Multimodal AI: From isolated information to Unified Understanding

Dino Ienco (dino.ienco@inrae.fr)
Research Director, PhD, HDR
INRAE, INRIA, EVERGREEN, UMR TETIS
Evergreen team - https://team.inria.fr/evergreen/

Outline

Introduction

- Introduction
- Multimodal AI model
- Field evolution

Outline

Introduction

- Introduction
- Multimodal AI model
- Field evolution

Some Technical details

- Model architectures
- Supervised vs Self-Supervised
- Foundation Model

Outline

Introduction

- Introduction
- Multimodal AI model
- Field evolution

Some Technical details

- Model architectures
- Supervised vs Self-Supervised
- Foundation Model

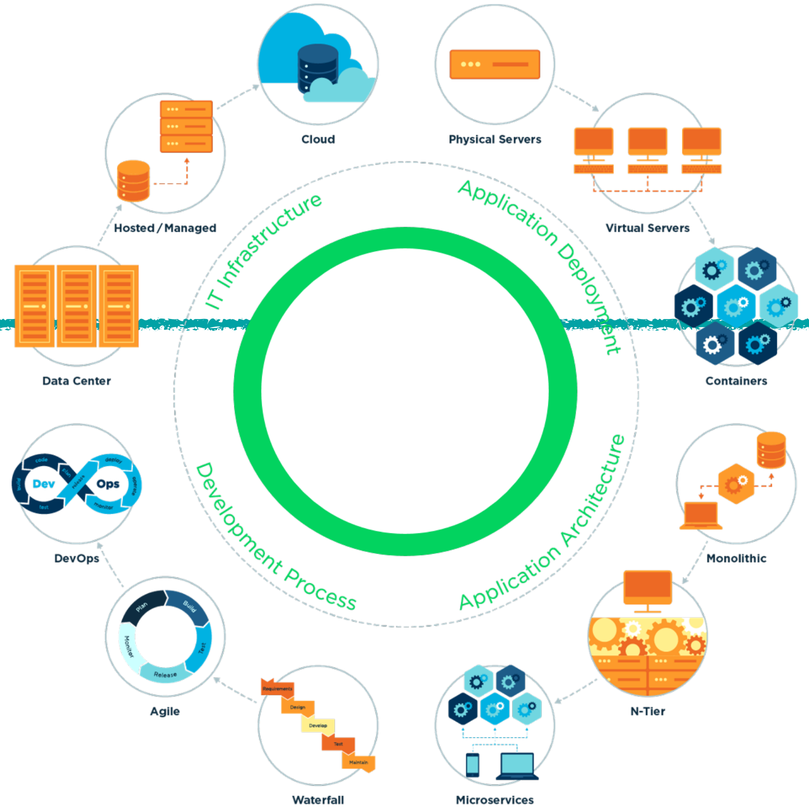
Challenges

- Research Challenges
- Societal Challenges

Introduction

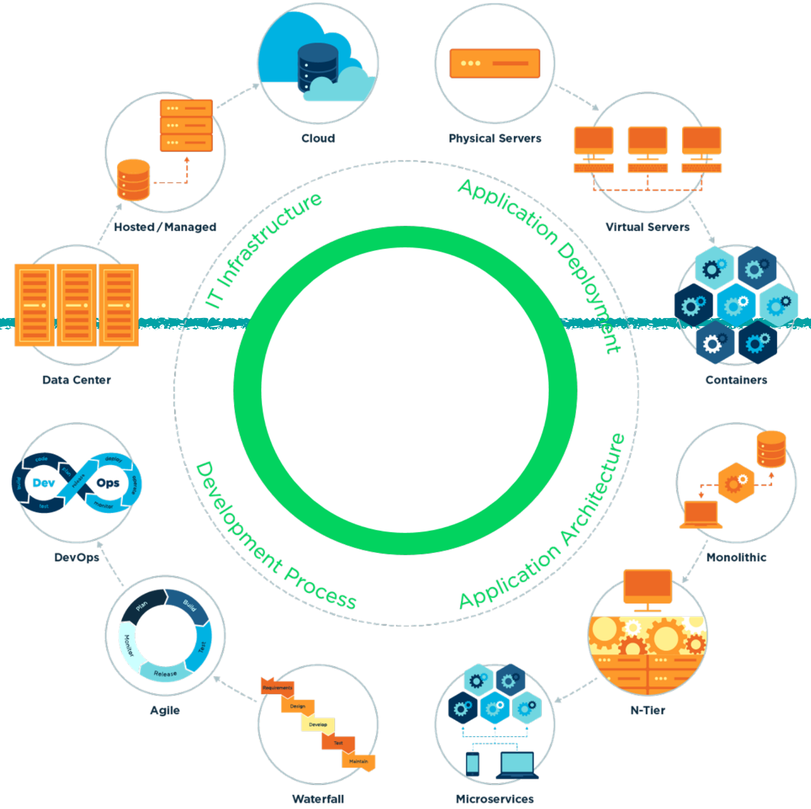
Introduction

- A **large diversity** of devices is continually sensing the environment.
- This results in a plethora of data collected under **different modalities** with complementary infos.



Introduction

- A **large diversity** of devices is continually sensing the environment.
- This results in a plethora of data collected under **different modalities** with complementary infos.
- The abundance of information poses **new challenges for** data-driven **AI (ML and CV)** techniques to, for instance:
  - Cope with different tasks (classification, understanding, generation, …)
  - Generalize to unseen data (or task) not seeing during training
  - Adapt to changing in evolving data



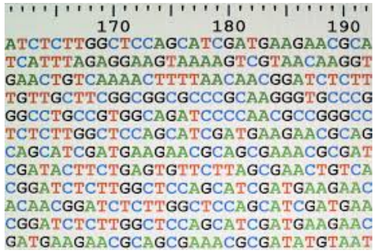
MULTIMODAL DATA

Introduction

Images data

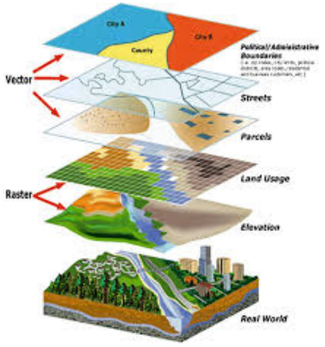
Sound data

Genomic data

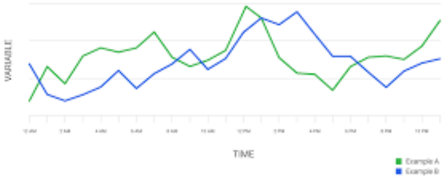


Text Data

Geospatial Data



Time Series data

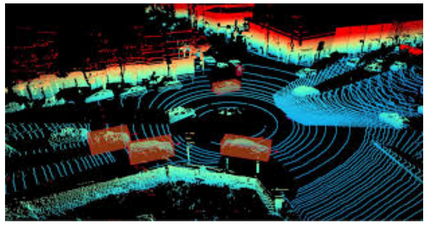


Tabular data

Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146

Social Media

Point Cloud (3D) data



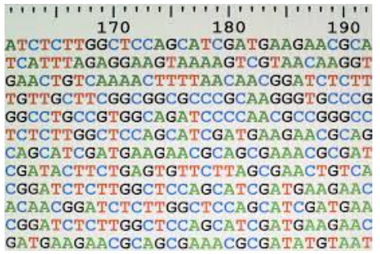
Video data

Introduction

Images data

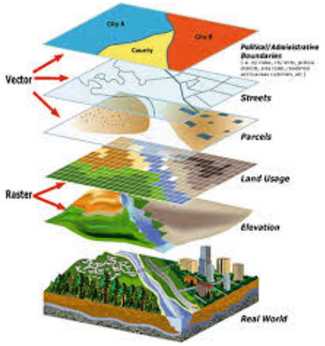
Sound data

Genomic data

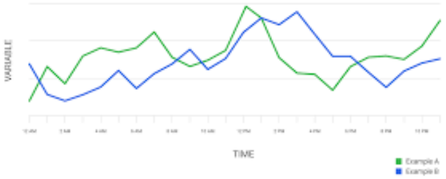


Text Data

Geospatial Data



Time Series data

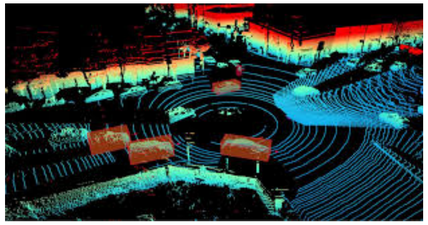


Tabular data

Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146

Social Media

Point Cloud (3D) data



Video data

Domains

- Medical
- Biological
- Remote Sensing
- Finance
- Human-Computer Interaction
- Robotics
- Autonomous driving

Multimodal AI model

Model → OUTPUT

Model → OUTPUT

Model → OUTPUT

Model → OUTPUT

Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146

Model → OUTPUT

Multimodal AI model

Model → OUTPUT

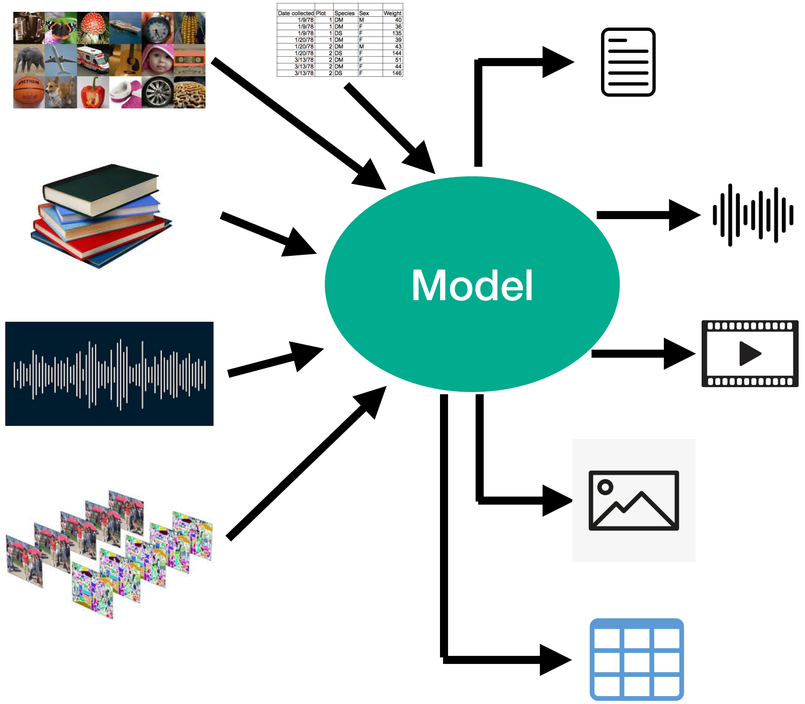
Model → OUTPUT

Model → OUTPUT

Model → OUTPUT

Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146

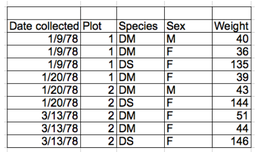
Model → OUTPUT

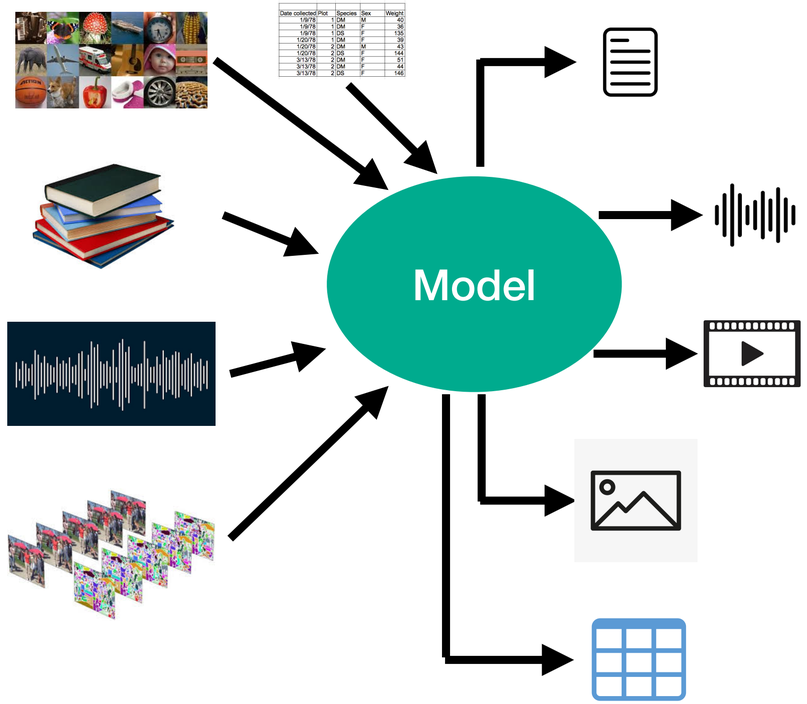


A multimodal AI model learns a **shared representation space** where modalities are aligned.
The model understands cross-modal relationships: image captioning, visual question answering, or text-to-image generation.

Multimodal AI model

Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146





A multimodal AI model learns a **shared representation space** where modalities are aligned.
The model understands cross-modal relationships: image captioning, visual question answering, or text-to-image generation.

Joint learning on data from multiple modalities

Unified model
for multiple type of data

Inter-Modality
“reasoning / understand.”

Generally based on **Transformer** archi.

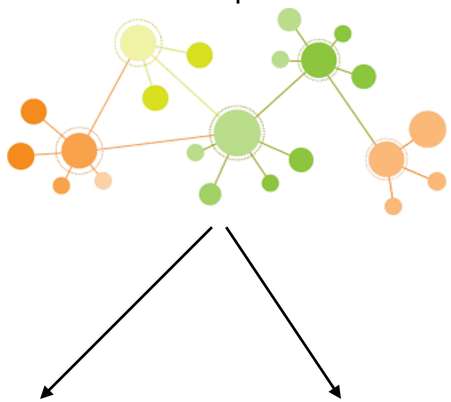
Trained on **massive amount** of multimodal **data**

Field evolution (i)

Before 2000

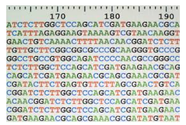
Each kind of data is isolated but connection between data and data structure

Graph



Social Media

Genomic data



Field evolution (i)

Before 2000

Each kind of data is isolated but connection between data and data structure

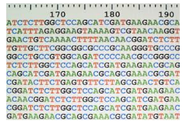
2000 - 2010

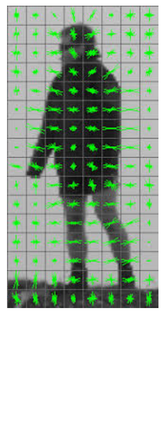
Specific data pre-processing to use the same ML methods (e.g. SVM)

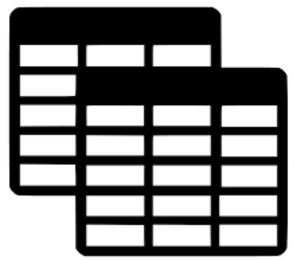
Graph

Social Media

Genomic data







Vector representation

Field evolution (i)

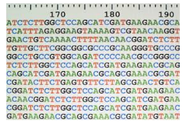
Before 2000

Each kind of data is isolated but connection between data and data structure

Graph

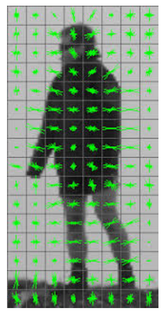
Social Media

Genomic data

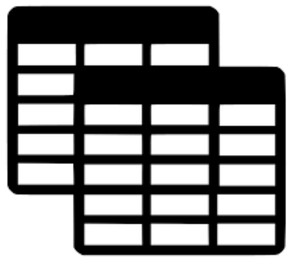


2000 - 2010

Specific data pre-processing to use the same ML methods (e.g. SVM)



Vector representation



2010 - 2015

Raise of Neural Network approaches (Deep Learning) and development of per modality architecture

Hand-Crafted Features → **Trainable Classifier (SVM, RF, NB, …)**

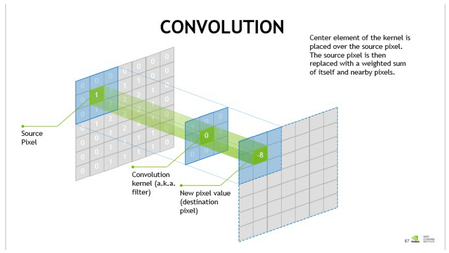
Trainable Feature Extractor → **Trainable Classifier**

Field evolution (ii)

2015 - 2020

Cross-fertilisation, in terms of approaches, across modality analysis

Convolutional Neural network (from Image)



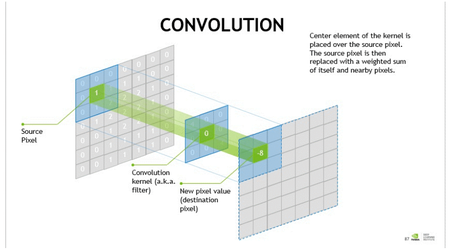
TEXT **GRAPH** **TIME SERIES**

Field evolution (ii)

2015 - 2020

Cross-fertilisation, in terms of approaches, across modality analysis

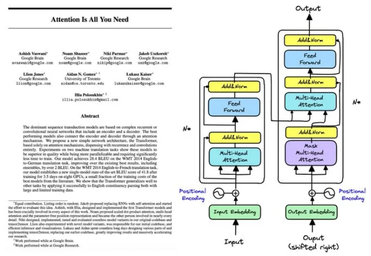
Convolutional Neural network (from Image)



TEXT **GRAPH** **TIME SERIES**

2017

Transformer model was proposed for NLP



Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Lukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* † illia.polosukhin@gmail.com

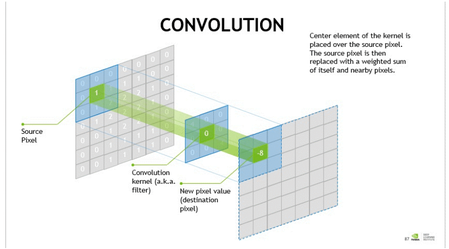
Abstract

Field evolution (ii)

2015 - 2020

Cross-fertilisation, in terms of approaches, across modality analysis

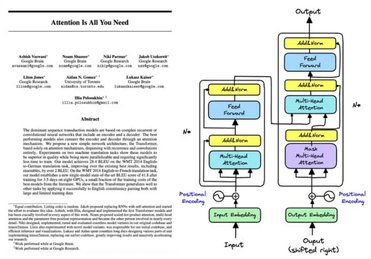
Convolutional Neural network (from Image)



TEXT **GRAPH** **TIME SERIES**

2017

Transformer model was proposed for NLP



2018

First Language models as known today (based on Transformers)

GPT-1 OpenAI

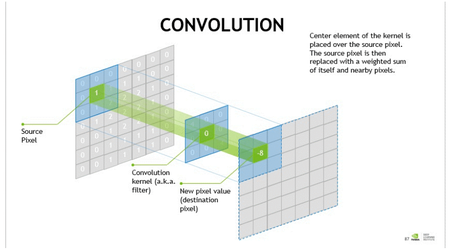
(The grandfather of ChatGPT)

Field evolution (ii)

2015 - 2020

Cross-fertilisation, in terms of approaches, across modality analysis

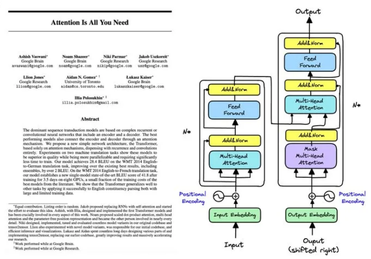
Convolutional Neural network (from Image)



TEXT **GRAPH** **TIME SERIES**

2017

Transformer model was proposed for NLP



2018

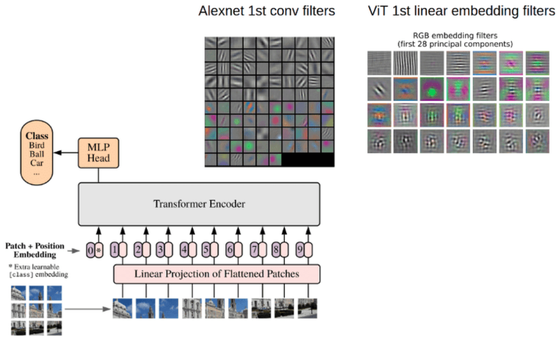
First Language models as known today (based on Transformers)

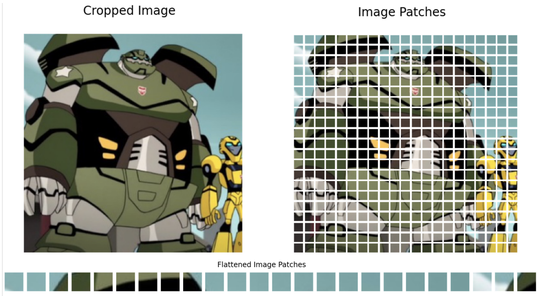
GPT-1 OpenAI

(The grandfather of ChatGPT)

2020

Vision Transformers

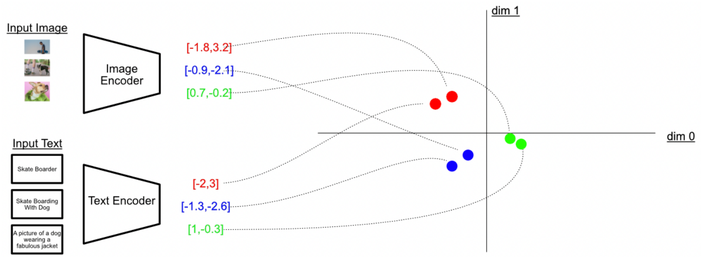




Field evolution (iii)

2021

VLM: Contrastive Language-Image Pre-Training (CLIP)



2021

Text2Image: DALL-E

2021

Foundation Model (FM) concept

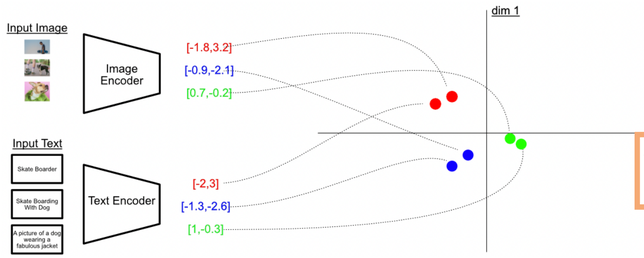
2021

Unimodal and Multimodal FM

Field evolution (iii)

2021

VLM: Contrastive Language-Image Pre-Training (CLIP)



2021

Text2Image: DALL-E

2021

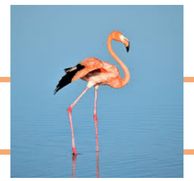
Foundation Model (FM) concept

2021

Unimodal and Multimodal FM

2022

VLM: Flamingo



2022

LLM: ChatGPT

ChatGPT

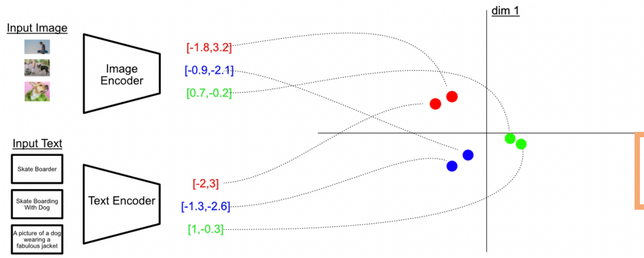
2023

VLM: LLAVA

Field evolution (iii)

2021

VLM: Contrastive Language-Image Pre-Training (CLIP)



2021

Text2Image: DALL-E

2021

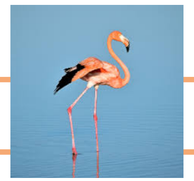
Foundation Model (FM) concept

2021

Unimodal and Multimodal FM

2022

VLM: Flamingo



2022

LLM: ChatGPT

ChatGPT

2023

VLM: LLAVA

2023

VLM: GPT4-V

2023

VLM: ImageBind (6 Modalities)

2024-today

Agentic AI, Diffusion Model, Omni & World Models, …

Claude

Qwen3-Omni

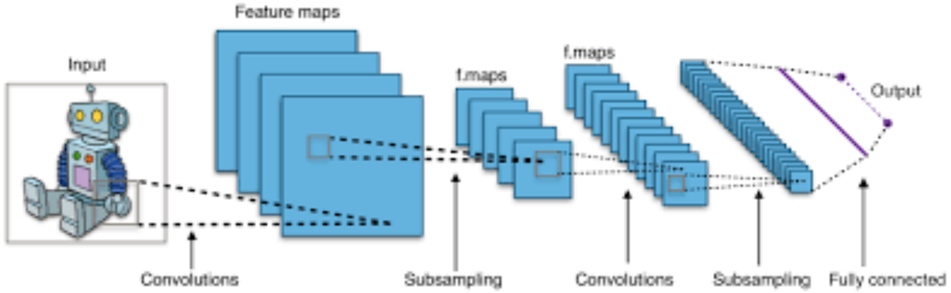
GPT-4o

Gemini

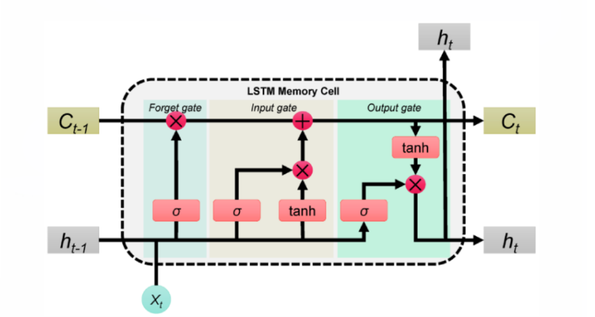
Some Technical Details

Unimodal architectures

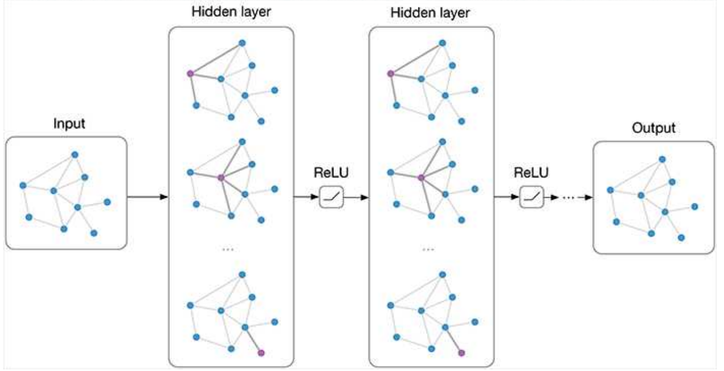
2010 - 2015



Convolutional Neural Network (**CNN**)
for image data



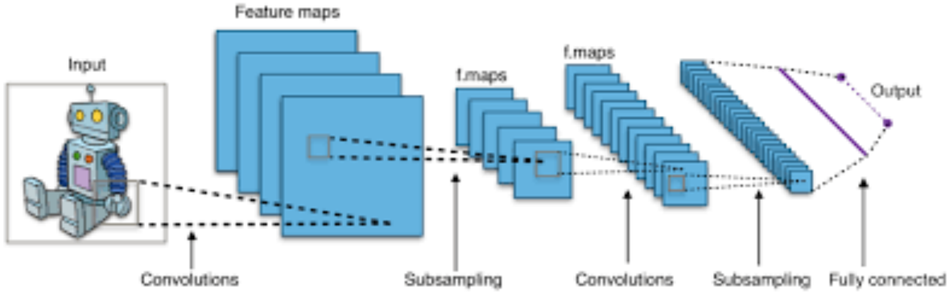
Recurrent Neural Network (**RNN**)
for time series and textual data



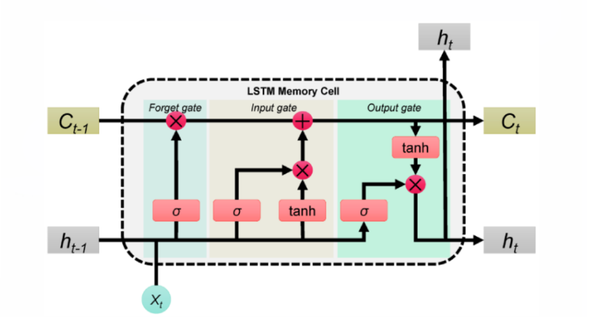
Graph Neural Network (**GNN**)
for structured data

Unimodal architectures

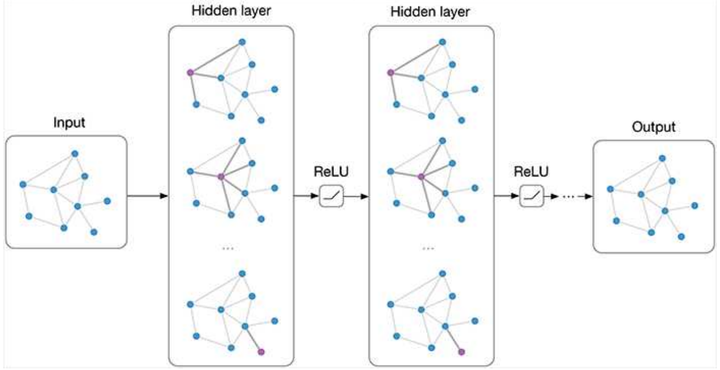
2010 - 2015



Convolutional Neural Network (**CNN**)
for image data



Recurrent Neural Network (**RNN**)
for time series and textual data

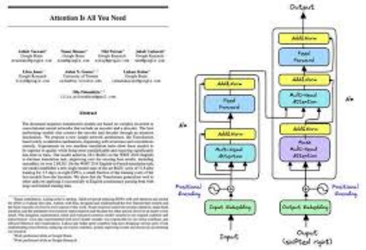


Graph Neural Network (**GNN**)
for structured data

2015 - 2020

Convolutional architecture starts to
be explored beyond image analysis:
- Text data
- Time Series Data
- Graph Data

CNN vs (Vision) Transformers

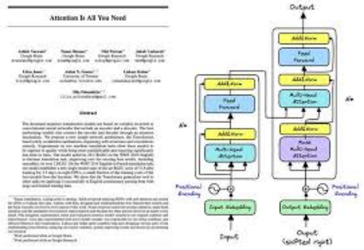


2017

Inductive Bias

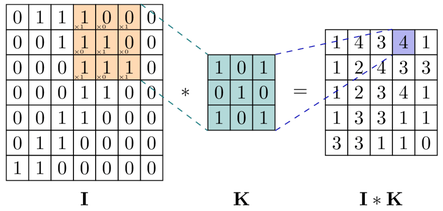
- Set of assumptions a machine learning model makes about the underlying relationship in input data.
- The apriori underlying a specific machine learning model

CNN vs (Vision) Transformers



2017

Inductive Bias

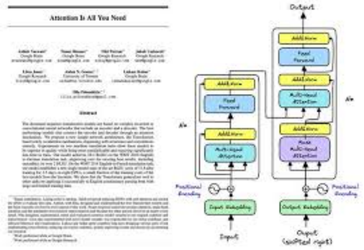


- Set of assumptions a machine learning model makes about the underlying relationship in input data.
- The apriori underlying a specific machine learning model

Convolutional Neural Networks:

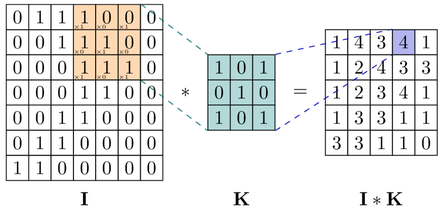
- Inspired by signal processing theory
- Learn local features from the signal
- Organize hierarchically the content of a signal
- Structured inductive bias to guide model learning

CNN vs (Vision) Transformers

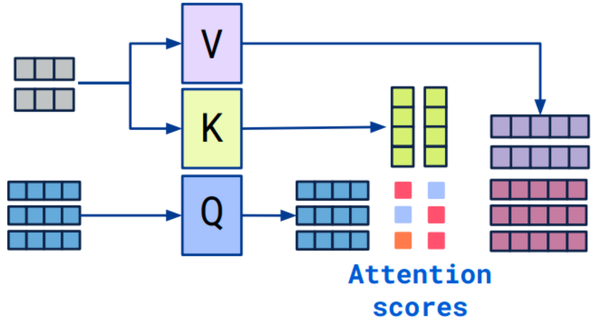


2017

Inductive Bias



- Set of assumptions a machine learning model makes about the underlying relationship in input data.
- The apriori underlying a specific machine learning model



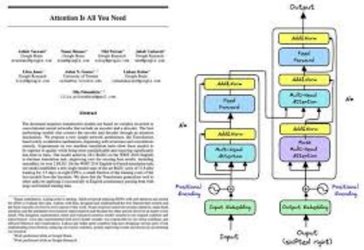
Convolutional Neural Networks:

- Inspired by signal processing theory
- Learn local features from the signal
- Organize hierarchically the content of a signal
- Structured inductive bias to guide model learning

Transformer:

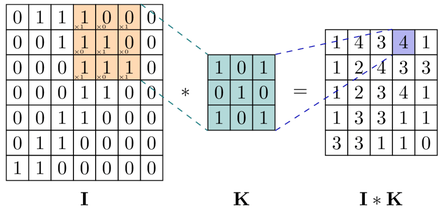
- Guided by empirical results in the DL era
- Allows early interactions between all the information in the signal
- Reduced inductive bias (apriori) related only to positional embedding and attention

CNN vs (Vision) Transformers

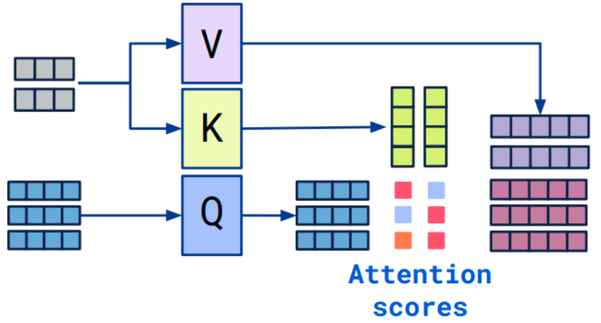


2017

Inductive Bias



- Set of assumptions a machine learning model makes about the underlying relationship in input data.
- The apriori underlying a specific machine learning model



Convolutional Neural Networks:

- Inspired by signal processing theory
- Learn local features from the signal
- Organize hierarchically the content of a signal
- Structured inductive bias to guide model learning

Transformer:

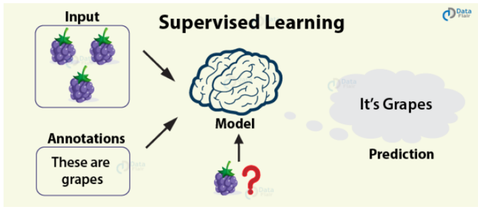
- Guided by empirical results in the DL era
- Allows early interactions between all the information in the signal
- Reduced inductive bias (apriori) related only to positional embedding and attention

+ Inductive bias → Data Efficient

- Inductive bias → Data Hungry

Supervised vs Self-Supervised

Supervised Learning
Characteristics:



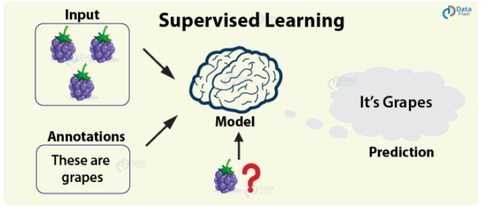
- Learn **relationships between input** data **and output** class
- Require **annotated data**
- Specific to a **particular task**

Output:
A (classification) **specialised model** for a particular task and problem

Supervised vs Self-Supervised

Supervised Learning

Characteristics:

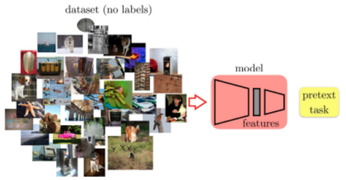


- Learn **relationships between input** data **and output** class
- Require **annotated data**
- Specific to a **particular task**

Output:

A (classification) **specialised model** for a particular task and problem

Self-Supervised Learning



Characteristics:

- Learn **relationships inside data**
- **No need of annotation** (aligned data for multimodal scenario)
- Target **general purpose** model

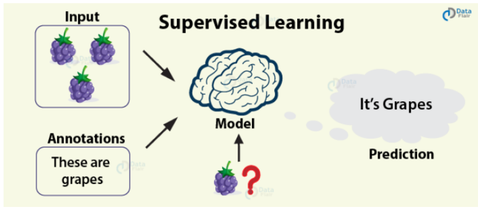
Output:

A **general purpose model** that can be successively used as it is or specialised towards a particular problem.

Supervised vs Self-Supervised

Supervised Learning

Characteristics:

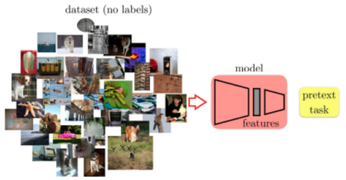


- Learn **relationships between input** data **and output** class
- Require **annotated data**
- Specific to a **particular task**

Output:

A (classification) **specialised model** for a particular task and problem

Self-Supervised Learning



Characteristics:

- Learn **relationships inside data**
- **No need of annotation** (aligned data for multimodal scenario)
- Target **general purpose** model

Output:

A **general purpose model** that can be successively used as it is or specialised towards a particular problem.

		Supervised (ResNet 101)	**Self-Supervised** (CLIP Vit-L)
Imagenet		76.2%	76.2%
ObjectNet		32.6%	72.3%

Supervised vs Self-Supervised

Supervised Learning

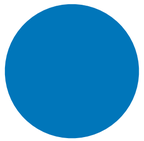
Self-Supervised Learning

Supervised vs Self-Supervised

Model Size

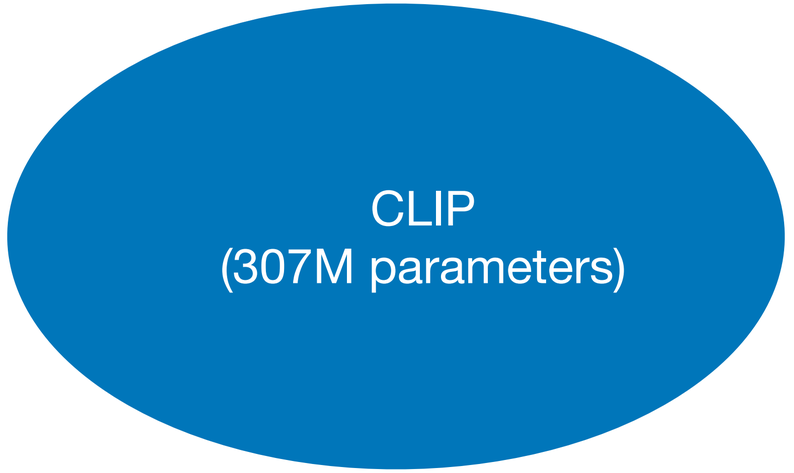
Supervised Learning

44.5M parameters

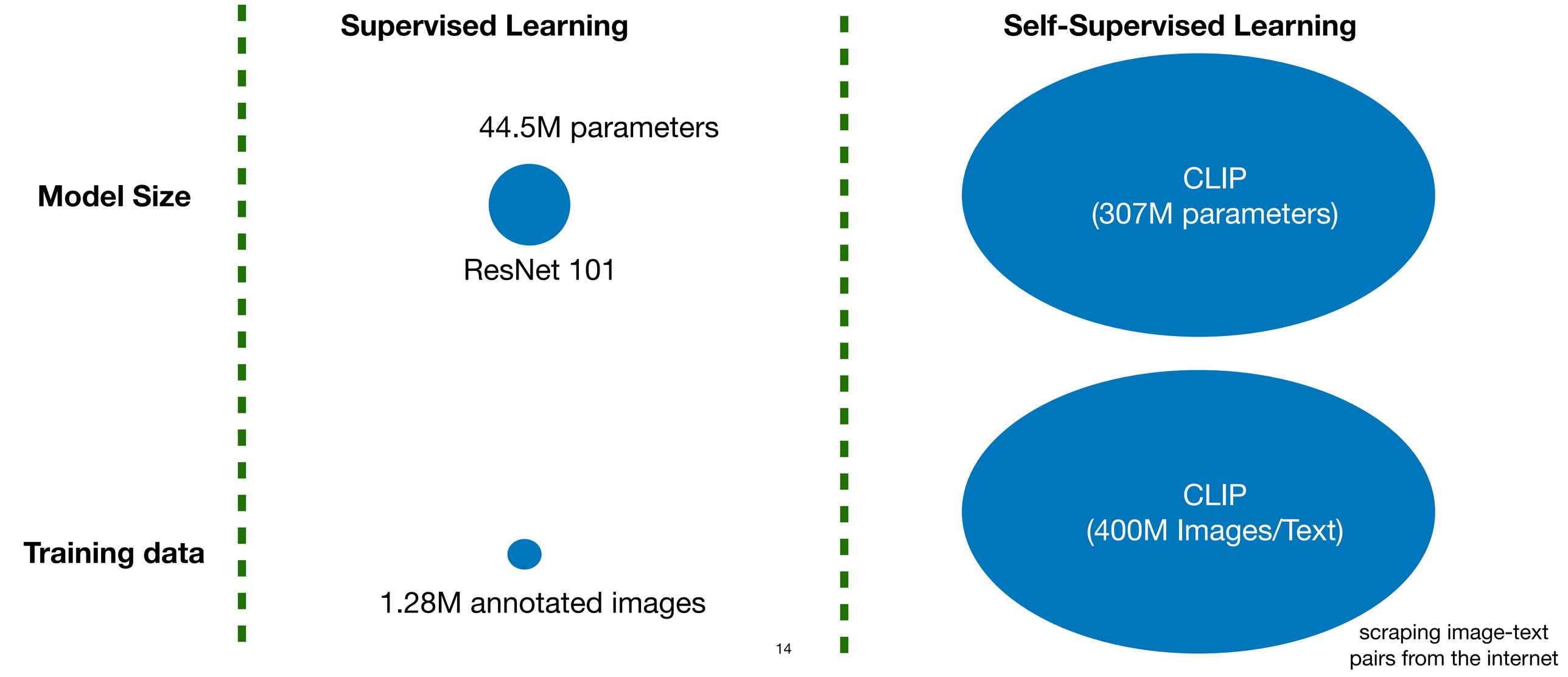


ResNet 101

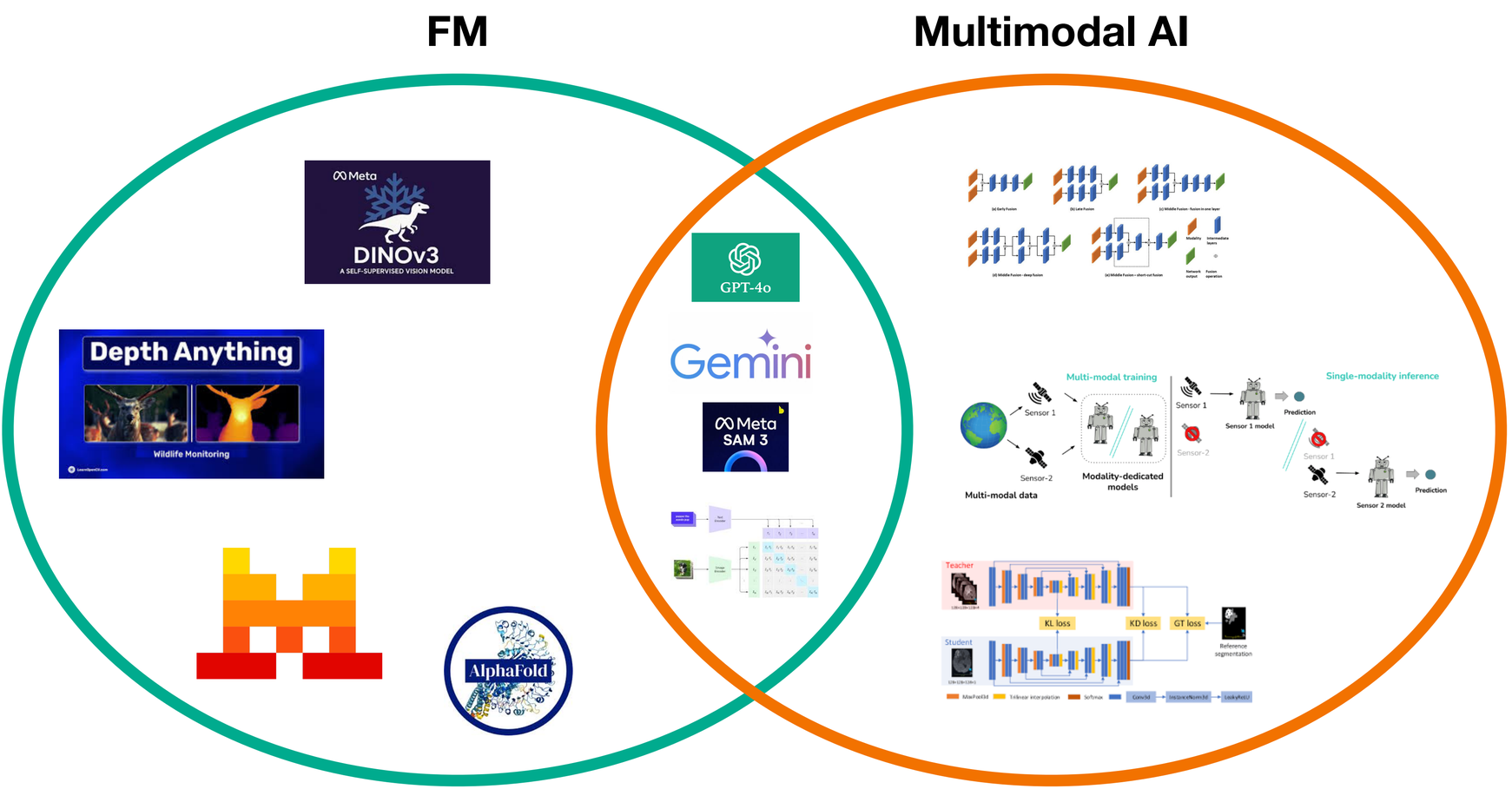
Self-Supervised Learning



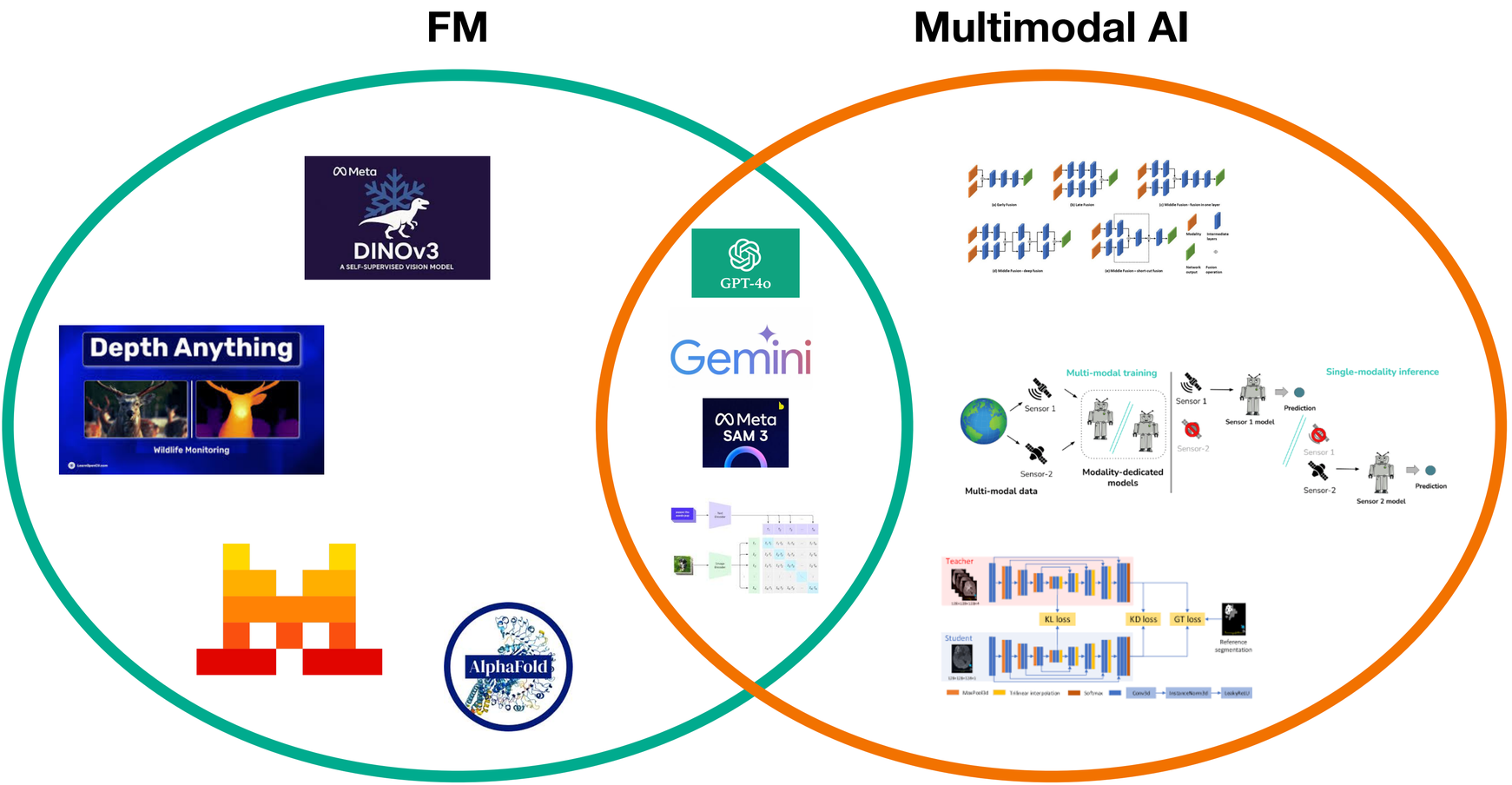
Supervised vs Self-Supervised



Foundational Model (FM)



Foundational Model (FM)



Foundation Models (FM):

- **Large-scale pre-trained** models
- Can be **unimodal** (GPT-3, BERT) or **multimodal** (GPT-4V, Gemini)
- Designed for **general purposes**, adaptable to specific tasks

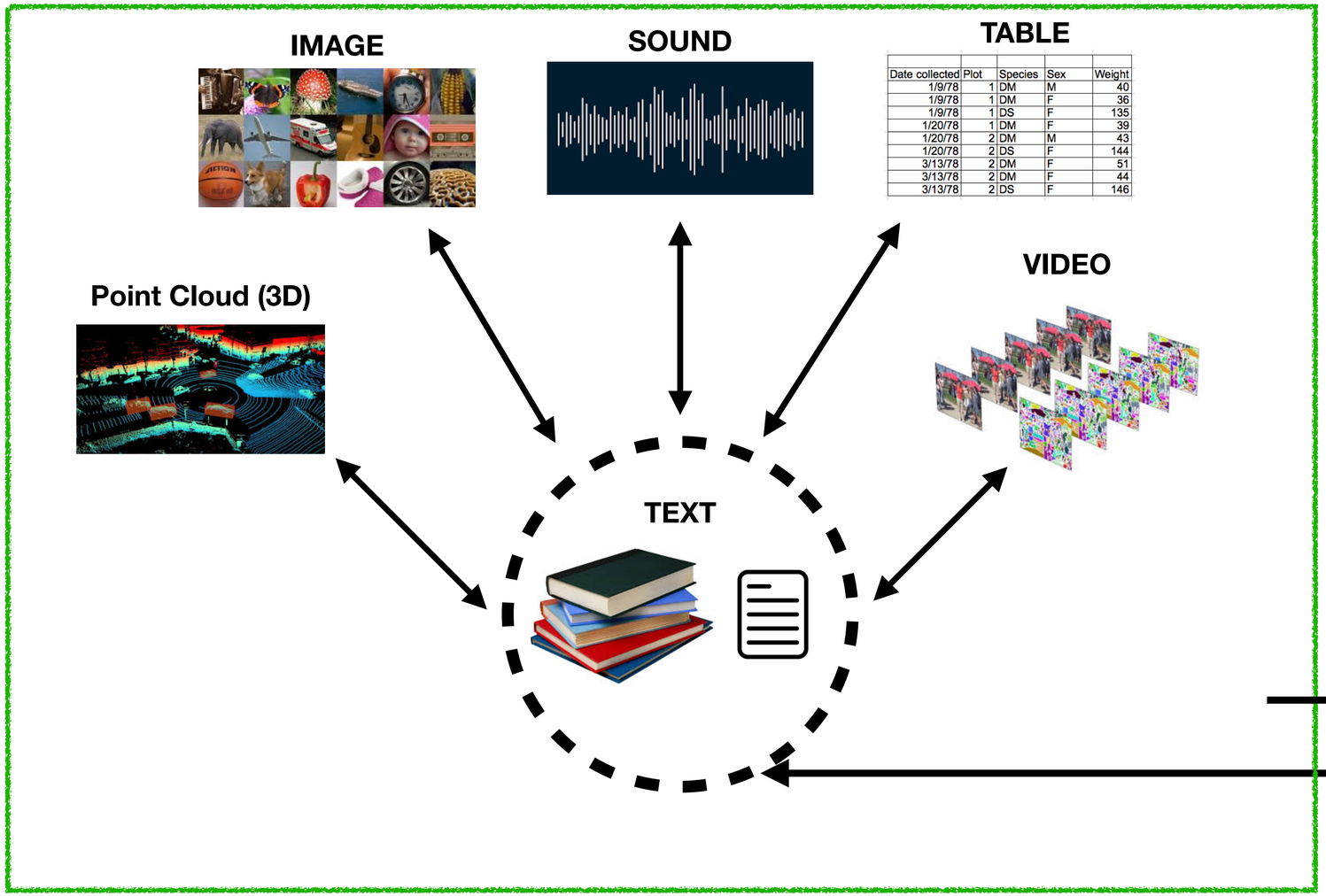
Multimodal AI ∩ FM:

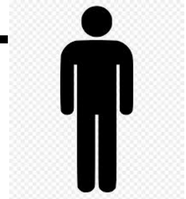
- Modern multimodal models (CLIP, GPT-4V, Gemini) **are Foundation Models**
- They combine: **scale** + **multimodality** + **generalization**
- But **not all** multimodal models **are FMs** (e.g., specialized medical imaging models)

Mutual Influence between FM and Multimodal AI, e.g.:

- Large **Scale pretraining** (FM)
- **Connecting infos** from different modalities (Multimodal AI)

Text as anchor modality





Text as anchor modality

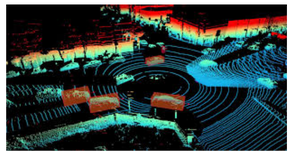
IMAGE

SOUND

TABLE

Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146

Point Cloud (3D)

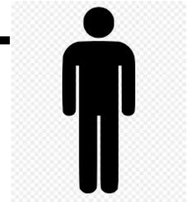


VIDEO

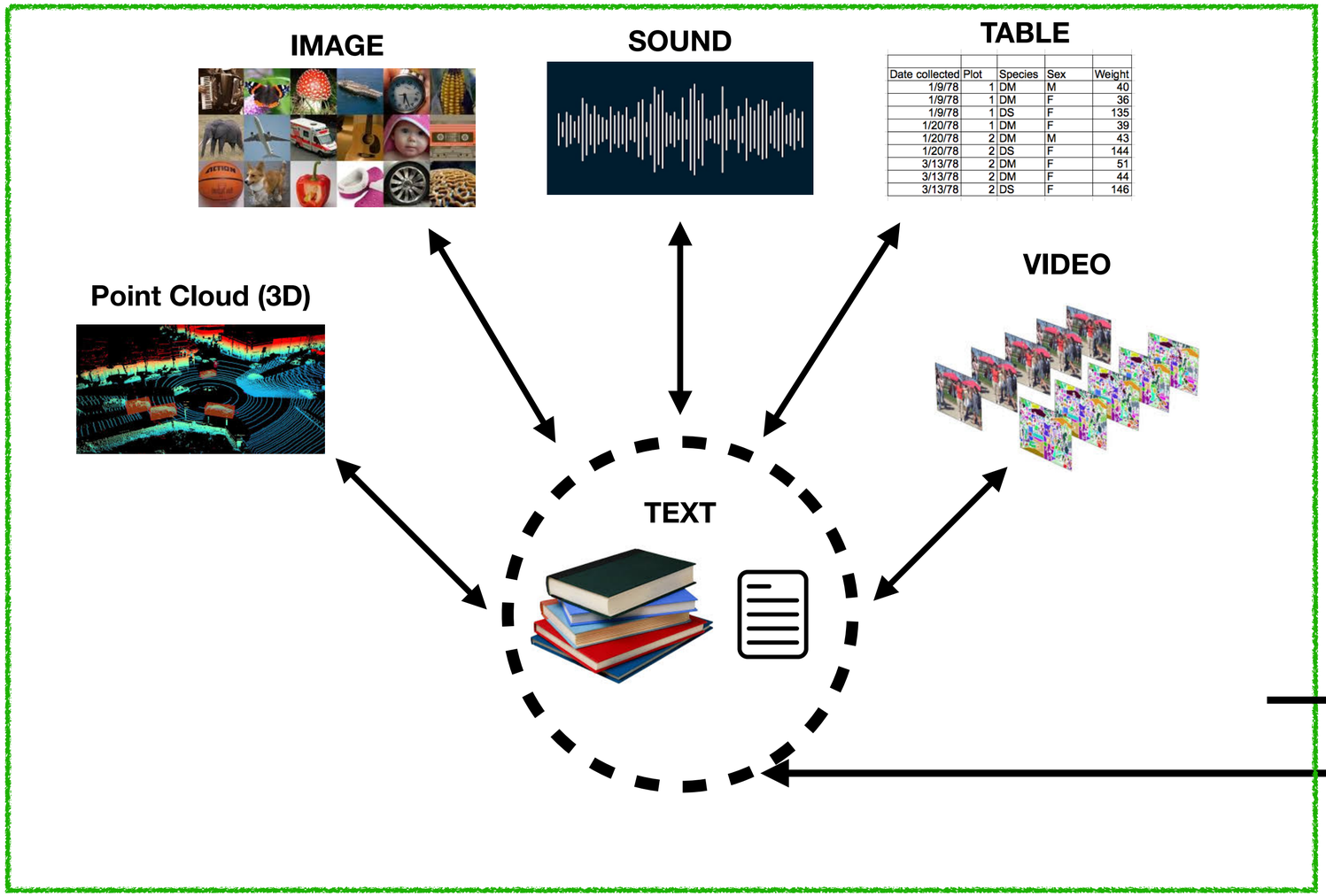
TEXT

Why Text (currently) is the pivot modality

TEXTUAL PROMPT



Text as anchor modality



Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146

Why Text (currently) is the pivot modality

- **Natural** Human-Machine Interaction/**Interface**
- Text enables **compositional reasoning** and complex instruction.
- Most multimodal architectures use text as the **common representation space** to project
- Most of Multimodal models are **extended LLM**

Challenges

Emerging Challenges

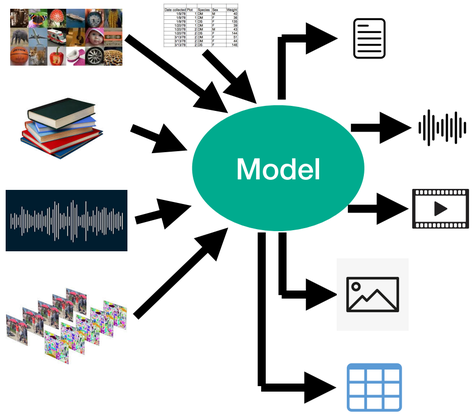
Hallucination: Generation of outputs that are plausible but inconsistent

Modern Multimodal AI models **are incline to hallucinate**

We can distinguish two main types of hallucinations

- **Factuality Hallucination**:
  Difference b/w generated content and verifiable real-world facts

- **Faithfulness hallucination:**
  Difference b/w generated content and user instructions



Bai et al. "**Hallucination of multimodal large language models: A survey**." *arXiv preprint arXiv:2404.18930* (2024).

Emerging Challenges

Hallucination: Generation of outputs that are plausible but inconsistent

Modern Multimodal AI models **are incline to hallucinate**

We can distinguish two main types of hallucinations

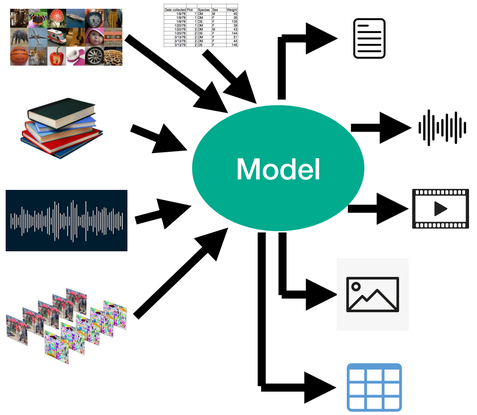
- **Factuality Hallucination**:
  Difference b/w generated content and verifiable real-world facts

- **Faithfulness hallucination:**
  Difference b/w generated content and user instructions

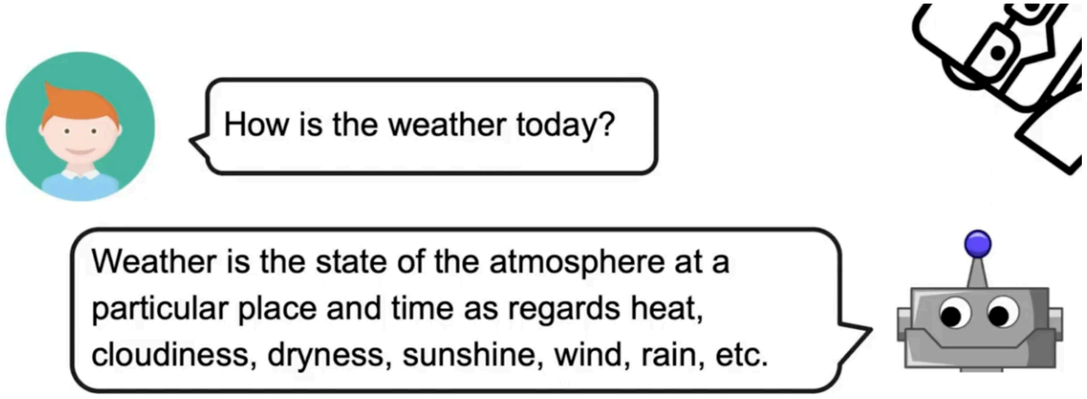
Hallucination **mitigation strategies** / techniques are necessary

Research activities are today devoted to **limit/mitigate hallucinations:**

- Data-centric Challenges - **Improve Data Quality**
- Enhance consistency via multimodal alignment
- Integrate mitigation strategies in the architecture design
- Enhance assessment and metric to verify hallucination tendency
- …



Bai et al. "**Hallucination of multimodal large language models: A survey**." *arXiv preprint arXiv:2404.18930* (2024).



Emerging Challenges

Data quality / Data validity / Bias (on internet/web):

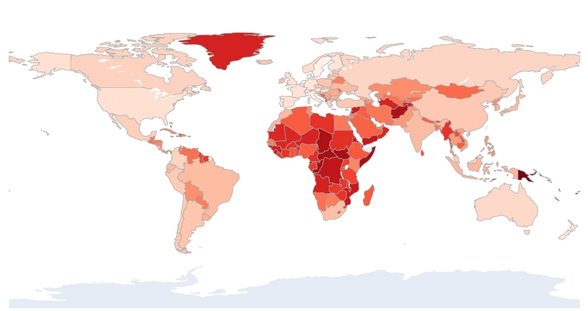
- Data contain **biases**
- Data **are finite**
- Data contains **knowledge about main/generic topic**, lacks of highly specialised content
- Data are biased **towards dominant languages** (e.g. English)
- Garbage in -> Garbage out

ML models are proxies of the data they have seen

Emerging Challenges

Data quality / Data validity / Bias (on internet/web):

- Data contain **biases**
- Data **are finite**
- Data contains **knowledge about main/generic topic**, lacks of highly specialised content
- Data are biased **towards dominant languages** (e.g. English)
- Garbage in -> Garbage out



Geographical bias in LLM

Moayeri et al. "**Worldbench: Quantifying geographic disparities in llm factual recall.**" ACM Conference on Fairness, Accountability, and Transparency. 2024.

ML models are proxies of the data they have seen

Article | Open access | Published: 24 July 2024

AI models collapse when trained on recursively generated data

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson & Yarin Gal

Nature **631**, 755–759 (2024) | Cite this article

636k Accesses | **452** Citations | **3607** Altmetric | Metrics

Ilya Sutskever
openAI cofounder
@NeurIPS24

OpenAI

Pre-training as we know it will end

Compute is growing:

- Better hardware
- Better algorithms
- Larger clusters

Data is not growing:

- We have but one internet
- **The fossil fuel of AI**

Research on Mitigation strategies is an **hot topic** today in the community

Emerging Challenges

Assessing **Emerging properties/capabilities:** Abilities or behaviors that arise unexpectedly from the model without being explicitly programmed or trained for

Wei et al. "**Emergent Abilities of Large Language Models**." Transactions on Machine Learning Research. 2022

Berti et al. "**Emergent abilities in large language models: A survey.**" arXiv preprint arXiv:2503.05788 (2025).

Emerging Challenges

Assessing **Emerging properties/capabilities:** Abilities or behaviors that arise unexpectedly from the model without being explicitly programmed or trained for

Wei et al. "**Emergent Abilities of Large Language Models**." Transactions on Machine Learning Research. 2022

Berti et al. "**Emergent abilities in large language models: A survey.**" arXiv preprint arXiv:2503.05788 (2025).

These capabilities:

- **Appear spontaneously** when models reach sufficient scale or complexity
- Were **not predictable from the model's performance** at smaller scales
- Go **beyond the sum of individual modalities** - the model can do things that wouldn't be possible by simply combining separate per modality systems

Emerging Challenges

Assessing **Emerging properties/capabilities:** Abilities or behaviors that arise unexpectedly from the model without being explicitly programmed or trained for

Wei et al. "**Emergent Abilities of Large Language Models**." Transactions on Machine Learning Research. 2022

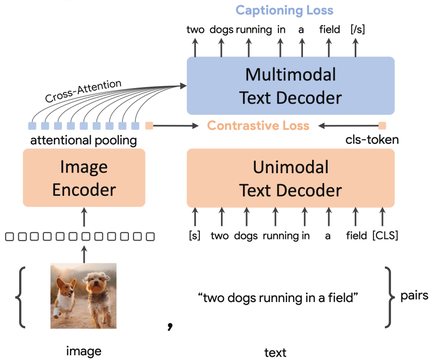
Berti et al. "**Emergent abilities in large language models: A survey.**" arXiv preprint arXiv:2503.05788 (2025).

These capabilities:

- **Appear spontaneously** when models reach sufficient scale or complexity
- Were **not predictable from the model's performance** at smaller scales
- Go **beyond the sum of individual modalities** - the model can do things that wouldn't be possible by simply combining separate per modality systems

An example

ZERO-SHOT TRANSFER



Zero-shot task transfer is defined as the **ability** of a multimodal model **to perform a new task without having been explicitly trained on** examples of that specific task.

Research on Identifying limits and domain scope (applicability) of current (large) Multimodal AI models

Emerging Challenges

Adoption of the current Multimodal AI models

Wang et al. "**Empowering edge intelligence: A comprehensive survey on on-device ai models.**" ACM Computing Surveys 57.9 (2025): 1-39.

Yao, Yuan, et al. "**Efficient GPT-4V level multimodal large language model for deployment on edge devices**." *Nature Communications* 16.1 (2025): 5509.

Emerging Challenges

Adoption of the current Multimodal AI models

Current Multimodal AI models:

- Require **large infrastructures** (HPC / Data Center / …)
- **Not well-suited for real-time** applications
- Generic knowledge, based **on publicly available data**

Wang et al. "**Empowering edge intelligence: A comprehensive survey on on-device ai models.**" ACM Computing Surveys 57.9 (2025): 1-39.

Yao, Yuan, et al. "**Efficient GPT-4V level multimodal large language model for deployment on edge devices.**" *Nature Communications* 16.1 (2025): 5509.

Emerging Challenges

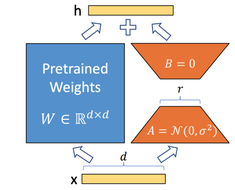
Adoption of the current Multimodal AI models

Current Multimodal AI models:

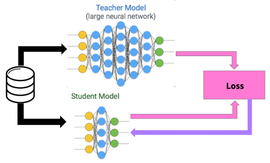
- Require **large infrastructures** (HPC / Data Center / …)
- **Not well-suited for real-time** applications
- Generic knowledge, based **on publicly available data**

Wang et al. "**Empowering edge intelligence: A comprehensive survey on on-device ai models.**" ACM Computing Surveys 57.9 (2025): 1-39.

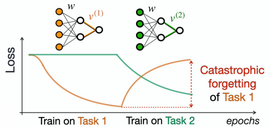
Yao, Yuan, et al. "**Efficient GPT-4V level multimodal large language model for deployment on edge devices.**" *Nature Communications* 16.1 (2025): 5509.



Parameter-Efficient Fine Tuning

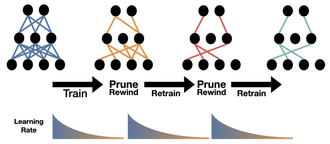


Knowledge Distillation



Catastrophic Forgetting

Model Quantization



Model Pruning

For real-world deployment, main challenges are related to:

- How to adopt model in **constrained-resources scenarios** (e.g. on device)
- Choose the **right model** (or model components) **for a particular scenario**
- Capitalize on the generic model as starting point for **task fine-tuning**

Emerging Challenges

Adoption of the current Multimodal AI models (4 Science)

Current Multimodal AI models are generic:

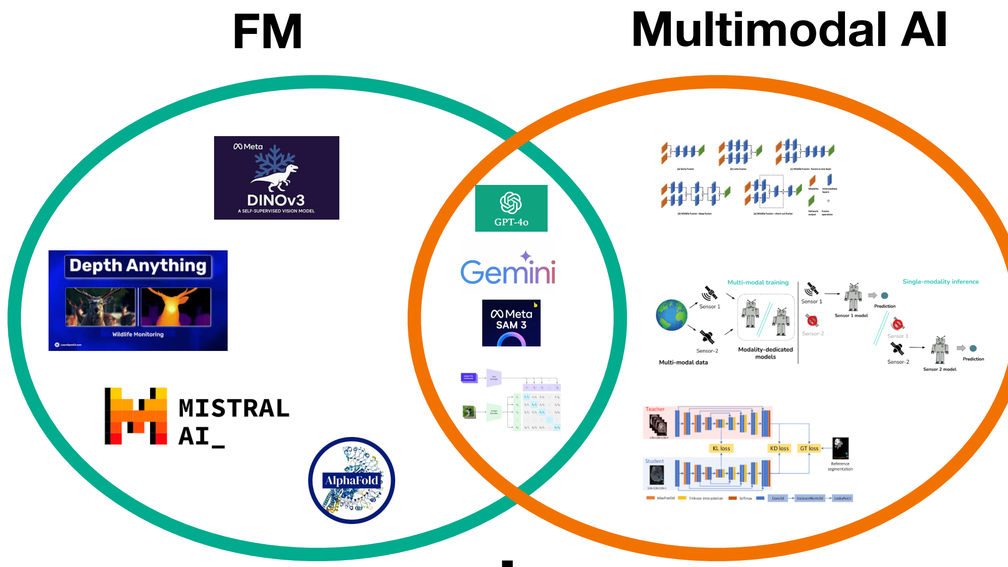
- Trained on **internet and public available knowledge** (text, image, video, …)
- **Lack** of deep domain expertise required for **scientific applications**
- **Lack understanding of scientific content** (notation, equations, table, molecular structures, …)
- **Limited Reasoning** and Planning capabilities

Emerging Challenges

Adoption of the current Multimodal AI models (4 Science)

Current Multimodal AI models are generic:

- Trained on **internet and public available knowledge** (text, image, video, …)
- **Lack** of deep domain expertise required for **scientific applications**
- **Lack understanding of scientific content** (notation, equations, table, molecular structures, …)
- **Limited Reasoning** and Planning capabilities



Can we (re-)use the models, or the recipes used to create them ?

? Expert knowledge
Physical laws
…

Domain Specific (Multimodal) Foundation Model

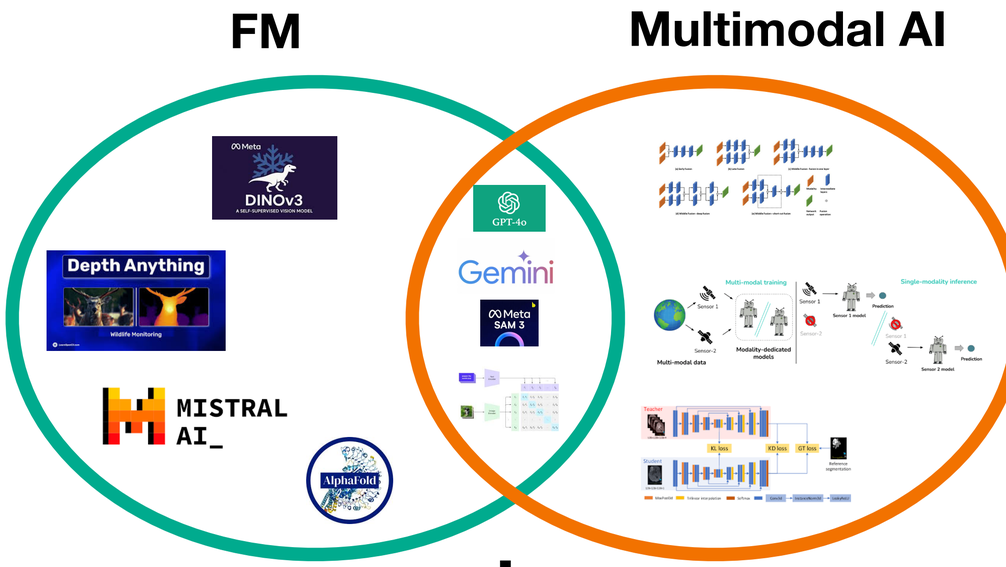
Emerging Challenges

Adoption of the current Multimodal AI models (4 Science)

Current Multimodal AI models are generic:

- Trained on **internet and public available knowledge** (text, image, video, …)
- **Lack** of deep domain expertise required for **scientific applications**
- **Lack understanding of scientific content** (notation, equations, table, molecular structures, …)
- **Limited Reasoning** and Planning capabilities

Research questions about using current Multimodal AI models as starting point for specific application domains or not



Can we (re-)use the models, or the recipes used to create them ?

? Expert knowledge
Physical laws
…

Domain Specific (Multimodal) Foundation Model

Bodnar, Cristian, et al. "**A foundation model for the Earth system.**" *Nature* (2025): 1-8.

Chen et al. "**An overview of domain-specific foundation model: key technologies, applications and challenges.**" Science China Information Sciences 69.1 (2026).

Technical/Emerging Challenges

Broad Societal challenges

- **Data/Model Sovereignty:** Dependence on foreign tech.
- **Digital divide:** Access inequality (compute infrastructure, know how, cultural biases).
- **Environmental impact:** Energy consumption, Sustainable AI development.
- **Labor support:** Limitation (or homogenisation) of the creative actions.
- **Timely Regulation and Laws:** Risk-based approach, transparency requirements (e.g. AI Act, Data Act, ..).
- …

Technical/Emerging Challenges

Broad Societal challenges

- **Data/Model Sovereignty:** Dependence on foreign tech.
- **Digital divide:** Access inequality (compute infrastructure, know how, cultural biases).
- **Environmental impact:** Energy consumption, Sustainable AI development.
- **Labor support:** Limitation (or homogenisation) of the creative actions.
- **Timely Regulation and Laws:** Risk-based approach, transparency requirements (e.g. AI Act, Data Act, ..).
- …

Technical/Emerging Challenges

Broad Societal challenges

- **Data/Model Sovereignty:** Dependence on foreign tech.
- **Digital divide:** Access inequality (compute infrastructure, know how, cultural biases).
- **Environmental impact:** Energy consumption, Sustainable AI development.
- **Labor support:** Limitation (or homogenisation) of the creative actions.
- **Timely Regulation and Laws:** Risk-based approach, transparency requirements (e.g. AI Act, Data Act, ..).
- …

Innovation pace and Adoption/Regulation pace are not always aligned

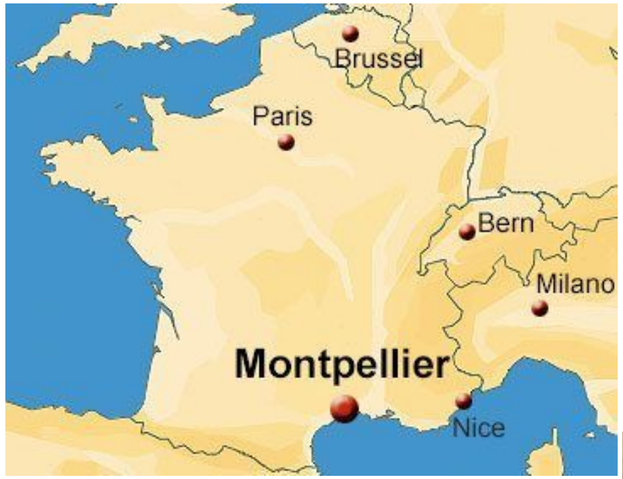
Research & Development

- New models every few months
- Capabilities evolving rapidly
- From research to product: months
- Driven by competition

Society & Governance

- Regulations: years cycle
- Workforce adaptation: years
- Educational reform: years
- Driven by consensus & caution

EVERGREEN Team



EVERGREEN Team: 5 permanents, 6 PhDs, 2 Post-doc, 2 Engineers, Visiting researchers (several per year)

Team Topics:
Machine Learning & Computer Vision for EO data with applications in Agriculture & Environment

R. Interdonato

R. Gaetano

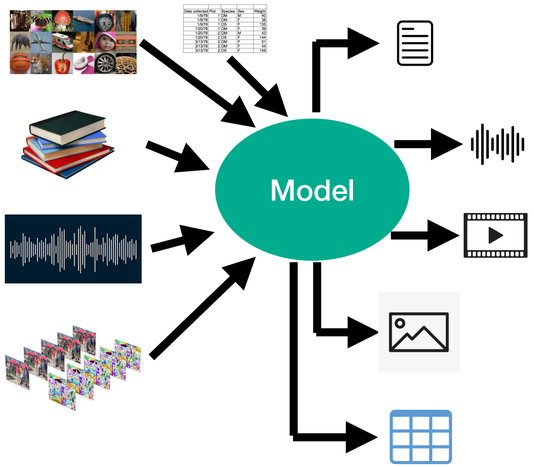
C. Fraga Dantas

D. Ienco

D. Marcos

Thank You for your attention

MULTIMODAL
DATA



tetis
TERRITOIRE ENVIRONNEMENT TÉLÉDÉTECTION
INFORMATION SPATIALE

INRAE

EVERGREEN

cirad

Inria