Pacific Northwest
NATIONAL LABORATORY

The Center for AI @PNNL

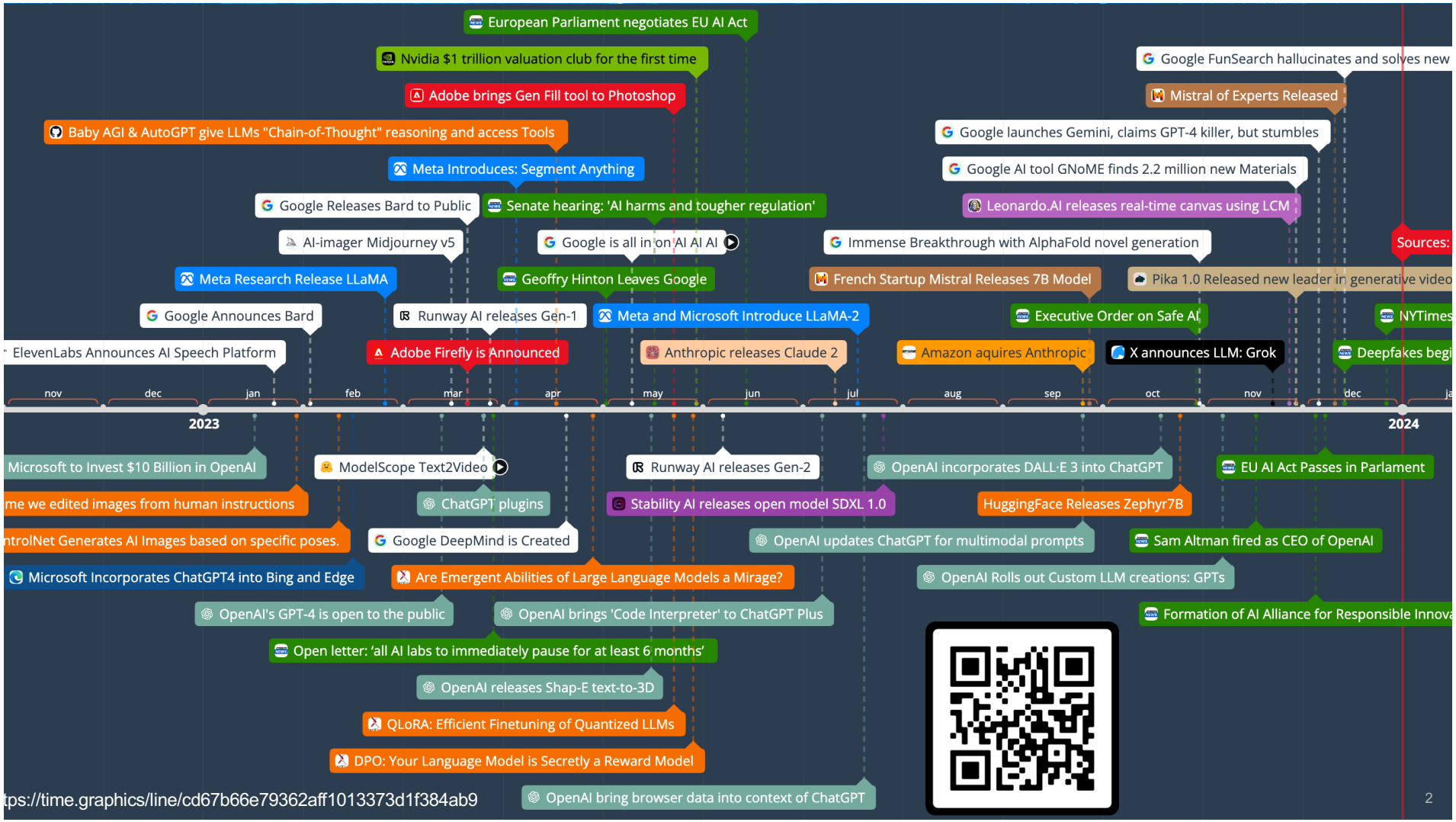
# The AI Summer is Here, will National Security Science and Engineering Bask in the Sunshine, or get Burned?

4/23/24

**Courtney D Corley, PhD**
Chief Scientist for AI
Director, Center for AI @ PNNL

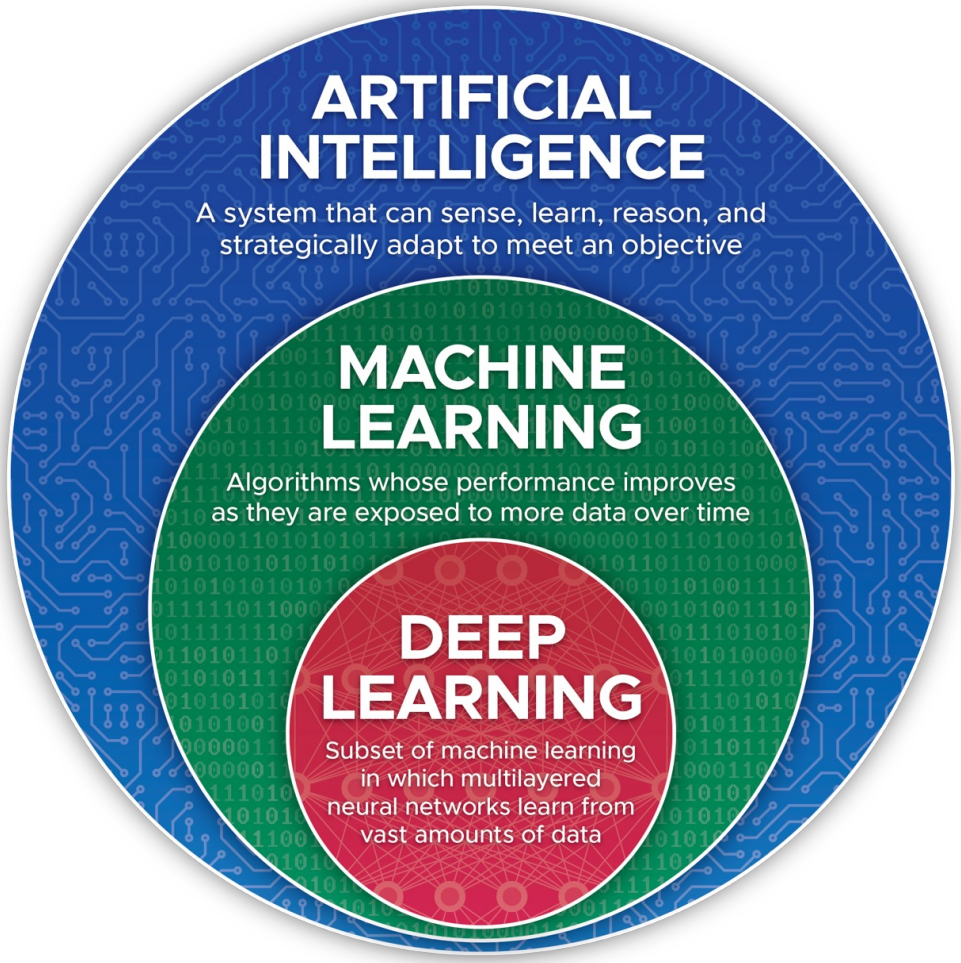
U.S. DEPARTMENT OF ENERGY BATTELLE

PNNL is operated by Battelle for the U.S. Department of Energy


European Parliament negotiates EU AI Act
Nvidia $1 trillion valuation club for the first time
Adobe brings Gen Fill tool to Photoshop
Baby AGI & AutoGPT give LLMs "Chain-of-Thought" reasoning and access Tools
Meta Introduces: Segment Anything
Google Releases Bard to Public
Senate hearing: 'AI harms and tougher regulation'
AI-imager Midjourney v5
Google is all in on AI AI AI
Meta Research Release LLaMA
Geoffry Hinton Leaves Google
Google Announces Bard
Runway AI releases Gen-1
Meta and Microsoft Introduce LLaMA-2
ElevenLabs Announces AI Speech Platform
Adobe Firefly is Announced
Anthropic releases Claude 2
Google FunSearch hallucinates and solves new
Mistral of Experts Released
Google launches Gemini, claims GPT-4 killer, but stumbles
Google AI tool GNoME finds 2.2 million new Materials
Leonardo.AI releases real-time canvas using LCM
Immense Breakthrough with AlphaFold novel generation
French Startup Mistral Releases 7B Model
Pika 1.0 Released new leader in generative video
Executive Order on Safe AI
Amazon aquires Anthropic
X announces LLM: Grok
Sources:
NYTimes
Deepfakes begi
nov
dec
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
2023
2024
Microsoft to Invest $10 Billion in OpenAI
ModelScope Text2Video
Runway AI releases Gen-2
OpenAI incorporates DALL·E 3 into ChatGPT
EU AI Act Passes in Parlament
me we edited images from human instructions
ChatGPT plugins
Stability AI releases open model SDXL 1.0
HuggingFace Releases Zephyr7B
ntrolNet Generates AI Images based on specific poses.
Google DeepMind is Created
OpenAI updates ChatGPT for multimodal prompts
Sam Altman fired as CEO of OpenAI
Microsoft Incorporates ChatGPT4 into Bing and Edge
Are Emergent Abilities of Large Language Models a Mirage?
OpenAI Rolls out Custom LLM creations: GPTs
OpenAI's GPT-4 is open to the public
OpenAI brings 'Code Interpreter' to ChatGPT Plus
Formation of AI Alliance for Responsible Innova
Open letter: 'all AI labs to immediately pause for at least 6 months'
OpenAI releases Shap-E text-to-3D
QLoRA: Efficient Finetuning of Quantized LLMs
DPO: Your Language Model is Secretly a Reward Model
tps://time.graphics/line/cd67b66e79362aff1013373d1f384ab9
OpenAI bring browser data into context of ChatGPT
2

Pacific Northwest
NATIONAL LABORATORY

# Artificial intelligence encompasses a broad range of advanced computing topics



Pacific Northwest NATIONAL LABORATORY

# Machine learning has a formal definition

"A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*" -- *Tom Mitchell, "Machine Learning" 1997*

| Task T | Experience E | Performance P |
|---|---|---|
| Image classification | 1 million labeled images | Classification accuracy |
| The game of Go | 30 million recoded moves | Winning Rate |
| Simulating fluid flow | Example solutions to fluid flow equations | Calculation Accuracy |
| Generate text | Gigabytes of natural language | Prediction of missing words |
| Detect bugs in code | Thousands of examples of source code with labeled bugs | Classification accuracy |
| Create artificial faces | 50,000 images of faces | The probability the generator creates a face that can't be distinguished from a real one |

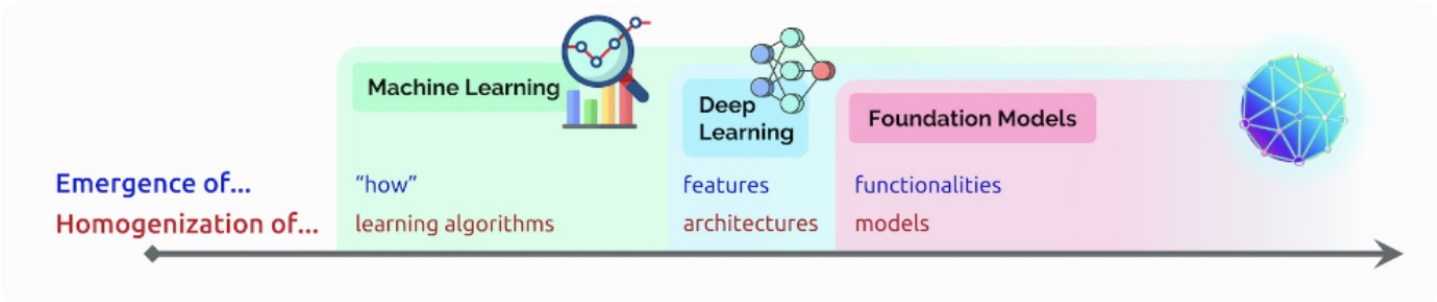
# What's new with foundation models? They wear many hats, not just one

### Narrow (Single-Purpose) AI

- Application specific
- Single/limited task
- Can't solve unfamiliar problems
- Need to build new models for each new task

### Foundation (Multi-Purpose) Models

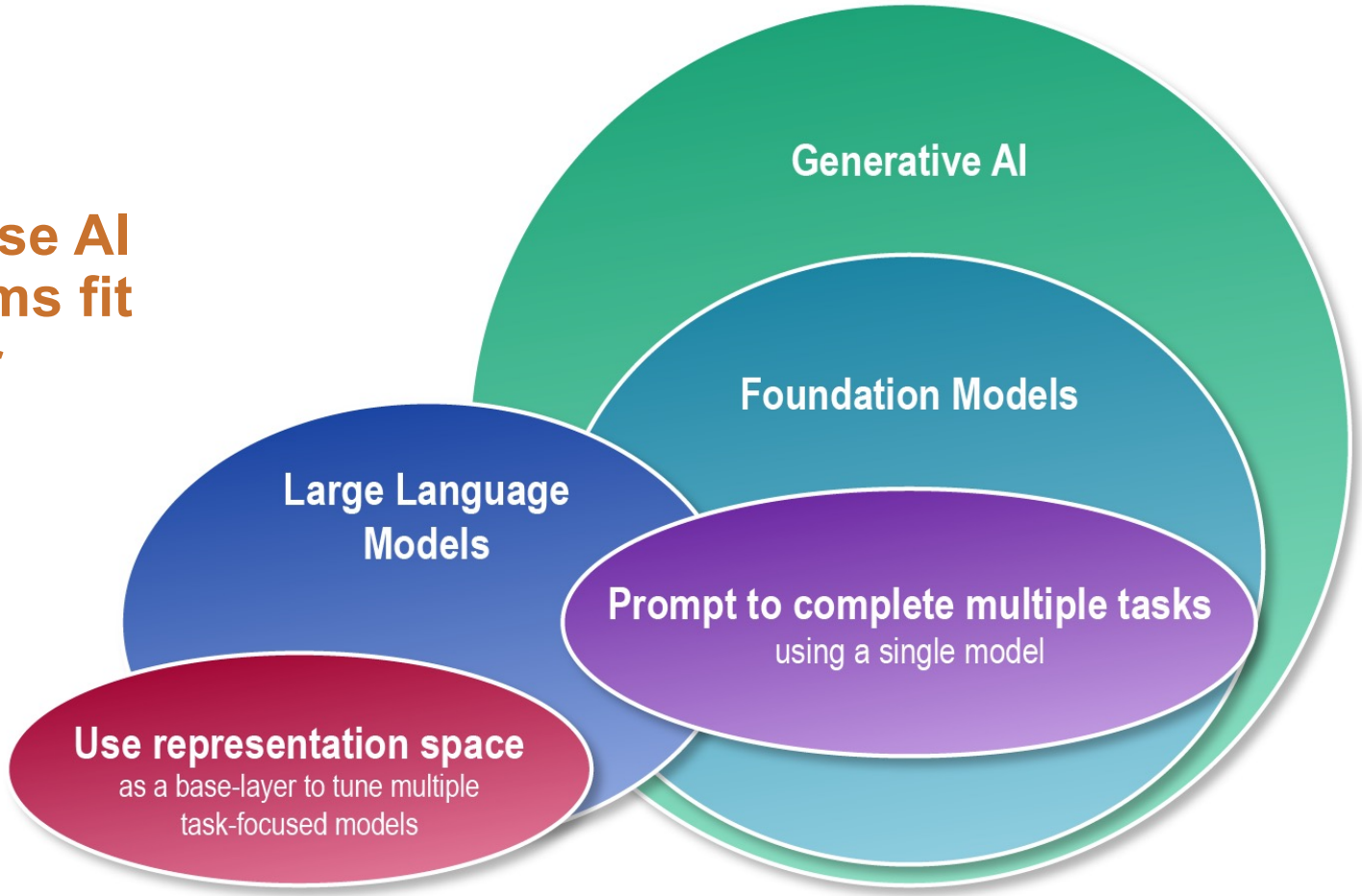
- Self-supervised learning on unlabeled data
- Generative
- Multipurpose (multiple tasks)
- Support new tasks and complex logic



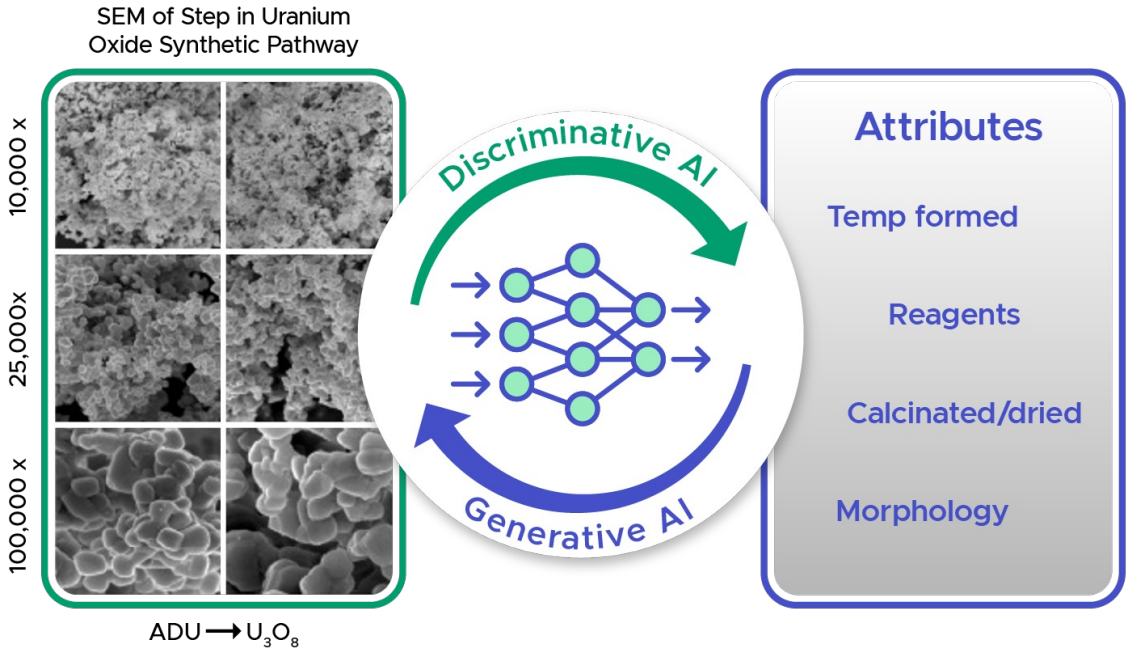

https://blogs.nvidia.com/blog/2022/03/25/what-is-a-transformer-model/

Pacific Northwest
NATIONAL LABORATORY

# How these AI paradigms fit together



# Toy example: generating plausible materials from a set of attributes as opposed to characterization



Pacific Northwest
NATIONAL LABORATORY

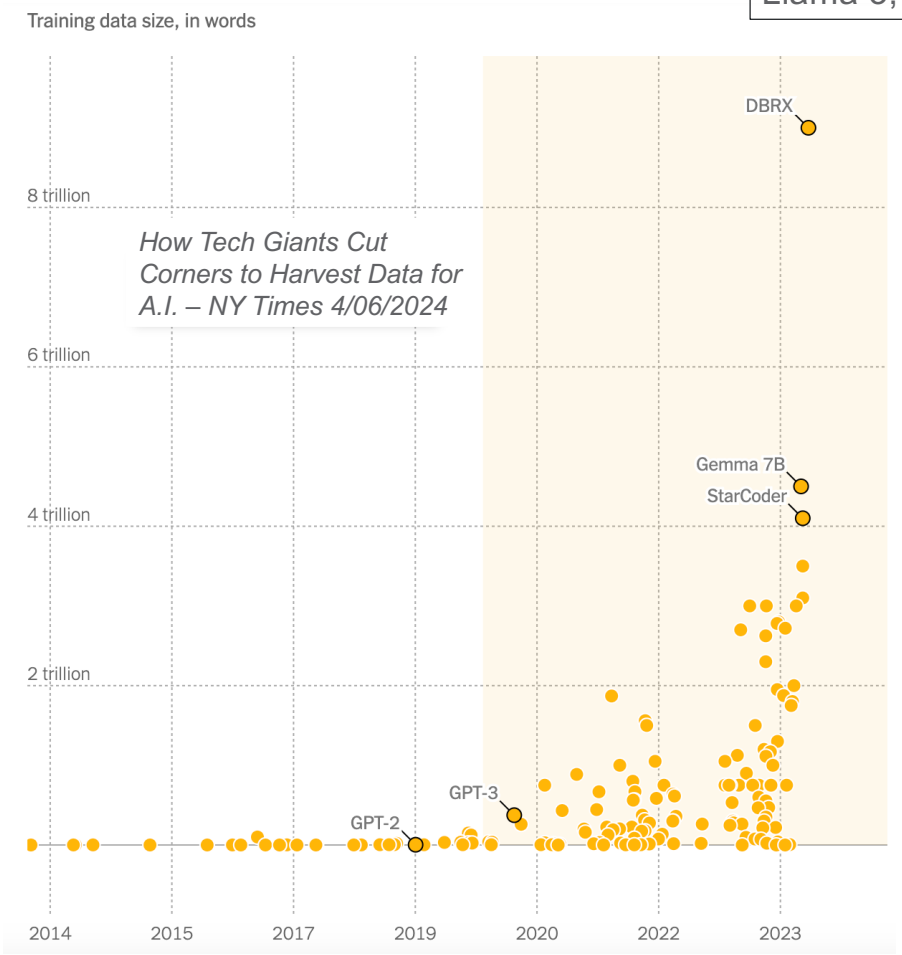
# Diffusion model

- Diffusion models can generate data similar to what it was trained on
- High quality video and image generation
- Text-to-image generation

Pacific Northwest
NATIONAL LABORATORY

# AI is data hungry, whether you have labels or not

Foundation models require beyond internet scale data




Source: https://www.nytimes.com/2024/04/06/technology/tech-giants-harvest-data-artificial-intelligence.html

# Medium to small AI models are relevant and vital for national security science and engineering

- Deep learning efficiently leverages large quantities of data to learn representations of classes
  - Deep learning *can* be applied to small data … but restrictions *will* apply …
- But what if the class you care about has few examples?
  - Ignore differences and just apply existing model (e.g., information retrieval)
  - Adapt a pre-trained models (e.g., transfer learning)
  - Train model to generalize to new classes (e.g., few-shot learning)

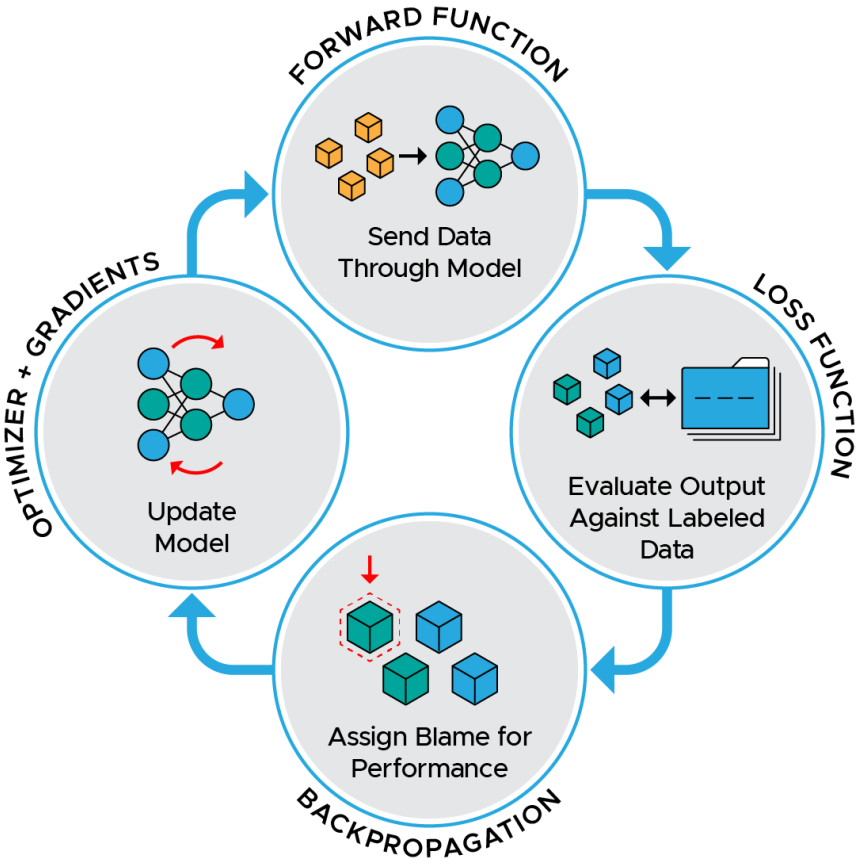
| Quantity | # Labels | # Data |
|---|---|---|
| **Supervised** | High | High |
| **Semi-supervised** | Low | High |
| **Self-supervised** | --- | High |
| **Transfer learning** | Low/Medium | Low/Medium |
| **Few/low/zero-shot** | Very low | Very low |


These models continue to train on PNNL's very first DGX-1, *ohmahgerd,* purchased in 2016

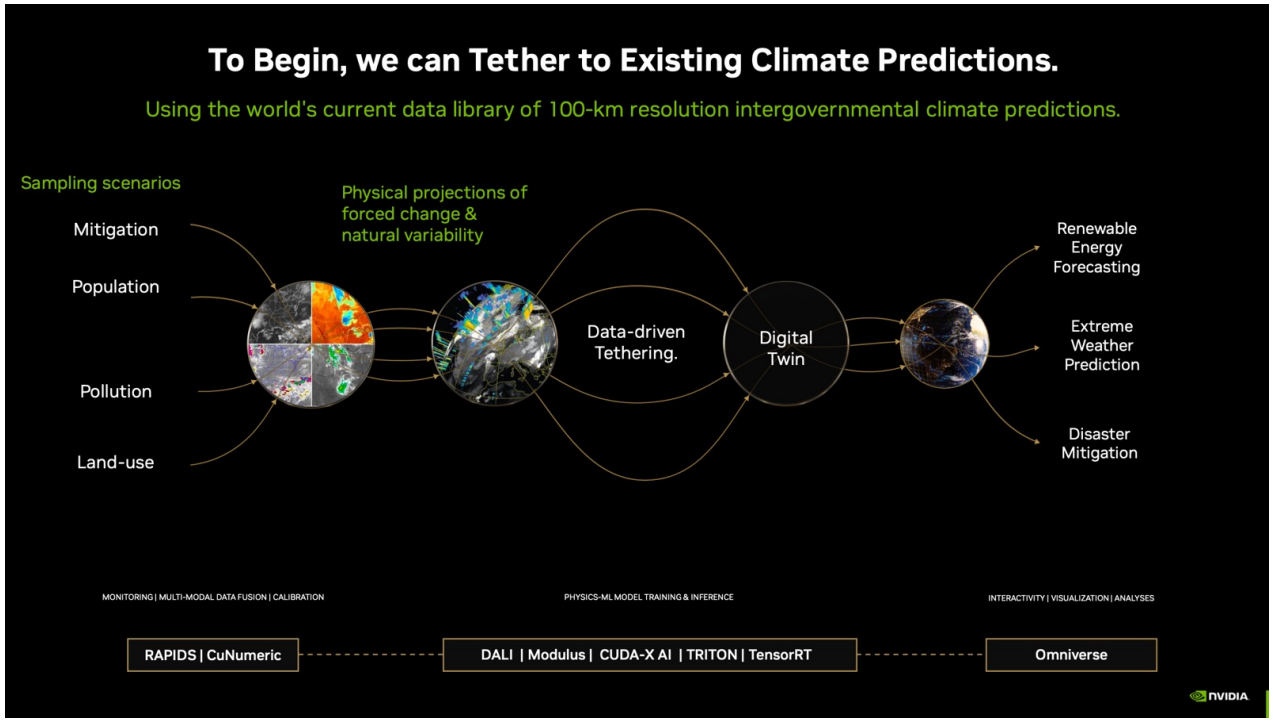
Pacific Northwest
NATIONAL LABORATORY

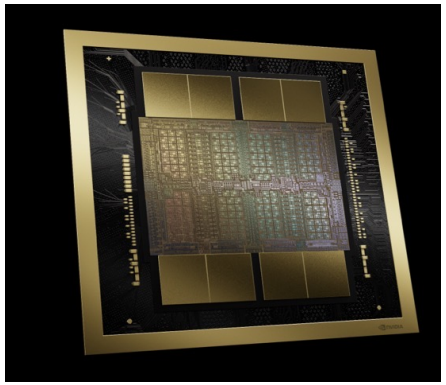
# One step up from everything is linear algebra

- Inside text are the steps that all models use
- Outside text are the terms when training a deep learning model, including transformer-based models



# Differentiable Modeling and Simulation
## *Is this the end of double precision architectures?*



NVIDIA Blackwell advertises 4-bit floating point interfaces w/ 20/40 petaFLOPS (dense/sparse)

Source: NVIDIA'S Earth -2: Digital Twins For Weather and Climate, Anandkumar, Kashinath, Pritchard

Pacific Northwest NATIONAL LABORATORY

# The Cost of AI Compute

- Commodity 8xH100 ~$250k
- Meta's $13B and 600,000 GPU purchase is nation state level investment
  - e.g., the unit cost of the U.S. Gerald R Ford aircraft carrier
  - Power? HVAC?

Llama-2 Cloud Deployment Costs (previous generation)

| Model | Instance Type | # of GPUs per replica | On-demand (per hour) | On-demand (per year) |
|---|---|---|---|---|
| Llama 7B | ml.g5.2xlarge | 1 | $ 1.52 | $13,315.20 |
| Llama 13B | ml.g5.12xlarge | 4 | $ 7.09 | $62,108.40 |
| Llama 70B quantized | ml.g5.48xlarge | 8 | $16.29 | $178,353.60 |
| Llama 70B | ml.p4d.24xlarge A100 | 8 | $32.77 | $287,065.20 |

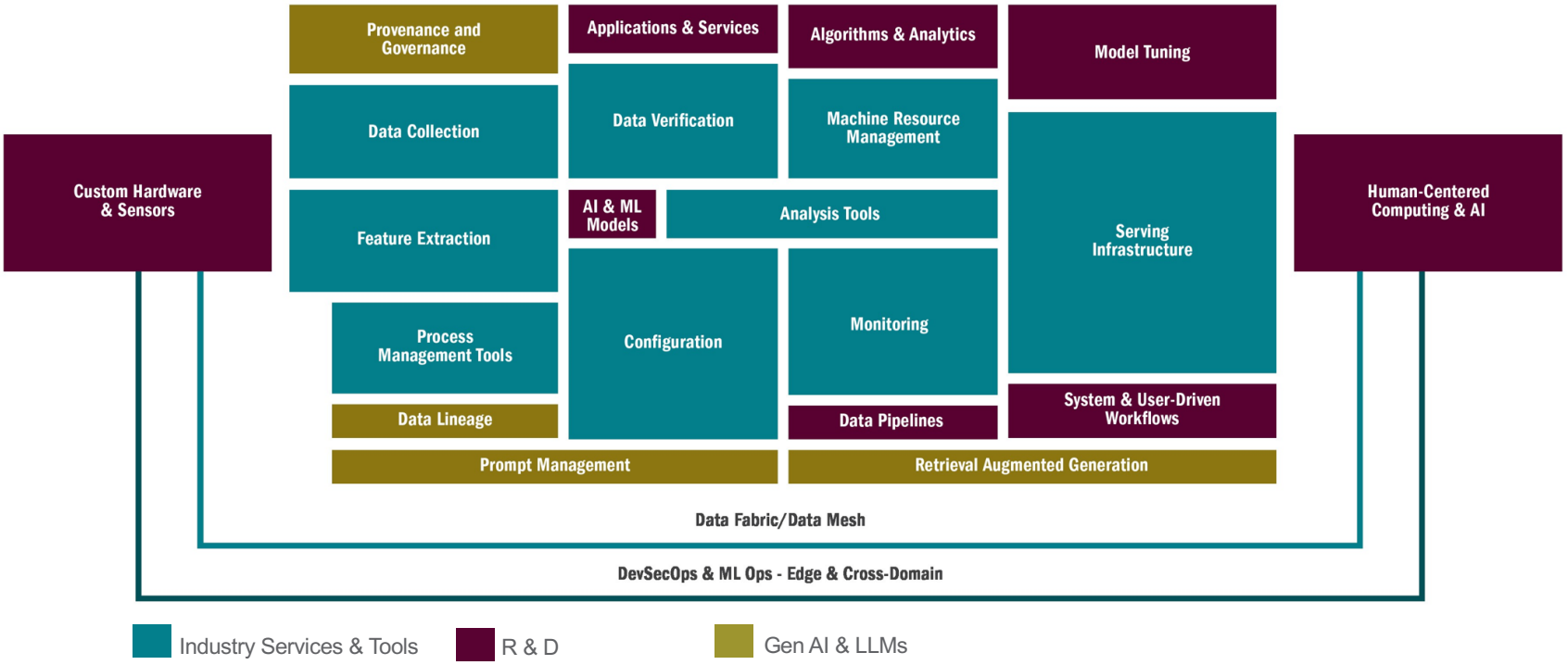
Llama-2 Cloud Pre-Training Costs (previous generation)

| Model | Instance Type (A100) | GPU Hours | AWS On-demand |
|---|---|---|---|
| Llama 7B | ml.p4d.24xlarge | 184,320 | $755,020.80 |
| Llama 13B | ml.p4d.24xlarge | 368,640 | $1,510,041.60 |
| Llama 70B | ml.p4d.24xlarge | 1,720,320 | $7,046,860.80 |
| DALL-E 2 | ml.p4d.24xlarge | 150,000 | $614,437.50 |
| ml.p4d.24xlarge 8GPUs $32.77/hr https://aws.amazon.com/ec2/instance-types/p4/ | | | |

https://web.archive.org/web/20171228190957/https://fas.org/sgp/crs/weapons/RS20643.pdf

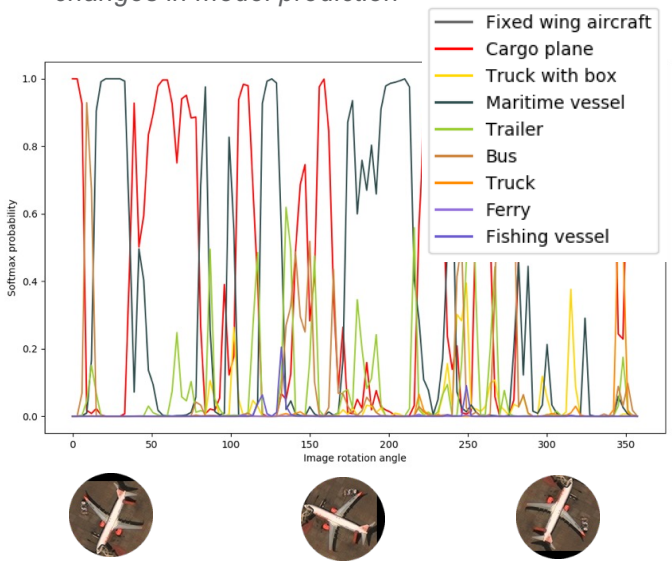
Pacific Northwest
NATIONAL LABORATORY

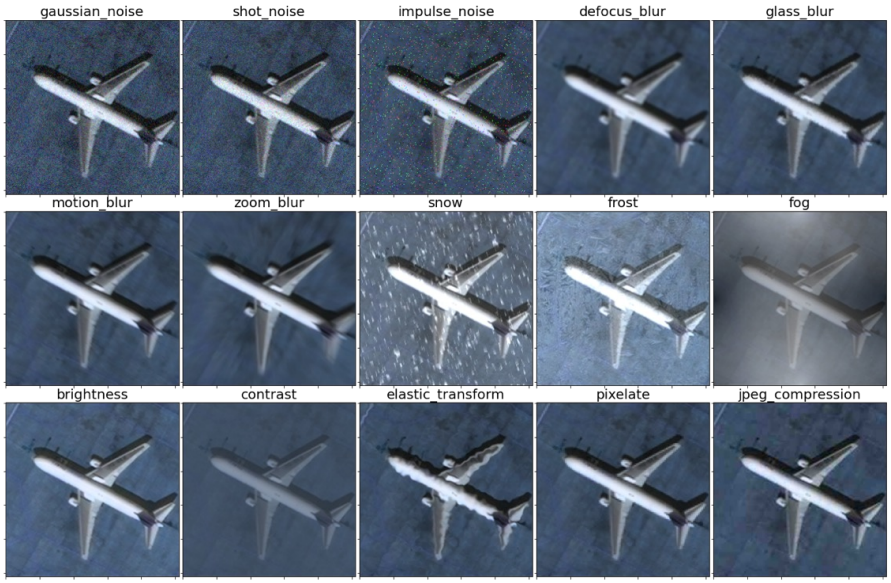
# AI requires services readily available via industry developed tools and platforms



# Even without manipulation, machine learning models lack robustness

*Natural changes in the data (such as the rotation of a plane) can result in dramatic and non-intuitive changes in model prediction*




*Environmental changes such as weather can cause a model to fail if it hasn't seen these conditions before*

# Hallucination (caveat emptor)

***erroneous*** *references, content, and statements, may be intertwined with correct information, and presented in a persuasive and confident manner, making their identification difficult* ***without close inspection and effortful fact-checking***

```
Human:> Can I get McDonalds at the SeaTac airport?

ChatGPT:> Yes, there is a McDonalds at the SeaTac airport,
located in the central terminal near gate C2. It is open from 5
a.m. to 10 p.m. daily.
```

Example of hallucination from: https://arxiv.org/pdf/2303.12712.pdf

# Reinforcement Learning with Human Feedback: Alignment and Guardrails

How do you create/ code a loss function for:

- What is *funny*?
- What is *ethical*?
- What is *safe*?

Don't Code it, Model It

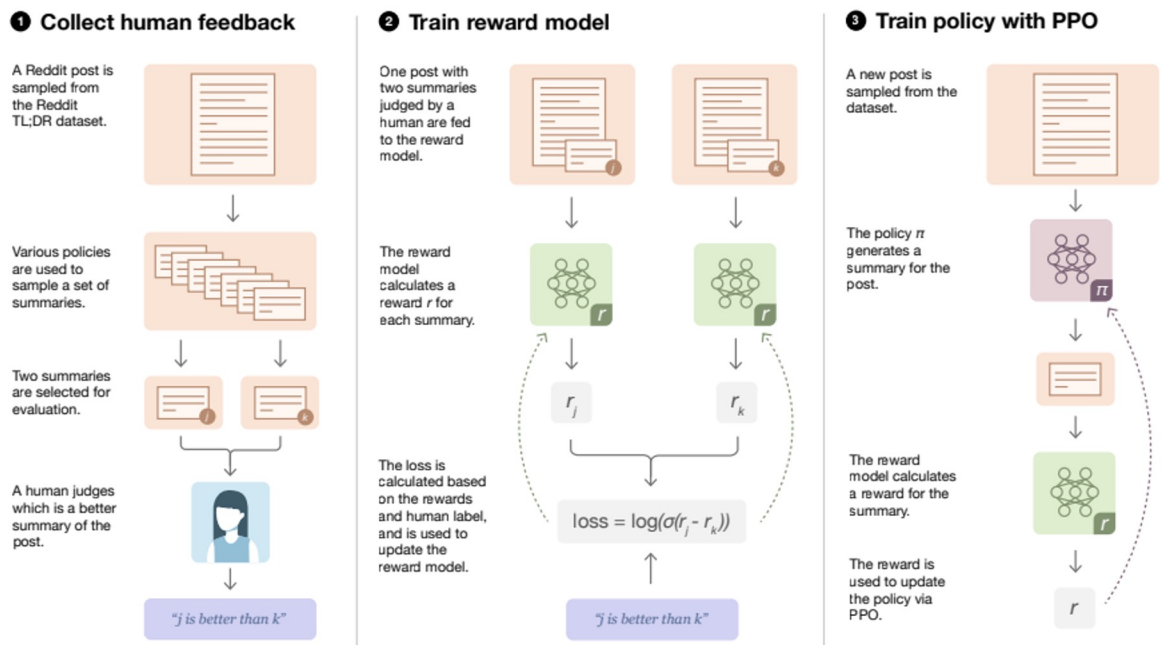


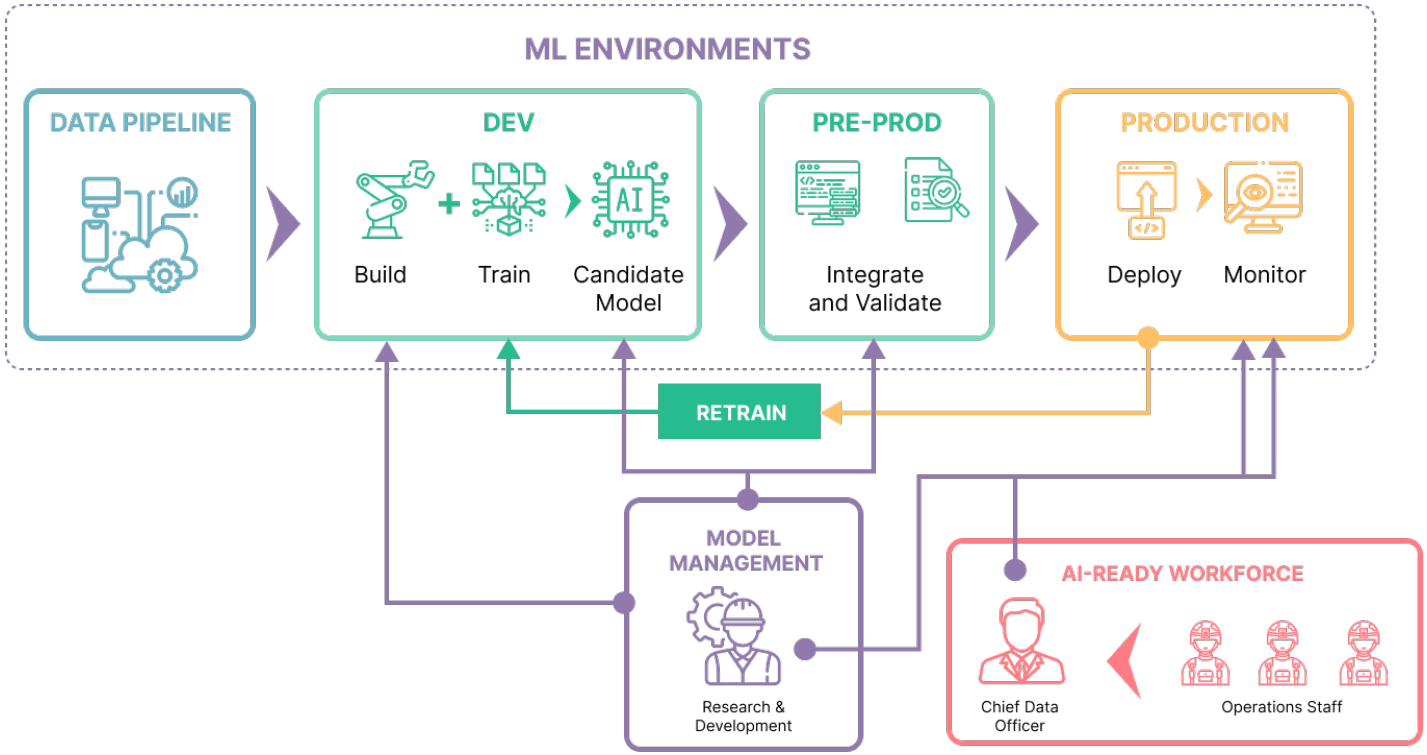

Figure 2: Diagram of our human feedback, reward model training, and policy training procedure.

Figure from: Stiennon, Nisan, et al. "Learning to summarize with human feedback." *Advances in Neural Information Processing Systems* 33 (**2020**): 3008-3021. Slide adapted from ICML 2023 Tutorial on Reinforcement Learning with Human Feedback

Pacific Northwest
NATIONAL LABORATORY

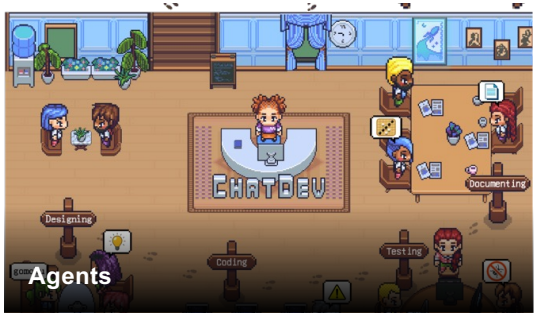
# AI requires fundamental changes to the workforce

## *Human-machine workflows and operating constructs*



# Bask in the sunshine, or get burned?

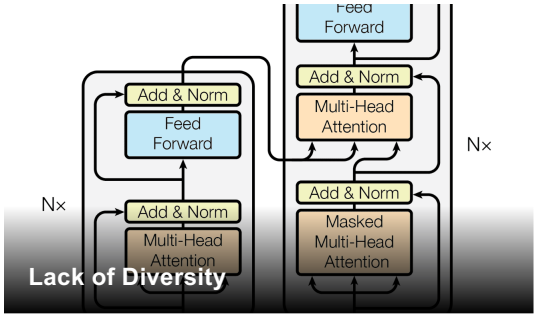
**The first and second derivatives are growing, if and when will they normalize?**



Agents

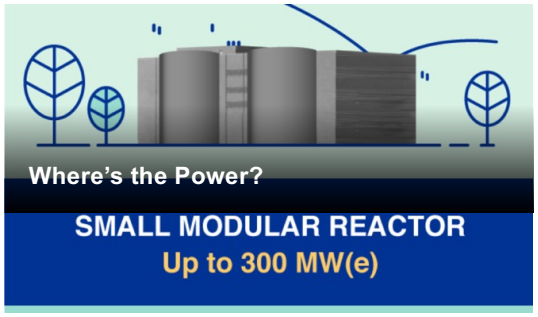
Visual Language Action

Open Source Catching Up



Lack of Diversity


Critical Workforce Gaps



Pacific Northwest
NATIONAL LABORATORY

# Thank you


Pacific Northwest
NATIONAL LABORATORY

**Court Corley, PhD**

Chief Scientist for AI
Director, Center for AI @ PNNL

NATIONAL SECURITY
DIRECTORATE

**court@pnnl.gov**