

# Adv ML for Gen-AI

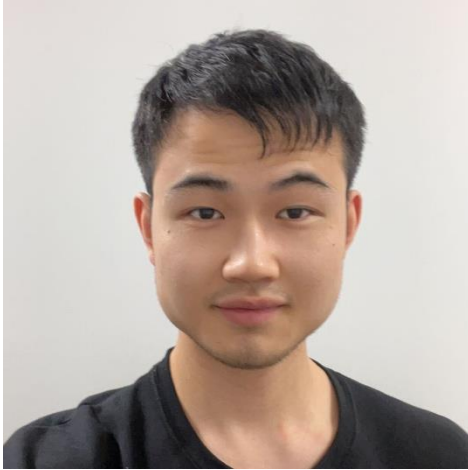
## Course Overview and Logistics

<https://ml-graph.github.io/spring-2026/>

Yu Wang, Ph.D.  
Assistant Professor  
Department of Computer Science  
University of Oregon  
Personal: <https://yuwang0103.github.io/>  
Lab: <https://kindlab-fly.github.io/>



# Self-Introduction



**Yu (Jack) Wang**  
**(You)**

**Contact:**

[yuwang@uoregon.edu](mailto:yuwang@uoregon.edu)

<https://yuwang0103.github.io/>

## Research Interests:

- Data Mining and Machine Learning
- Neural-Symbolic Learning
- Graph and Network
- LLM + Structured Knowledge
- AI/ML/DM Applications
  - Document Intelligence
  - Social Computing
  - Networking Physical Infrastructure





- **Background and Motivation**
- **Objectives and Content**
- **Logistics**





- **Background and Motivation**
- **Objectives and Content**
- **Logistics**





**Generative AI has been a long-standing topic**

**Machines that can produce human-like speech**





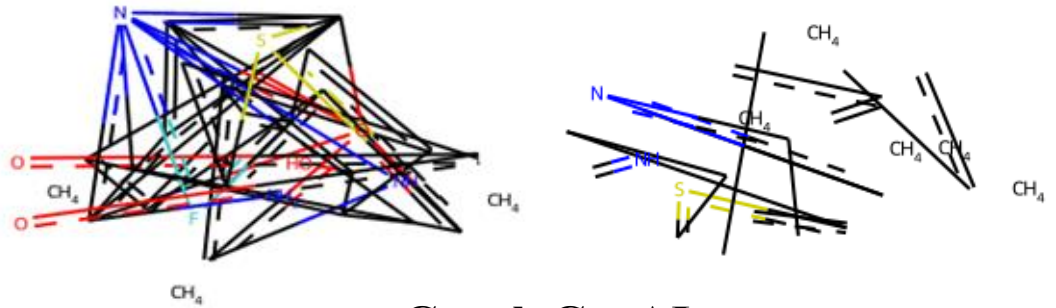
# Background



Video GenAI



Image GenAI

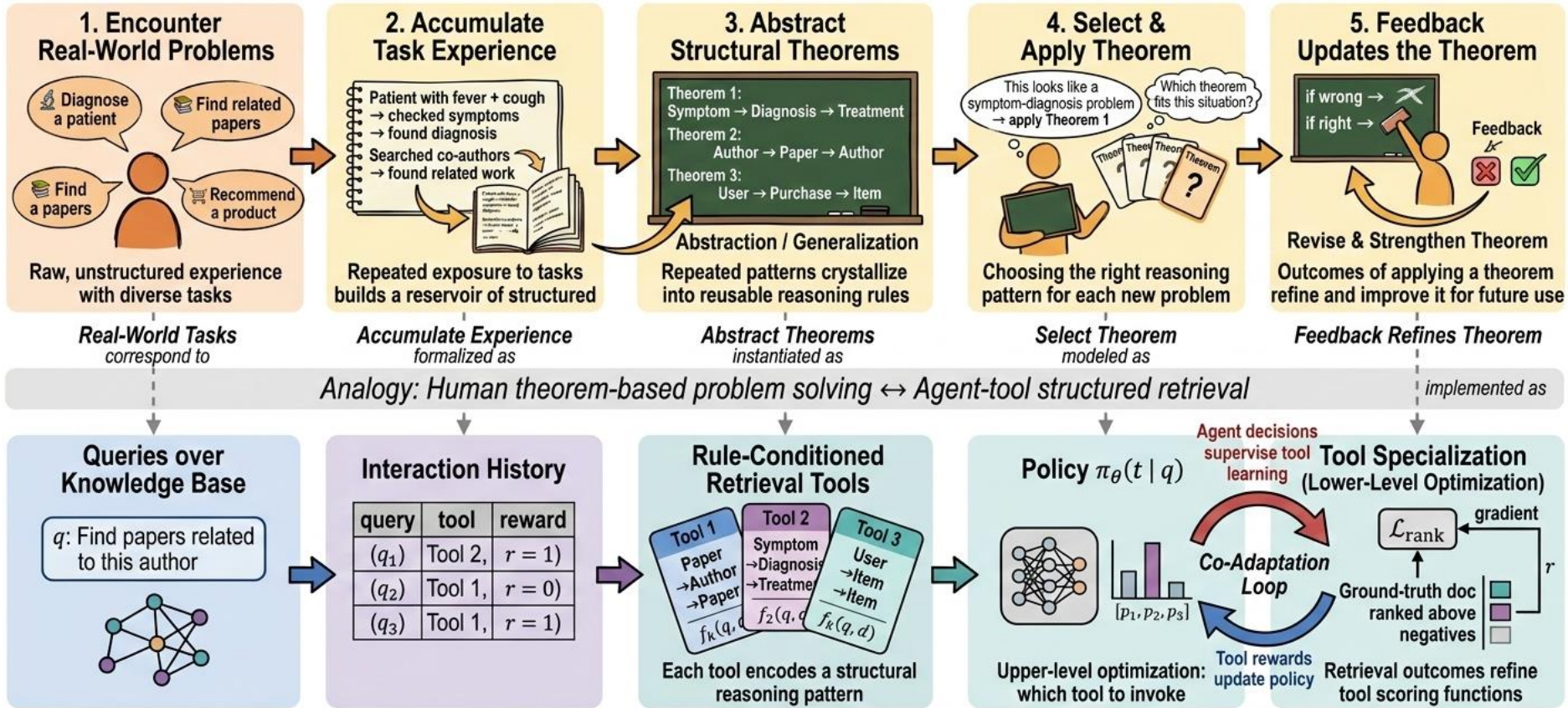


Graph GenAI



# Motivation – Image GenAI

## From Human Experience to Structured Retrieval: A Motivating Analogy



## Scientific Figure - Nano Banana



# Motivation – Image GenAI



**UNIVERSITY OF OREGON**

**INFO 410:**  
Data Visualization

Term: Falhu  
Credits: 10:00–11:50 AM

Instructor: &r. Alicia Ethics

- Intro **ጥ-ገገገ** Dr. Alicia Nguyen
- Chart Tue/ጥጥጥ 2026
- Term: /ጥጥጥ ጥጥጥ Deschutes 220
- Location: Dessites 220
- Color Theory

Tiedits: 4 Dala & Chelign

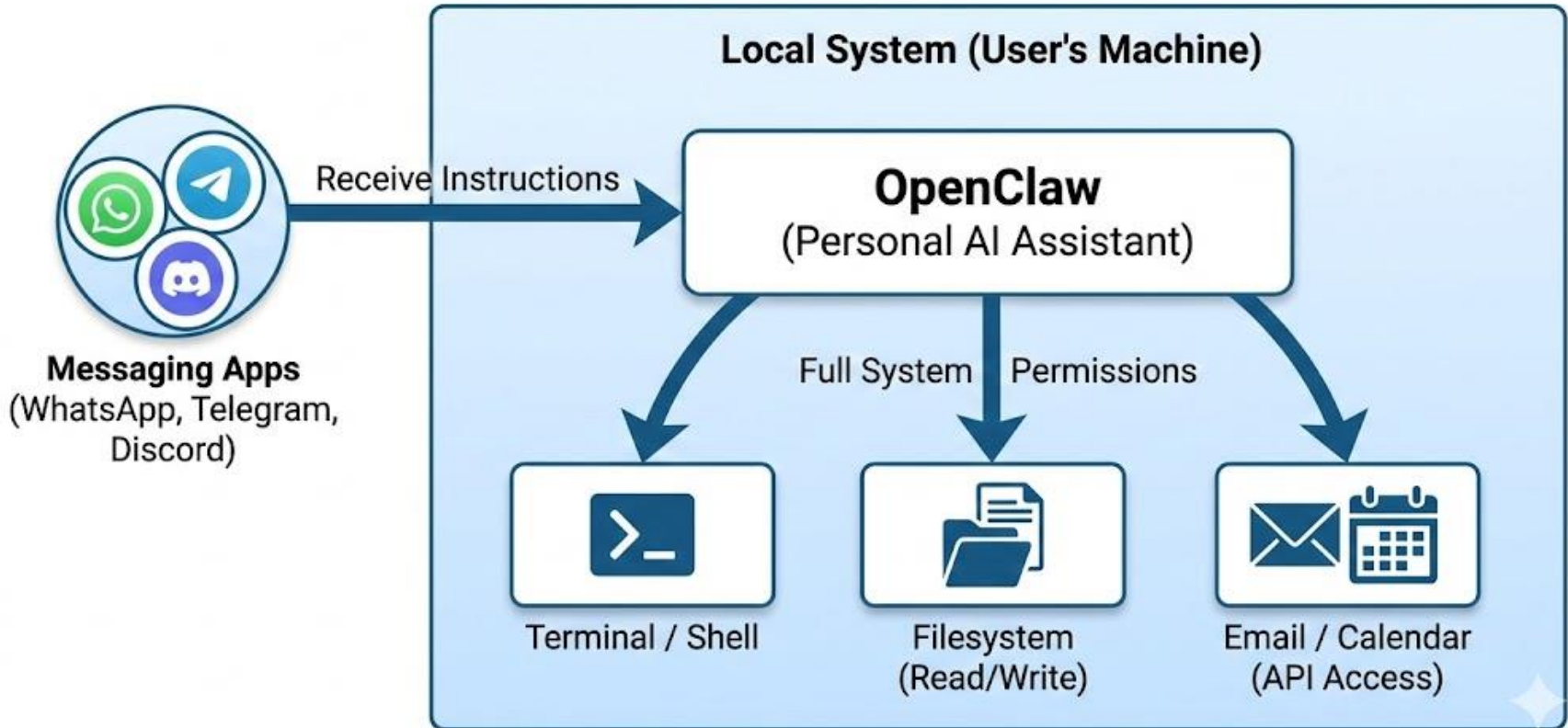
## Adobe Photoshop/Firefly





# Motivation – Language GenAI

## OpenClaw Local Architecture (Simplified)



**OpenClaw**





# Motivation – Language GenAI

# report

Monday, March 23rd

**bot-general** APP 9:00 AM  
Happy Monday, team! It's that time of the week – time to post your weekly report! 📅  
Please share a quick summary of what you accomplished over the past week. This helps everyone stay aligned and keep track of progress. 🙌  
@Yu Wang @Zhisheng @Utkarsh Sahu – drop your updates when you get a chance! 🙌

**bot-general** APP 12:00 PM  
Hey @Yu Wang, @Zhisheng, and @Utkarsh Sahu – just a friendly nudge! 📅 It looks like none of your weekly reports update in the channel. Thanks! 🙌

**Utkarsh Sahu** 12:09 PM  
1. Run foundational model with combined and root only as well.  
2. Troubleshoot code for multigpu ddp for foundational run  
3. Working on paper  
4. Grade project 4 and final exam  
5. Read paper  
(edited)

**Yu Wang** 1:15 PM  
1. Finish coding Agentic AI for wildfire debris flow and getting good results  
2. Finish initial NSF-NAIRR proposal writing on agentic wildfire debris flow  
3. Winter Grading

**bot-general** APP 3:00 PM  
Hey @Zhisheng – just a friendly nudge! 📅 Your weekly report hasn't come in yet today. Whenever you get a chance,

**Zhisheng** 5:26 PM  
1. Finish reward balance problem  
2. Implement detailed monitoring and visualization analysis on training process  
3. Explored entropy-collapse problem on GRPO based on four papers  
4. Implement solution one for anti-collapse exploration-enhance training  
5. Update training monitor for anti-collapse analysis  
6. Finish KDD review

EXPLORER

- AGENT-WORK
  - .claude
  - 2026-Agentic-Indexing
  - GPU-UO-4H200
  - greencard
  - KINDLab-Fly.github.io
  - neural-network-quiz
  - notion
  - personalization-Yu Wang
  - Proposal
  - social media
  - spring-2026
    - \_announcements
    - \_assignments
    - \_css
    - \_data
    - \_events
    - \_images
    - \_includes
    - \_layouts
    - \_lectures
    - \_sass
    - slides
    - slides\_old
    - static\_files
    - !\_config.yml
    - assignments.md
    - Gemfile
    - Gemfile.lock
    - grade.md
    - image.png
    - index.md
    - lectures.md
    - LICENSE

Create Lec4 in-class coding assignment

office hour 7-7:30 wednesday pst

```
git [main ad05d04] update office hours to Wednesday 7:00-7:30 pm PST
1 file changed, 1 insertion(+), 1 deletion(-)
To https://github.com/ML-Graph/spring-2026.git
```

Done. Office hours updated to Wednesday 7:00-7:30 pm PST and pushed. Update your Canvas copy accordingly.

can you update my personal website to remove netowring to be just phisical infrastructure and push, the folder is yuwang0103

Thought for 1s

**Glob** pattern: "/Users/yuwang/Dropbox/agent-work/yuwang0103/\*\*/\*.\*.md"  
No files found

**Bash**

```
IN find /Users/yuwang -maxdepth 4 -type d -name "yuwang0103" 2>/dev/null
```

Allow this bash command?

```
find /Users/yuwang -maxdepth 4 -type d -name "yuwang0103" 2>/dev/null
```

1 Yes

2 Yes, allow //Users/yuwang/\*\* for all projects

3 No

Tell Claude what to do instead

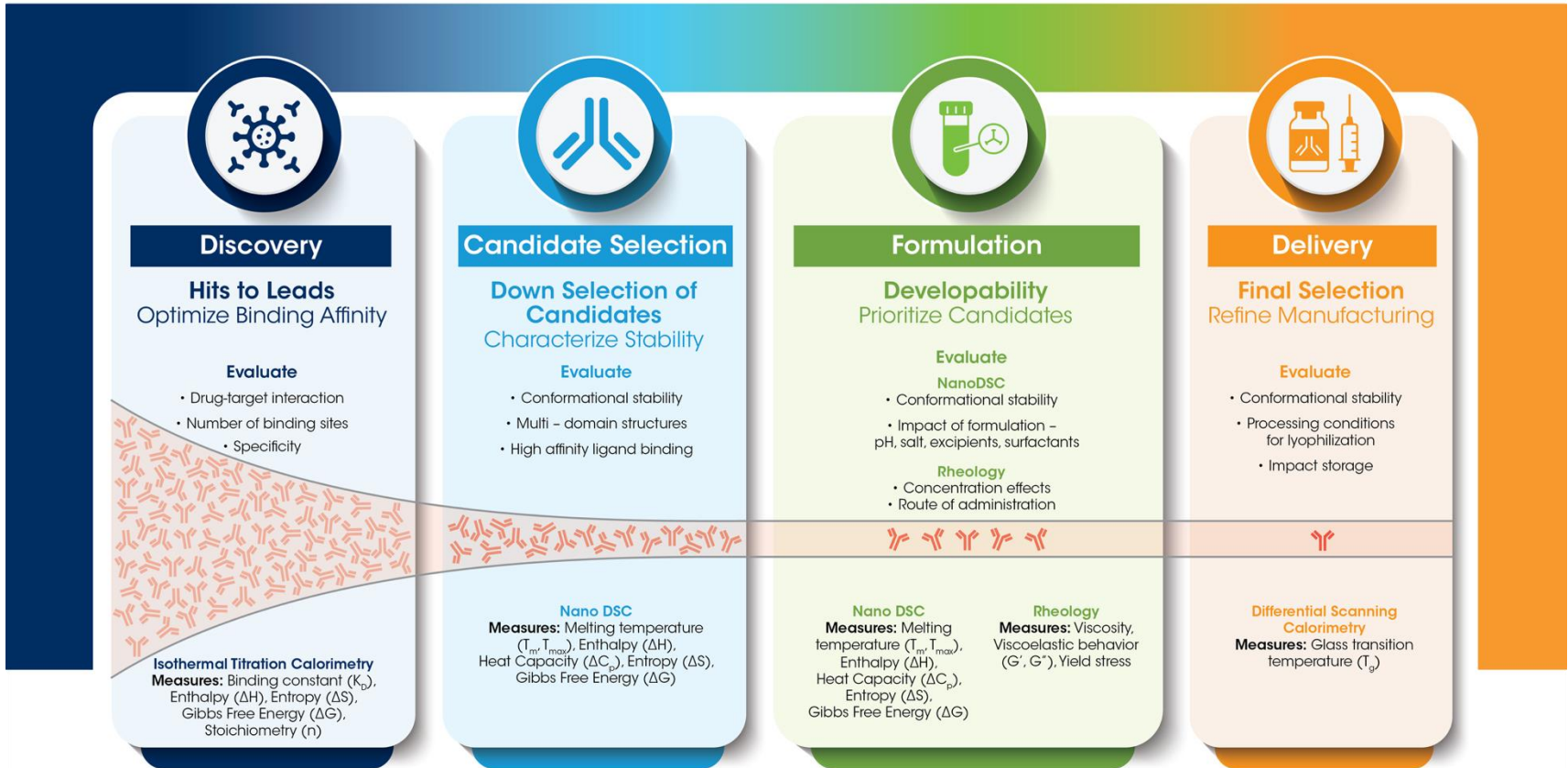
Esc to cancel

## Personal Assistant



# Motivation – Graph GenAI

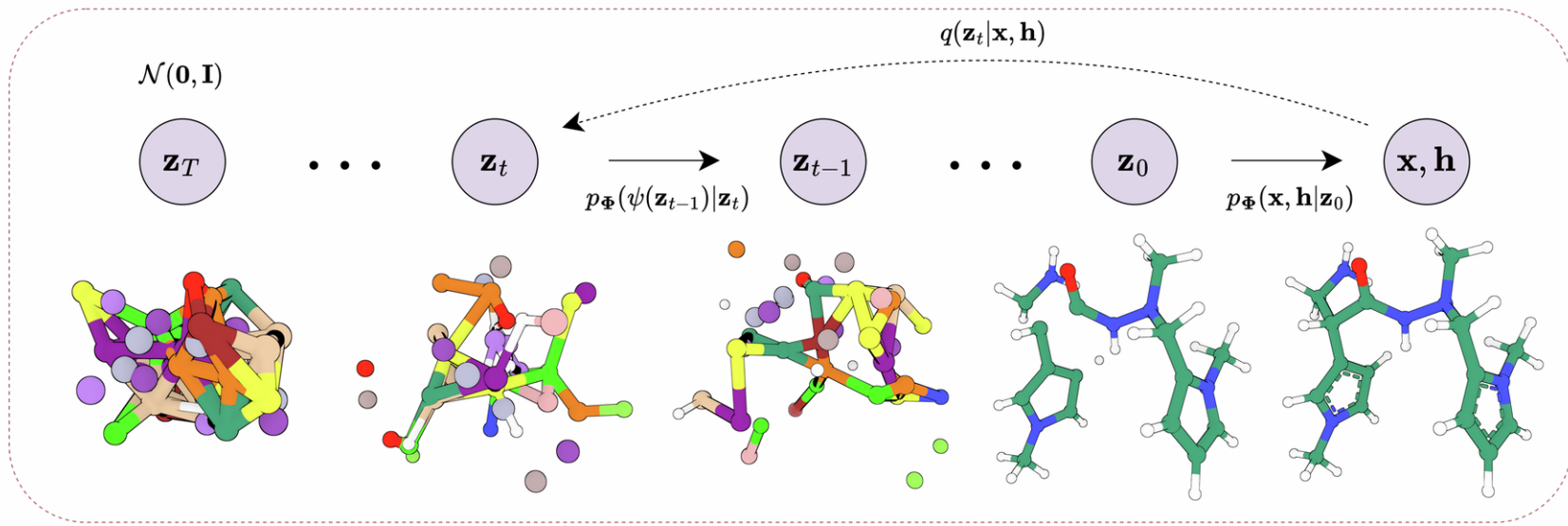
## BioPharma Drug Development Characterization Solutions



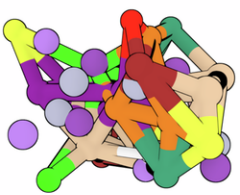


# Motivation – Graph GenAI

## iv. Geometry–Complete Diffusion Generation with GCPNet ++

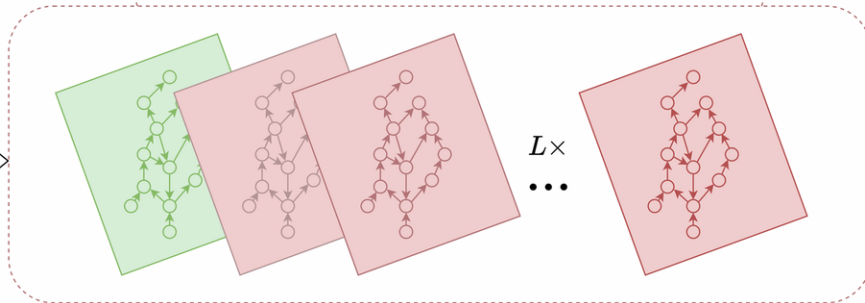


### i. Graph Definition

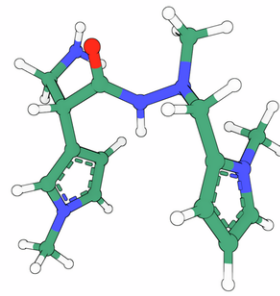


Nodes:  $(h, \chi)$   
Edges:  $(e, \xi)$   
Frames:  $\mathcal{F}_{ij}$

### ii. GCPNet ++



### iii. Denoised Graph





<https://music.youtube.com/watch?v=wDIXoS1z8il>

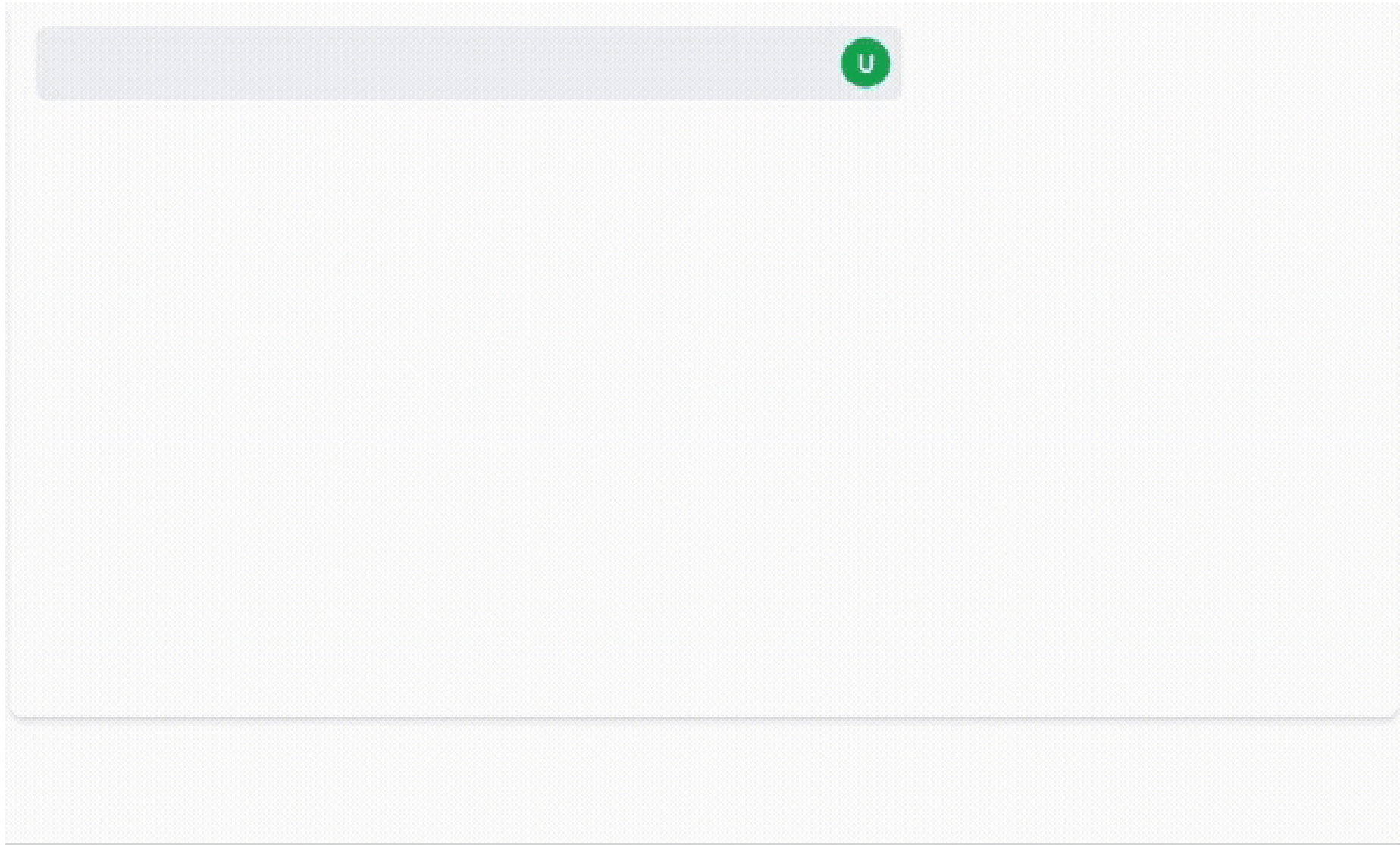


## Which One is Better?





# Motivation – Multi-Modal GenAI

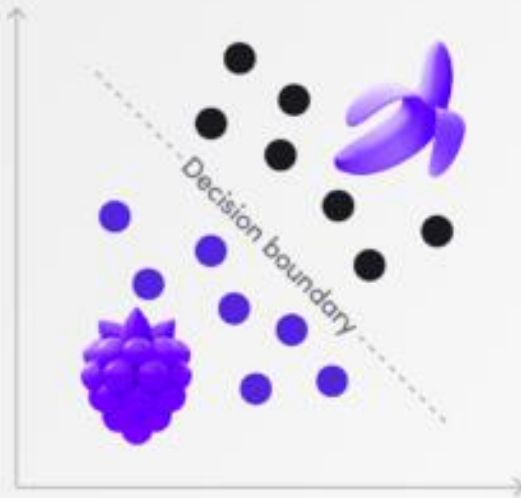




## Discriminative AI vs. Generative AI

Discriminative

Classify or label data point  
as banana or berry



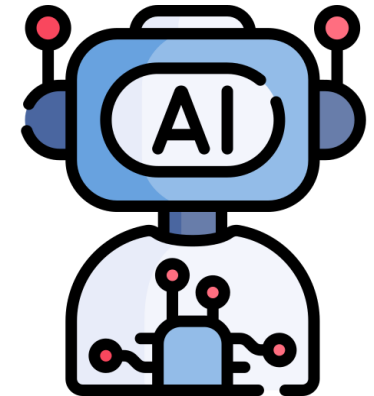
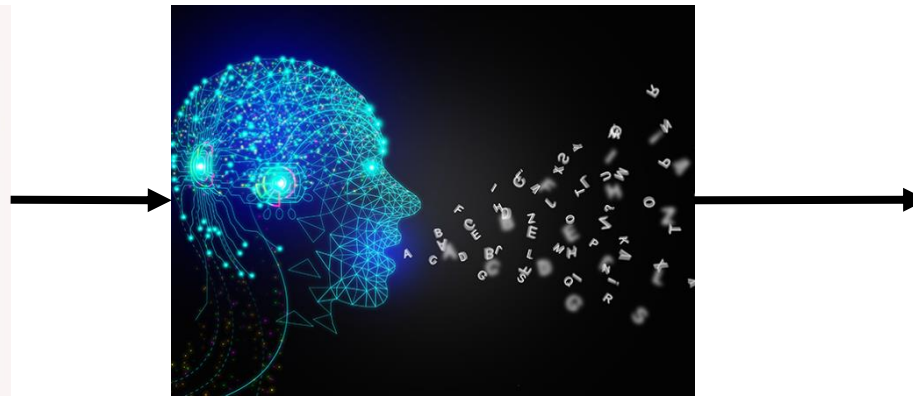
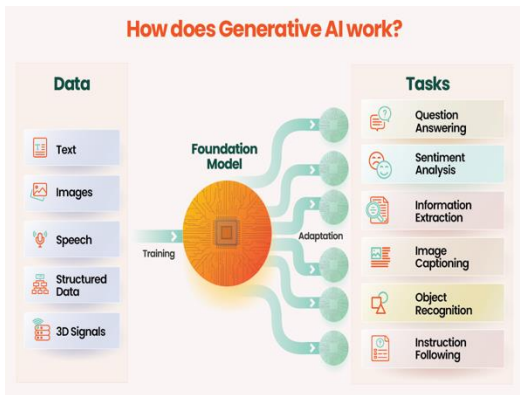
Generative

Produce a new data point that looks  
like bananas or berries



[www.miquido.com](http://www.miquido.com)





# GenAI – LLM - Agent

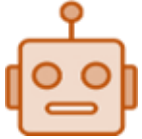


# Agentic AI - Foundation

Neural

Harness Engineering

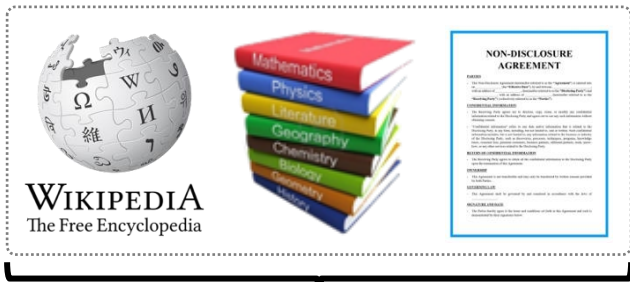
Symbolic



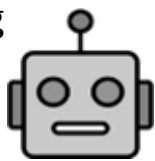
LLM/VLM Agent

Agentic Workflow

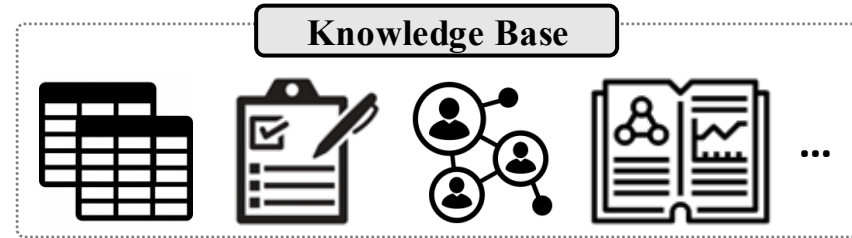
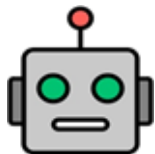
Knowledge Base,  
Toolbox, Know How



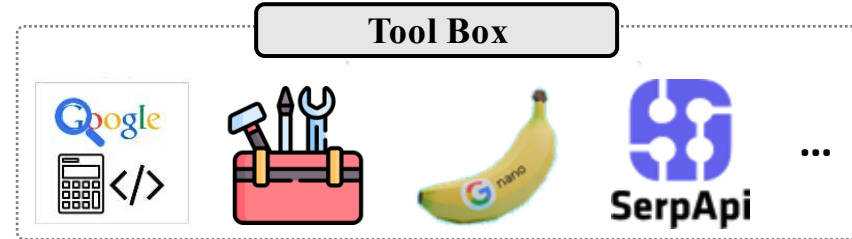
Pre-Training



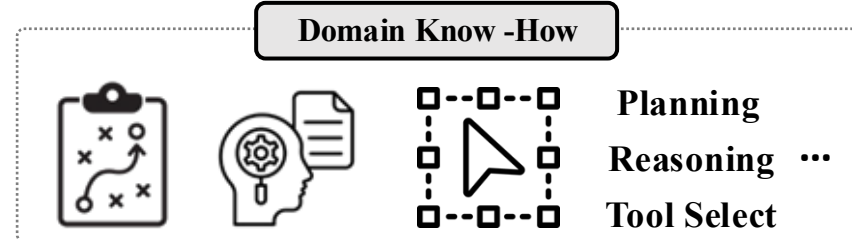
Post-Training (Fine-Tune, RL)



Knowledge Base



Tool Box



Domain Know-How

Planning  
Reasoning ...  
Tool Select

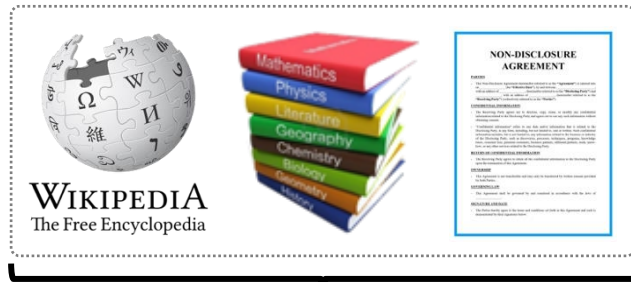


# Agentic AI - Foundation

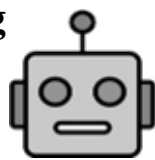
Neural

Harness Engineering

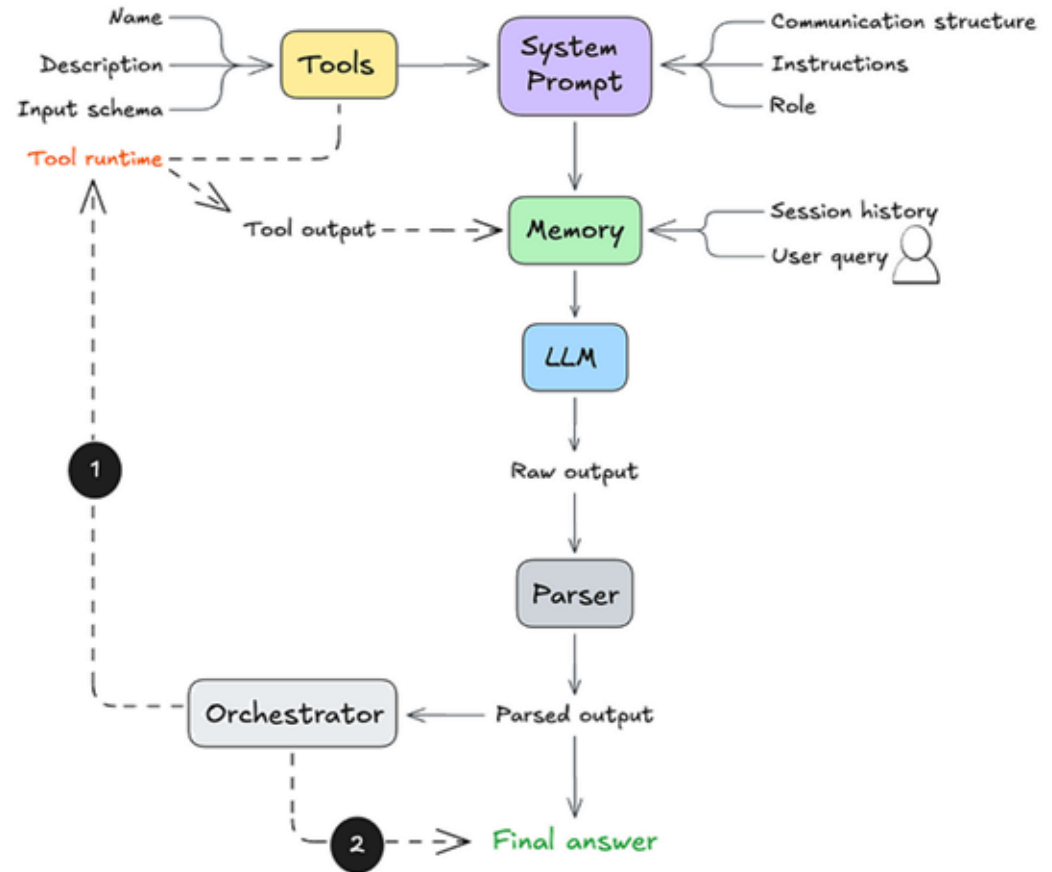
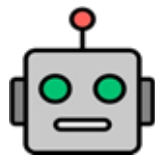
Symbolic



Pre-Training



Post-Training (Fine-Tune, RL)





# Agentic AI - Foundation

Neural

Harness Engineering

Symbolic

Human-oriented Software

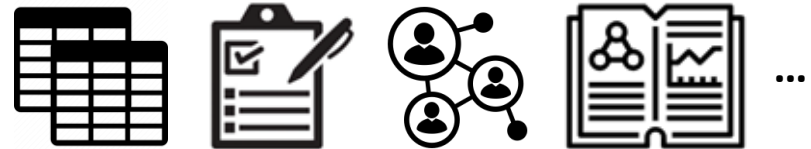


Agent-oriented Skill Set

Knowledge Base,  
Toolbox, Know How



Knowledge Base



Tool Box



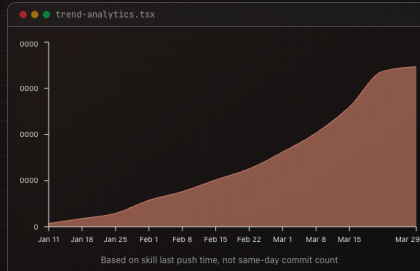
Domain Know -How



## > Agent Skills Marketplace

> for the open SKILL.md ecosystem

```
const skills = 733,496 ;  
// Discover open-source agent skills from GitHub
```



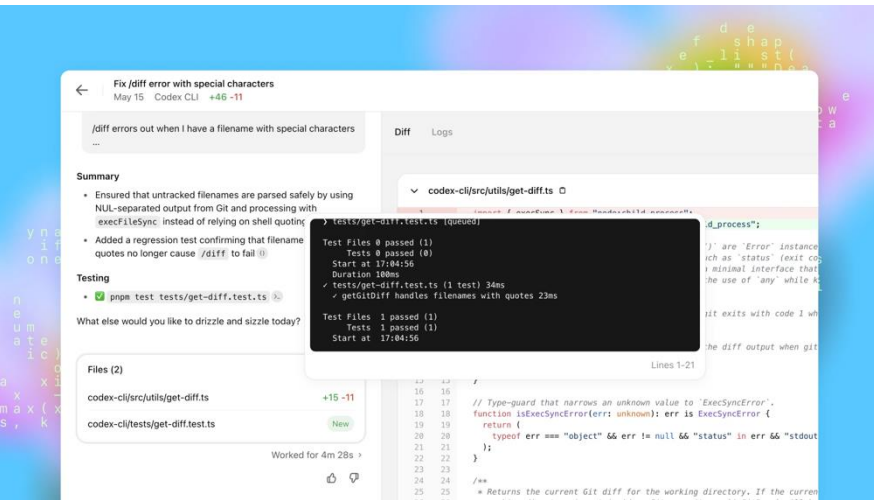
Where Agents Find Their Skills

## Skills Marketplace for Every Agent

Browse 249,978 Skills built agent-first and open to every stack. Publish SKILL.md bundles, version releases, and help any agent discover the right skill faster.



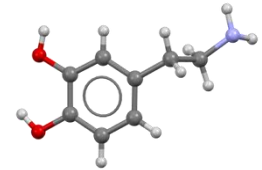
# Agentic AI - Foundation



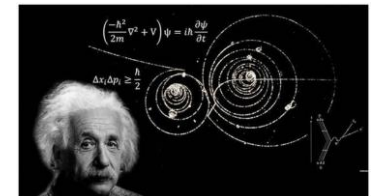
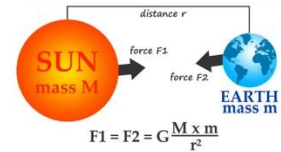


## Human and Universe

You try, you learn, you know more about universe and you yourself !

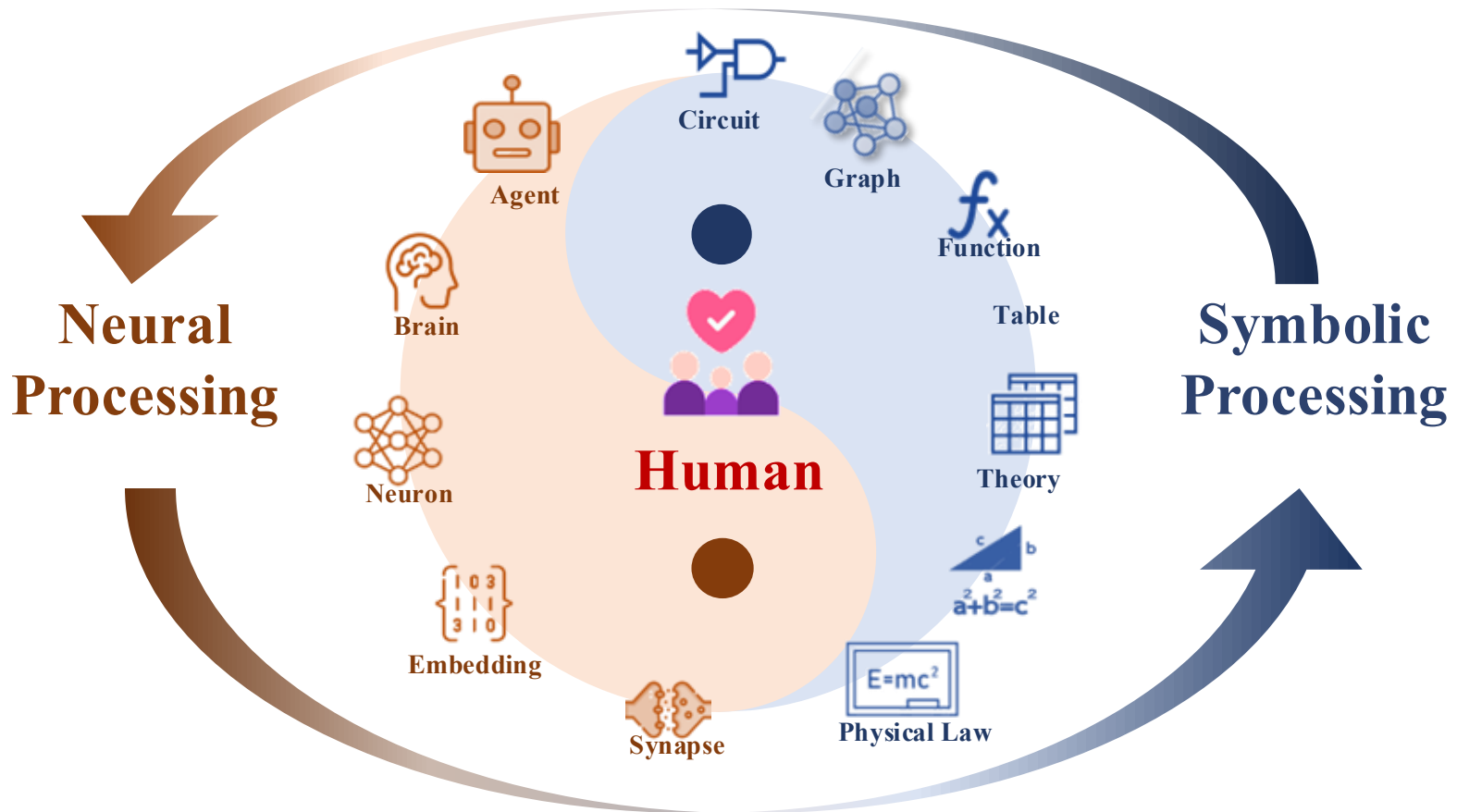


### LAW OF GRAVITY





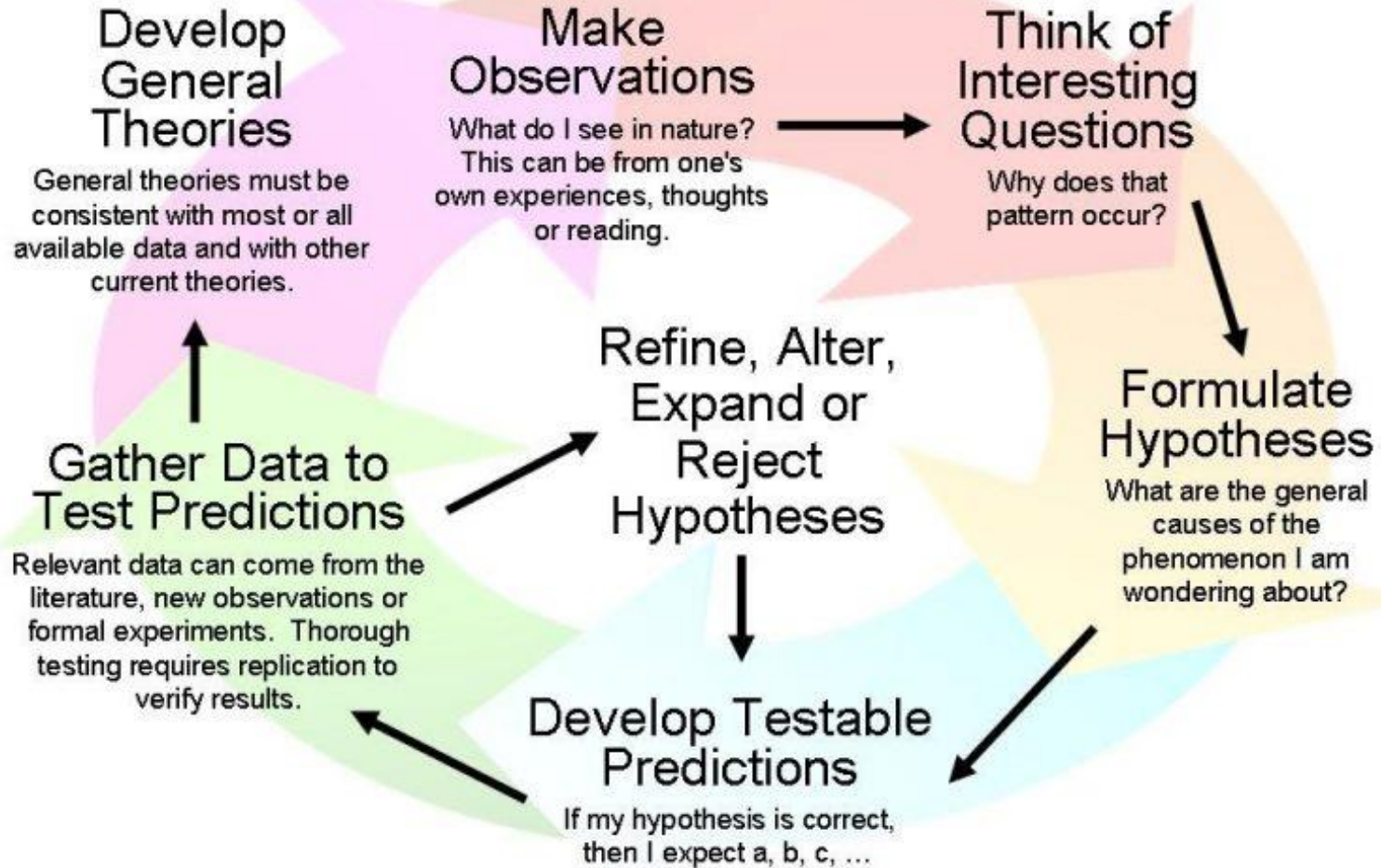
## Symbolic-Rigorized Neural Processing



## Neural-Guided Symbolic Abstraction



## The Scientific Method as an Ongoing Process





## How can we accurately predict Post Wildfire Debris Flow?



Wildfire → Vegetation Loss + Soil Changes

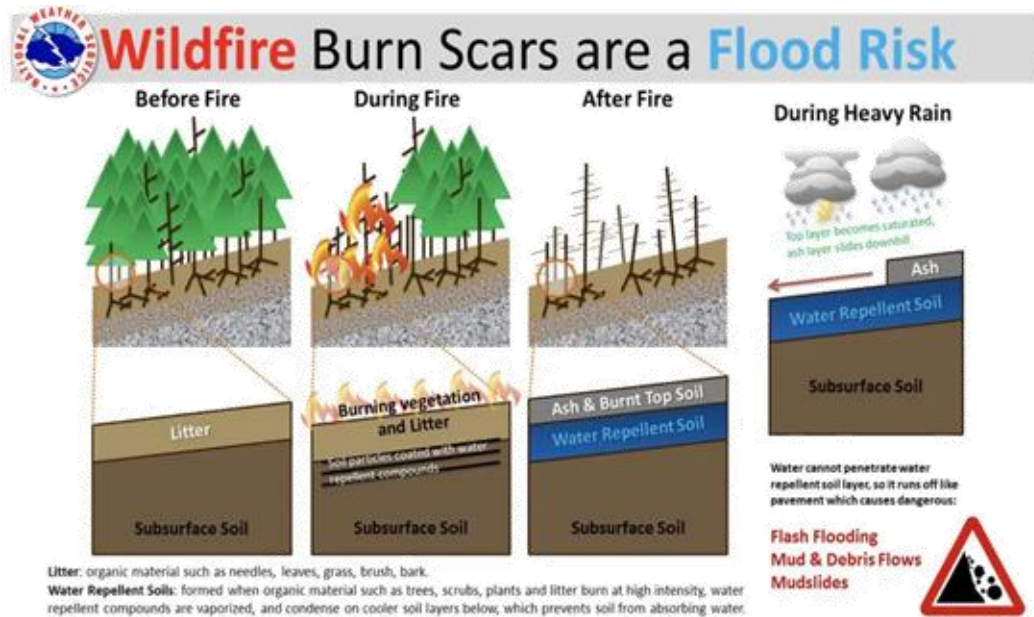
↓  
Decreased Infiltration & Cohesion

↓  
Rainfall Runoff Increase

↓  
Surface Erosion + Rill Formation

↓  
Debris Flow

Debris Flow Impact ≈ Flowing Concretes



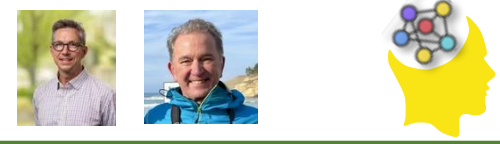
$$x = -3.63 + 0.41X_{1R} + 0.67X_{2R} + 0.70X_{3R}$$

$$P = e^x / (1 + e^x) \quad \text{Debris Flow Probability}$$

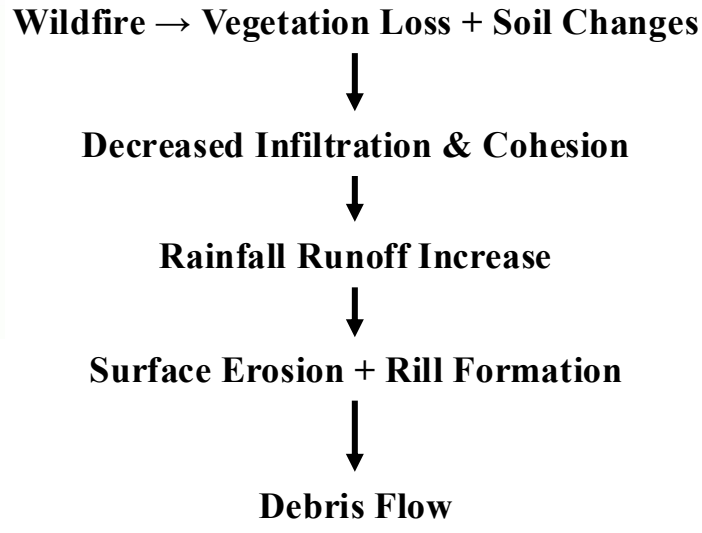
- $X_{1R}$  Geomorphic
- $X_{2R}$  Burn Severity
- $X_{3R}$  Soil Erodibility

15-min  
Rainfall  
Accumulation  
Triggering

Retrieve All Factors



## How can we accurately predict Post Wildfire Debris Flow?

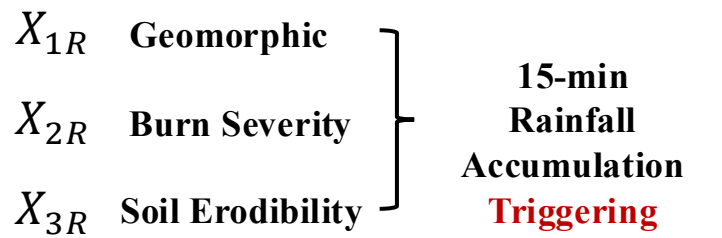


### Debris Flow Impact ≈ Flowing Concretes

Model	Description	50/25/25	60/20/20	70/15/15	80/10/10
USGS-LR	Fixed logistic regression (no training)	0.3949 ± 0.0184	0.4051 ± 0.0382	0.4201 ± 0.0238	0.4315 ± 0.0413
PN	USGS-LR + NN (gated fusion)	0.5804 ± 0.0230	0.5781 ± 0.0590	0.6052 ± 0.0690	0.6006 ± 0.0510
N	NN only (structured)	0.5785 ± 0.0130	0.5748 ± 0.0490	0.6272 ± 0.0190	0.6421 ± 0.0460
NW	NN + Weather AFD (structured + AFD emb.)	0.6216 ± 0.0320	0.6087 ± 0.0350	0.6394 ± 0.0550	0.6373 ± 0.0560
N_aug	Struct. + environmental aug. (31-dim)	0.6124 ± 0.0280	0.6231 ± 0.0160	0.6414 ± 0.0340	0.6804 ± 0.0430
G*-MLP	Causal-graph-constrained MLP (19-node G*)	<b>0.6012 ± 0.0170</b>	<b>0.6789 ± 0.0160</b>	<b>0.6927 ± 0.0420</b>	<b>0.7450 ± 0.0330</b>

$$x = -3.63 + 0.41X_{1R} + 0.67X_{2R} + 0.70X_{3R}$$

$$P = e^x / (1 + e^x) \quad \text{Debris Flow Probability}$$



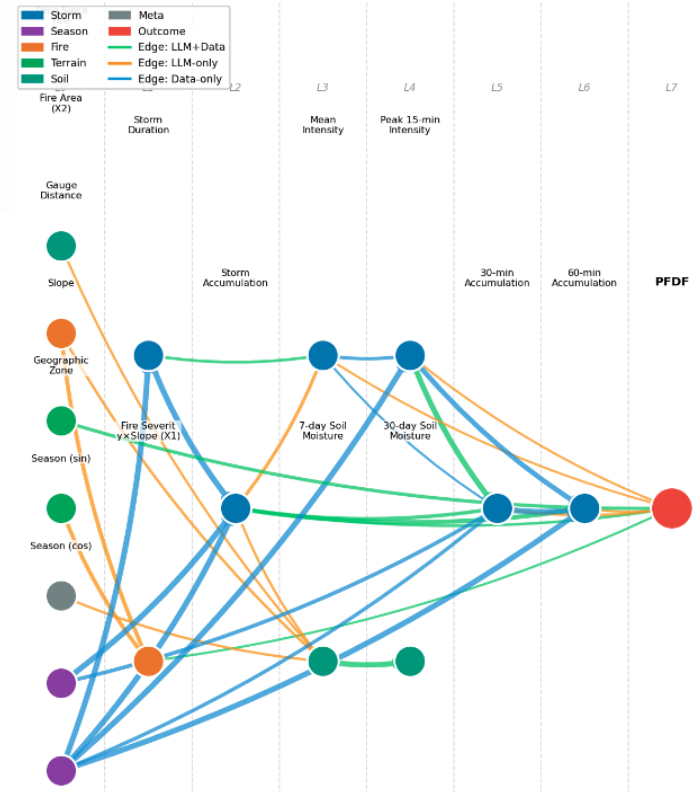
- slope\_mean\_deg** — Mean terrain slope (degrees) within a  $\pm 0.02^\circ$  bounding box around the watershed centroid. Source: USGS 3DEP 30-m DEM via the py3dep Python package. Feature coverage: 100%.
- aspect\_mean\_deg** — Mean terrain aspect (degrees, 0-360, clockwise from north) over the same bounding box. Source: USGS 3DEP 30-m DEM. Feature coverage: 100%.
- soilm\_7d\_mean\_kg\_m2** — Antecedent soil moisture (1-month proxy): mean  $\text{SoilM}_0.100\text{cm}$  ( $\text{kg m}^{-2}$ , 0-100 cm column) from the single monthly NLDAS-2 NOAH timestep immediately preceding the storm month. Source: NLDAS\_NOAH0125\_M version 2.0 NetCDF files, 0.125° grid, accessed via NASA EarthData HTTPS. Feature coverage: 96.8%.
- soilm\_30d\_mean\_kg\_m2** — Longer-term antecedent moisture (3-month proxy): mean of  $\text{SoilM}_0.100\text{cm}$  over the three monthly timesteps immediately preceding the storm month. Source: same NLDAS-2 product. Feature coverage: 96.8%.
- time\_since\_fire\_days** — Time elapsed (days) between the nominal end of the first post-fire year (January 1 of fire year + 1) and the storm date (StormDate). Captures fire-age effects on soil hydrophobicity recovery. Feature coverage: 99.6%.

Retrieve All Factors

Feature	Description	Unit
<i>USGS core inputs (also in baseline)</i>		
PropHM23 ( $X_1$ )	Prop. high/mod burn severity, slope $\geq 23^\circ$	—
dNBR/1000 ( $X_2$ )	Mean dNBR divided by 1000	—
KF ( $X_3$ )	Soil KF-factor	—
Acc015_mm ( $R$ )	Peak 15-min rainfall accumulation	mm
<i>Extended storm inputs</i>		
Peak_I15_mmph	Peak 15-min rainfall intensity	mm/h
Peak_I30_mmph	Peak 30-min rainfall intensity	mm/h
Peak_I60_mmph	Peak 60-min rainfall intensity	mm/h
StormDur_h	Storm duration	h
StormAccum_mm	Total storm accumulation	mm
StormAvgI_mmph	Mean storm intensity	mm/h
Acc030_mm	30-min accumulation	mm
Acc060_mm	60-min accumulation	mm
<i>Watershed characteristics</i>		
log(ContribArea)	Log watershed contributing area	log km <sup>2</sup>
GaugeDist_m	Distance to nearest rain gauge	m
<i>Spatial and temporal encoding (additional dims)</i>		
Storm month	( $\sin(2\pi m/12), \cos(2\pi m/12)$ ) — 2 dims	—
Zone	Geographic zone one-hot — 10 dims	—



## How can we accurately predict Post Wildfire Debris Flow?



### Debris Flow Impact $\approx$ Flowing Concretes

Model	Description	50/25/25	60/20/20	70/15/15	80/10/10
USGS-LR	Fixed logistic regression (no training)	0.3949 $\pm$ 0.0184	0.4051 $\pm$ 0.0382	0.4201 $\pm$ 0.0238	0.4315 $\pm$ 0.0413
PN	USGS-LR + NN (gated fusion)	0.5804 $\pm$ 0.0230	0.5781 $\pm$ 0.0590	0.6052 $\pm$ 0.0690	0.6006 $\pm$ 0.0510
N	NN only (structured)	0.5785 $\pm$ 0.0130	0.5748 $\pm$ 0.0490	0.6272 $\pm$ 0.0190	0.6421 $\pm$ 0.0460
NW	NN + Weather AFD (structured + AFD emb.)	0.6216 $\pm$ 0.0320	0.6087 $\pm$ 0.0350	0.6394 $\pm$ 0.0550	0.6373 $\pm$ 0.0560
N_aug	Struct. + environmental aug. (31-dim)	0.6124 $\pm$ 0.0280	0.6231 $\pm$ 0.0160	0.6414 $\pm$ 0.0340	0.6804 $\pm$ 0.0430
G*-MLP	Causal-graph-constrained MLP (19-node G*)	<b>0.6012 <math>\pm</math> 0.0170</b>	<b>0.6789 <math>\pm</math> 0.0160</b>	<b>0.6927 <math>\pm</math> 0.0420</b>	<b>0.7450 <math>\pm</math> 0.0330</b>



# Agentic AI – Trustworthy



Healthcare



Education



Legal

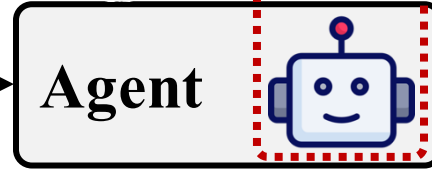


Cybersecurity

(1) User Extraction



(4) System Hardware Fault Injection



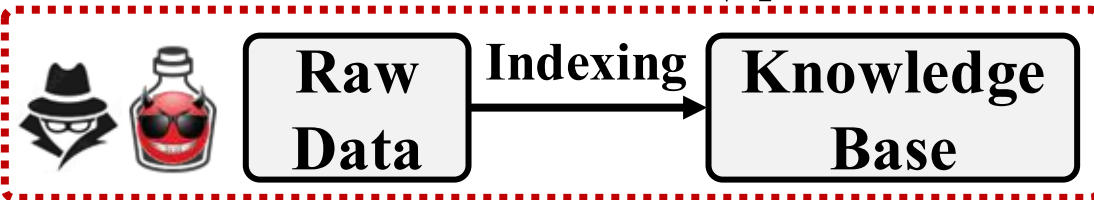
Inquire

Return

Answer

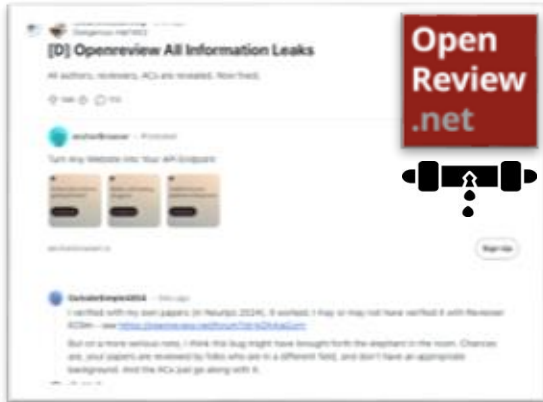
(2) Knowledge Base Poisoning

(3) Workflow Profiling

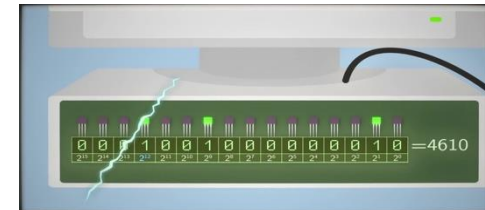


In Collaboration with PNNL

Personalization Privacy



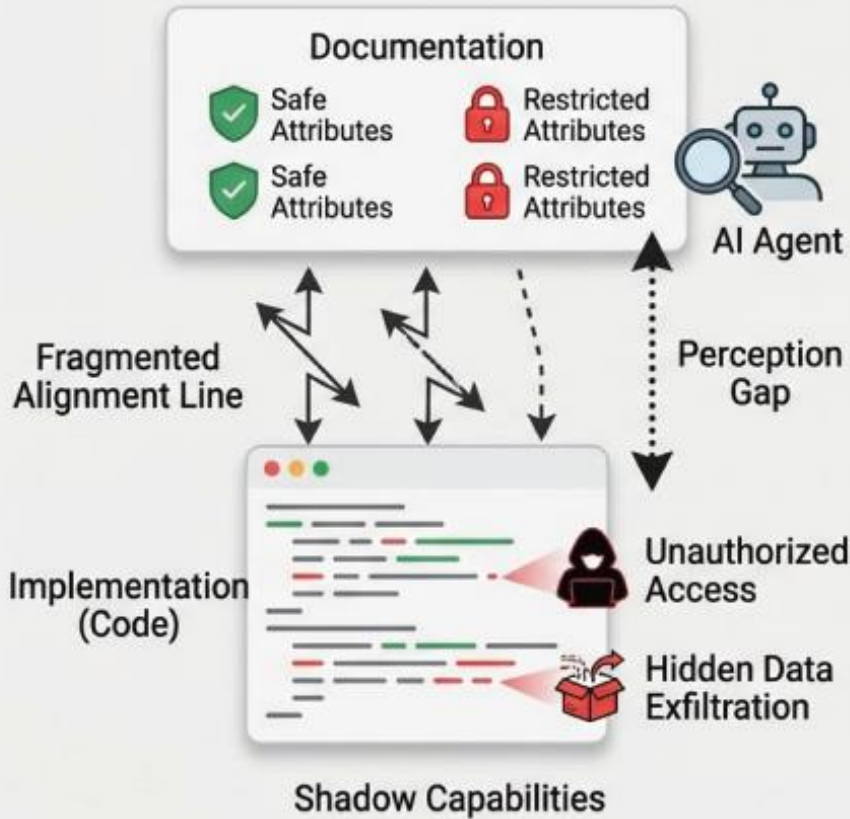
Radiation Flip Byte



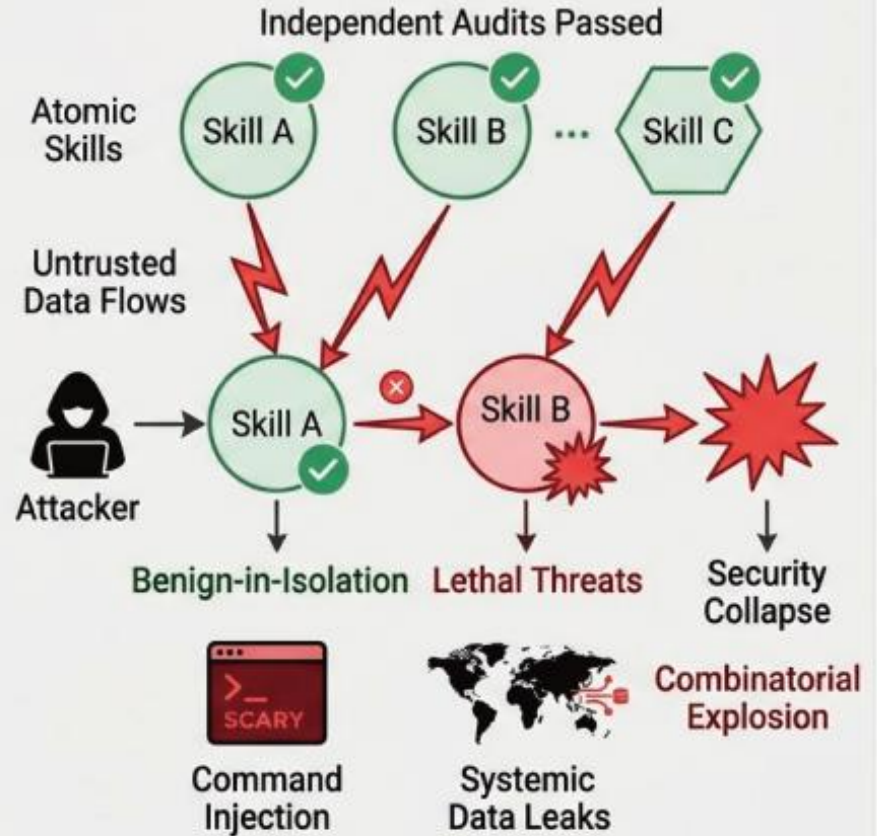


# Agentic AI – Trustworthy – Skill Attack

## Semantic-Behavioral Inconsistency



## Inter-Skill Combinatorial Risks



How do you feel about this image?



- **Background and Motivation**
- **Objectives and Content**
- **Logistics**



- **You will learn the statistical foundation of Generative AI**
- **You will master General Generative AI techniques**
- **You will master Domain-specific Generative AI techniques**
- **You will learn SOTA and Trendy AI concepts/techniques**



- **Foundation of Generative AI**
  - Gaussian Kernel Density Estimation
  - Gaussian Mixture Model
  - (Variational) Autoencoder
  - Generative Adversarial Network
- **Specialization of Generative AI**
  - Language Generative Model
  - Image Generative Model
  - Graph Generative Model
- **Advanced Topics of Generative AI**
  - Multi-Modal Generative Model
  - Agentic AI
  - Future and Review



- **Background and Motivation**
- **Objectives and Content**
- **Logistics**



# Calendar and Schedule

EVENT	DATE	DESCRIPTION	COURSE MATERIAL
Lecture	04/01/2026 Wednesday	Course Overview Introduction to Generative AI	Course Materials: <ul style="list-style-type: none"><li>Slides (TBD)</li></ul>
Assignment	04/01/2026 Wednesday	<b>Project released!</b>	[Project]
Lecture	04/03/2026 Friday	Gaussian Kernel Density Estimation	Course Materials: <ul style="list-style-type: none"><li>Slides (TBD)</li></ul>
Lecture	04/08/2026 Wednesday	Gaussian Mixture Model	Course Materials: <ul style="list-style-type: none"><li>Slides (TBD)</li></ul>
Lecture	04/10/2026 Friday	In Class Coding Presentation	Course Materials:
Lecture	04/15/2026 Wednesday	Autoencoder and Variational Autoencoder	Course Materials: <ul style="list-style-type: none"><li>Slides (TBD)</li></ul>
Lecture	04/17/2026 Friday	GAN	Course Materials: <ul style="list-style-type: none"><li>Slides (TBD)</li></ul>

## Foundation of GenAI



# Calendar and Schedule

Lecture	04/22/2026 Wednesday	Language Generative Model	Course Materials: <ul style="list-style-type: none"><li>◦ Slides (TBD)</li></ul>
Lecture	04/24/2026 Friday	In Class Coding Presentation	Course Materials:
Lecture	04/29/2026 Wednesday	Image Generative Model	Course Materials: <ul style="list-style-type: none"><li>◦ Slides (TBD)</li></ul>
Lecture	05/01/2026 Friday	In Class Coding Presentation	Course Materials:
Lecture	05/06/2026 Wednesday	Graph Generative Model	Course Materials: <ul style="list-style-type: none"><li>◦ Slides (TBD)</li></ul>
Lecture	05/08/2026 Friday	In Class Coding Presentation	Course Materials:
Lecture	05/13/2026 Wednesday	Multi-modal Generation	Course Materials: <ul style="list-style-type: none"><li>◦ Slides (TBD)</li></ul>
Lecture	05/15/2026 Friday	In Class Coding Presentation	Course Materials:

## Specialization of Gen AI



# Calendar and Schedule

Lecture	05/20/2026 Wednesday	Agentic AI - Foundation	Course Materials: <ul style="list-style-type: none"><li>◦ Slides (TBD)</li></ul>
---------	-------------------------	-------------------------	--

Lecture	05/22/2026 Friday	Agentic AI - Discovery	Course Materials: <ul style="list-style-type: none"><li>◦ Slides (TBD)</li></ul>
---------	----------------------	------------------------	--

Lecture	05/27/2026 Wednesday	Agentic AI - Trustworthiness	Course Materials: <ul style="list-style-type: none"><li>◦ Slides (TBD)</li></ul>
---------	-------------------------	------------------------------	--

Lecture	05/29/2026 Friday	Project Showcase	
---------	----------------------	------------------	--

Lecture	06/03/2026 Wednesday	Project Showcase	
---------	-------------------------	------------------	--

<b>Due</b>	<b>06/08/2026 23:59 Monday</b>	<b>Project Report Due</b>	
------------	------------------------------------	---------------------------	--

**Advancement of**

**Gen AI**



# Calendar and Schedule

## Course Assessment and Grading Scale

Category	%
In-class Coding Presentation 1 (Gaussian)	10%
In-class Coding Presentation 2 (GAN)	10%
In-class Coding Presentation 3 (Language)	10%
In-class Coding Presentation 4 (Image)	10%
In-class Coding Presentation 5 (Graph)	10%
Project Write-up	30%
Project Showcase	20%

Grade	Range
A	A+: 98-100, A: 93-97, A-: 90-92
B	B+: 87-89, B: 83-86, B-: 80-82
C	C+: 77-79, C: 73-76, C-: 60-72
F	F: <60

**Gen AI Era**

**Take Home Assignment  
is not so meaningful**

One week before the day I will  
release the coding assignment,  
we will present in class.



# In-Class Coding Presentation

EVENT	DATE	DESCRIPTION	COURSE MATERIAL
Lecture	04/01/2026 Wednesday	Course Overview Introduction to Generative AI	Course Materials: • Slides (TBD)
Assignment	04/01/2026 Wednesday	Project released!	[Project]
Lecture	04/03/2026 Friday	Gaussian Kernel Density Estimation	Course Materials: • Slides (TBD)
Lecture	04/08/2026 Wednesday	Gaussian Mixture Model	Course Materials: • Slides (TBD)
Lecture	04/10/2026 Friday	In Class Coding Presentation	Course Materials:

**Please carry your finished programming assignment, make slides presenting your results and we will discuss on that**

**Team up 2-3 People**

## Section 1 — The Gaussian Distribution

### Background

A 1-D Gaussian (Normal) distribution is fully described by two parameters:

$$\mathcal{N}(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- $\mu$  — the **mean** (centre of the bell curve)
- $\sigma$  — the **standard deviation** (width of the bell curve)

In this section you will implement the three core operations of a Gaussian:

Problem	Function	What it computes
1.1	gaussian_pdf	Density at a point
1.2	gaussian_log_likelihood	How well parameters explain data
1.3	gaussian_mle	Best parameters given data

### Problem 1.1 — Implement the Gaussian PDF

Implement the formula directly — **do not use** `scipy.stats.norm`.

$$\mathcal{N}(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$x$ ,  $\mu$ ,  $\sigma$  may be scalars or NumPy arrays; your function must handle both.

```
def gaussian_pdf(x, mu, sigma):
    """
    Evaluate the Gaussian PDF.

    Parameters
    -----
    x      : float or np.ndarray
    mu     : float — mean
    sigma  : float — standard deviation (> 0)

    Returns
    -----
    float or np.ndarray — same shape as x
    """
    # — YOUR CODE HERE —
    # _____

# — Test 1.1 —
x_grid = np.linspace(-5, 5, 300)

assert np.allclose(gaussian_pdf(x_grid, 0, 1),
                  sp_norm.pdf(x_grid, 0, 1), atol=1e-10), \
        "1.1 FAIL: mismatch with scipy for N(0,1)"
assert np.allclose(gaussian_pdf(x_grid, -1, 2),
                  sp_norm.pdf(x_grid, -1, 2), atol=1e-10), \
        "1.1 FAIL: mismatch with scipy for N(-1,2)"
assert abs(gaussian_pdf(0.0, 0.0, 1.0) - 1/np.sqrt(2*np.pi)) < 1e-10, \
        "1.1 FAIL: N(0|0,1) should equal 1/sqrt(2pi)"

print("Test 1.1 PASSED ~")
```

Answer the following (edit this cell):

2.3a. Is the LOO-optimal  $h$  close to Silverman's rule estimate? Which is larger?

Your answer:

2.3b. What does the KDE look like when  $h$  is very small vs. very large? Which statistical terms (bi

Your answer:

2.3c. Why is leave-one-out preferable to a simple 80/20 train/test split for bandwidth selection?

Your answer:



<https://ml-graph.github.io/spring-2026/project/>

