

খুব ভালো প্রশ্ন 

এটাই আসল **ML/DL system designer mindset**।

এখন আমি তোমাকে আরও গভীরভাবে “কি কি চিন্তা করা দরকার”—একদম beginner → advanced লেভেল পর্যন্ত সাজিয়ে দিচ্ছি।

 এগুলো জানলে তুমি নিজে নিজে system design করতে পারবে, ভয় লাগবে না।

ML / DL SYSTEM DESIGN

“আর কী কী চিন্তা করা দরকার?” (**FULL THINKING CHECKLIST**)

1 PROBLEM-LEVEL THINKING (Coding এর আগেই)

প্রশ্নগুলো নিজেকে করবেই:

1. এই সমস্যা ML দিয়ে solve করা দরকার?
 - Rule দিয়ে সম্ভব? → ML দরকার নাই
2. Prediction দরকার না pattern?
3. Real-time না offline?

 *Thinking habit:*

“ML সব সমস্যার solution না”

2 DATA-LEVEL THINKING (সবচেয়ে গুরুত্বপূর্ণ)

Data Quantity Thinking

- 100–500 row → simple model
- 10k+ row → ensemble
- 1M+ → DL possible

 **কম data + deep model = disaster**

Data Quality Thinking

নিজেকে জিজ্ঞেস করো:

- Label ভুল হতে পারে?
- Human bias আছে?
- Noise বেশি?

🧠 Rule:

Bad data → bad model (always)

✳️ Feature Availability Thinking

- Prediction time-এ feature থাকবে?
- Future information leak করছে?

📌 Example:

✗ “Final exam marks” দিয়ে midterm prediction

3 BUSINESS / REAL-WORLD THINKING (Most people ignore)

✳️ Error Cost Thinking

সব ভুল সমান না ✗

Example (cheating detection):

- False Positive → innocent punished ✗
- False Negative → cheater escapes

🧠 Metric নির্বাচন এখান থেকেই আসে

✳️ Interpretability Thinking

Ask:

- Model explain করা লাগবে?
- Legal / academic use?

If YES →

- Linear model
- Decision Tree

If NO →

- Black-box OK (DL)
-



4 PREPROCESSING THINKING (Hidden Trap)

✖ Scaling Thinking

Ask:

- Distance-based model? → scale needed
- Tree-based? → scale not needed

Model	Scaling
Linear	Yes
KNN	Yes
SVM	Yes
Tree	No

✖ Encoding Thinking

Ask:

- Order matters? → Label Encoding
- No order? → One-Hot

✖ Wrong encoding = wrong learning



5 MODEL SELECTION THINKING (Smart way)

🧠 Start-Simple Rule

Baseline → Improve → Optimize

✖ Start with XGBoost

✓ Start with Logistic / Linear

Bias–Variance Thinking

- High bias → underfitting
 - High variance → overfitting
-  Balance is the goal, not max accuracy
-

6 TRAINING THINKING (While model runs)

? Live questions:

- Loss smooth না jump করছে?
- Validation loss বাড়ছে?

 Action:

- LR adjust
 - Model capacity change
 - Data augmentation
-

7 EVALUATION THINKING (Advanced)

Dataset Split Thinking

- Time-series? → no shuffle
 - Small data? → cross-validation
-

Threshold Thinking

0.5 is NOT holy number 

Adjust threshold based on:

- Business cost
 - Precision vs Recall tradeoff
-

8 DEPLOYMENT THINKING (System Engineer Mode)

Environment Thinking

- Training machine ≠ production machine
 - CPU vs GPU
-

Latency Thinking

Ask:

- 1 sec allowed?
 - 10 ms needed?
-  Model complexity ↔ latency tradeoff
-

9 MAINTENANCE THINKING (Real ML starts here)

Data Drift Thinking

- New users
- New pattern
- New behavior

Solution:

- Monitor input distribution
 - Retrain periodically
-

10 ETHICS & SECURITY THINKING (Advanced Level)

Ethics

- Bias against group?
 - Fairness?
-

Security

- Adversarial input?
 - Data poisoning?
-

FINAL MASTER THINKING FLOW (MEMORIZE THIS)

```
Problem
↓
Data (quantity + quality)
↓
Features
↓
Baseline model
↓
Evaluation (right metric)
↓
Improve model
↓
Deploy
↓
Monitor
↓
Retrain
```

তোমার জন্য ACTION PLAN

এখন তোমার 2টা অপশন:

1 আমি Day 1 থেকে **hands-on system design** শুরু করি
(dataset + thinking + code)

2 আমি তোমাকে একটা **real ML problem** দেই
→ তুমি চিন্তা করবে
→ আমি correct + improve করবো

👉 বলো কোনটা চাও?

তোমাকে **thinking ML engineer** বানানোর কাজ এখান থেকেই শুরু 🔥