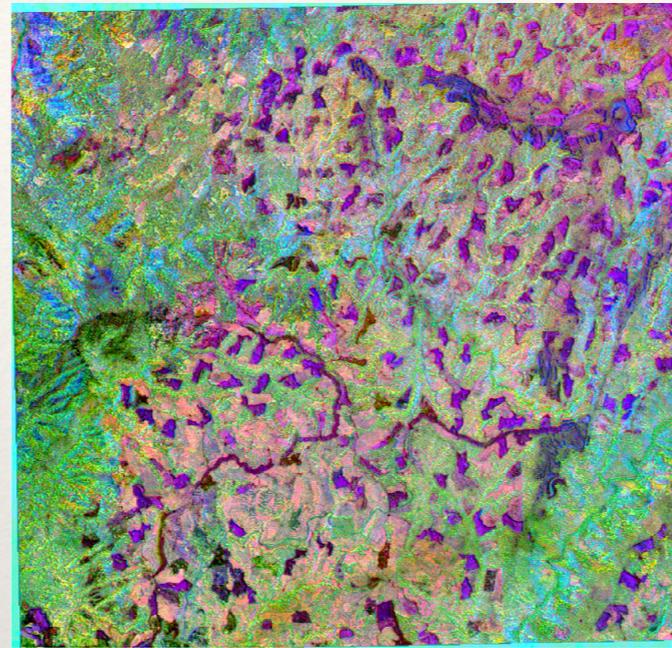
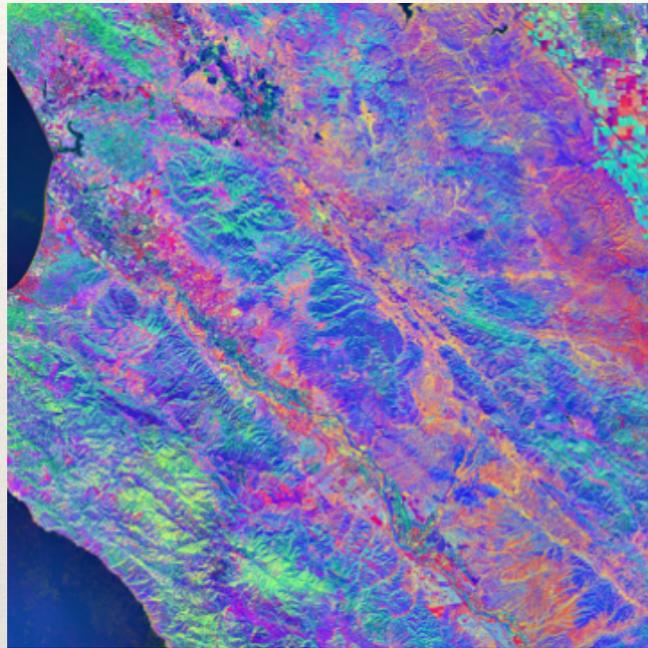
# The Bitter Lesson for Satellite Data: life's too short

*(to do anything other than self supervised learning)*

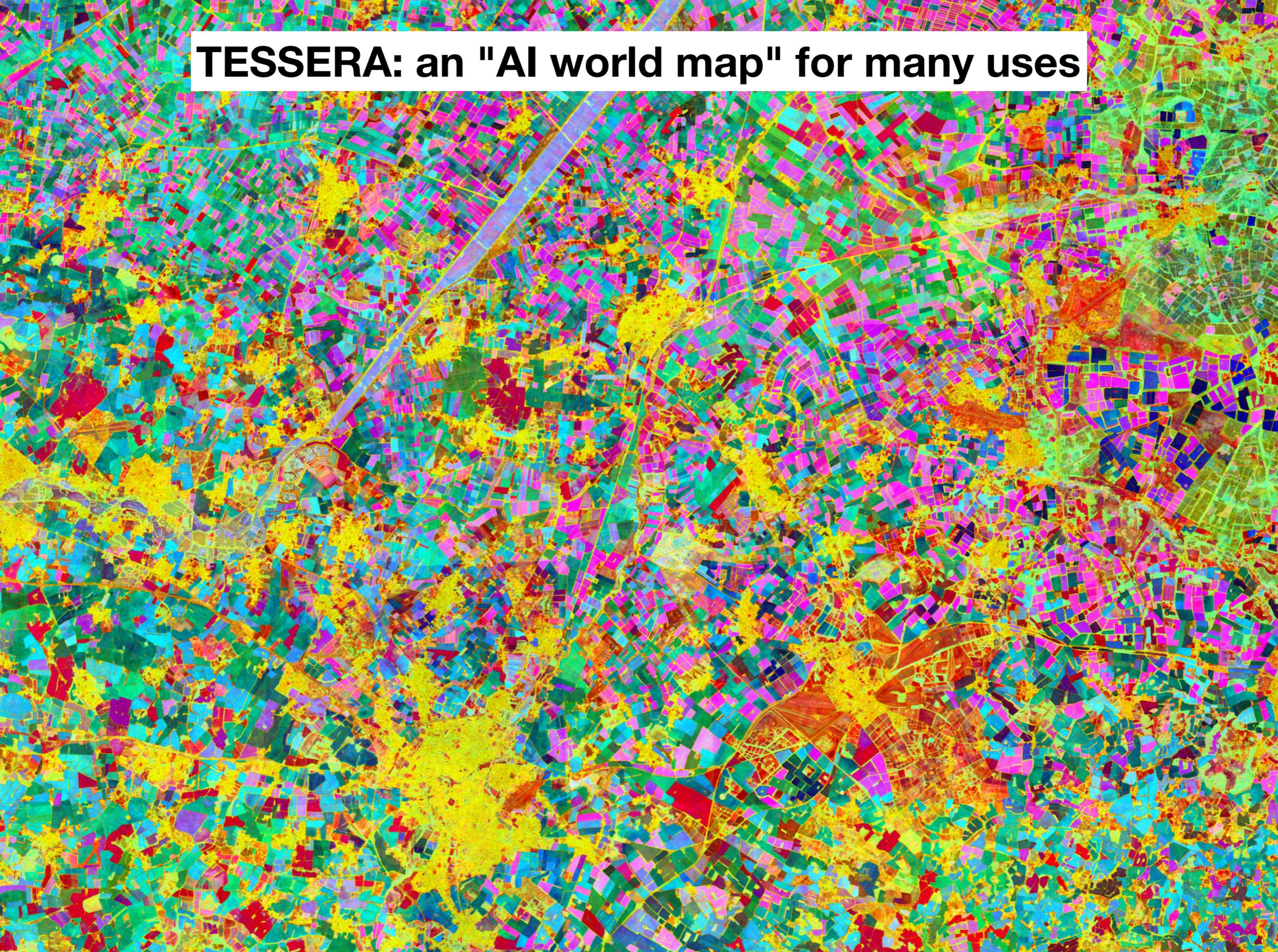


# TESSERA: an "AI world map" for many uses

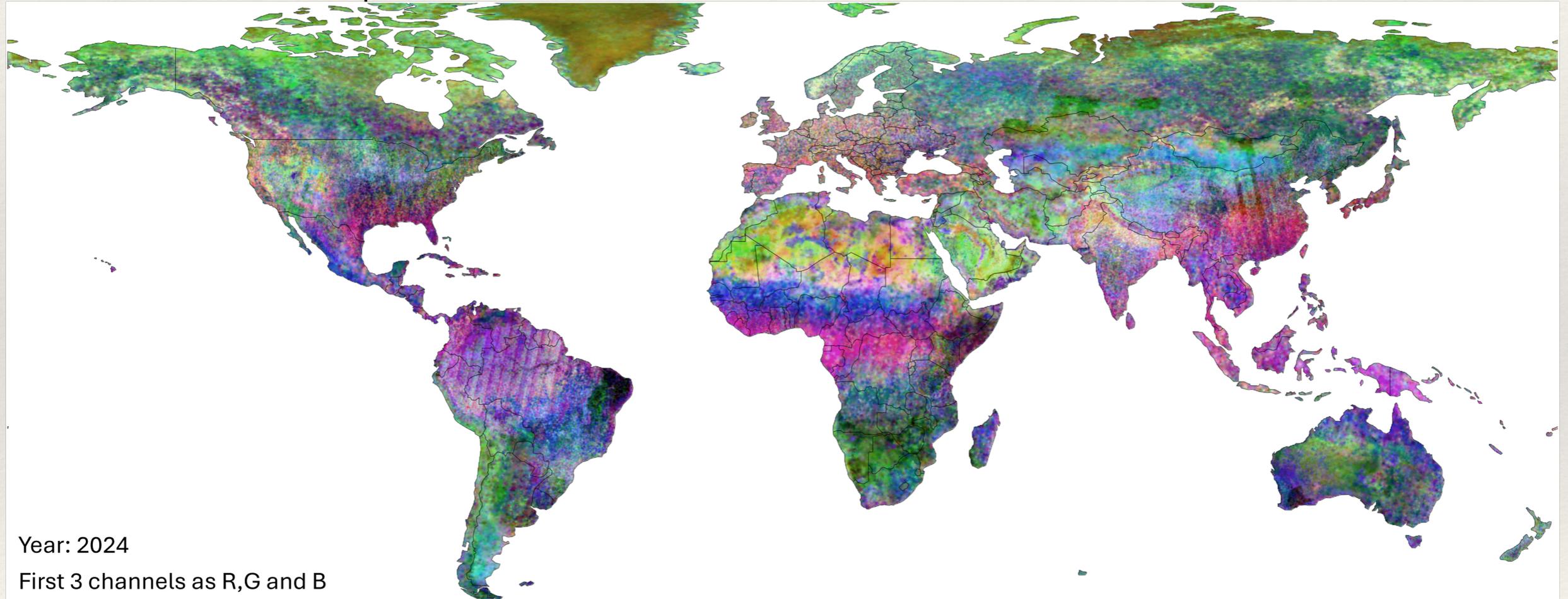
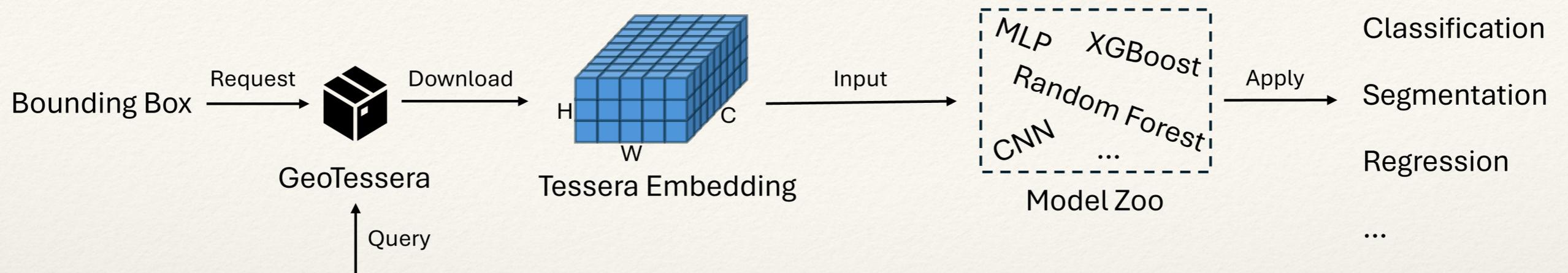
We generate 128-dimension images encoding the **annual spectral-temporal evolution** for 10m<sup>2</sup> tiles



**TESSERA: an "AI world map" for many uses**



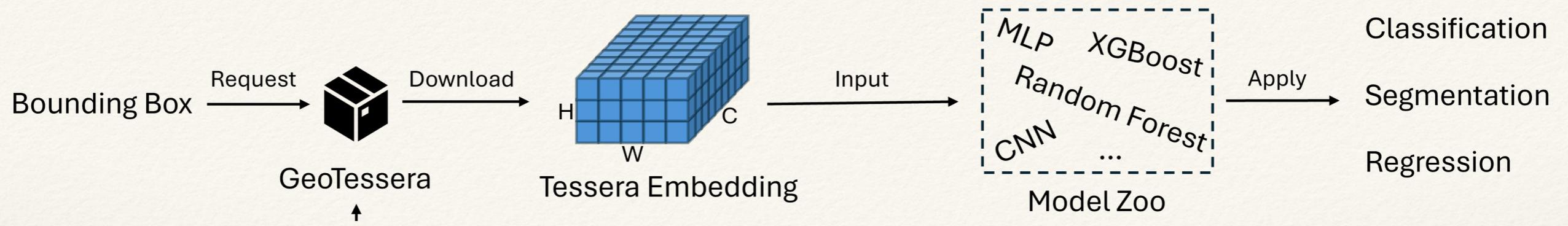
# Let's see TESSERA in action!



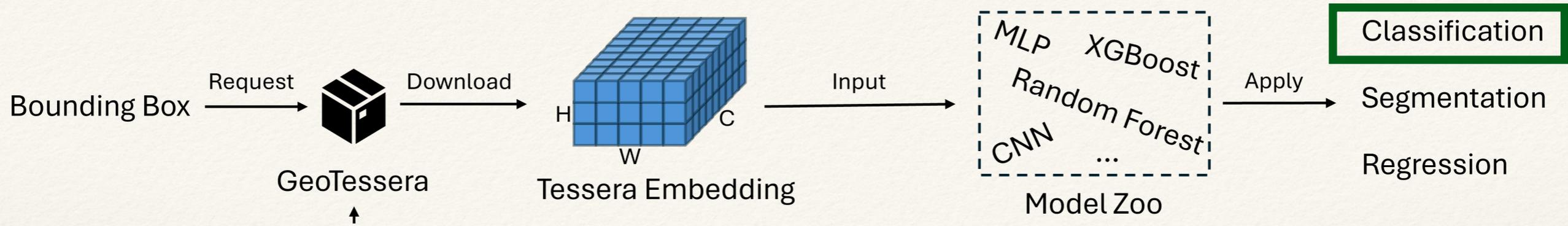
*"TESSERA: A Foundation Model for Earth Observation", Z. Feng et al  
<https://arxiv.org/abs/2506.20380>*



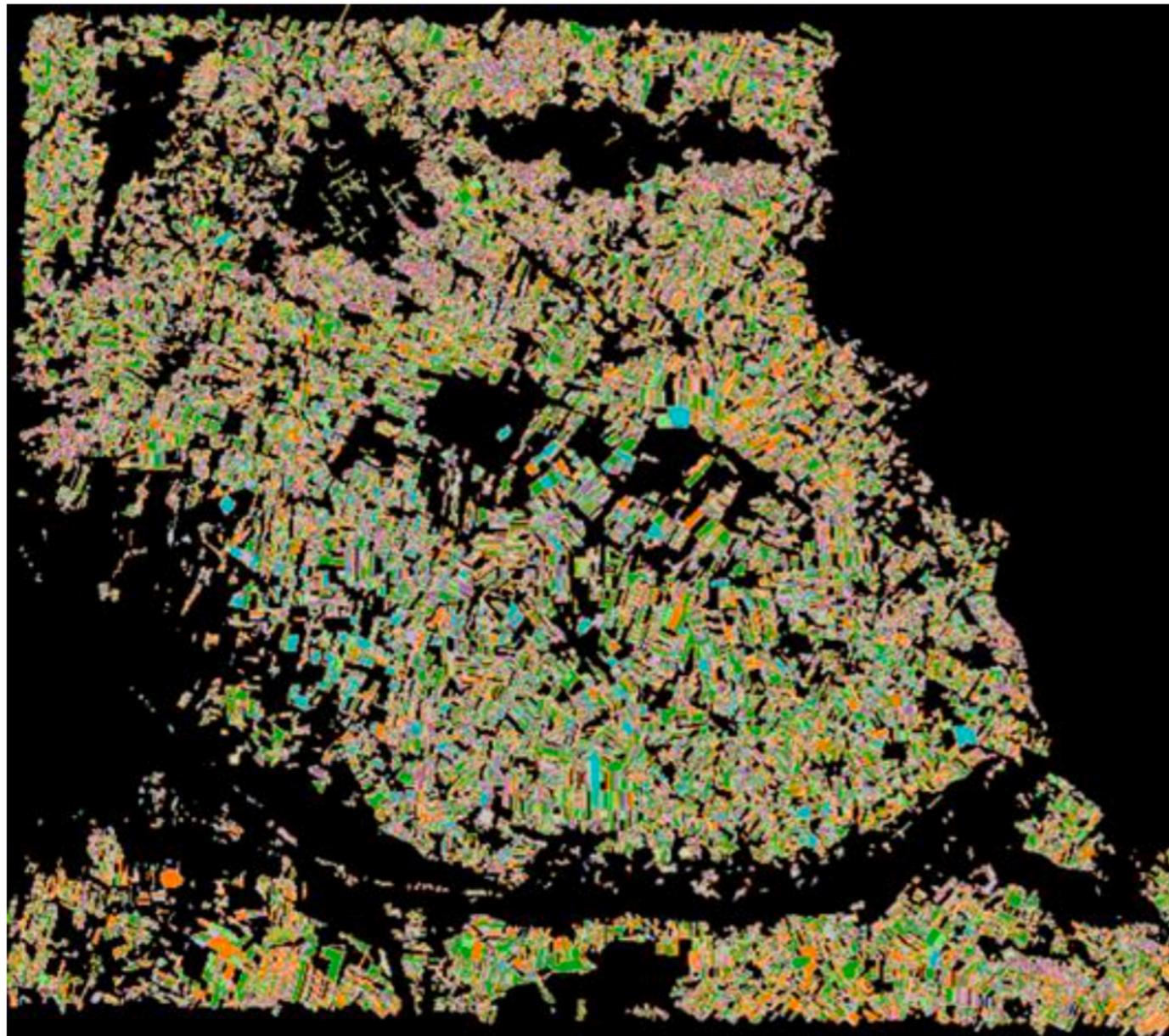
# TESSERA: same map tiles, lots of downstream uses



# TESSERA: same map tiles, lots of downstream uses



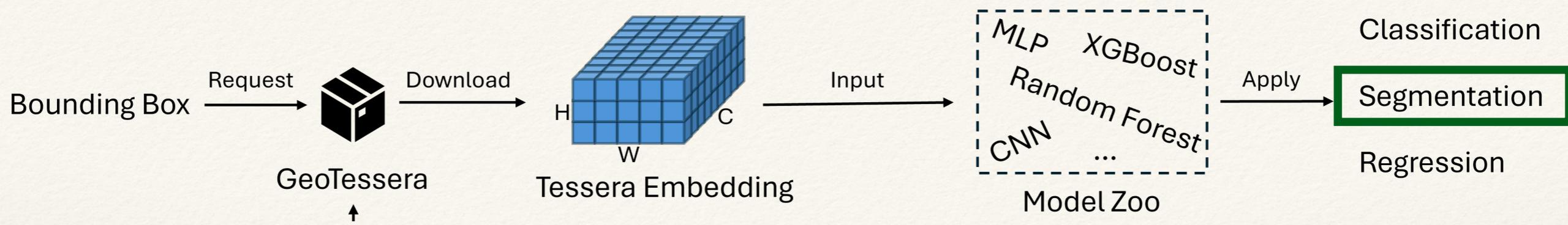
## INVEKOS Austrian Crop Dataset



Name	Number of Samples	Percent age
Legume	2227	3.31%
Soy	5892	8.76%
Summer Grain	2475	3.68%
<b>Winter Grain</b>	<b>24914</b>	<b>37.05%</b>
Corn	6902	10.27%
Sunflower	207	0.31%
Mustard	1734	2.58%
Potato	2514	3.74%
Beet	1257	1.87%
Squash	2019	3.00%
Grapes	222	0.33%
Tree Fruit	347	0.52%
Cover Crop	1418	2.11%
Grass	2349	3.49%
Fallow	4484	6.67%
Other (Plants)	8220	12.23%
Other (Non Plants)	57	0.08%
<b>Total</b>	<b>67238</b>	<b>100%</b>

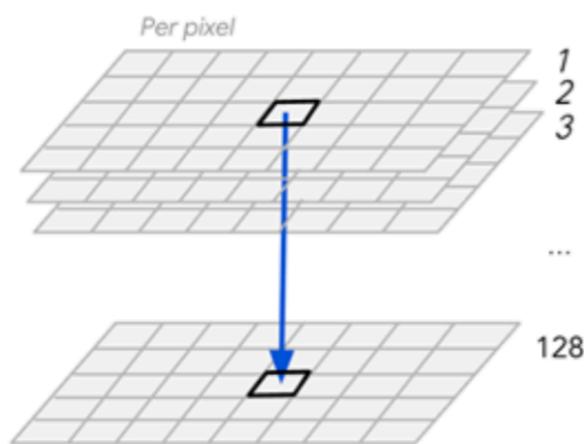
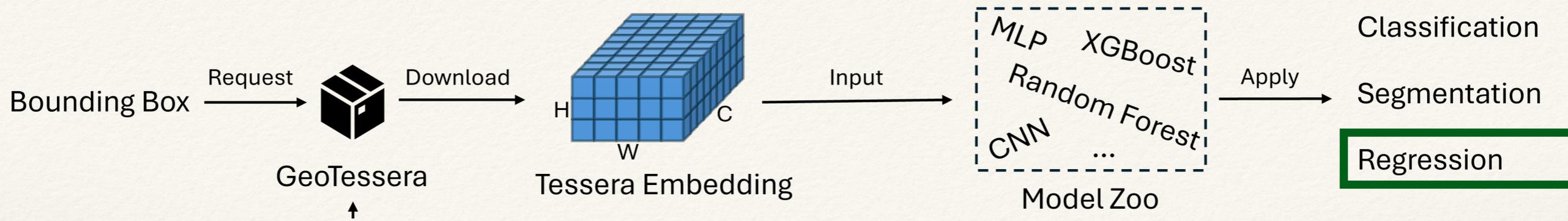


# TESSERA: same map tiles, lots of downstream uses

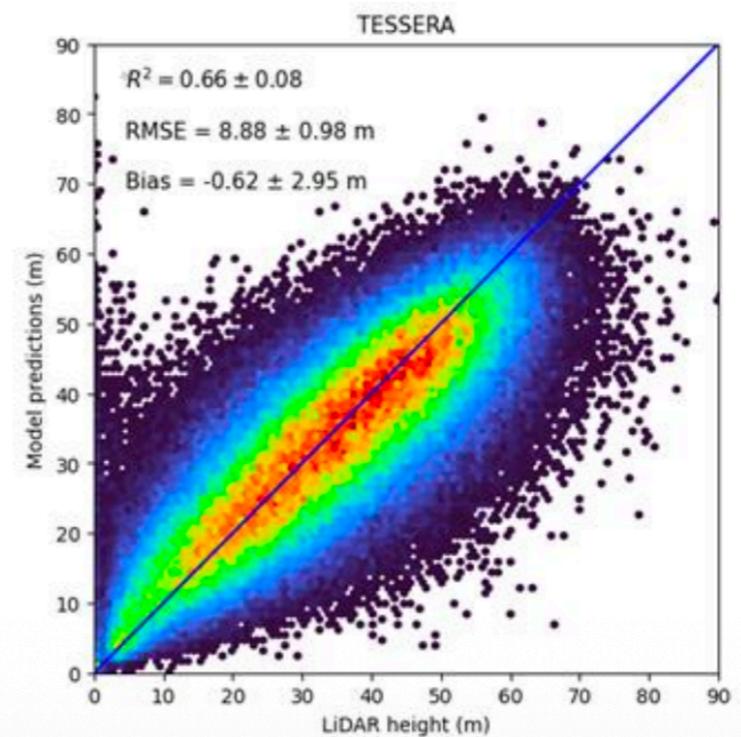
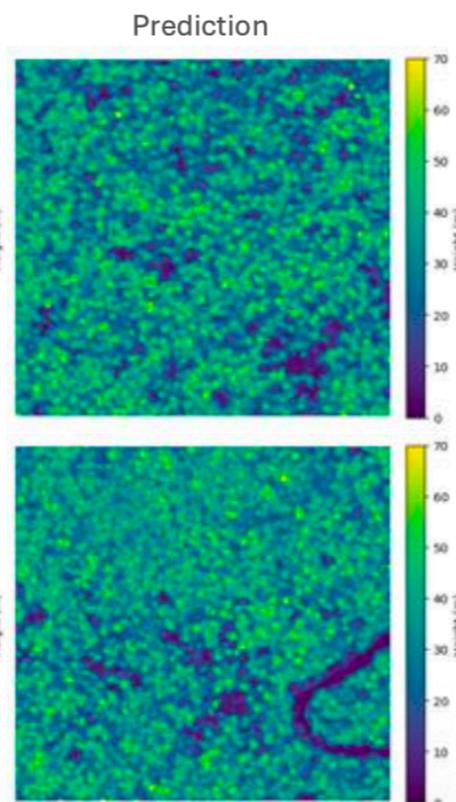
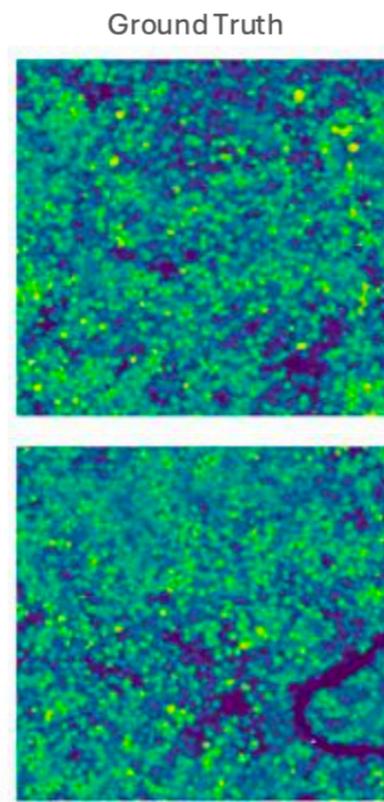
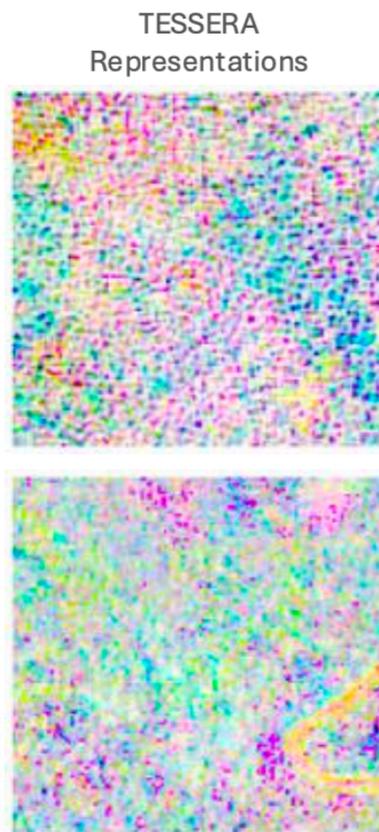
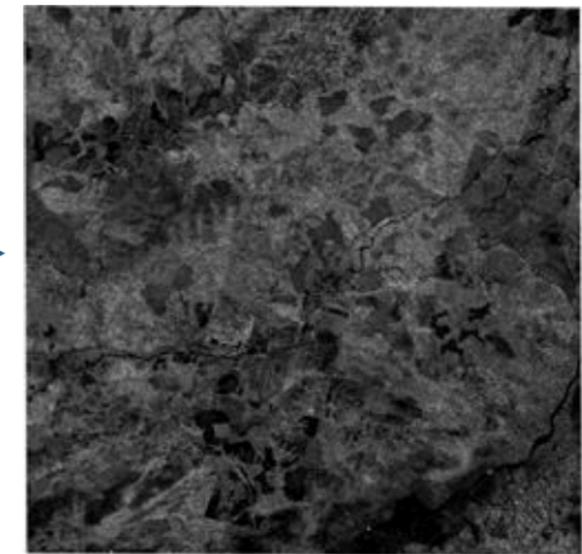




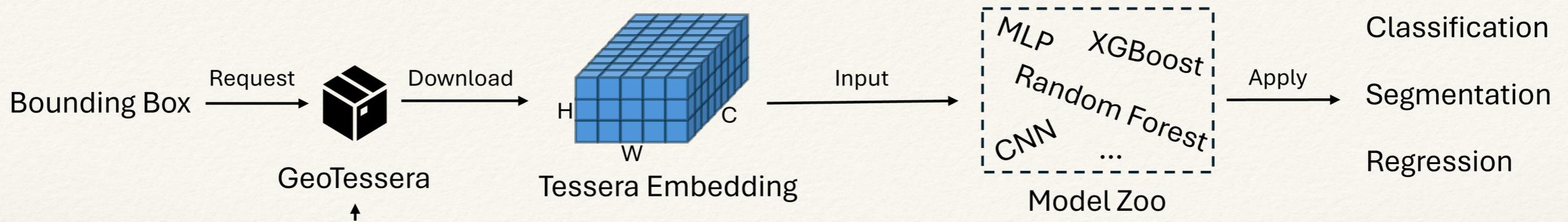
# TESSERA: same map tiles, lots of downstream uses



## Predicting Forest Height



# TESSERA: growing a worldwide user community



**1) Label Efficient:** just a few hints are enough to get started.

*lots of energy  
to train it once*

**TESSERA**

*open, global,  
reproducible, free*

*very little energy  
to re-use multiple times*

**Download  
Tiles for ROI**

*can run offline  
a mobile phone*

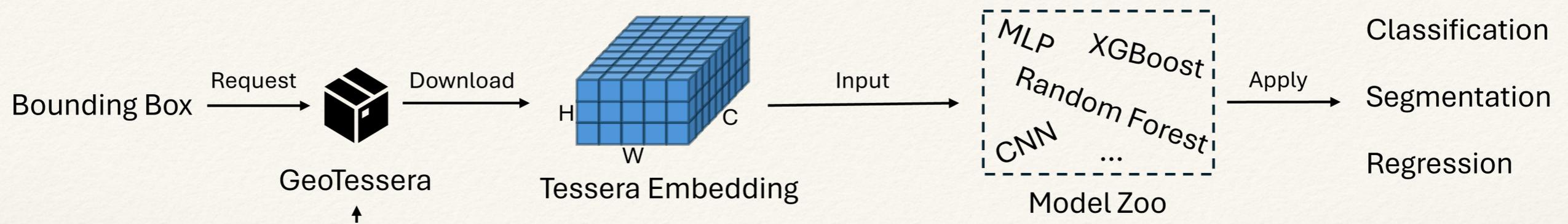
**Simple  
Model**

*no GPU  
required!*

**Validation  
Pipeline**

*fewer labels  
required*

# TESSERA: growing a worldwide user community



## 1) Label Efficient: just a few hints are enough to get started.



*lots of energy  
to train it once*

*very little energy  
to re-use multiple times*

**TESSERA**



**Download  
Tiles for ROI**



**Simple  
Model**



**Validation  
Pipeline**

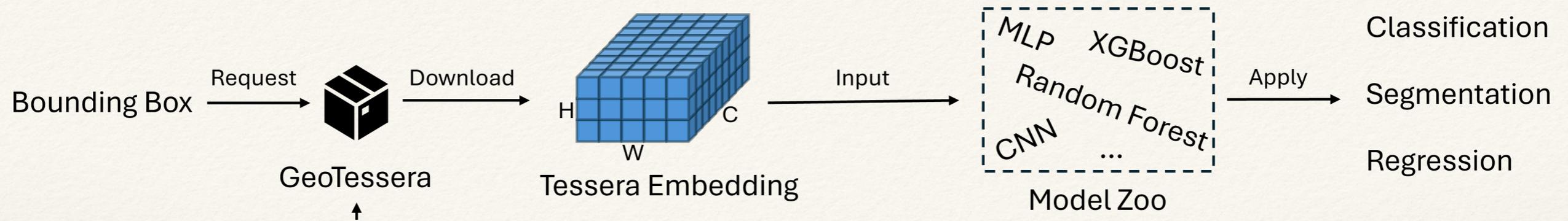
*open, global,  
reproducible, free*

*can run offline  
a mobile phone*

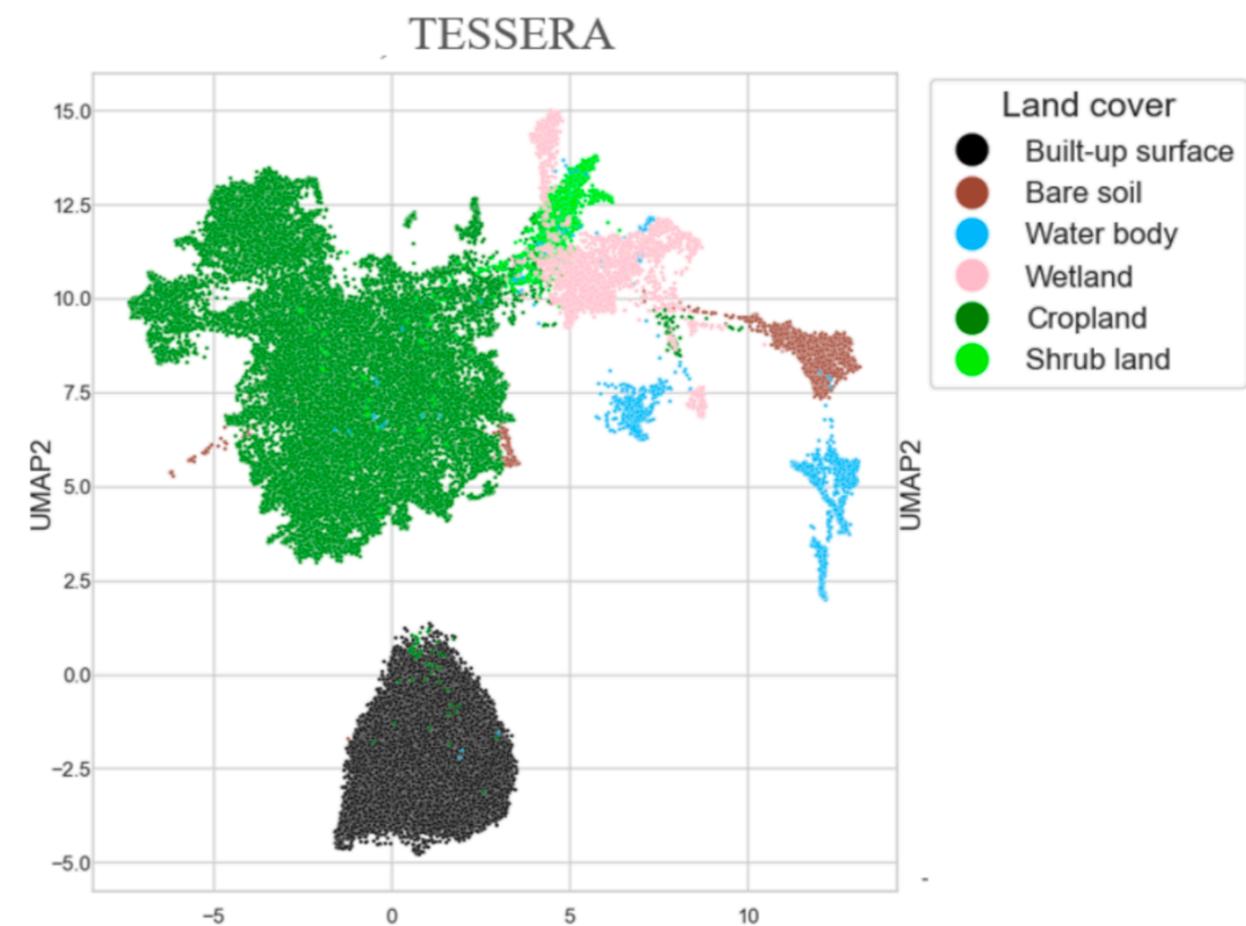
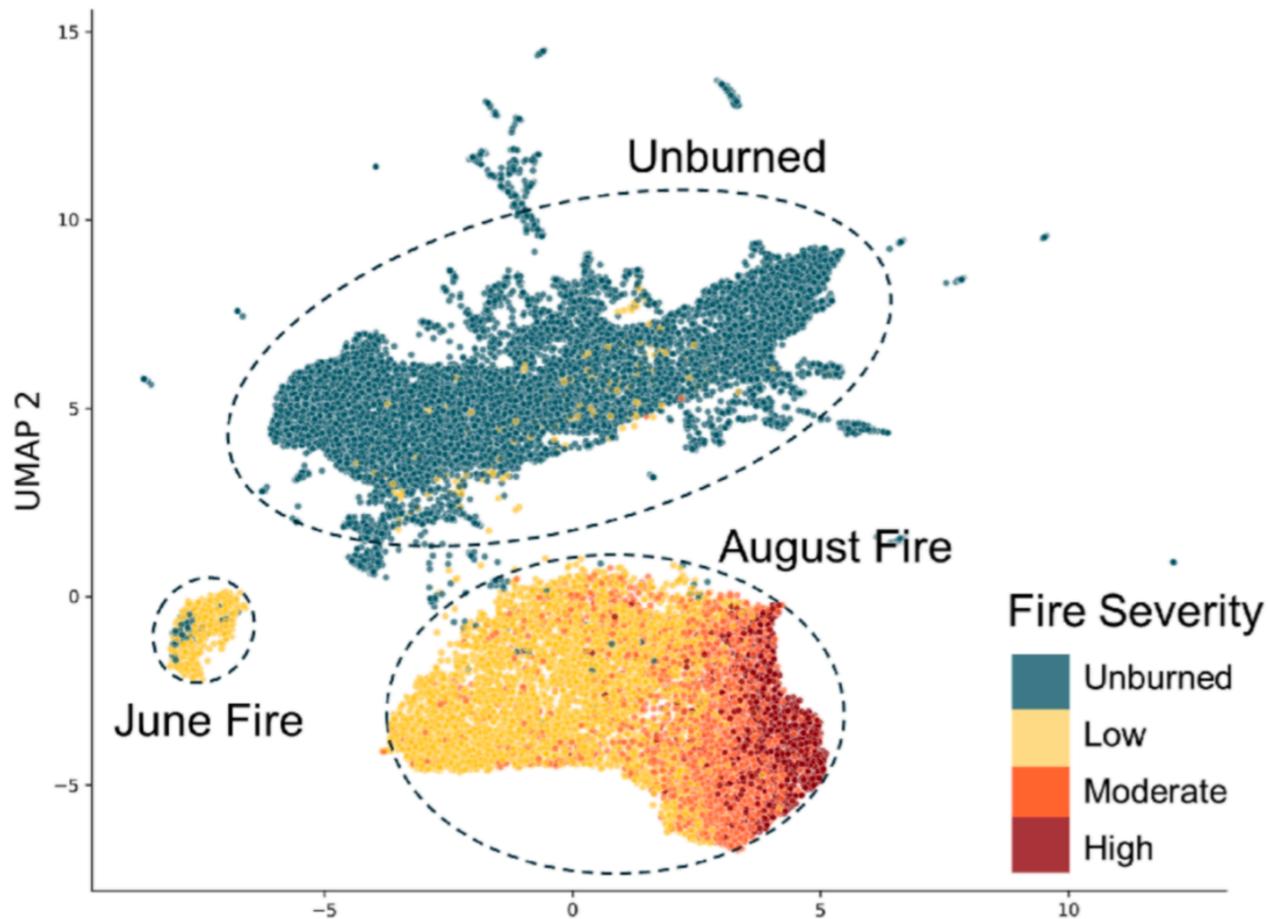
*no GPU  
required!*

*fewer labels  
required*

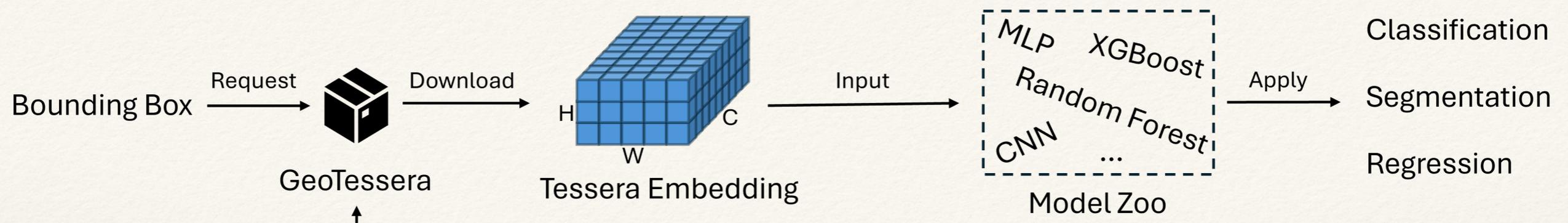
# TESSERA: growing a worldwide user community



- 1) Label Efficient:** just a few hints are enough to get started.
- 2) Emergent Learning:** embeddings cluster unexpectedly.



# TESSERA: growing a worldwide user community

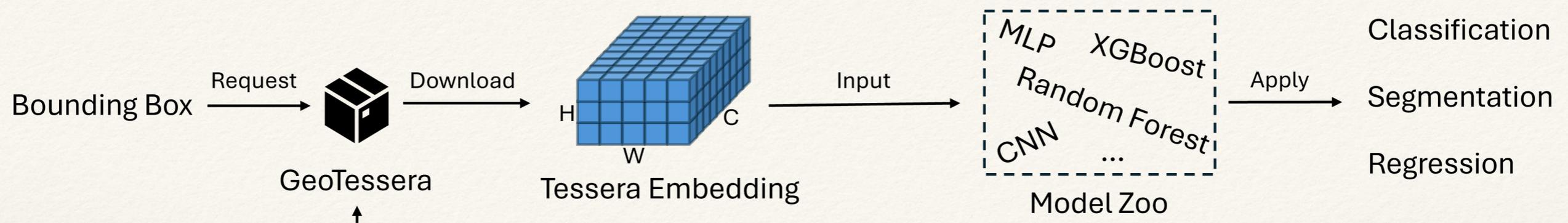


- 1) **Label Efficient:** just a few hints are enough to get started.
- 2) **Emergent Learning:** embeddings cluster unexpectedly.
- 3) **Global Model:** regional fine-tuning makes things worse!

The screenshot shows the GitHub interface for the repository 'ucam-eo / geotessera'. The 'Issues' tab is active, showing 20 open issues. The top issue is '[Embedding Request] Microsoft AI for Good / 2017-2025 / Kenya' with 2 comments. Other issues include requests for HU, UNODC, and Xiamen, China.

Issue Title	Author	Comments
[Embedding Request] Microsoft AI for Good / 2017-2025 / Kenya	satejsoman	2
[Embedding Request] HU / 2022 - 2025 / Zambezia Province (MOZ)	eichfussl	1
[Embedding Request] UNODC / 2023 2024 2025 / Andean region	caalbarracinb	1
[Embedding Request] Universität Trier / [2017, 2018, 2019, 2024] / Central England (Manchester)	franzschug	3
[Embedding Request] ORG / 2024 / Xiamen, China	dongGIS-geo	9

# TESSERA: growing a worldwide user community



- 1) **Label Efficient:** just a few hints are enough to get started.
- 2) **Emergent Learning:** embeddings cluster unexpectedly.
- 3) **Global Model:** regional fine-tuning makes things worse!

**Easy-to-use Python/OCaml libraries**

**No GPUs needed** for many simple tasks

**Growing community** of worldwide users helping

**Cloud platform independent** and on-prem / offline

**Sovereign training and inference** via torchgeo

**Coding agent friendly** with Claude skills



# How will TESSERA change the world?

v1 Can we identify activity anywhere?

Interactively!

v2 Can we spot changes in realtime?

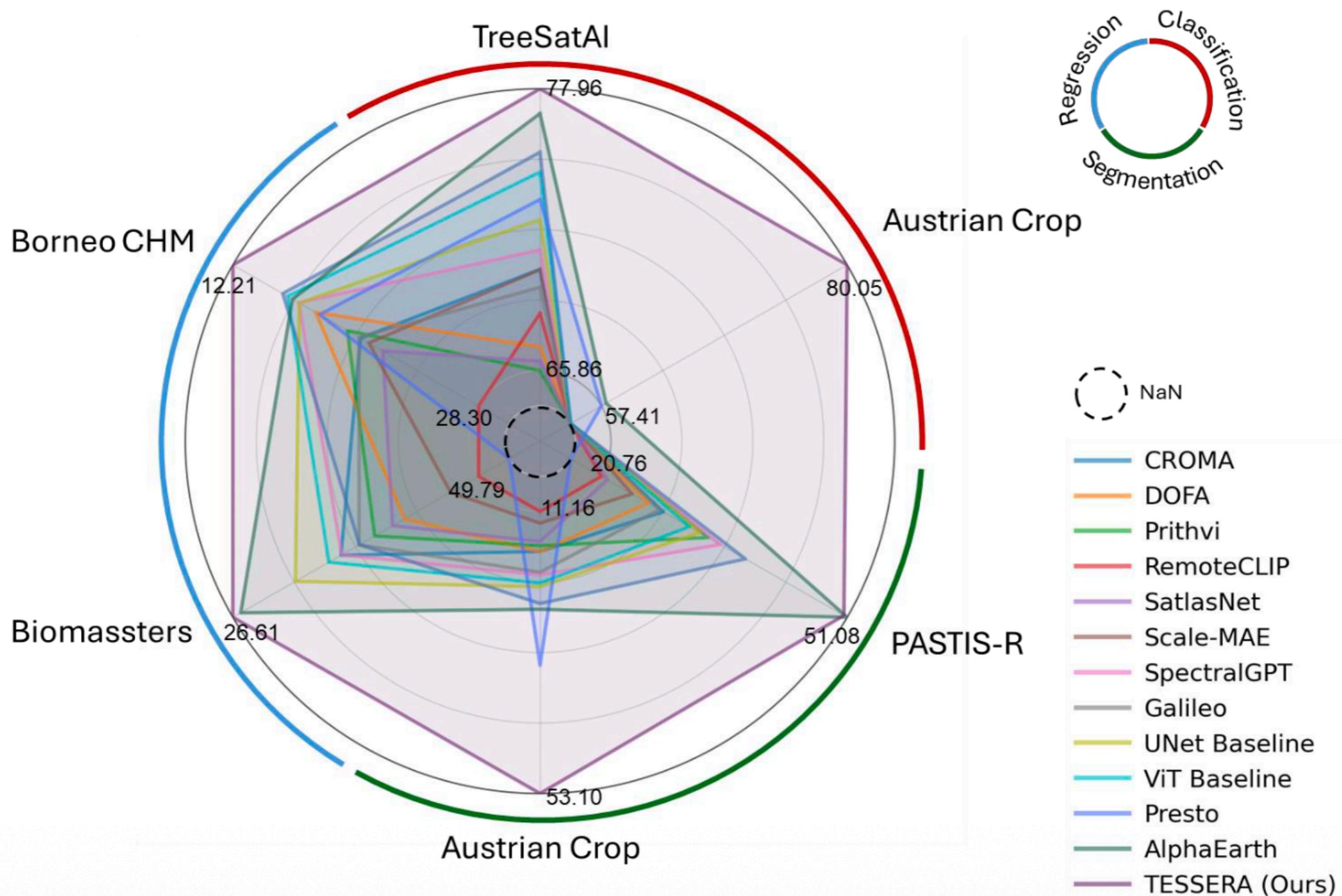
Mega scaling!

v3 Can we predictively model the planet?

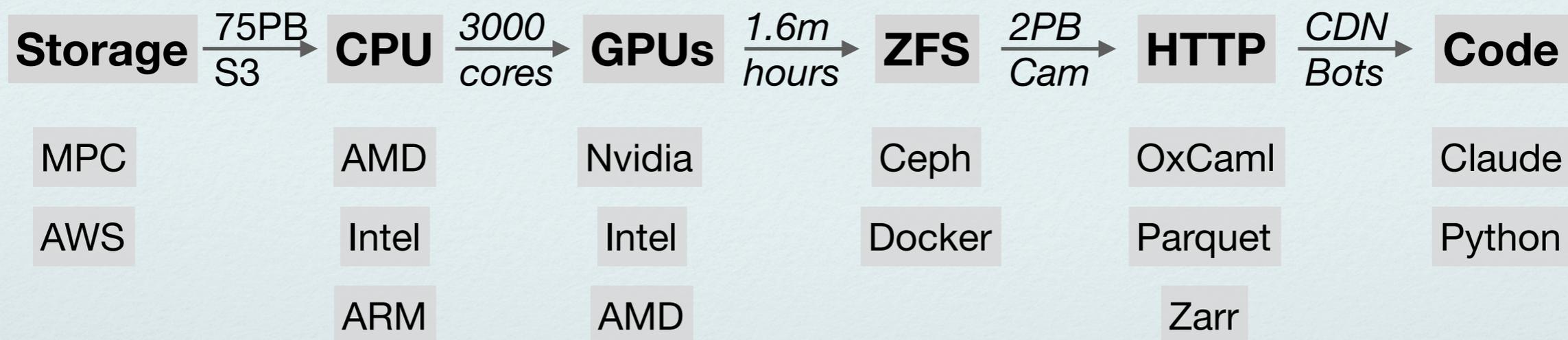
Emergent reasoning!

# Scaling the model size: v1 to v2 (at a University)

v1 is currently SOTA, but it can be even better!



# Scaling the model size: v1 to v2 (at a University)

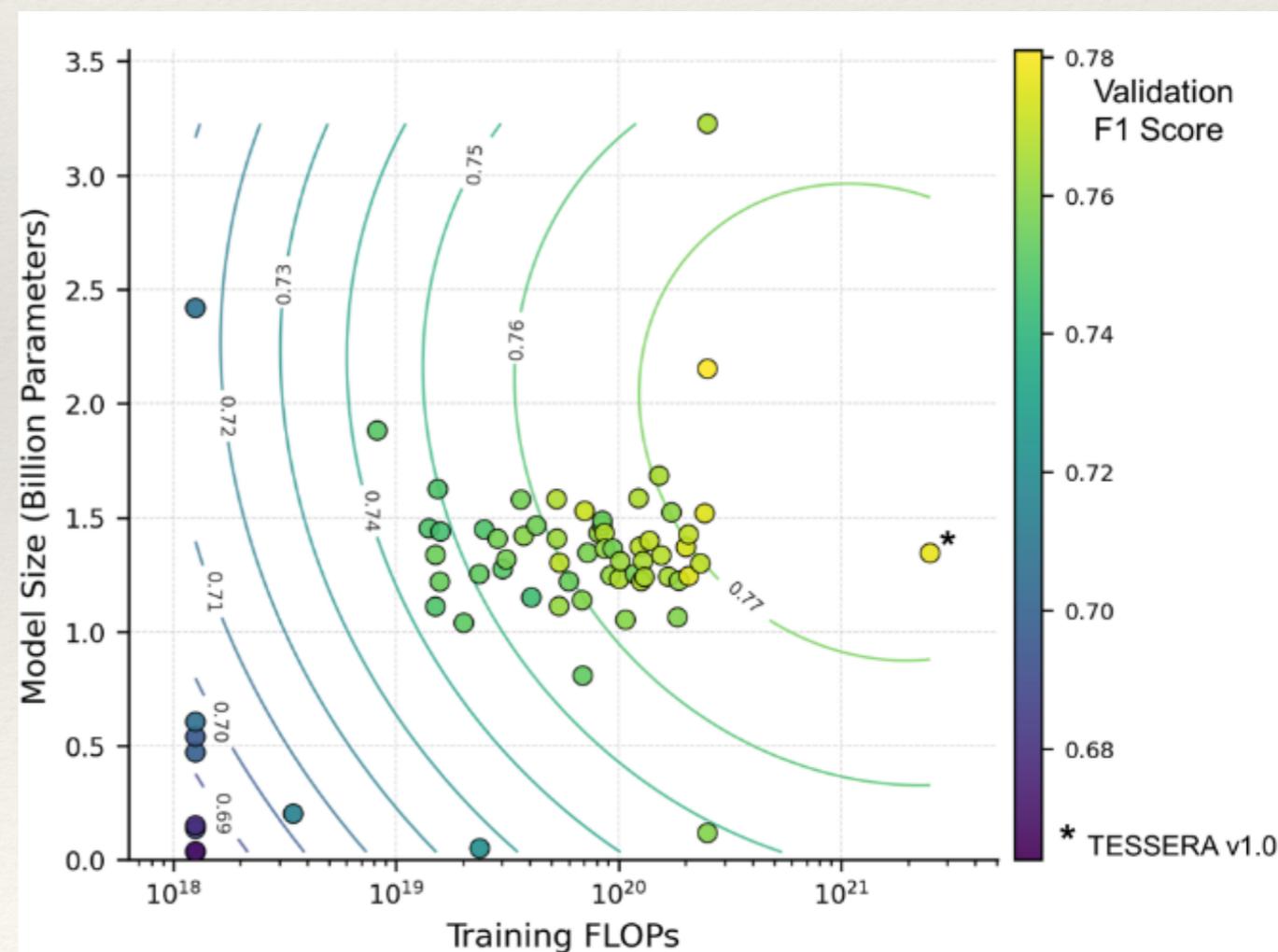


v1 took 6,200 GPU hours on AMD MI300Xs and used 1/400th of a single year.

We have vast amounts of training data not being used in v1!

**Scaling experiments show log(n) improvements in v2.**

Going from 1bn- $\rightarrow$ 10bn parameters:  
**if we can find GPUs+storage!**



# Spotting planetary changes in real time

Unlabelled Self-Supervised Learning



*Petabytes*



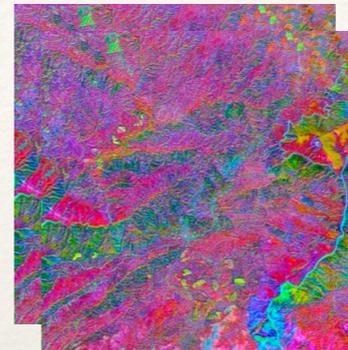
**Satellite Data**

*Terabytes*



**Model Pretraining**

*Gigabytes*

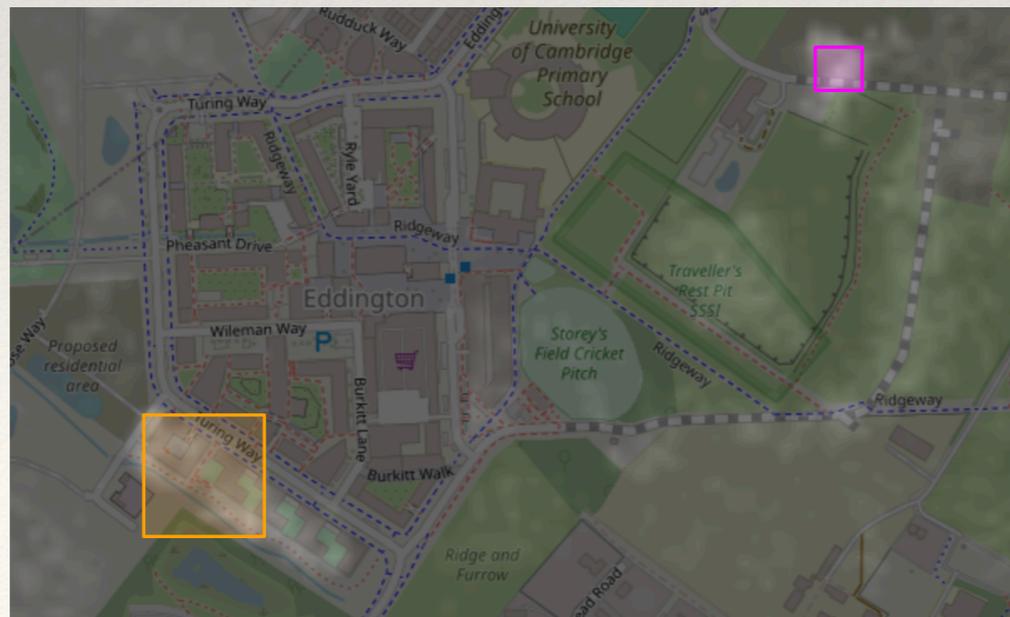


**Annual Map Tiles**

**Train model to predict next year**

2017  
2018  
2019

**Unsupervised!**



# How will TESSERA change the world?

v1 Can we identify activity anywhere?

Interactively!

v2 Can we spot changes in realtime?

Mega scaling!

**v3 Can we predictively model the planet?**

Emergent reasoning!

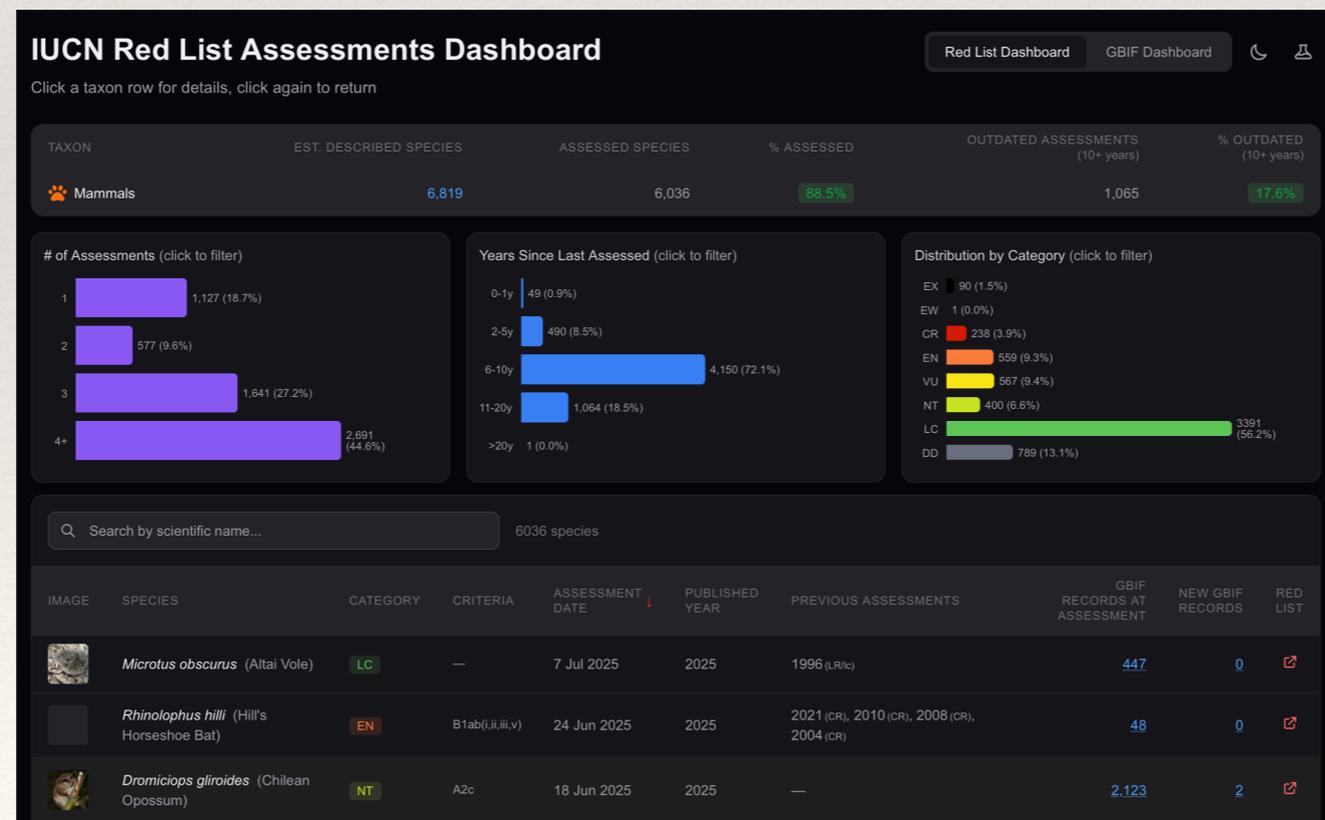
# What if we could get a billion geolabels?



We've spent two years building up a giant global label database from public and private datasets, using LLMs and coding agents.

Global Biodiversity Information Facility (3.6bn observations).  
IUCN Red List of Endangered Species, iNaturalist, OpenStreetMap, ...  
15m academic papers fulltexts via the University Library

The debug dashboards we built for this are currently accidentally going viral in the ecology community!



# But how do do LIVE programming?!



Scientists continuously collect new data and refine modeling and analysis approaches. But models and analyses as reported in publication and scientific artifact repositories are frozen in time.

Not good enough.

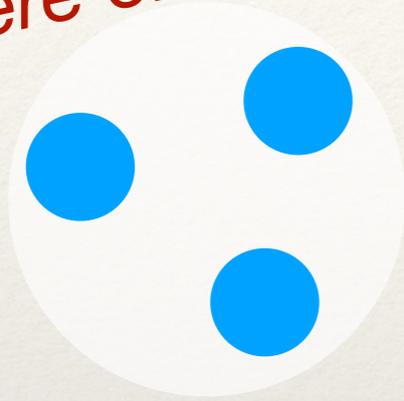
## Design + impl. criteria for Fairground's *orchestration language*

**Fairground** is formally a **single massive live program** being collaboratively edited by thousands of contributors (humans + safely sandboxed AIs).

1. To support **interoperability**, we need a robust and customizable **foreign function interface (FFI)**, **foreign data interface (FDI)**, and **type system**.
2. To support **reproducibility** and **liveness**, we need to **control dependencies and side effects**.
3. To support **liveness at scale**, we need **incremental + parallel + distributed + gap-free type checking and program execution**.

# A Hilbert curve to describe a planet

"Hedgehogs were spotted here on Friday!"  
x6 billion



Observational



# A DAG to describe a planet

"Hedgehogs were spotted here on Friday!"  
x6 billion



**Observational**

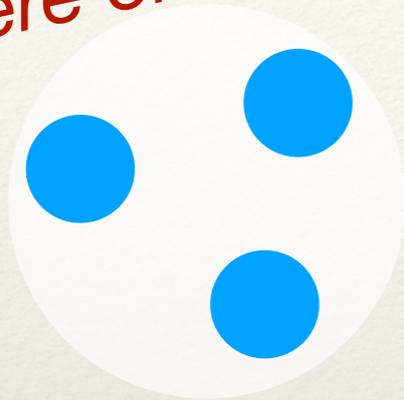
"Hedgehogs thrive in brambles and hedgerows in western Europe" x15m papers

**Causal**

Association  
Intervention  
Counterfactual

# A program to describe a planet

"Hedgehogs were spotted here on Friday!"  
x6 billion



**Observational**

"Hedgehogs thrive in brambles and hedgerows in western Europe" x15m papers

**Causal**

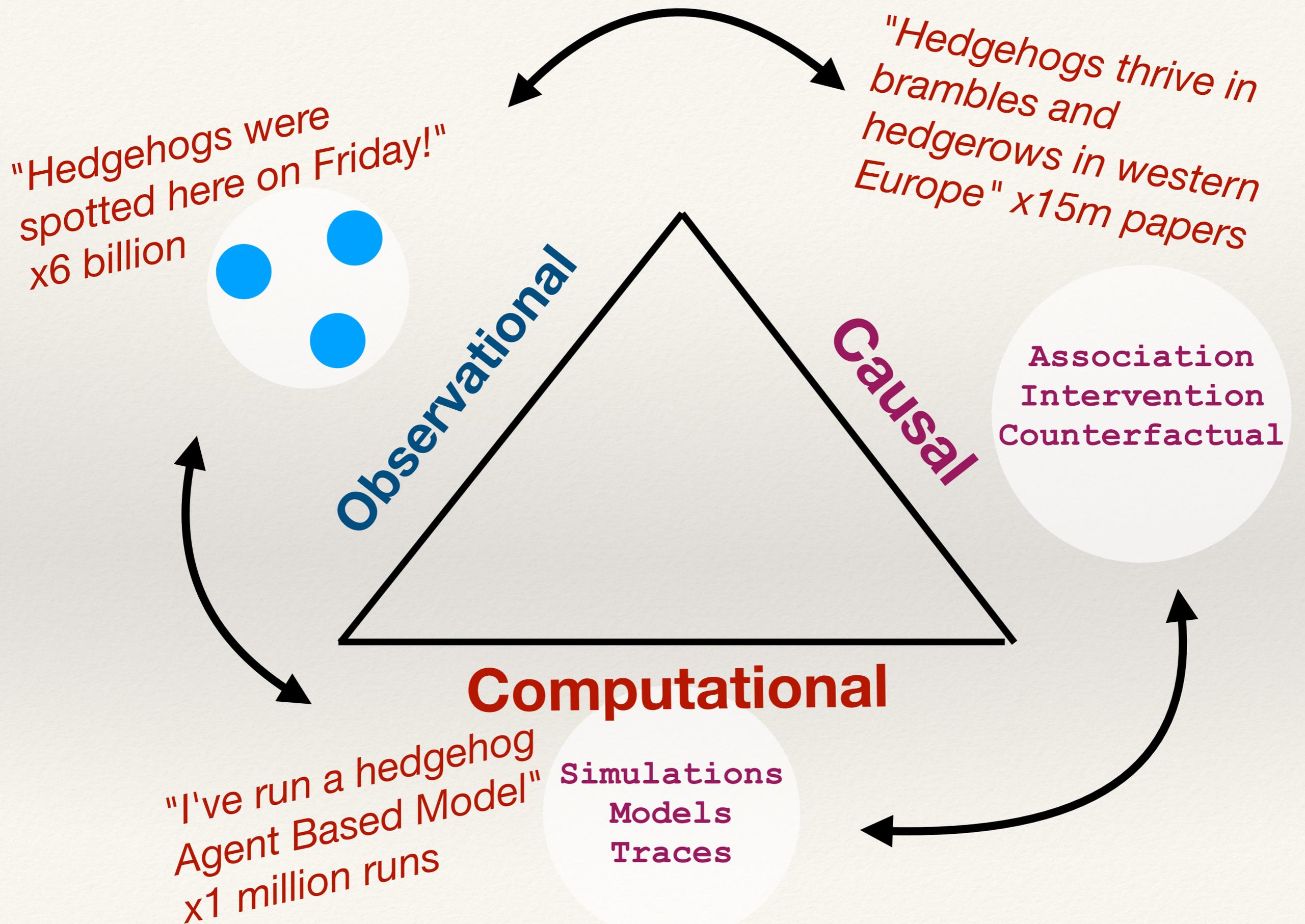
Association  
Intervention  
Counterfactual

**Computational**

"I've run a hedgehog Agent Based Model"  
x1 million runs

Simulations  
Models  
Traces

# A (functional?) program to describe a planet



# The GAIA foundation model of the planet

