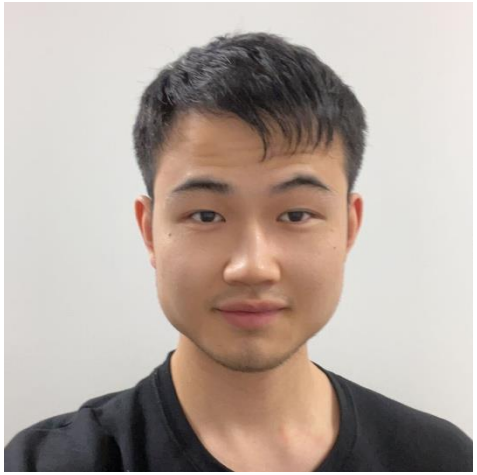
Adv ML for Gen-AI

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

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

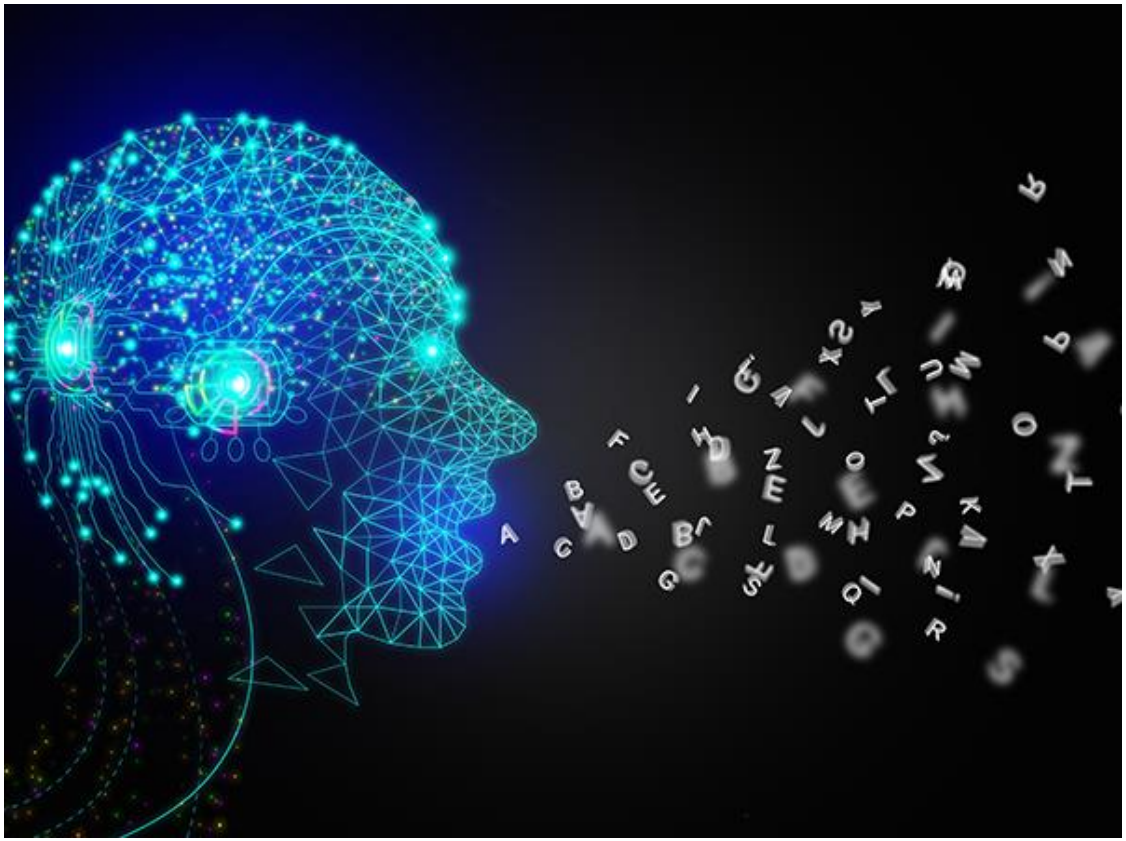
Outline

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

Outline

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

Background



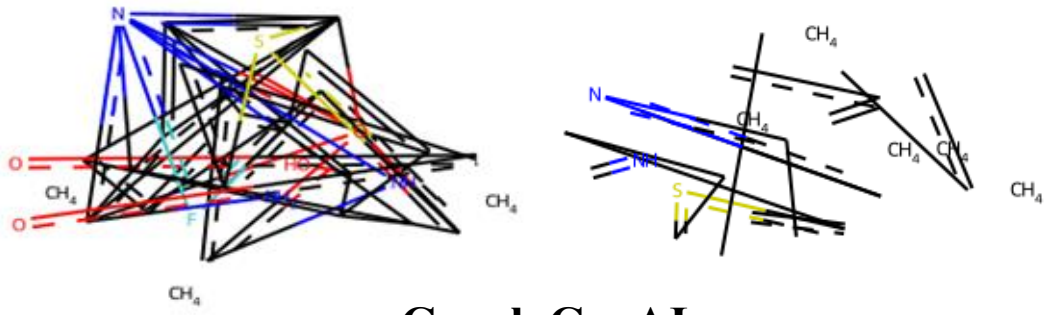
Generative AI has been a long-standing topic

Machines that can produce human-like speech

Background

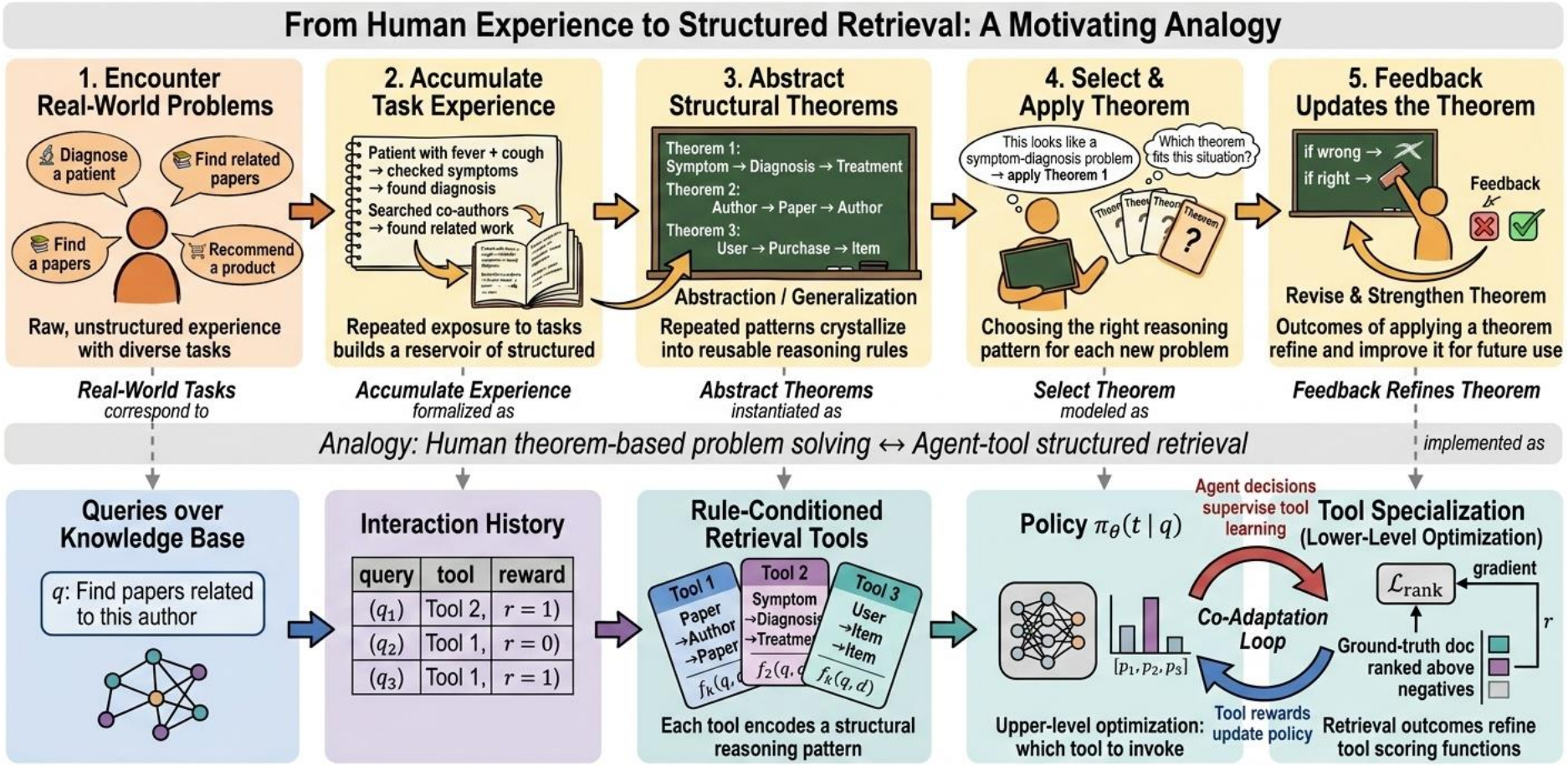
Video GenAI

Image GenAI



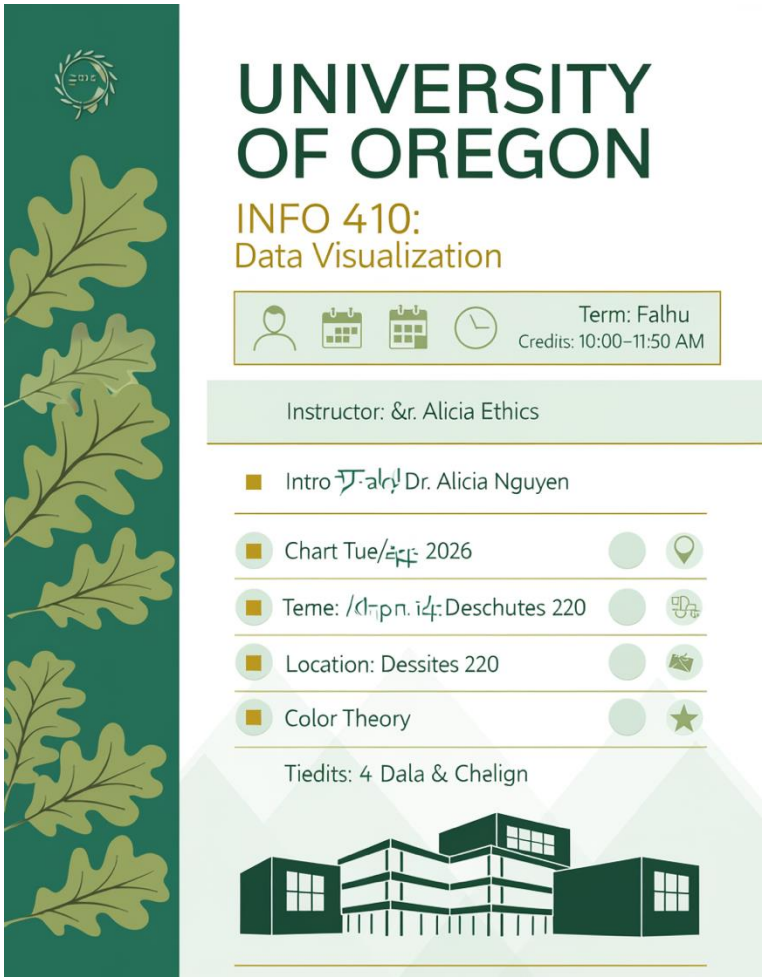
Graph GenAI

Motivation – Image GenAI



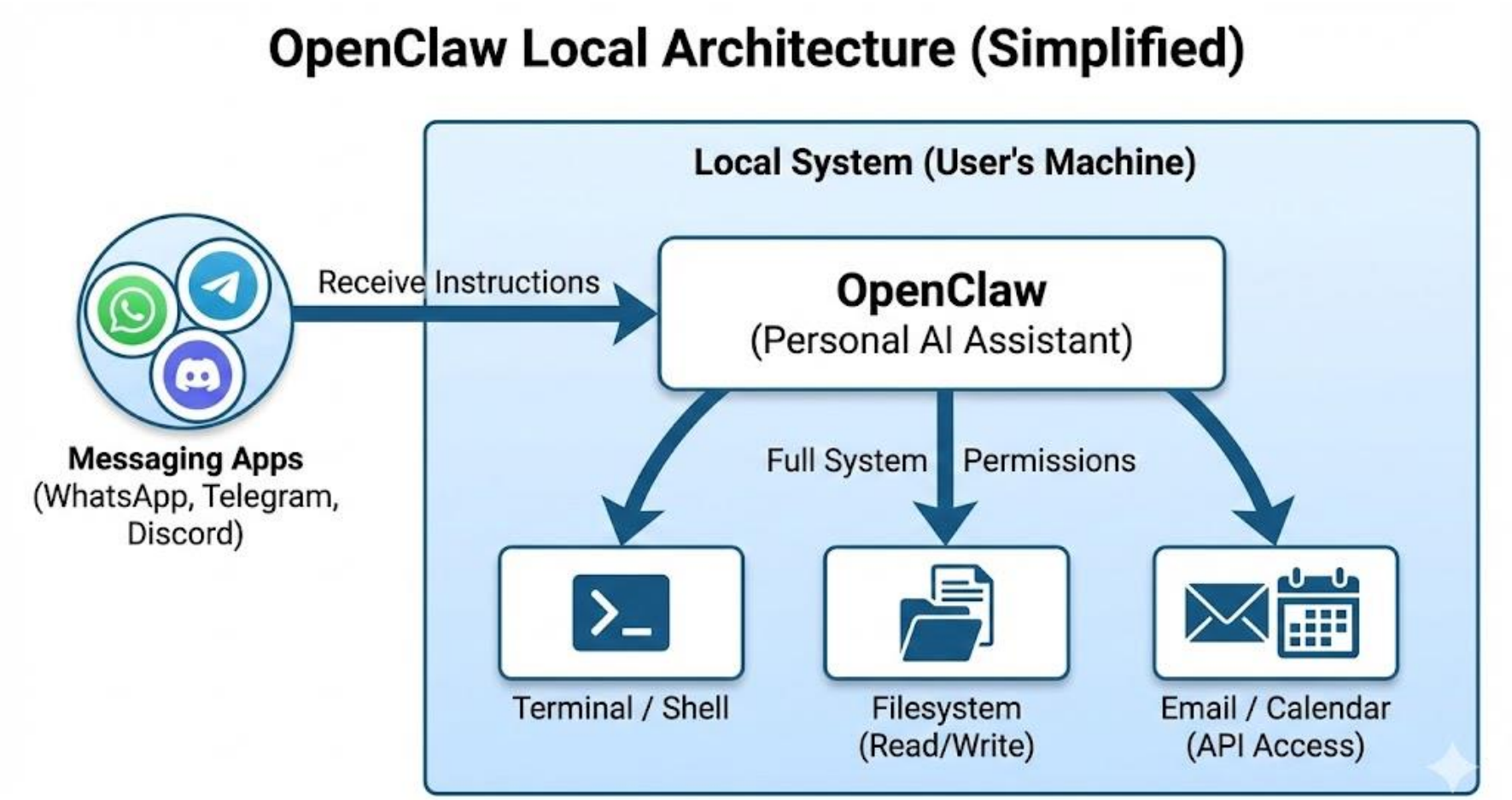
Scientific Figure - Nano Banana

Motivation – Image GenAI



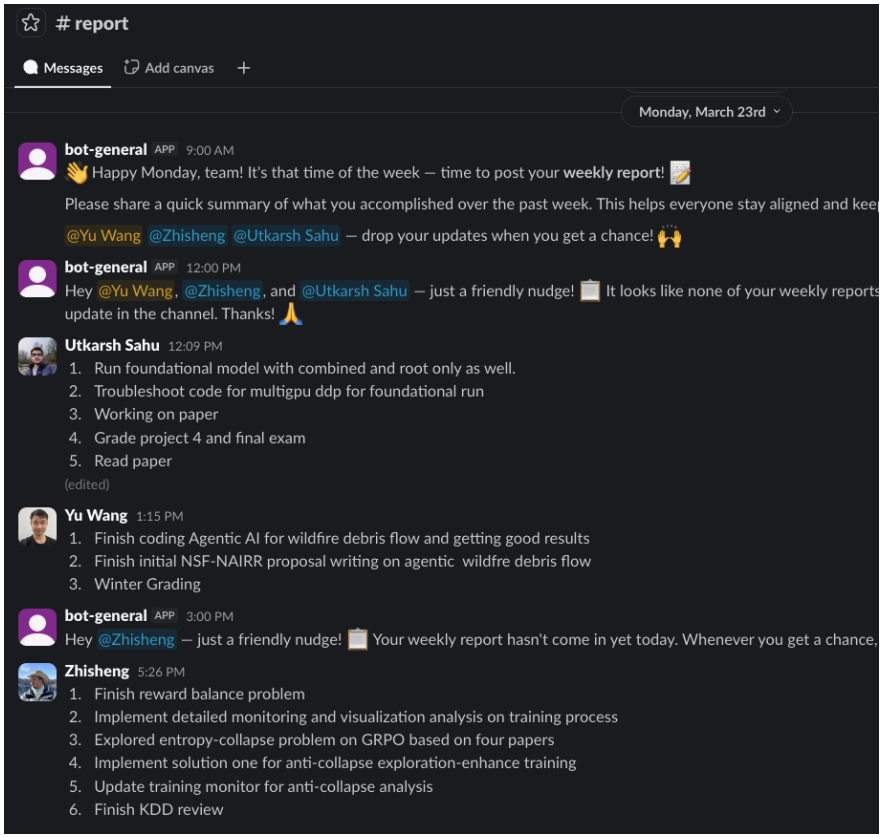
Adobe Photoshop/Firefly

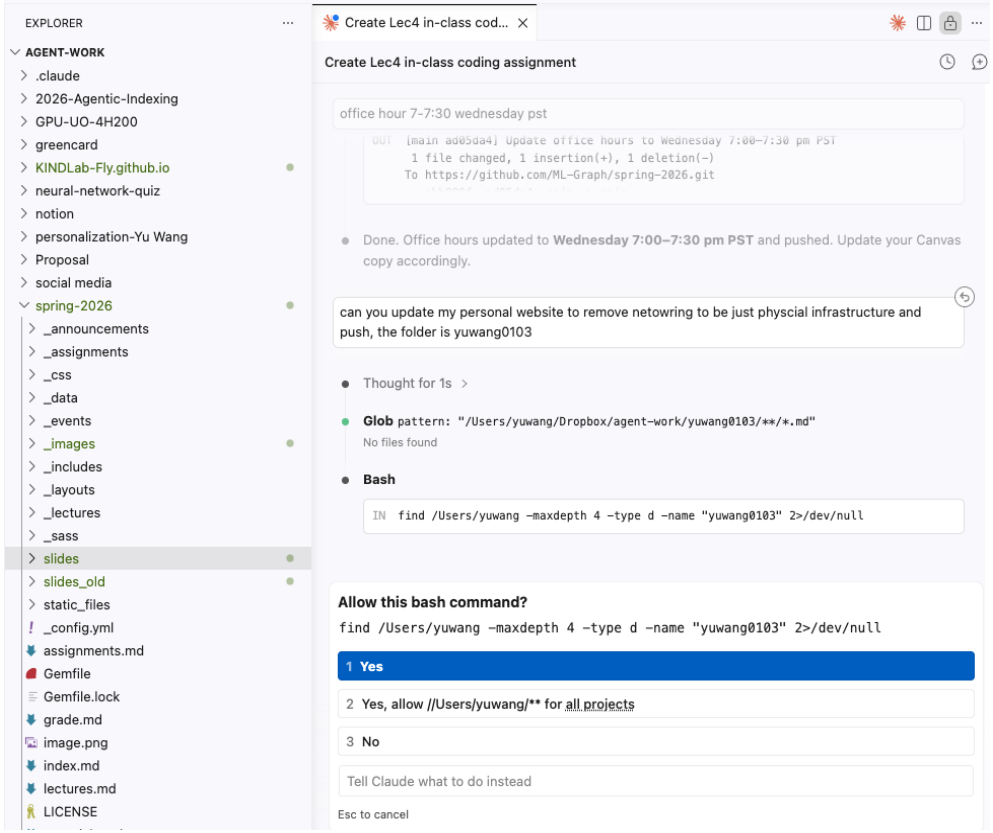
Motivation – Language GenAI



OpenClaw

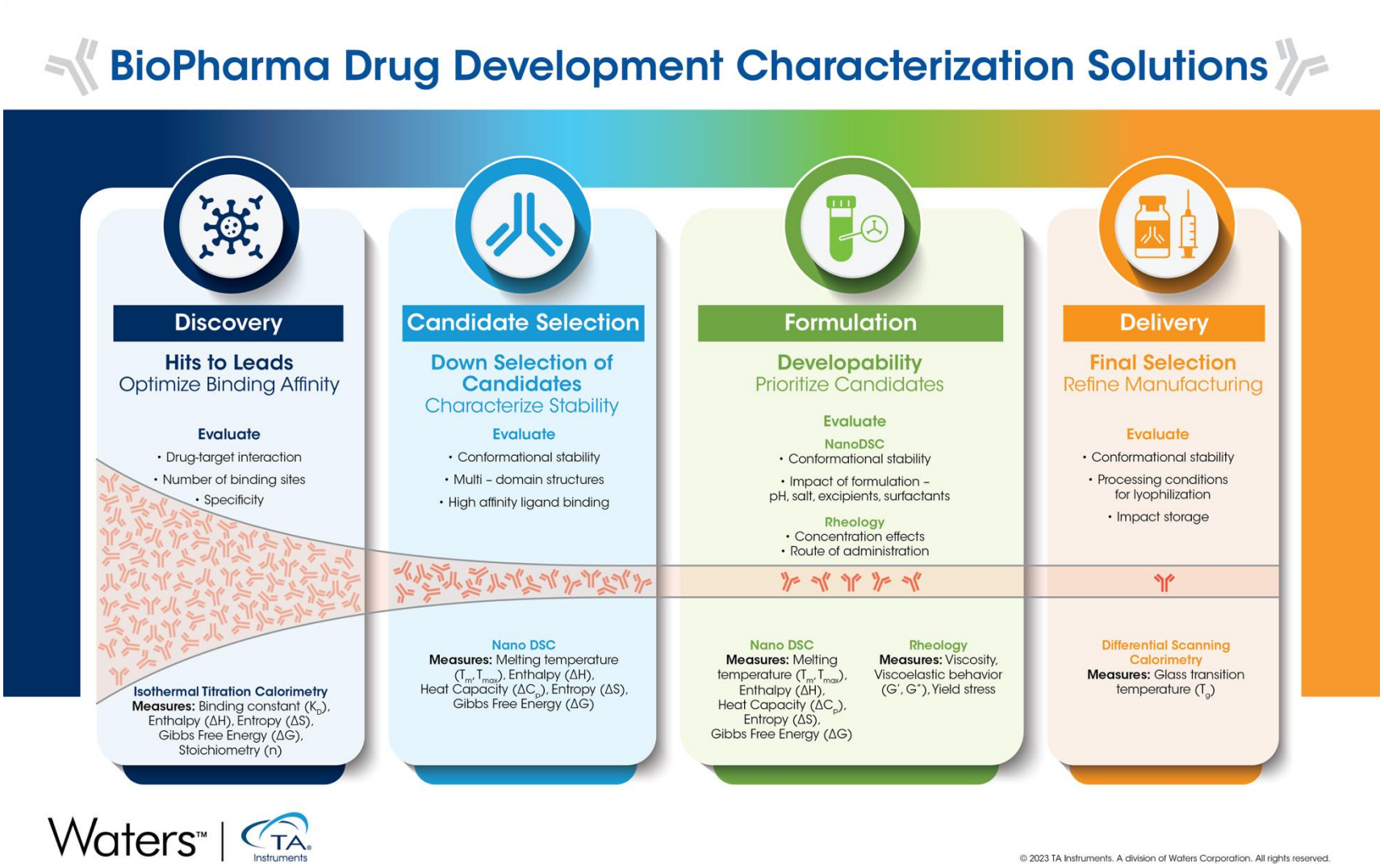
Motivation – Language GenAI



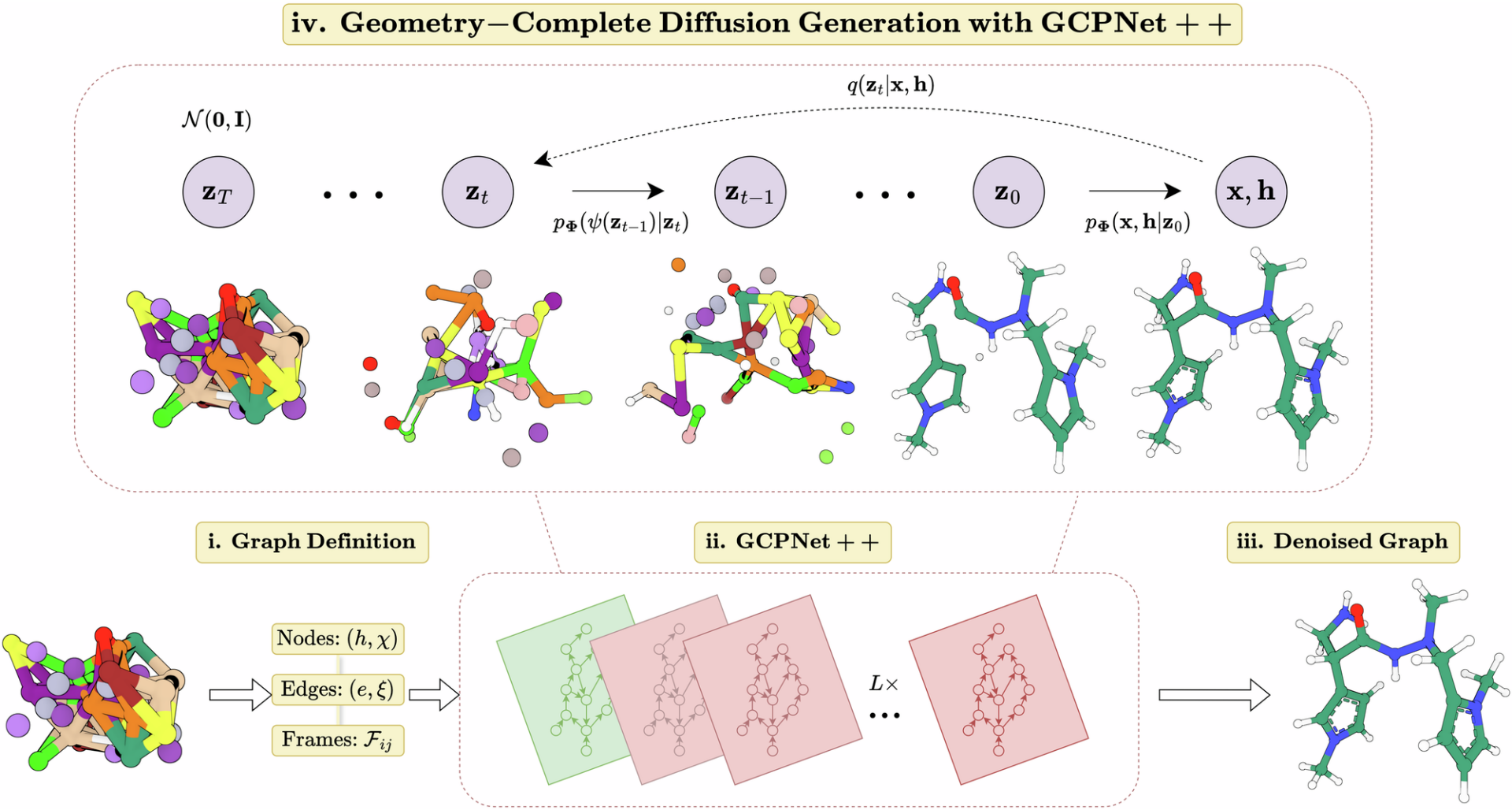


Personal Assistant

Motivation – Graph GenAI



Motivation – Graph GenAI



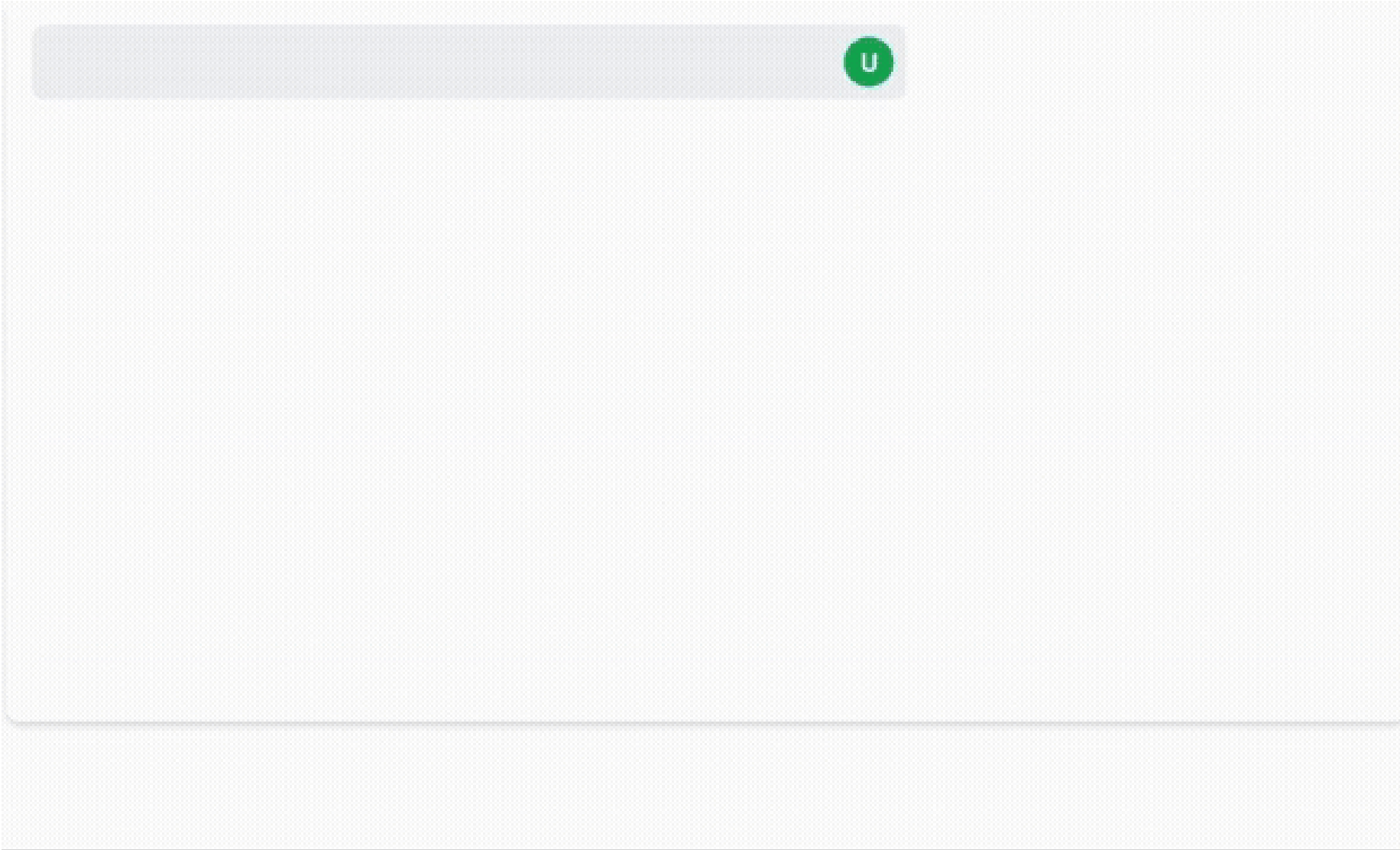
Motivation – Music GenAI

https://music.youtube.com/watch?v=wDIXoS1z8iI

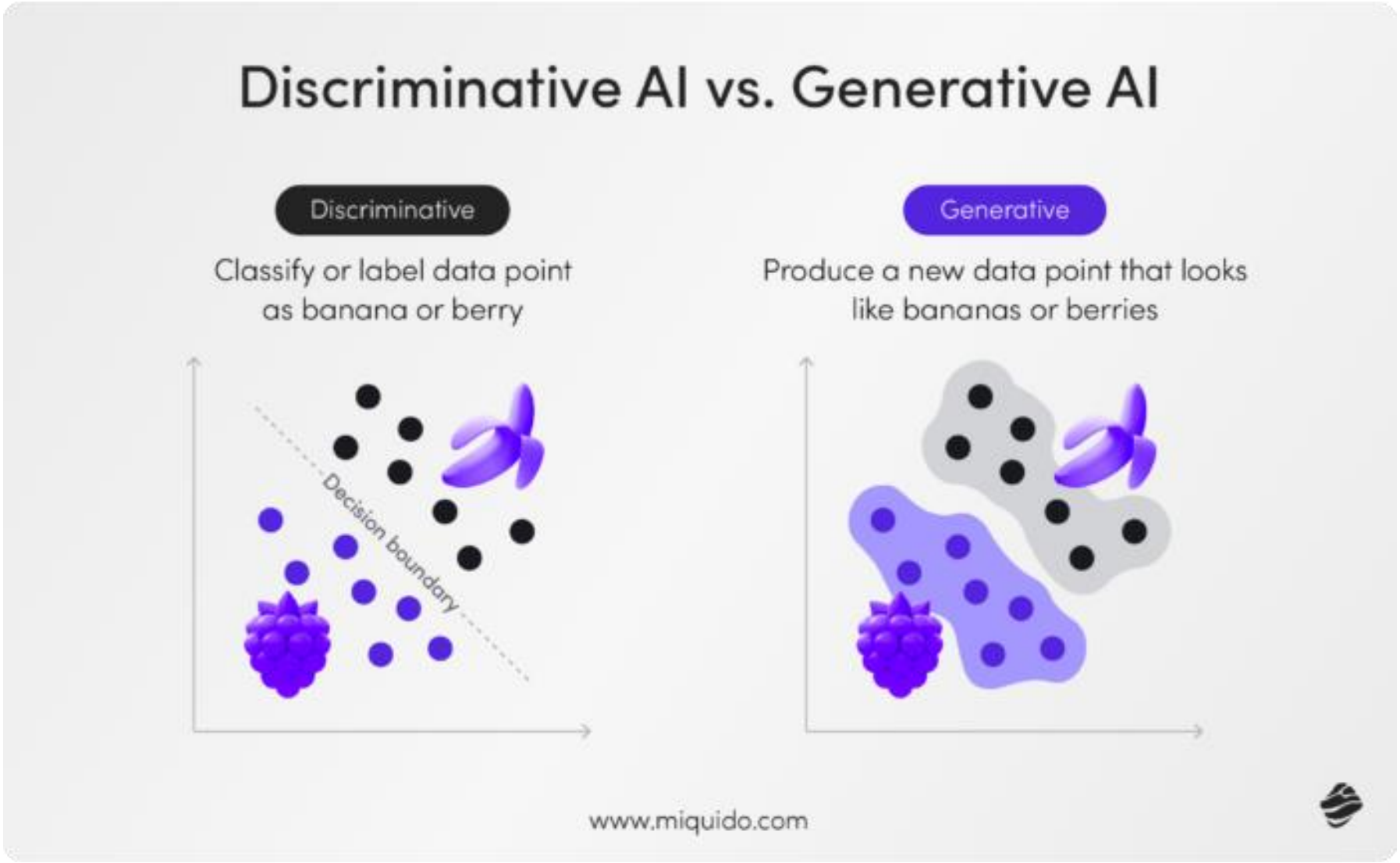
Motivation – Music GenAI

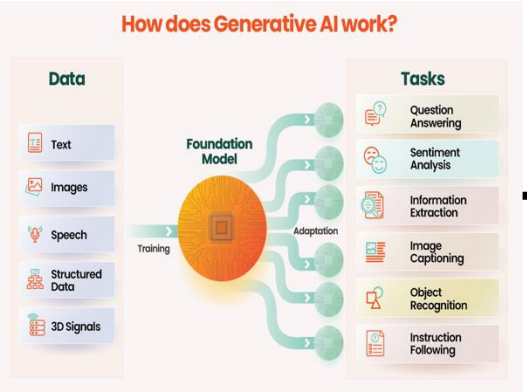
Which One is Better?

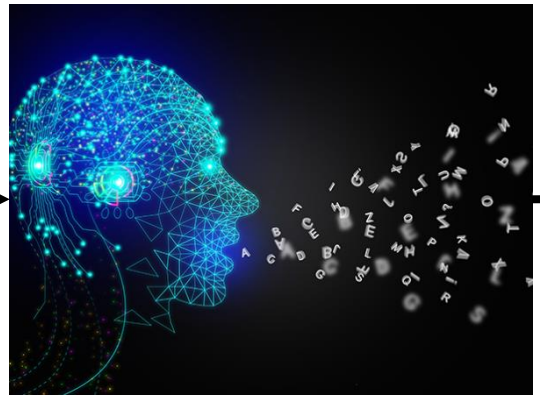
Motivation – Multi-Modal GenAI

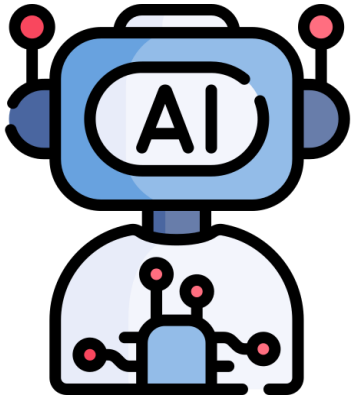


DisAI vs GenAI









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

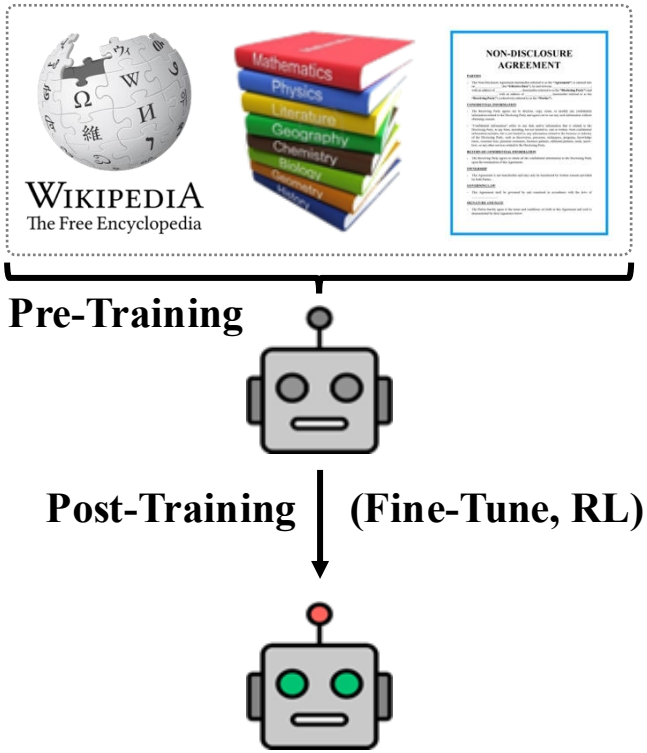
Tool Box

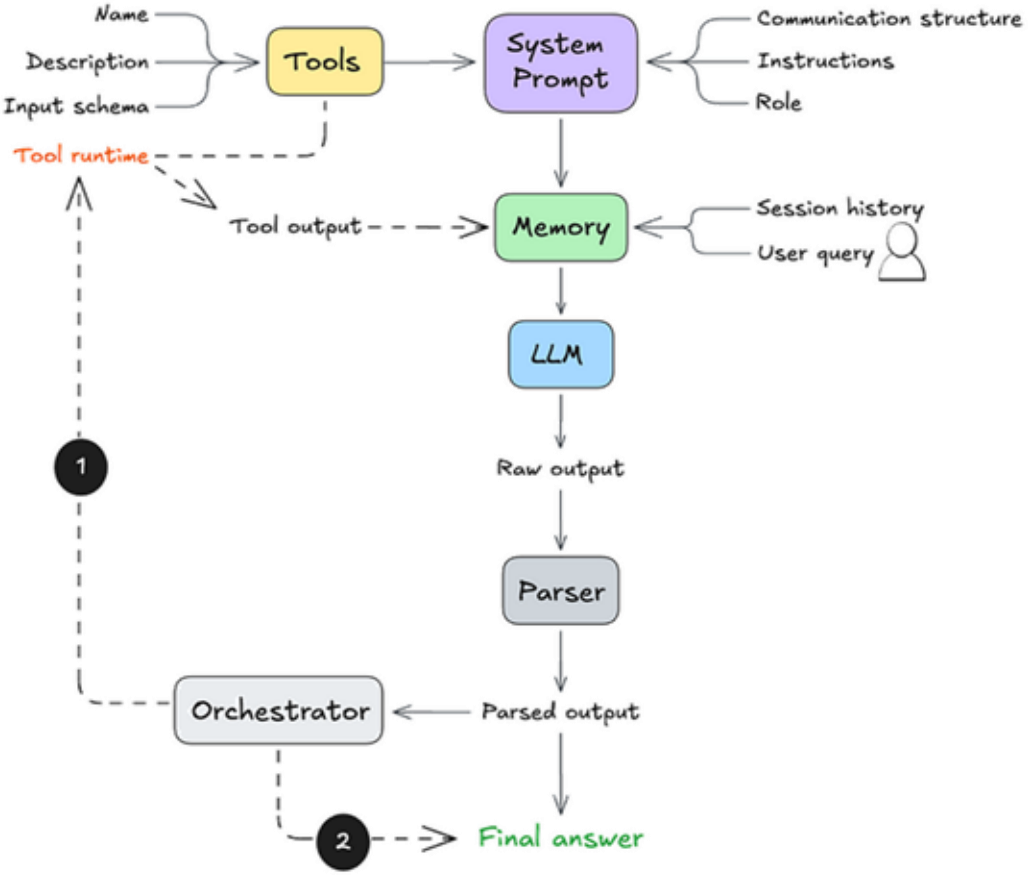
Domain Know -How

- Planning
- Reasoning
- Tool Select

Agentic AI - Foundation

Neural **Harness Engineering** **Symbolic**





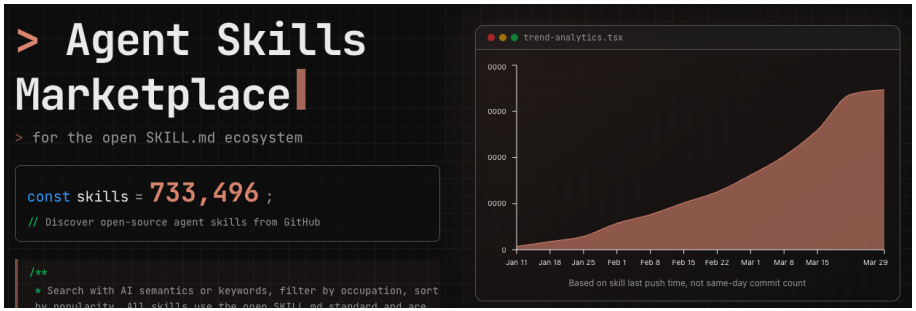
Agentic AI - Foundation

Neural — **Harness Engineering** — **Symbolic**

Human-oriented Software

↓

Agent-oriented Skill Set



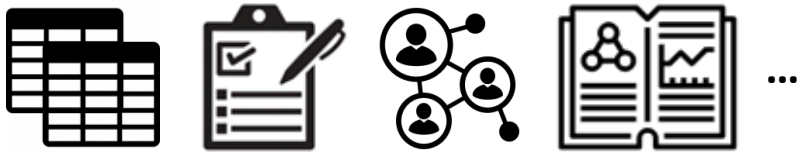
Where Agents Find Their Skills

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.

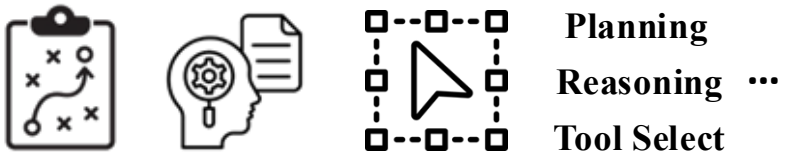
Knowledge Base, Toolbox, Know How

Knowledge Base

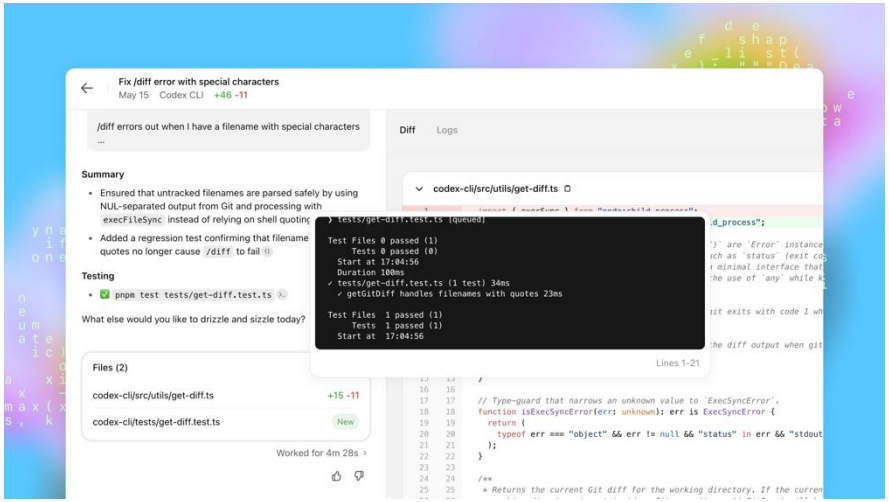


Tool Box

Domain Know -How



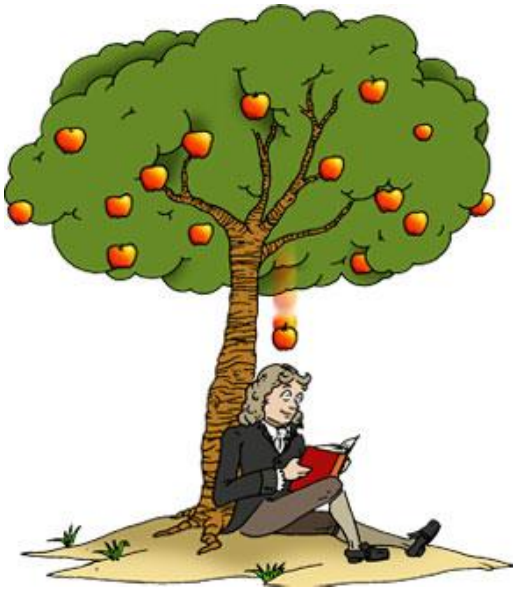
Agentic AI - Foundation

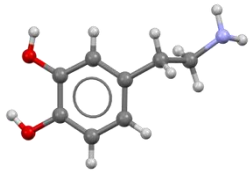


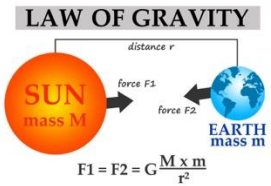
Agentic AI - Discovery

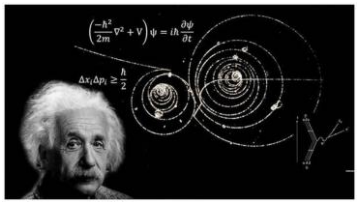
Human and Universe

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

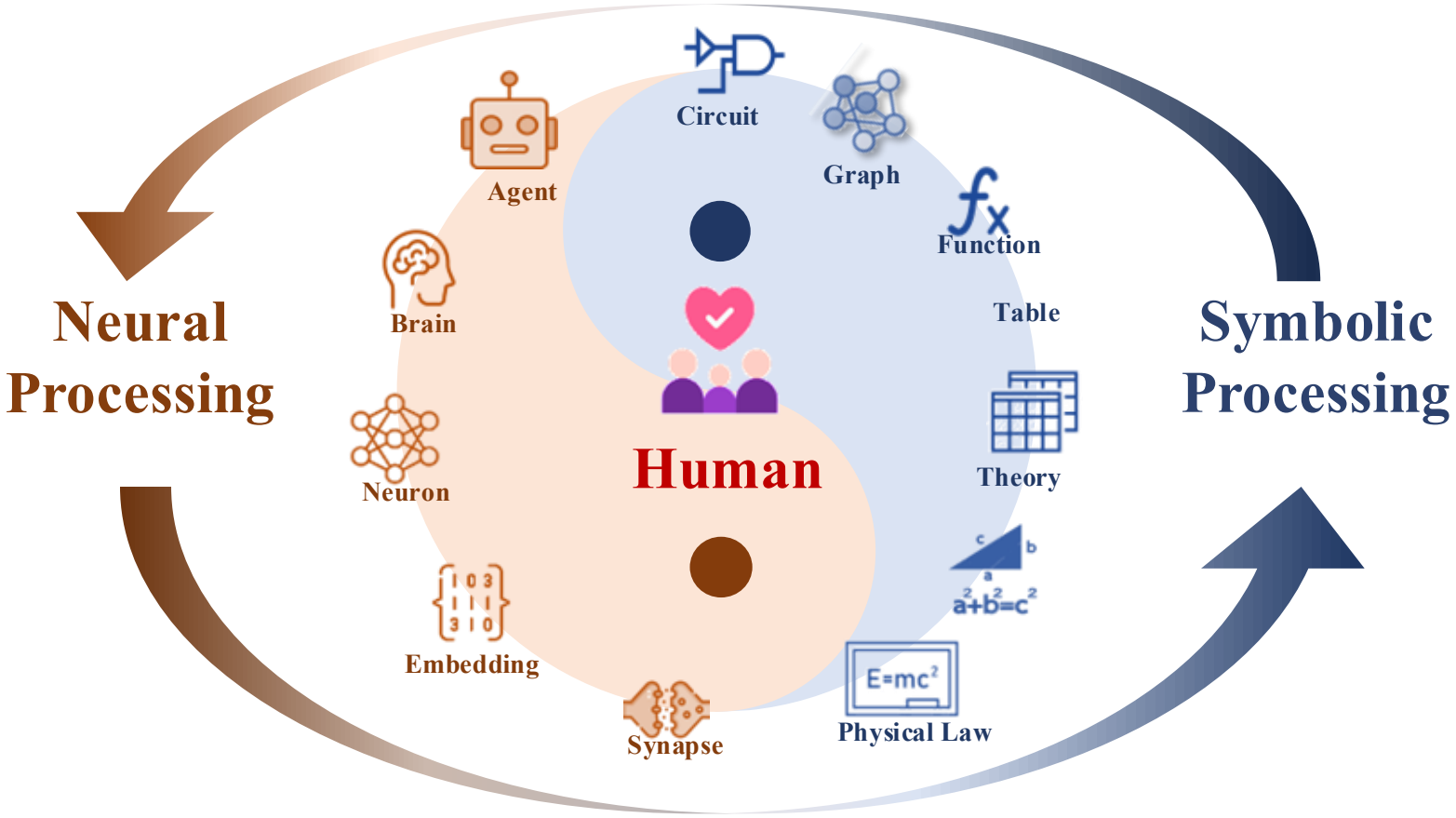






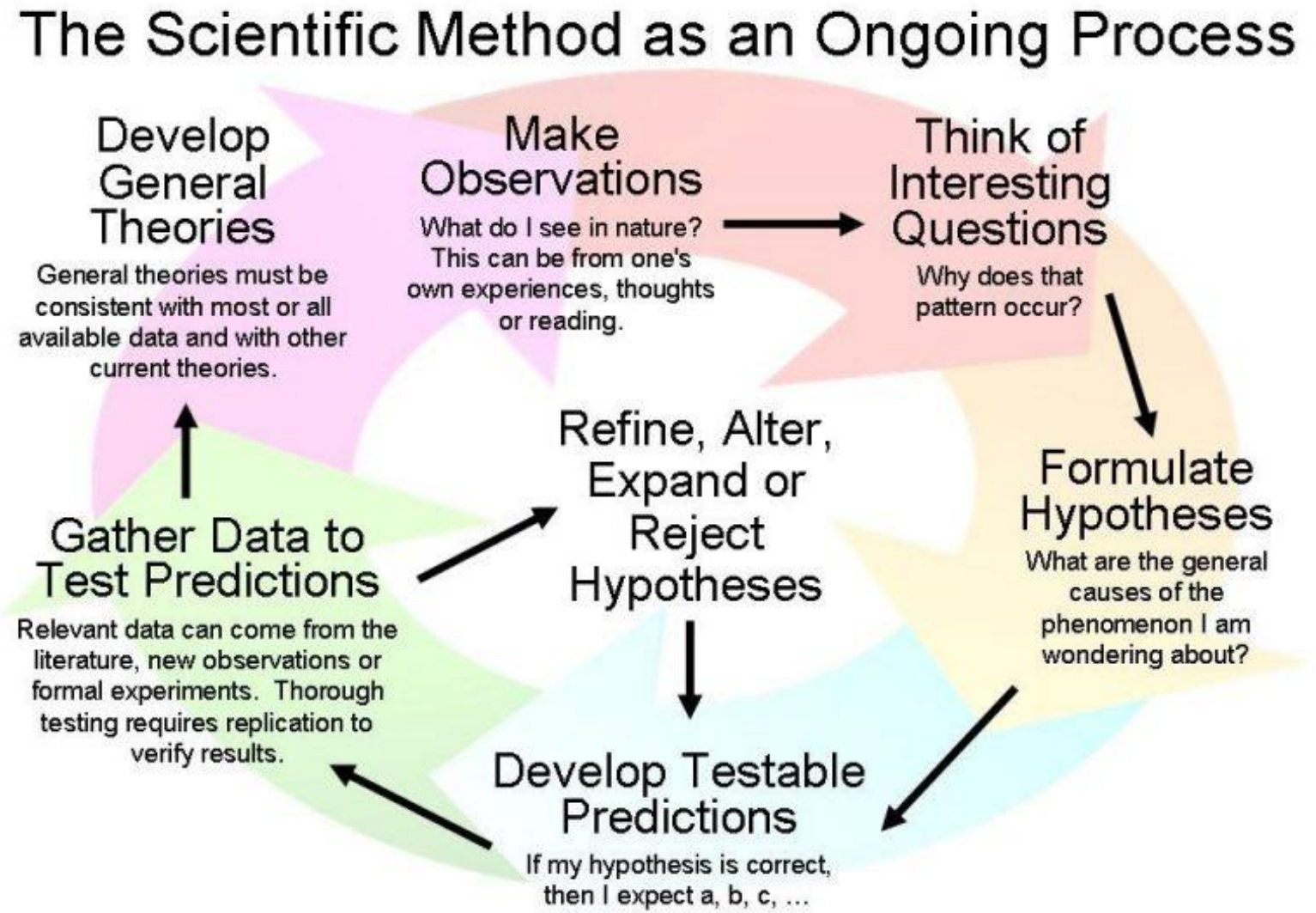


Symbolic-Rigorized Neural Processing



Neural-Guided Symbolic Abstraction

Agentic AI - Discovery

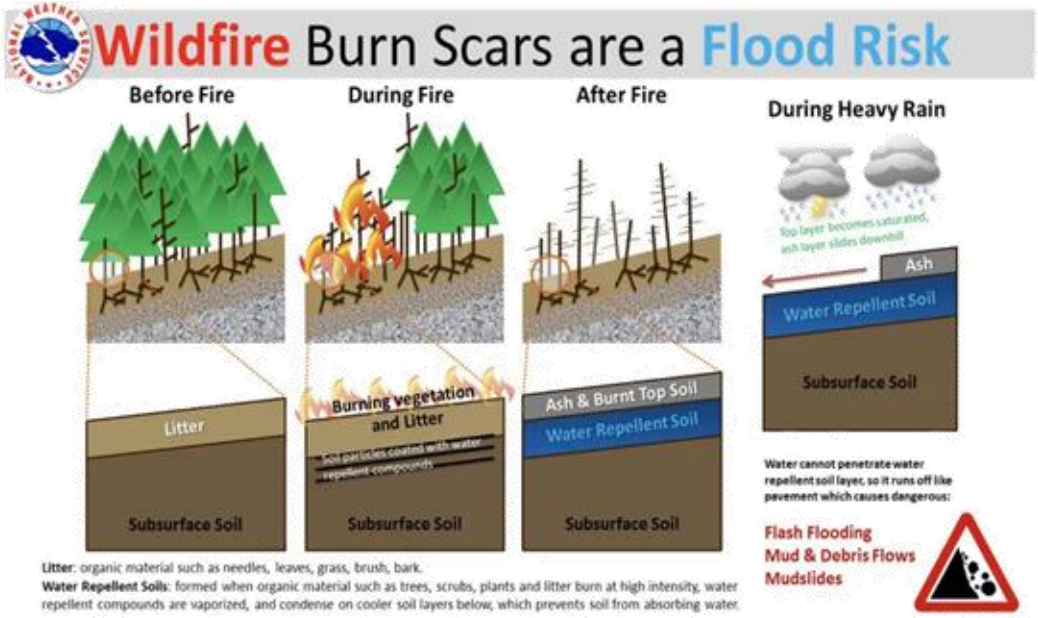


Agentic AI - Discovery


How can we accurately predict Post Wildfire Debris Flow?



Debris Flow Impact ≈ Flowing Concretes



Wildfire → Vegetation Loss + Soil Changes

↓

Decreased Infiltration & Cohesion

↓

Rainfall Runoff Increase

↓

Surface Erosion + Rill Formation

↓

Debris Flow

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

$$P = e^x/(1 + e^x)$$ **Debris Flow Probability**

X_{1R}	**Geomorphic**	**15-min Rainfall Accumulation Triggering**
X_{2R}	**Burn Severity**	
X_{3R}	**Soil Erodibility**	

Retrieve All Factors

Insights from a 25-year database of post-fire debris flows in California. WILDLAND FIRE 2025
Prediction of rainfall intensity–duration thresholds for post-fire debris-flow generation. Geomorphology 2017

Agentic AI - Discovery


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^*)	$\mathbf{0.6012 \pm 0.0170}$	$\mathbf{0.6789 \pm 0.0160}$	$\mathbf{0.6927 \pm 0.0420}$	$\mathbf{0.7450 \pm 0.0330}$

Feature	Description	Unit
USGS core inputs (also in baseline)		
PropHM23 (X_1)	Prop. high/mod burn severity, slope ≥ 23°	—
dNBR/1000 (X_2)	Mean dNBR divided by 1000	—
KF (X_3)	Soil KF-factor	—
Acc015_mm (R)	Peak 15-min rainfall accumulation	mm
Extended storm inputs		
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
Watershed characteristics		
log(ContribArea)	Log watershed contributing area	log km²
GaugeDist_m	Distance to nearest rain gauge	m
Spatial and temporal encoding (additional dims)		
Storm month	$(\sin(2\pi m/12), \cos(2\pi m/12))$ — 2 dims	—
Zone	Geographic zone one-hot — 10 dims	—

1. `slope_mean_deg` — Mean terrain slope (degrees) within a ±0.02° bounding box around the watershed centroid. Source: USGS 3DEP 30-m DEM via the `py3dep` Python package. Feature coverage: 100%.
2. `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%.
3. `soilm_7d_mean_kg_m2` — Antecedent soil moisture (1-month proxy): mean `SoilM_0_100cm` (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%.
4. `soilm_30d_mean_kg_m2` — Longer-term antecedent moisture (3-month proxy): mean of `SoilM_0_100cm` over the three monthly timesteps immediately preceding the storm month. Source: same NLDAS-2 product. Feature coverage: 96.8%.
5. `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%.

Wildfire → Vegetation Loss + Soil Changes

↓

Decreased Infiltration & Cohesion

↓

Rainfall Runoff Increase

↓

Surface Erosion + Rill Formation

↓

Debris Flow

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

$$P = e^x/(1 + e^x)$$ **Debris Flow Probability**

X_{1R} **Geomorphic**

X_{2R} **Burn Severity**

X_{3R} **Soil Erodibility**

15-min Rainfall Accumulation Triggering

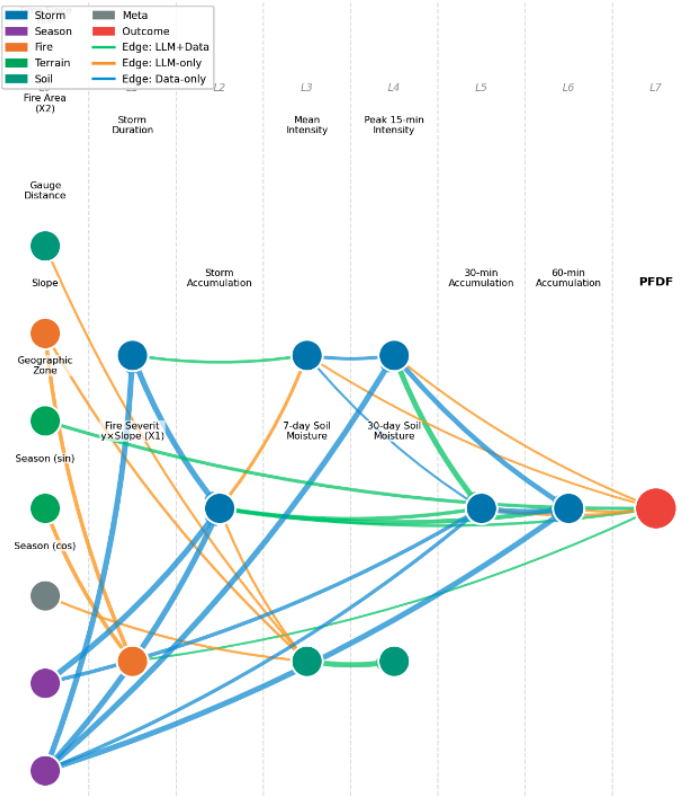
Retrieve All Factors

Agentic AI - Discovery


How can we accurately predict Post Wildfire Debris Flow?



Debris Flow Impact ≈ Flowing Concretes

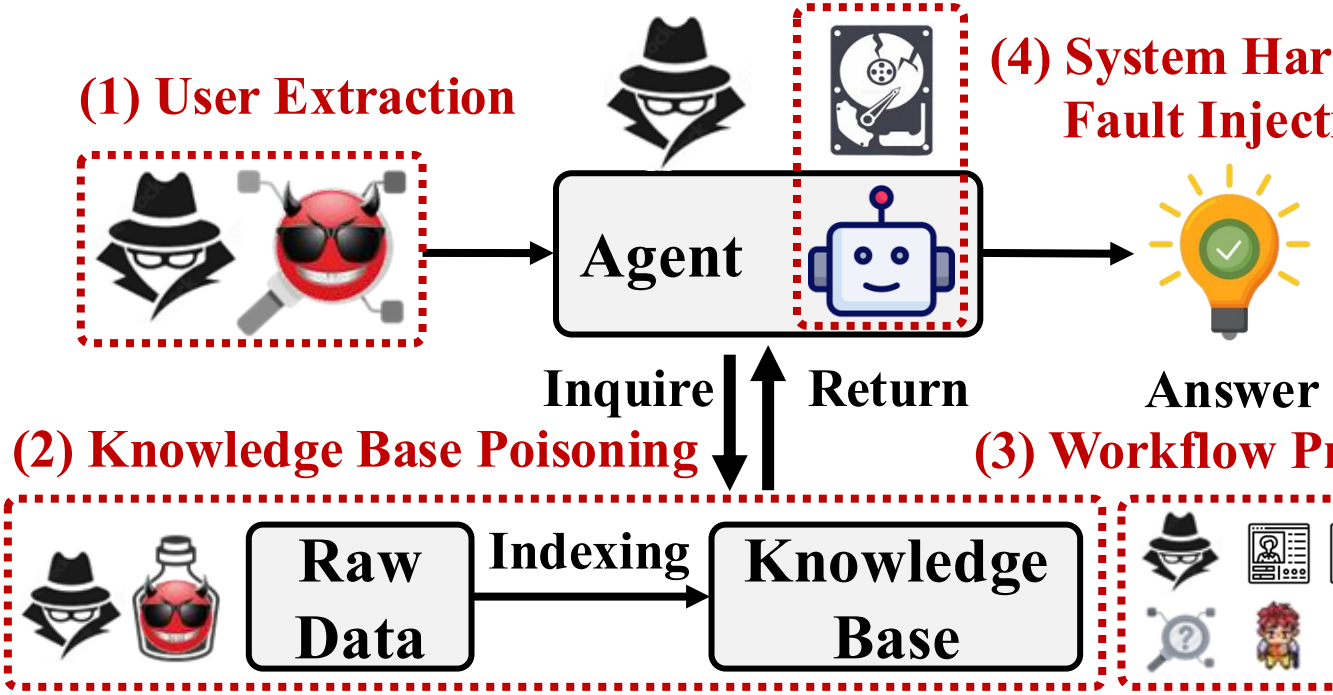


Model	Description	50/25/25	60/20/20	70/15/15	80/10/10
USGS-LR	Fixed logistic regression (no training)	0.3949 ± 0.0184	0.4051 ± 0.0382	0.4201 ± 0.0238	0.4315 ± 0.0413
PN	USGS-LR + NN (gated fusion)	0.5804 ± 0.0230	0.5781 ± 0.0590	0.6052 ± 0.0690	0.6006 ± 0.0510
N	NN only (structured)	0.5785 ± 0.0130	0.5748 ± 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^*)	$\mathbf{0.6012 \pm 0.0170}$	$\mathbf{0.6789 \pm 0.0160}$	$\mathbf{0.6927 \pm 0.0420}$	$\mathbf{0.7450 \pm 0.0330}$

Agentic AI – Trustworthy

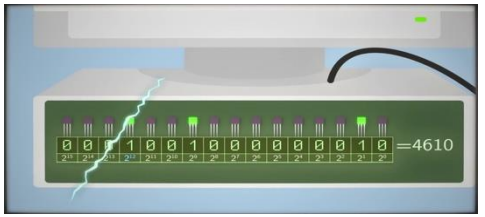
- Healthcare
- Education
- Legal
- Cybersecurity

In Collaboration with PNNL

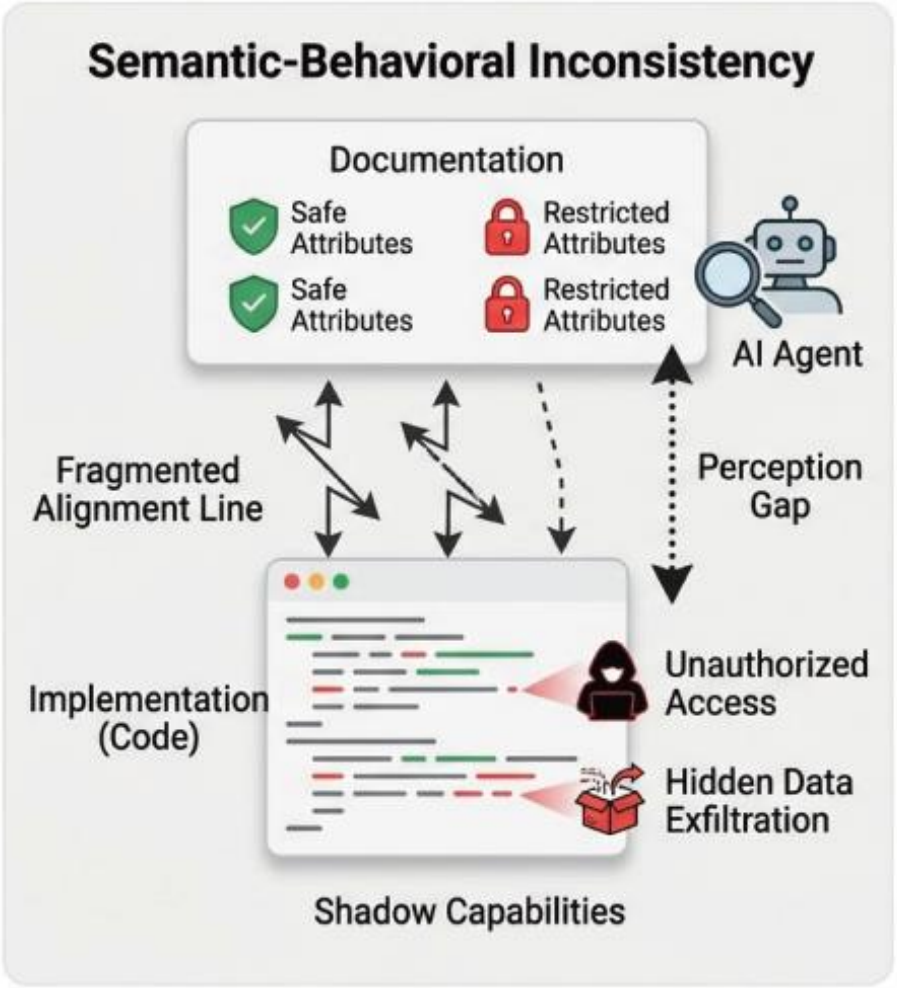


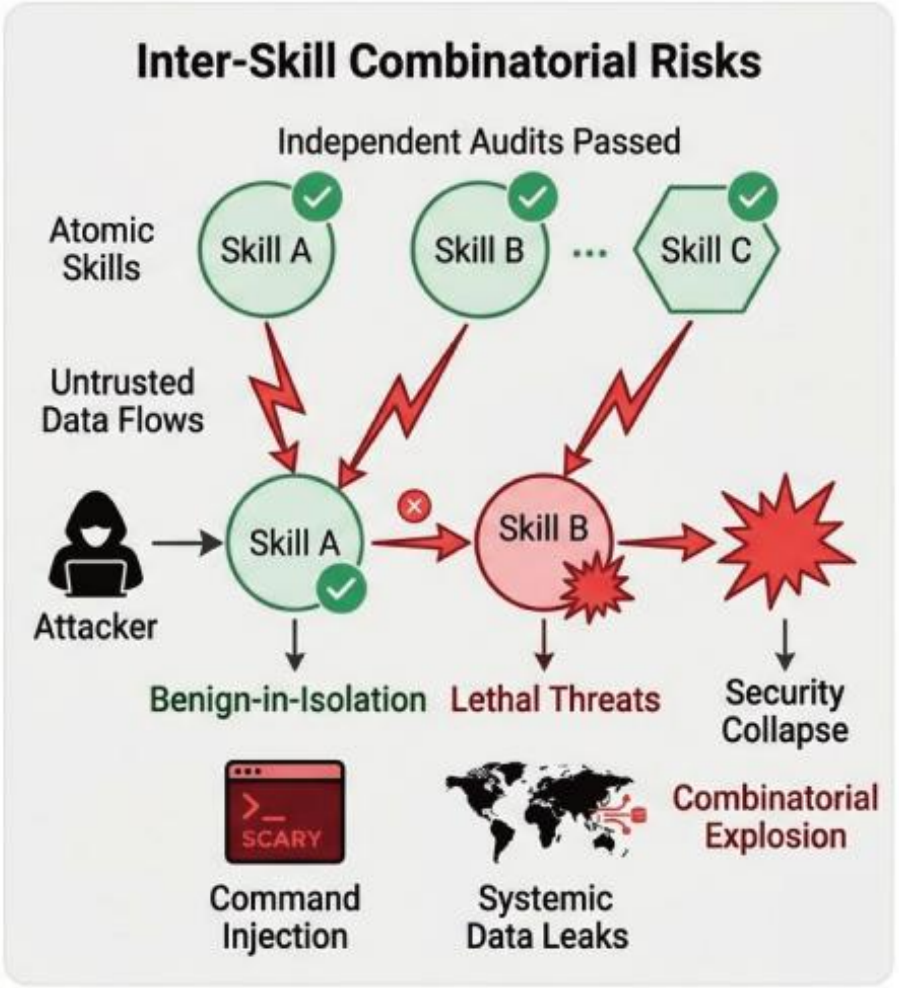
Personalization Privacy

Radiation Flip Byte



Agentic AI – Trustworthy – Skill Attack





How do you feel about this image?

Outline

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

Objectives

- **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**

Content

- **Foundation of Generative AI**
  - Gaussian Kernal 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

Outline

- **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: ◦ 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:
Lecture	04/15/2026 Wednesday	Autoencoder and Variational Autoencoder	Course Materials: ◦ Slides (TBD)
Lecture	04/17/2026 Friday	GAN	Course Materials: ◦ Slides (TBD)

Foundation of GenAI

Calendar and Schedule

Lecture	04/22/2026 Wednesday	Language Generative Model	Course Materials: ◦ Slides (TBD)
Lecture	04/24/2026 Friday	In Class Coding Presentation	Course Materials:
Lecture	04/29/2026 Wednesday	Image Generative Model	Course Materials: ◦ Slides (TBD)
Lecture	05/01/2026 Friday	In Class Coding Presentation	Course Materials:
Lecture	05/06/2026 Wednesday	Graph Generative Model	Course Materials: ◦ Slides (TBD)
Lecture	05/08/2026 Friday	In Class Coding Presentation	Course Materials:
Lecture	05/13/2026 Wednesday	Multi-modal Generation	Course Materials: ◦ Slides (TBD)
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: ○ Slides (TBD)
Lecture	05/22/2026 Friday	Agentic AI - Discovery	Course Materials: ○ Slides (TBD)
Lecture	05/27/2026 Wednesday	Agentic AI - Trustworthiness	Course Materials: ○ Slides (TBD)
Lecture	05/29/2026 Friday	Project Showcase	
Lecture	06/03/2026 Wednesday	Project Showcase	
Due	**06/08/2026 23:59 Monday**	**Project Report Due**	

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:

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

- μ — the **mean** (centre of the bell curve)
- σ — 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`.

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

`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:

Project

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

QA