

Frugal AI solutions for India from IITK



IIT Kanpur
AI for India

Nitin Saxena
(joint works with Nisheeth, Sruti, Aditya, Sayak)

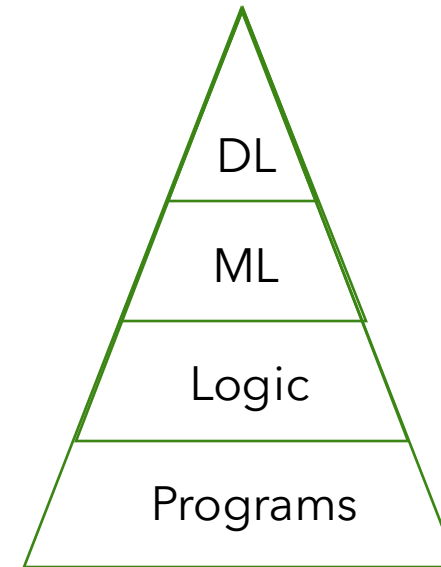
Coordinator, Center for Developing Intelligent Systems +
Dean, Wadhvani School of AI & Intelligent Systems +
N.Rama.Rao.Chair Professor (HAG), CSE +
Founding Director, CDIS Private Limited

IIT Kanpur



What is machine intelligence?

- **Simple programs**
 - If the password matches the database entry, let the person in
- **Logical reasoning**
 - If someone tries to buy a ticket, make sure there are tickets available before charging them
- **ML** (Machine Learning)
 - If your database row has this mean/average, you collect it
- **DL** (Deep Learning)
 - Just give me labelled data and get out of my way



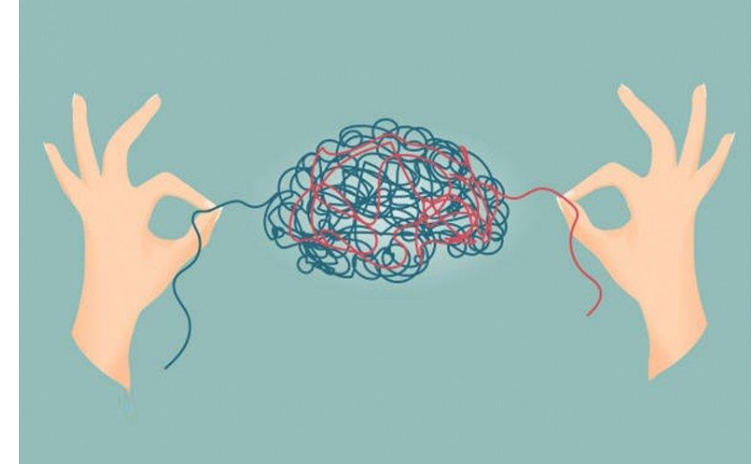
AI – a layperson's definition



- Any computing system that
 - does something that machines couldn't do before.
- *Corollary: the AI effect*
 - Anytime machines start doing something *intelligent, en masse*, it is no longer considered AI!
- Examples of AI systems that are *no longer* considered AI
 - Chess solvers
 - OCR (Optical character recognition)
 - Automated assembly lines
 - Automated CNC lathes (Computer Numerical Control)

AI and complexity

- The AI effect: **if you understand it, it isn't AI.**
- **Consequence:** apply AI to everything we don't understand how to do.
 - AI for health-care
 - AI for mental-health
 - AI for music-writing
 - AI for governance
- Apparent logic:
 - We don't understand AI
 - We don't understand how to do X →
 - AI will understand how to do X.

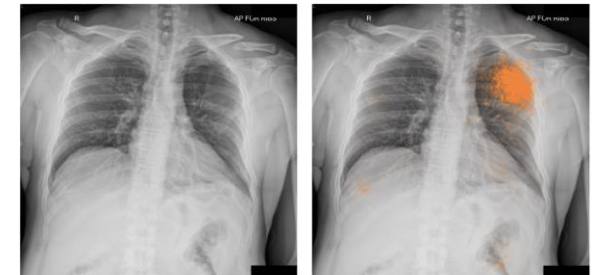


(a) Three samples in criminal ID photo set S_C .



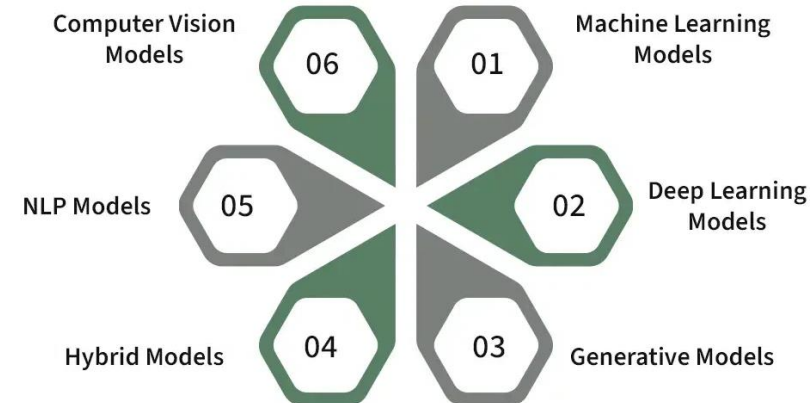
(b) Three samples in non-criminal ID photo set S_n .

Figure 1. Sample ID photos in our data set.

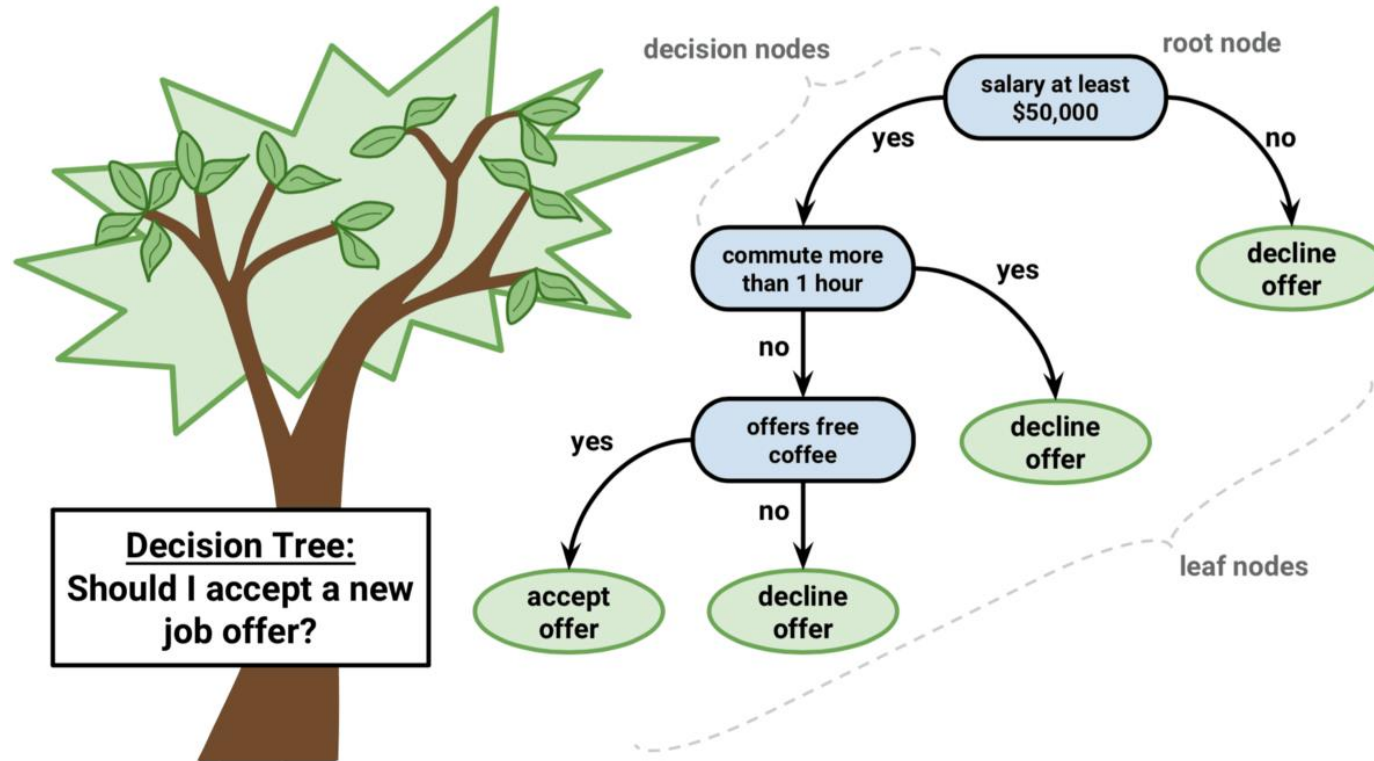


What does the AI see?

- It sees **models** of the world
- Models can be simple or **COMPLEX**
- Simple programs see simple models
- **COMPLEX** programs see **COMPLEX** models
- All models are wrong, some are useful.
- Understanding and fixing simple models is easier than doing this for complex models.



Example 1 →



Simple model: Decision tree

Example 2 →

```
import numpy as np
import pandas as pd
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense

data = {
    'feature1': [0.1, 0.2, 0.3, 0.4, 0.5],
    'feature2': [0.5, 0.4, 0.3, 0.2, 0.1],
    'label': [0, 0, 1, 1, 1]
}

df = pd.DataFrame(data)

X = df[['feature1', 'feature2']].values
y = df['label'].values

model = Sequential()

model.add(Dense(8, input_dim=2, activation='relu'))
model.add(Dense(1, activation='sigmoid'))

model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])

model.fit(X, y, epochs=100, batch_size=1, verbose=1)

test_data = np.array([[0.2, 0.4]])

prediction = model.predict(test_data)
predicted_label = (prediction > 0.5).astype(int)
print(f"Predicted label: {predicted_label[0][0]}")
```

Activation functions introduce non-linearity into the network enabling it to learn and model complex data patterns.

Common activation functions include:

Sigmoid: $\sigma(x) = \frac{1}{(1+e^{-x})}$

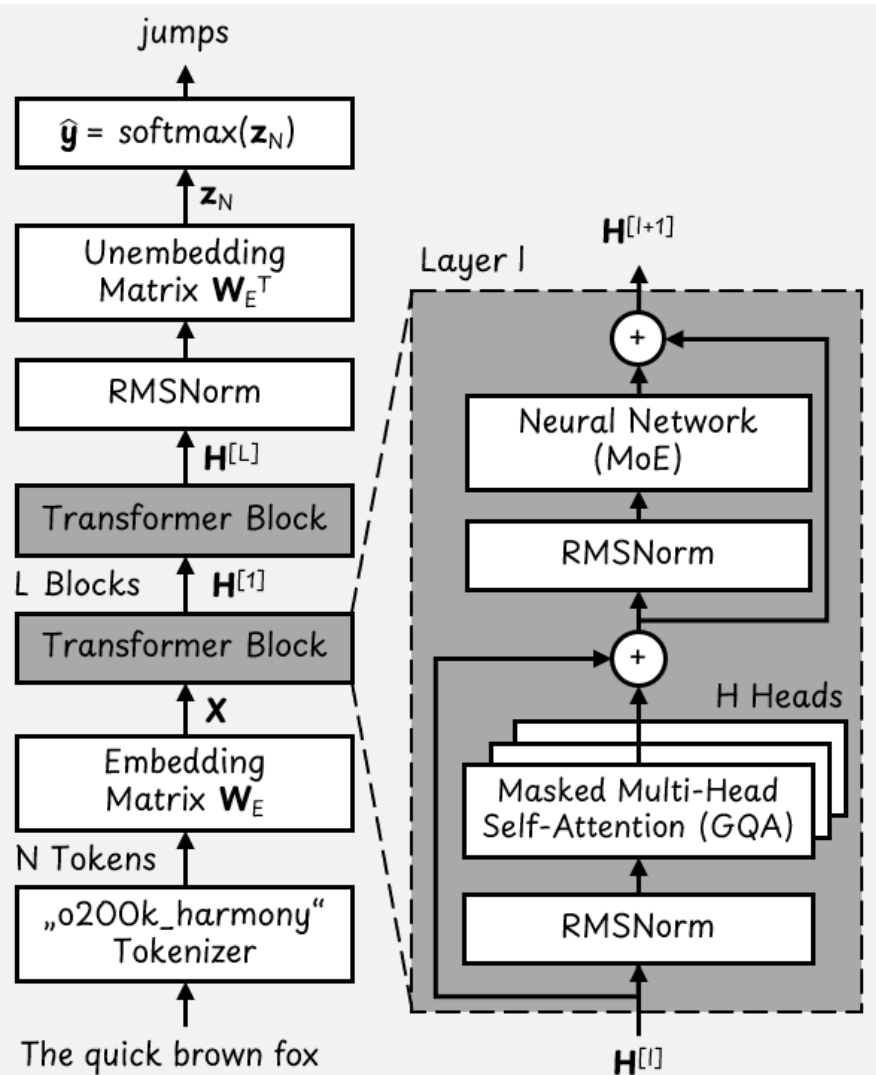
Hyperbolic tangent: $\tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$

Rectified Linear Unit: $\text{ReLU}(x) = \max(0, x)$

Complex model: Neural Network

Example 3 →

Highly-Complex model: LLM Architecture



Tokenizer: converts the input sentence into a list of tokens.

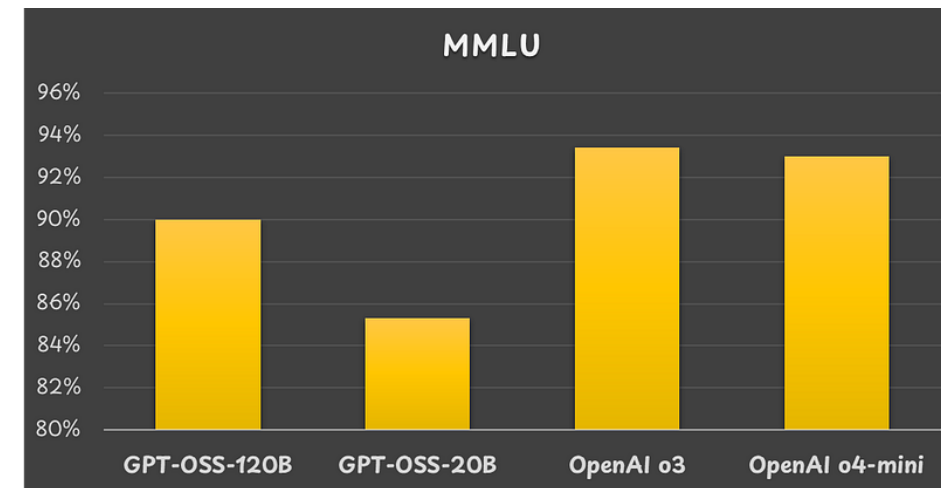
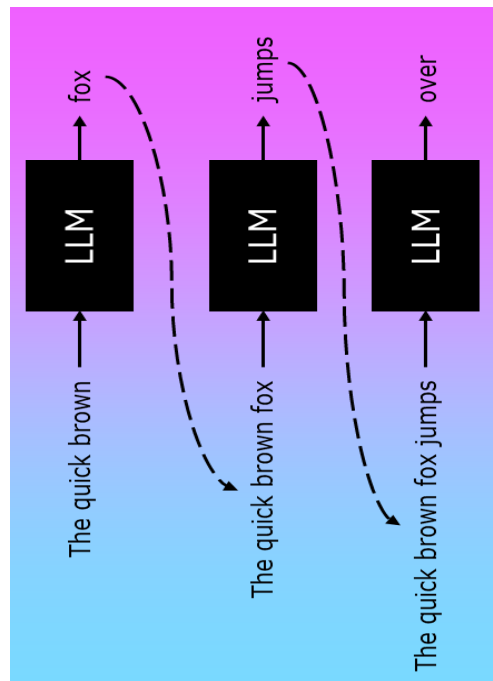
Embedding: transforms the list of tokens into a dense matrix.

Self-attention mask: tells the model which words to pay attention to and which ones to skip over.

MoE (mixture of experts): the router selects multiple experts with different weights.

Unembedding: transforms the embedding vectors back into token space.

Softmax: converts each of the probable tokens (=logits) into **probabilities**.



Massive Multitask Language Understanding (MMLU) benchmark results

Could AI be trusted?

- AI systems *cannot* be evaluated using software engineering quality testing
 - Governance of AI systems must recognize this basic fact
- Modern AI systems can sometimes fail in unexpected ways
- AI systems must demonstrate that they do what they claim to do, to show they are trustworthy

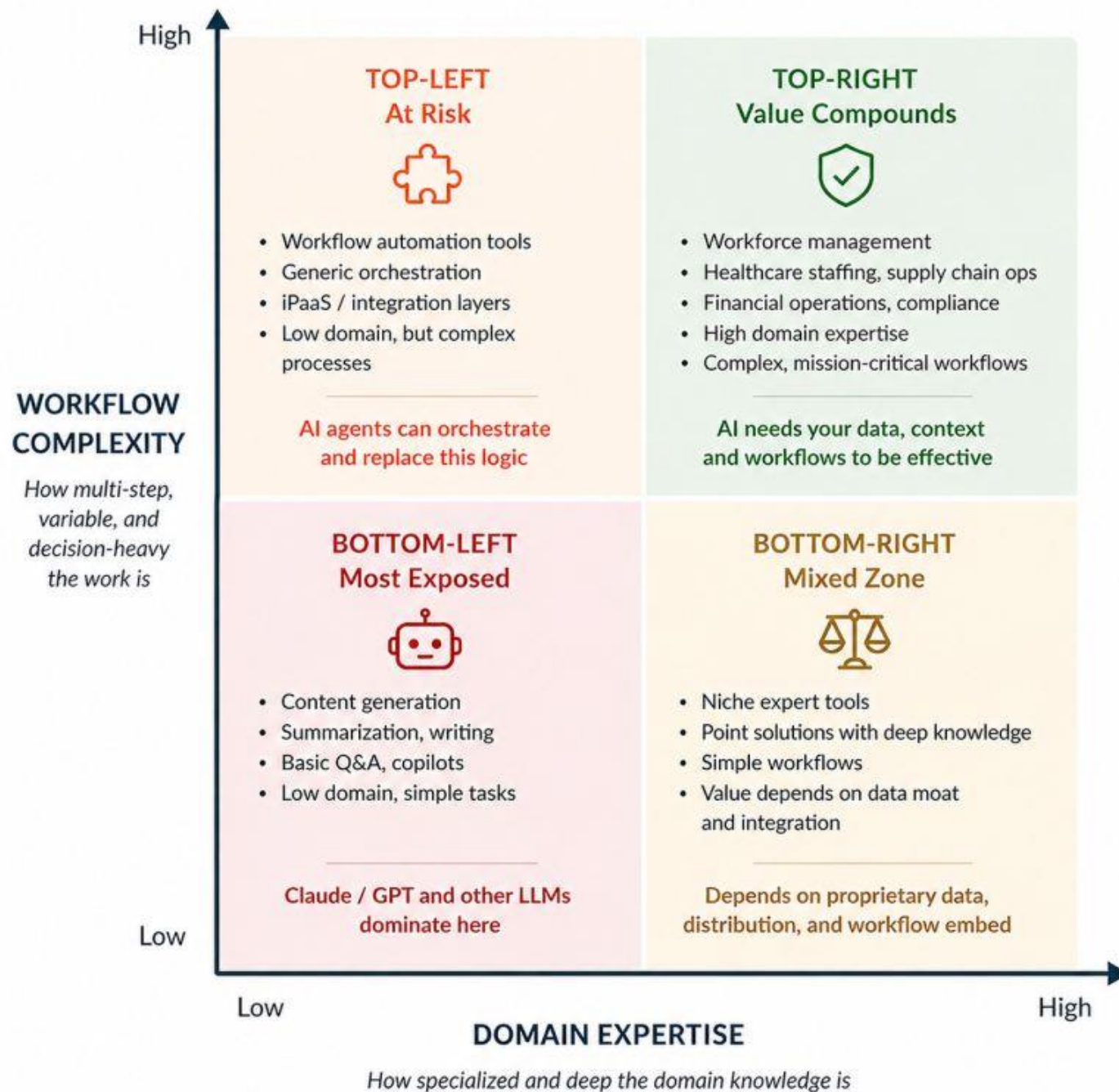
AI will not necessarily be bad or good.

All we know for sure is that it will be alien and it will be fallible.

-Yuval Noah Harari



Could AI eat Software?



There's a take going around that AI will "eat software."
That's directionally right but dangerously ...more

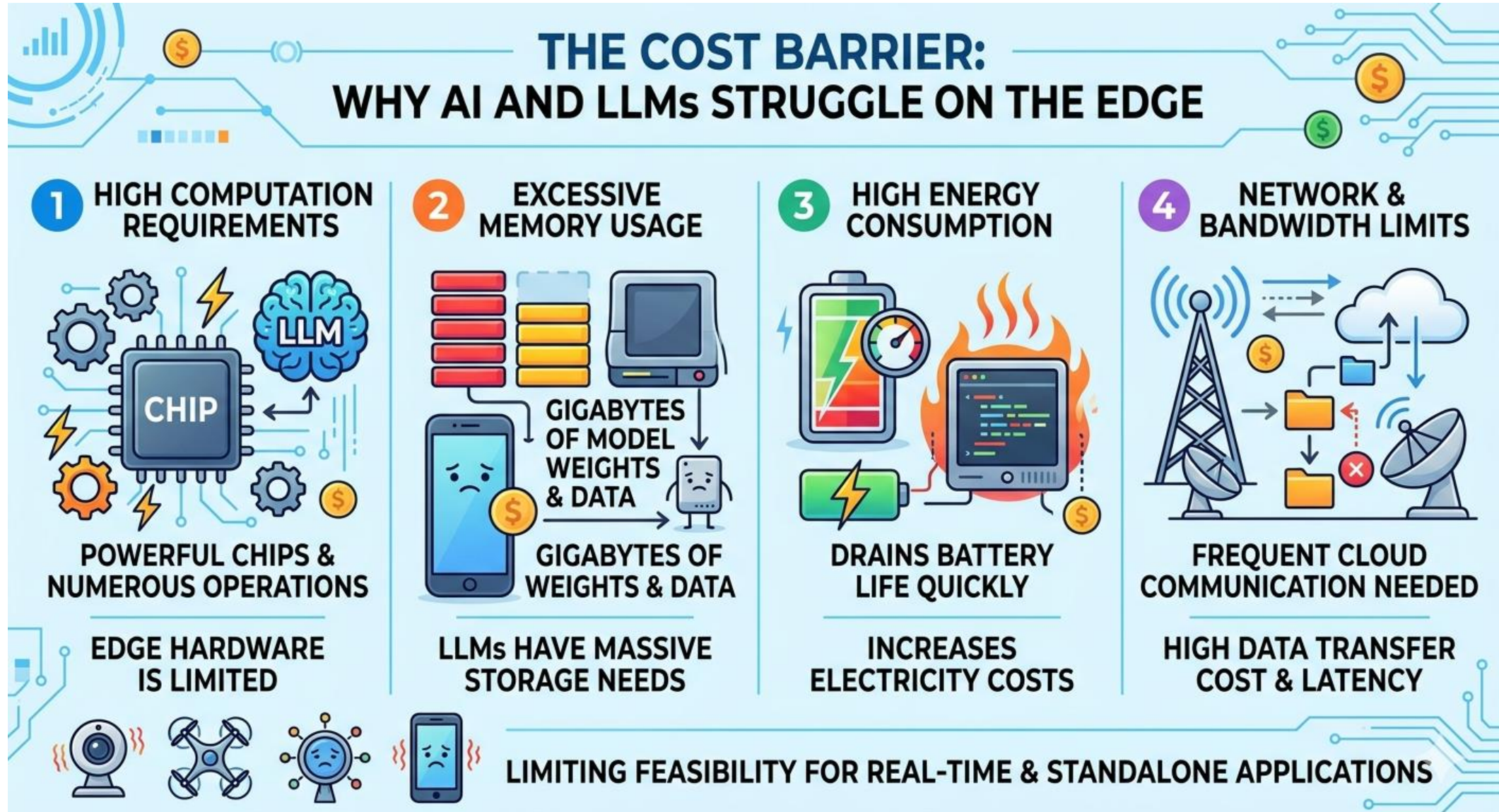


Suresh Vitalin • 2nd
Chief Product Officer & EVP at UKG | HC...

1d • 🌐

+ Follow

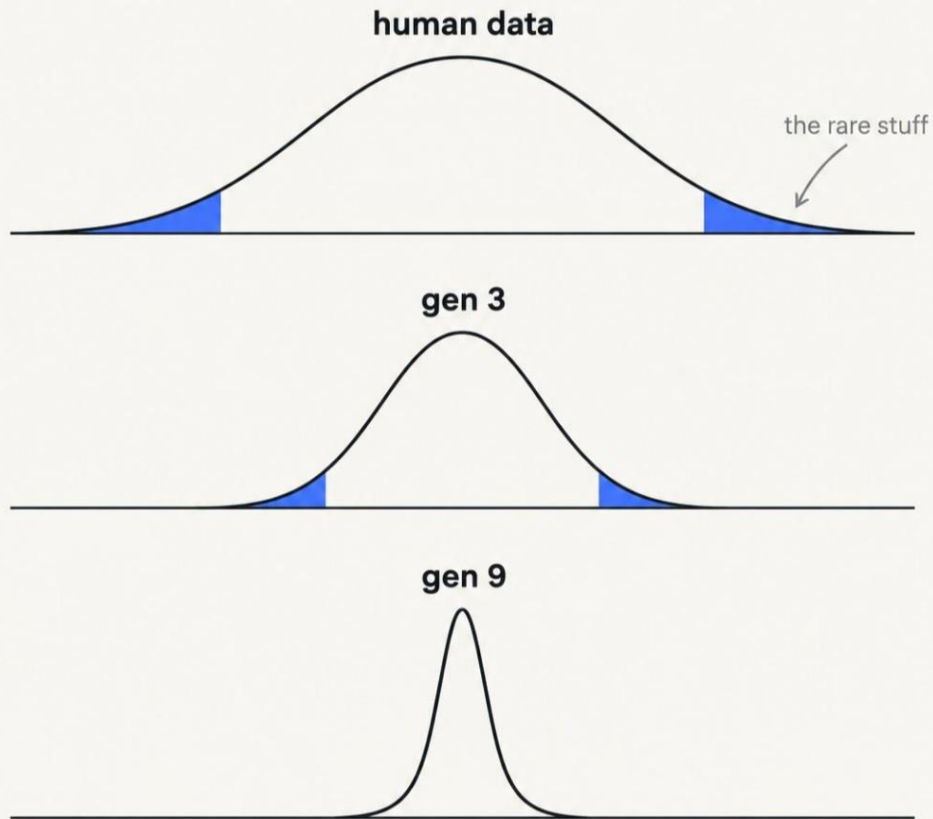
Could AI reduce cost?



Could AI have ideas?



what AI forgets when it learns from AI



model keeps the average. loses the tails.

- 1) "AI novelty" is **human novelty**.
- 2) "AI learning from AI" gives only the **average idea**.



Strategic thinking
needed to use the
AI tool effectively.



Shiraz Shah • 3rd+

Finding viruses that prevent chro...
21h • 🌐

+ Follow

Stop using frontier AI for research, engineering, business, anything that's your moat.

Why? Because AI labs are under pressure and may inadvertently run with your ideas.

This happened to us. It happened to Buckmaster when OpenAI solved Navier-Stokes.

What's the solution?

Buy a 16GB gaming GPU and run Qwen 3.8 27b on your own hardware. If you don't know how, I'll show you how. Let me know below.

<https://lnkd.in/eH7M8Gqd>

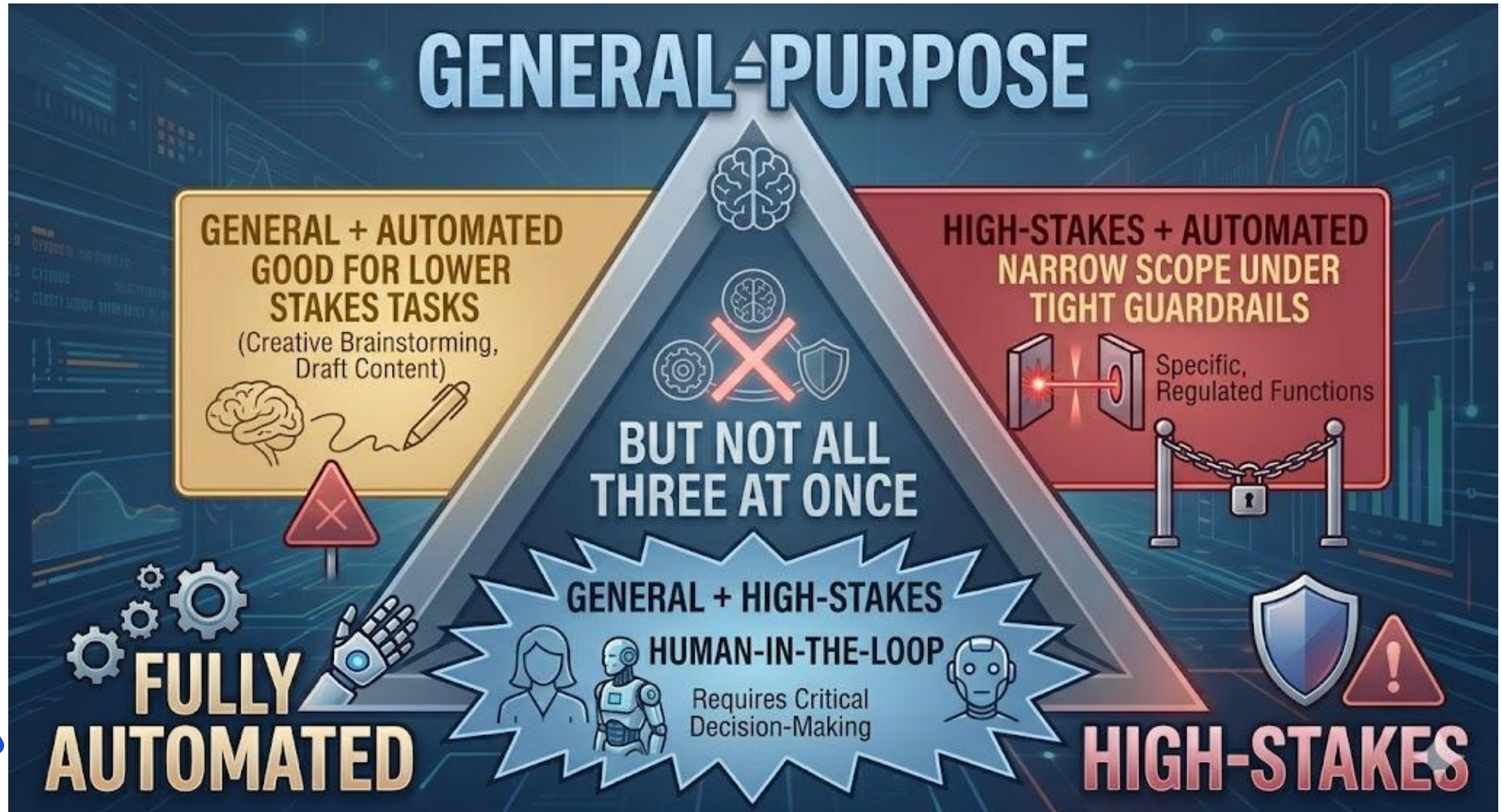
Mario Rodríguez Mestre, Section of Microbiology, UCPH



**Did Anthropic's A.I.
Really Make a Scientific
Discovery on Its Own?**

nytimes.com

The AI Agents Trilemma





What's IIT Kanpur doing?



WSAIS: Wadhvani School of AI & Intelligent Systems

- Department of Intelligent Systems (DIS)
 1. Center for Developing Intelligent Systems (CDIS)
 2. Center for Cybersecurity and Cybersecurity of Cyber-physical Systems (C3i)
 3. Center for Robotics (CR)
 4. Center for AI Policy and Outreach (CAPO).
 - Sustainable Cities AI-CoE (ARF)
 - Cybersecurity CoE (C3iHub)
- 

Use-cases:
AI for ...

- Sensing
- Deciding
- Acting



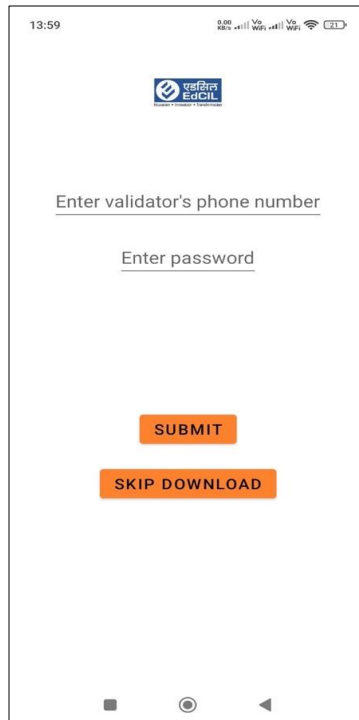
AI for sensing

Face Recognition App (Edge-AI, High-stakes)

- Runs on local smartphone, no internet.
- Uses TensorFlow-Lite for lightweight, fast inference.
- Face data processed locally.
- Reduces cloud usage and latency.

Work Flow (FR App)

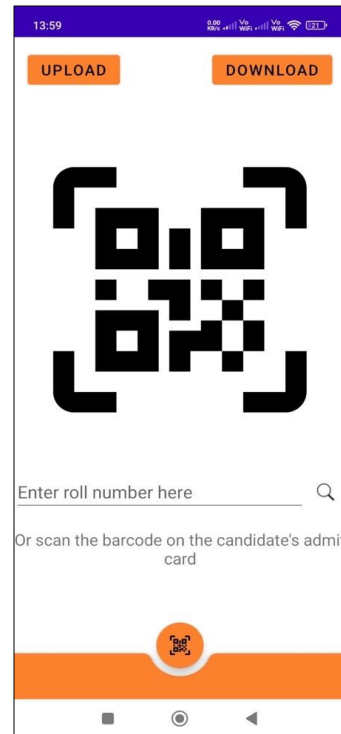
Log In Screen



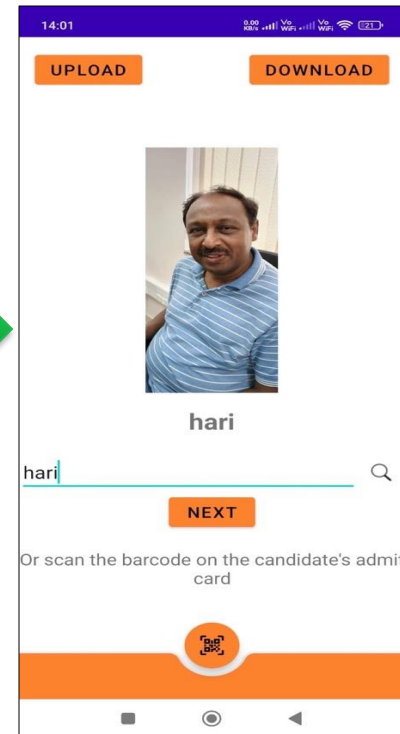
START

Enter login credentials to download images from server and click on "SUBMIT".

Application Interface

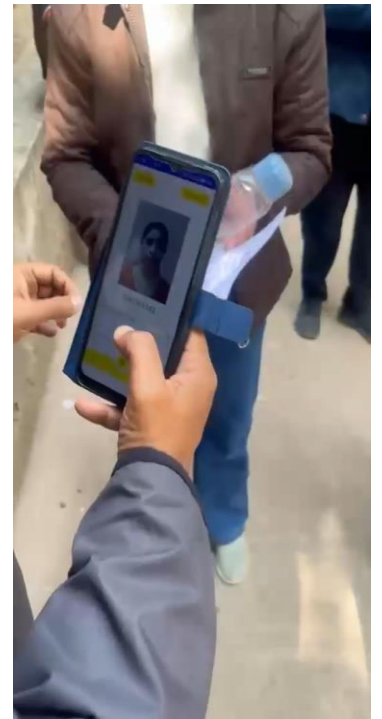


Scan the barcode on candidate's admit card or manually enter candidate's roll/ name.



After candidate's application image is displayed, click on "NEXT" to capture the live image.

Live Process

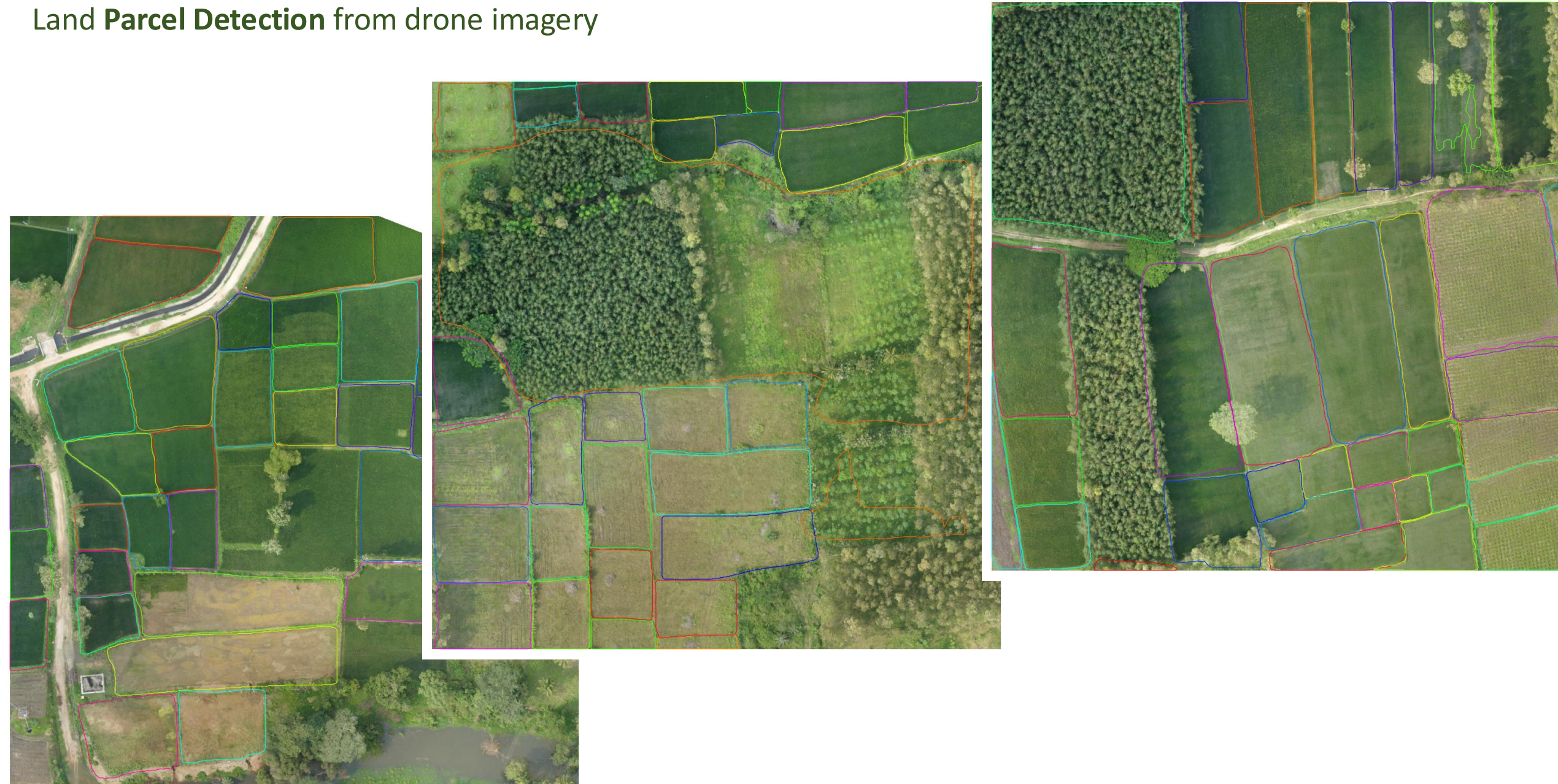


The captured live image of the candidate will be compared with the application image and a match or mismatch status will be indicated.

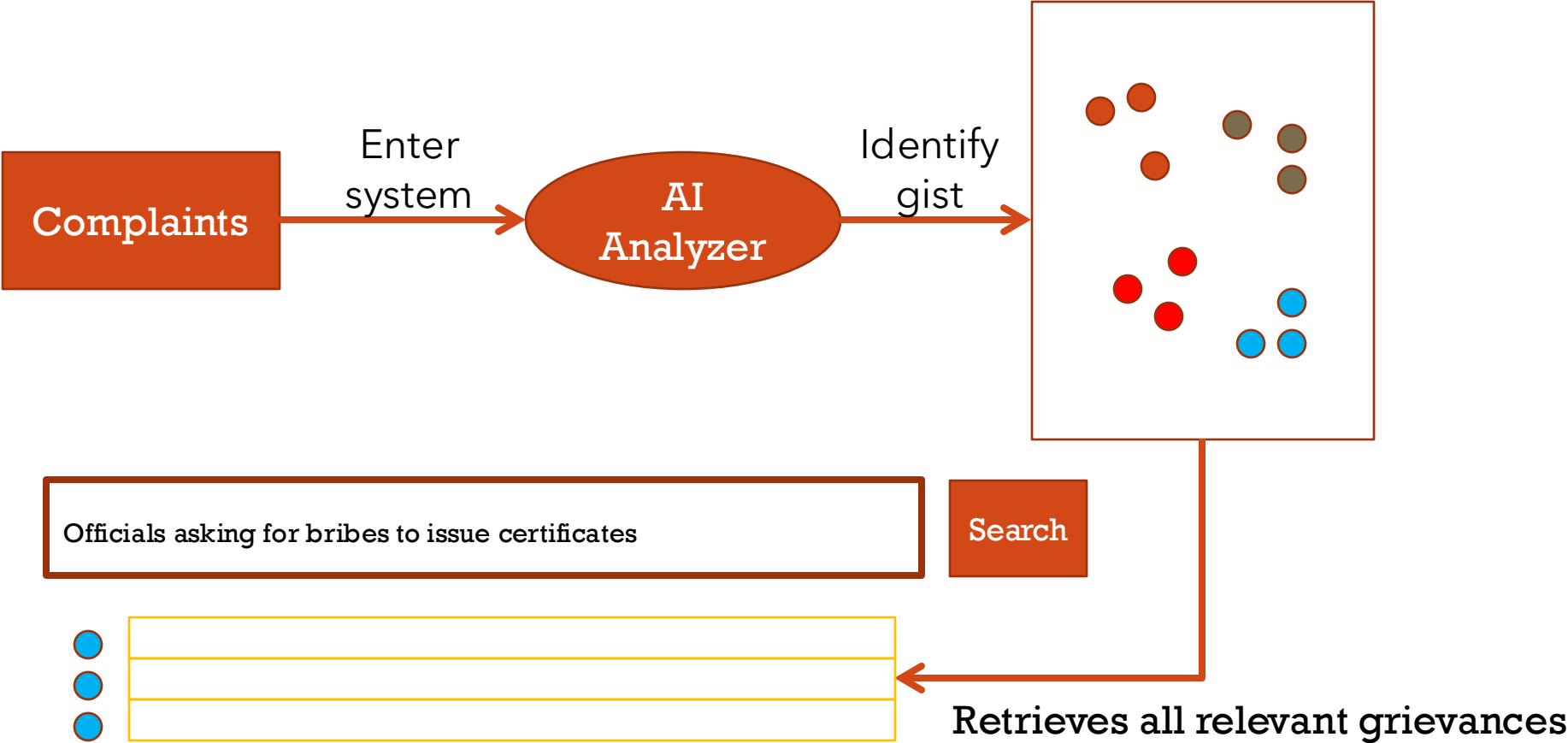
END

Automated Land Parcel Detection (General-purpose)

Land Parcel Detection from drone imagery



Sensing citizens' intent (General-purpose)



Deployed for all central government departments and ministries since Feb 2023



Detecting structure within documents (General-purpose)

RAJRATAN GLOBAL WIRE LIMITED
Regd. Office: Rajatan House, T-3, Meera Park, Dhule Market, Indore-492003, Madhya Pradesh
India. Tel. : +91-261-254401 Fax: +91-261-254404
Factory: 200-B, Sector-1, Phulwara-491775, Dist. Dist. Madhya Pradesh, India
Tel. : +91-260-254405, 254407 Fax: +91-260-254407
GSTIN No. 23AABCR4530Q128
PAN No. AABCR4530Q
CIN No. L2700RP1988PLCO04716

SUPPLY CUM TAX INVOICE

Details of Supplier:
Place of Supply: MAHARASHTRA
Details of Receiver / Shipped to: GANGOTRI CORP. LLP
GROUND FLOOR, GALA NO 4
SNEHA IND ESTATE, EKVIRA GAS GODOWN,
MAHARASHTRA
PIN No. 401108
Buyers GSTIN No. 27AARFG1763H1Z3
Shipping Address:

Details of Receiver:
Ack No.: 162110494814881
Ack Date: 31-03-2021 4:09:00 PM
IRN: G1861402c41187046ac5861baac996a08b
Date of Supply: 31/03/2021
Transporter: PITHAMPUR BOMBAY ROADWAYS
Vehicle No.: MP09GH8921
P.O. No.: EMAIL
Date: 05-03-2021
Destination: BHAYANDAR
State: MAHARASHTRA
Invoice No.: SIFA20212917
Invoice Date: 31/03/2021
State: MADHYA PRADESH
State Code: 23
DO. No.: SDFACD1469
Date: 09-03-2021

Shipments:

Name of Product / Service	HSN / ACS	Quantity in Nos.	Quantity in MT	Rate / MT	Amount
BLACK WIRE 1.83MM	72171020	COILS: 25	2.225	21900.00	58428.60
BLACK WIRE 1.50 MM	72171020	COILS: 25	2.257	71800.00	162052.60
BLACK WIRE 1.63MM	72171020	COILS: 32	2.915	71800.00	209297.00
PATENTED BASE WIRE 2.60MM	72171030	COILS: 13	2.368	65100.00	154156.80

Taxable Value: 583905.00
CGST 18.00%: 105102.90
IGST 18.00%: 105102.90
Round Off: 0.41
GST Payable on Reverse Charge: No
Remarks: Certified that the particulars given above are true and correct and the amount indicated represents the price actually charged and that there is no flow of additional consideration directly or indirectly from the buyer.
Payment Terms: 1. Interest will be charged @16%, if invoice is not paid by due date.
2. All Disputes shall be subject to Jurisdiction of Indore (M.P.) Courts.
3. Interest on Late Payment will attract to GST.

For: RAJRATAN GLOBAL WIRE LIMITED
AUTHORISED SIGNATORY

Rajasthan State Health Assurance
Swasthya Bhawan, Tilak Marg, C-Scheme, Jaipur
Date: 30-01-2021 05

Challan

Patient / Beneficiary Details

Registration Number	Patient Name	Case Type	Normal
128012110515304	Amar Singh	Normal	

Booked Package Details

S No.	Package Name	Package Code	Package Category	Package Rate
1	Medical Treatment (Acute MI) (Self)	W1100024	Tertiary	INR15,000.00
Total				INR15,000.00

Remaining Amount

Secondary Wallet	Tertiary Wallet
INR24,750.00	INR203,250.00

Note: Your medical treatment is totally free of cost to you under Bhamashah Swasthya Bima Yojana.

Signing Authority: 
Hospital Name: Swasthya Bhawan And Trauma center

Satisfaction Certificate

Name of Patient: Amar Singh
Mobile No: 

I/We hereby declare that I/We have taken medical treatment at aforesaid hospital and hospital has not charged me for any expenses made during for the medical treatment. I/We further giving feedback based on my/our experience as mentioned below.

Feedback: 

Signature of Patient: 
Signature of Patient of Thumb Impression: 

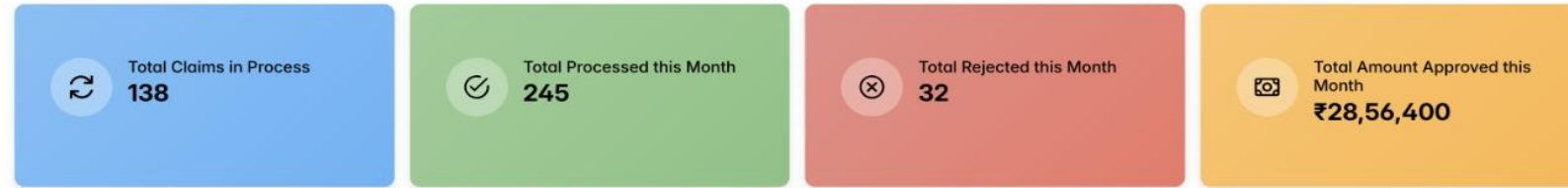
Note: Patient's are advised to mandatorily take a duplicate copy of the Invoice from the hospital.



AI for deciding

ECHS (Edge-AI)

- ECHS (Ex-Servicemen Contributory Health Scheme), with *document labeling* as one module.
- Classifies documents *locally*, minimizing server compute.
- OCR extracts text from PDFs/ images; *lightweight preprocessing* ensures efficiency.
- Keyword-based *labeling* reduces heavy ML model requirements, with optional scoring for accuracy.



Enter claim ID to search

or

Select date range Select City

Min Amount Max Amount

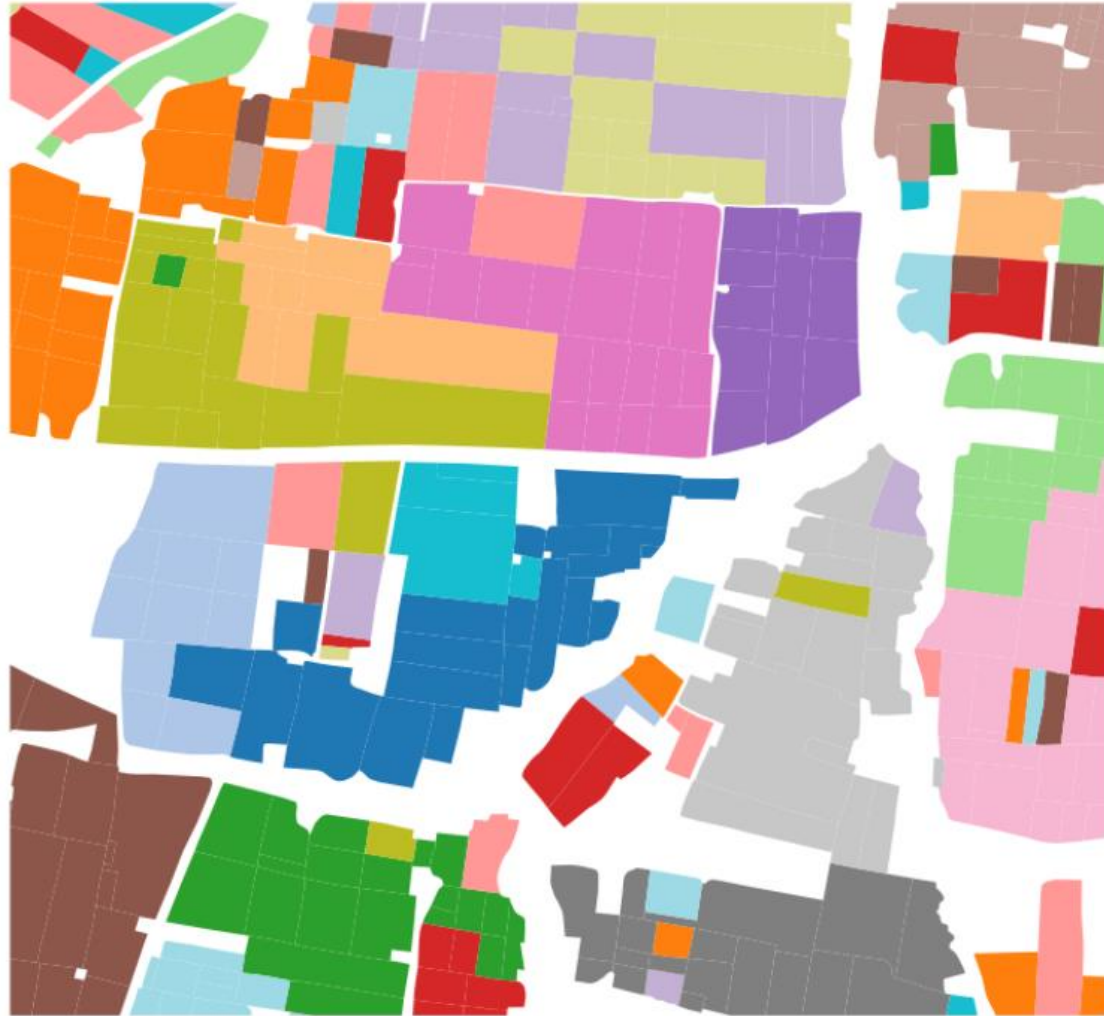
← Authorized Claim Range: ₹4,00,000 - ₹8,00,000

Show Anomalies Only

Claim Status	Name	Age & Gender	Location	Date	Amount	Intimation ID	Card ID
✓ Verified	RAMESH KUMAR SHARMA	81 yrs, Male	New Delhi	17/01/2025	Not specified	35489082	DL2000003791954
✓ Verified	GANGA BHADUR GURUNG	63 yrs, Male	Syangja District	15/01/2025	Not specified	35472639	NE000007549559
⚠ Anomaly Detected (1)	LAKSHMI CHARAN PATNAIK	75 yrs, Male	KABISURYANAGAR	08/01/2025	Not specified	35351667	000006429471

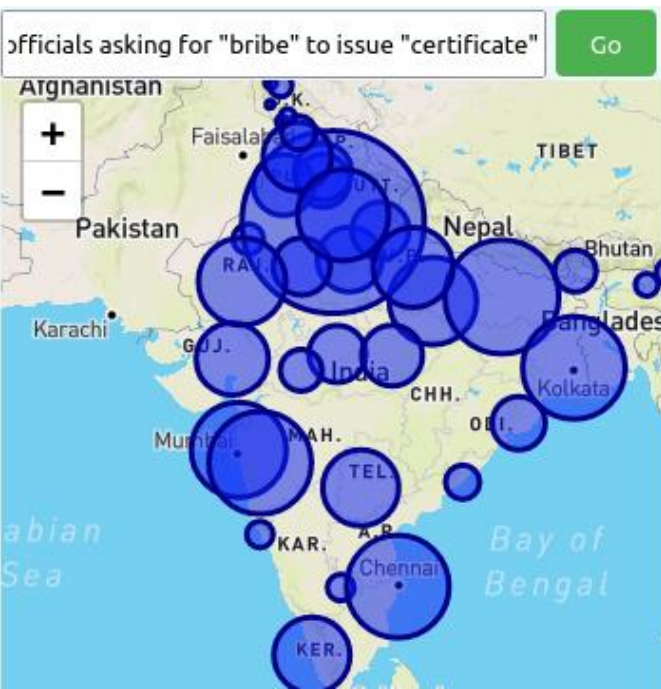
Automated Land Consolidation (General-purpose)

Land **Consolidation** using breadth-first search based on administrative criteria



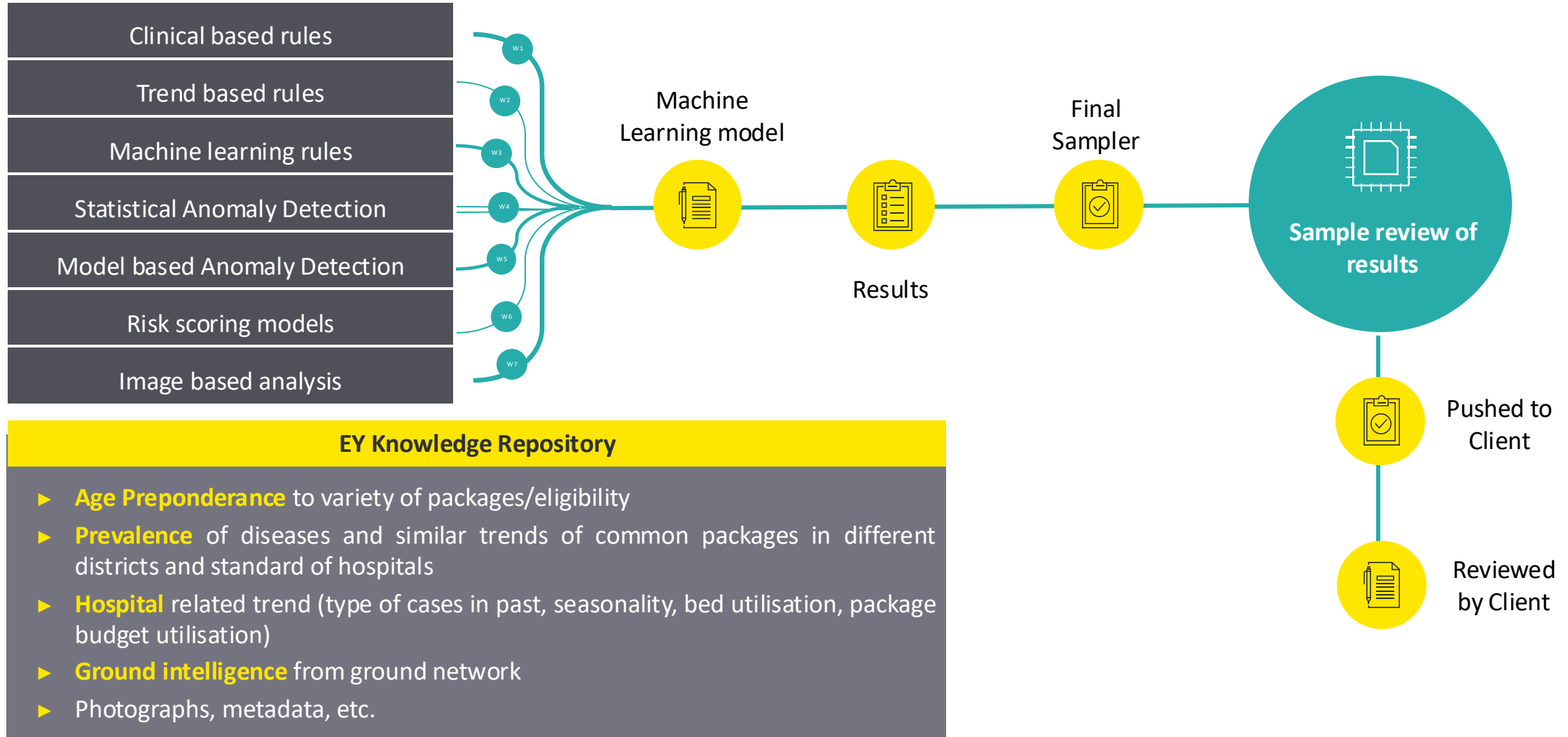
Root-cause analysis (General-purpose)

- Automated **detection of topics** in grievance streams
 - Hierarchical labeling of topics using bags of words



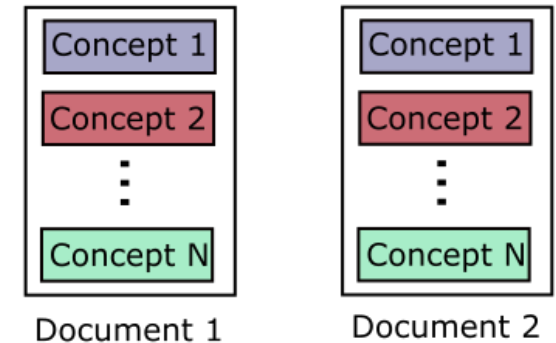
NHA: Health insurance fraud detection (General-purpose)

EY solution has more than 50 rules which are broadly bifurcated in following categories:

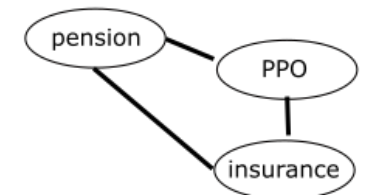


Cybersecurity assessments: AI potential

- In the last five years, LLMs have shown great promise
- Reading documents and extracting information, answering questions based on documents (RAG), etc.
- Reduces tedium on people
- Allows for objectivity especially when there is ambiguity
- Reduces effort, more frequency



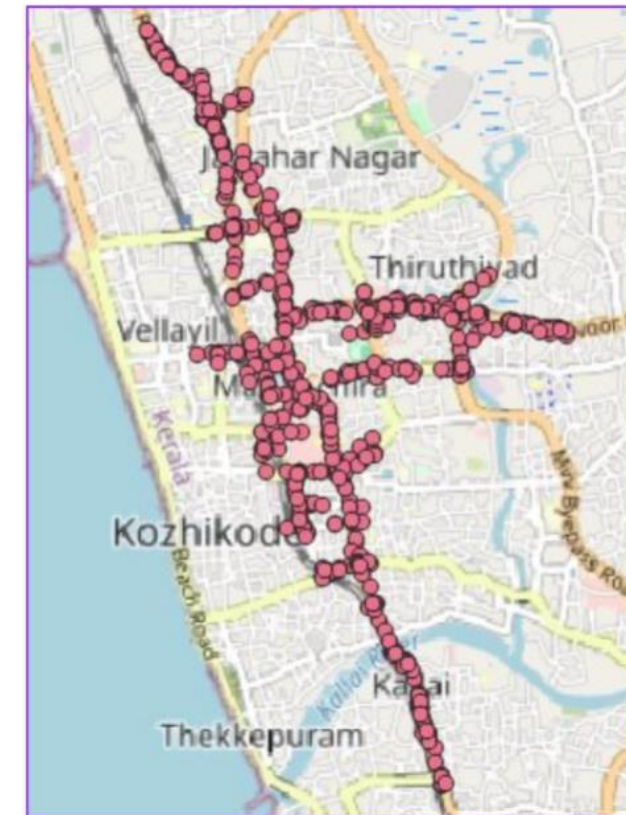
Use AI to understand documents as sets of concepts





Accident Zone Identification System

- AI identifies accident-prone 'black spots' on national highways
- Fuses road geometry, traffic, weather and historical crash data
- GIS visualization for highway planners and engineers
- Directly prevents deaths through pre-emptive road engineering for 1.4 lakh km
- MoU signed with Ministry of Road Transport & Highways (MoRTH) – national-level deployment



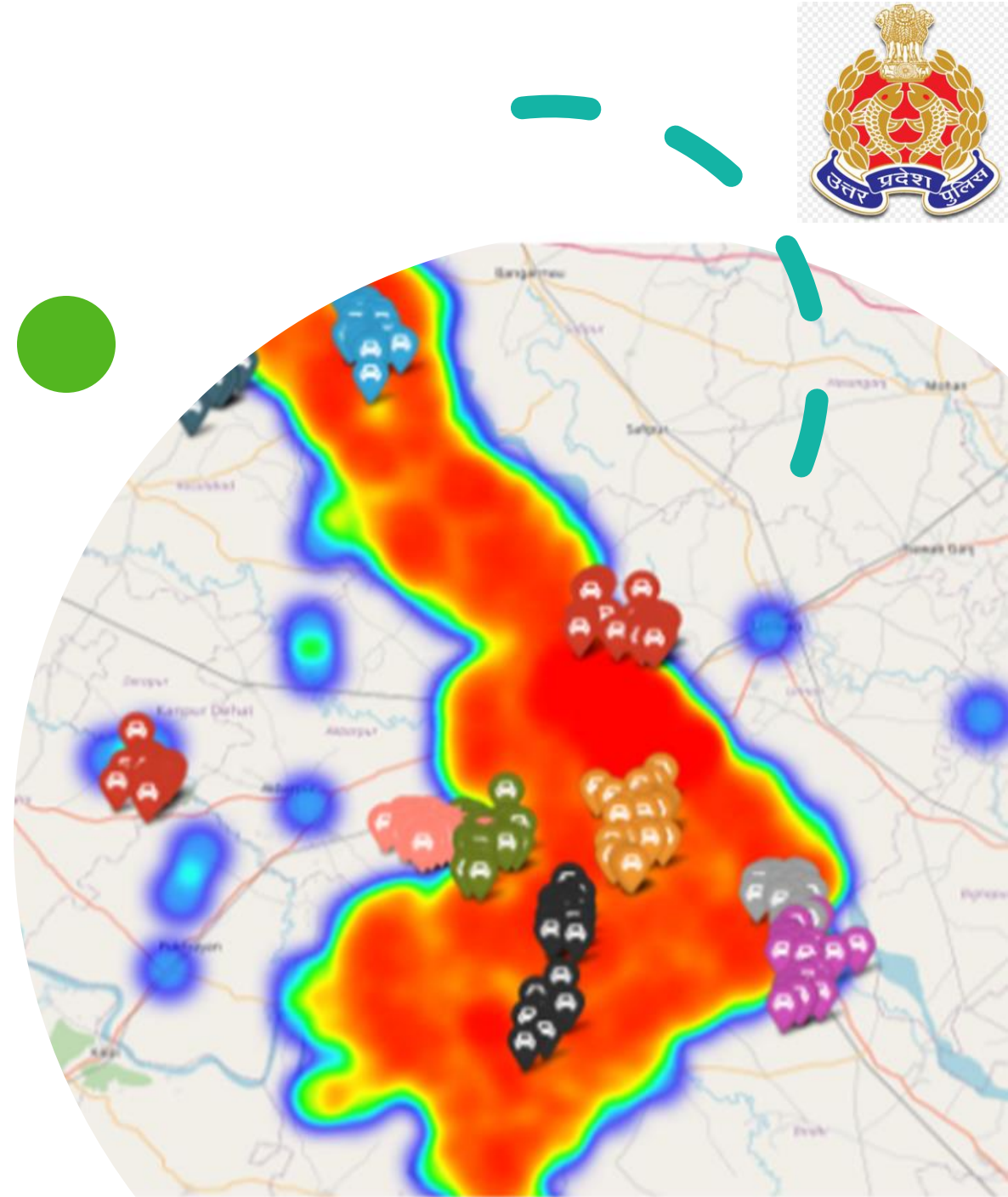
High-density cluster in Kozhikode

AI for acting

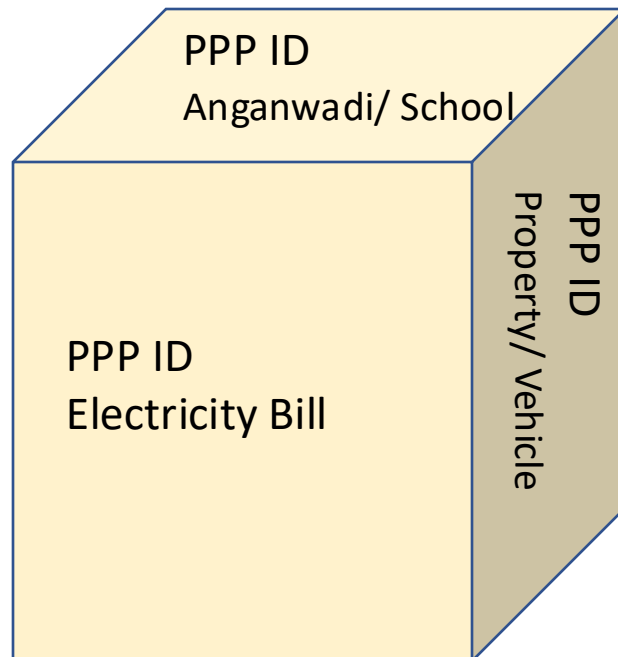


Predictive patrolling

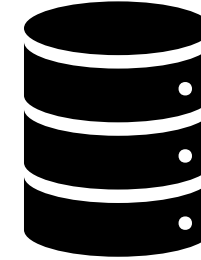
- UP112 has about 700 patrol vehicles that collectively serve a population of about 220 million people
 - UP112 currently receives 50000 calls each day, of which about 10% result in an emergency response
- Average first response times - 20 minutes
 - Average distance driven per patrol vehicle per day - 70 kms
- A **predicting patrolling** system reduces response times by 20% and fuel consumption by 25%
 - Similar idea being used in *allocating* the field engineers in a Bengaluru Internet Provider!



Income assessment



HRMS (Salary labels)



- Use regression modelling to **predict income** levels of households
- **Validate** against held out sample from HRMS, labor, contracts
- **Out-of-sample** validation with **survey** using 2200 respondents from all districts



Dashboard



Category Analysis



Spatial Analysis



Root Cause Analysis



History



Saved

Dashboard / Home

Home

LOG O



Fresh

12,321

+55% than last week



Normal

24,231

+55% than last week



Spam

2,300

+3% than last month



Repeat

3,462

-2% than yesterday



+5% than yesterday

Select Ministry

ALL



From

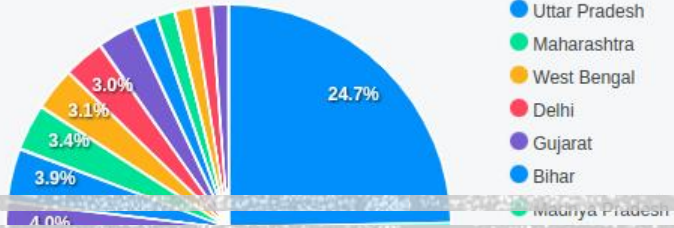
21/07/2023



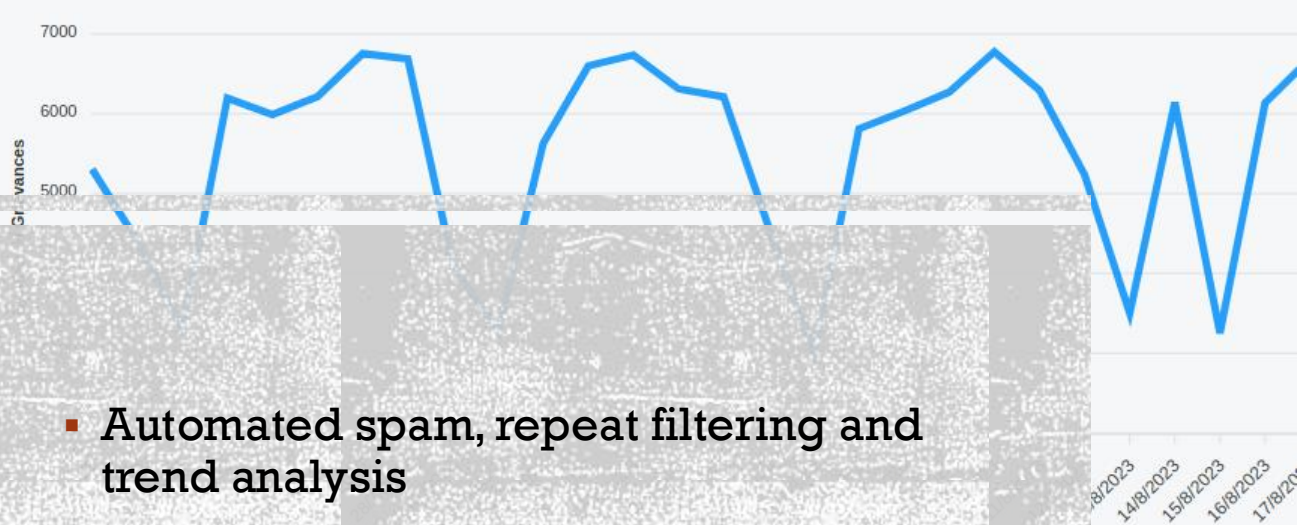
To

21/08/2023

Top 20-States / UT's Grievances



Daily Grievances



IGMS 2.0

- Automated spam, repeat filtering and trend analysis



Nitin Saxena, nitin@iitk.ac.in
Nisheeth Srivastava, nsrivast@iitk.ac.in

CSE & WSAIS & CDIS & CDIS Pvt Ltd & AIRAWAT CoE
IIT Kanpur

<https://iitk.ac.in/cdis/>

Center for Developing Intelligent Systems

Academia



C.D.I.S.

Market



- TRL 1 – basic principles observed
- TRL 2 – technology concept formulated
- TRL 3 – experimental proof of concept
- TRL 4 – technology validated in lab
- TRL 5 – technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies)
- TRL 6 – technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)
- TRL 7 – system prototype demonstration in operational environment
- TRL 8 – system complete and qualified
- TRL 9 – actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies; or in space)

Quick-fire Question?

Q.1.

Which of the following best describes a systemic limitation of current Large Language Models (LLMs)?

- A. They lack a large enough training dataset.
- B. They cannot access the internet in real-time.
- C. They generate text based on probabilistic patterns rather than actual conceptual understanding.
- D. They are incapable of writing functional software code.

Answer: C.

LLMs predict the next most likely token based on statistical patterns, meaning they can mimic reasoning without actually "understanding" the underlying logic.

Quick-fire Question?

Q.2.

Which of these tasks represents a current *boundary* where AI excels practically, but struggles systemically?

- A. Summarizing a 50-page corporate financial report.
- B. Fact-checking a politically sensitive historical event.
- C. Inventing a completely new mathematical theorem from scratch.
- D. Replicating human empathy in grief counseling.

Answer: B (and arguably D).

While AI can practically summarize or synthesize text/math perfectly (A/C), it struggles with objective fact-checking due to training data biases and hallucinations, and lacks genuine emotional capacity.

Quick-fire Question?

Q.3.

If an AI system achieves 99% accuracy on a narrow medical diagnostic task, is it considered safe to deploy autonomously without human supervision?

- A. True.
- B. False.

Answer: B.

The remaining 1% often includes edge cases, systemic biases, or "hallucinations" that require a human-in-the-loop to evaluate context, ethics, and safety boundaries.

Group Discussion Questions?

Topic 1: The "Fluency vs. Factuality" Paradox

Prompt:

Current generative AI can write a flawless, highly persuasive essay that is factually completely incorrect (hallucination).

- i. How do we design workflows that leverage AI's incredible speed and fluency without falling into the trap of trusting its accuracy blindly?
- ii. Who bears the responsibility when an AI-generated error causes real-world harm?

Group Discussion Questions?

Topic 2: Moving the Goalposts of "Intelligence"

Prompt:

Every time AI masters a task previously thought to require human intelligence (like playing chess, passing the bar exam, or writing poetry), critics argue, "That's just math, not true intelligence."

- i. Are we unfairly shifting the definition of intelligence?
- ii. What is one capability an AI could demonstrate that would finally convince you it possesses genuine understanding, rather than sophisticated pattern matching?

Group Discussion Questions?

Topic 3: The Economic Bottleneck

Prompt:

Training and running frontier AI models requires massive data centers, millions of gallons of water for cooling, and immense electrical power.

- i. Is the physical and environmental cost of AI a hard systemic limitation to its scalability?
- ii. Will the future of AI be bottlenecked by infrastructure and physics rather than software engineering?

Group Discussion Questions?

Topic 4: Could AI solve science problems?

Prompt:

The social media is flooded with examples of open questions, in mathematics, that have been solved using LLM.

- i. Does an AI truly "understand" math, or does it just guess the next correct symbol?
- ii. Will human math skills become less valued as AI gets better at proofs?