BP801T · B.Pharm Semester VIII · Theory Notes

Ethical Considerations and Translational Applications of AI in Pharmacy

Unit II — Regulatory Framework & Explainable AI (6 hours)

Unit II (6 hours)

Pills & Python Notes

From-zero teaching standard: every technical term is defined before use.

Syllabus coverage: Unit II (6 hours).

PILLS & PYTHON

Pharmacy • Data • AI • Healthcare

Instagram @pillsandpython · pillsandpython.com · YouTube @PillsAndPython

CHAPTER 6: Regulatory Frameworks for AI in Pharmacy

1. Learning Objectives

By the end of this chapter, you will be able to:

- Explain what a regulatory framework is, and why AI used in pharmacy needs regulation.
- Describe the **risk-based approach** to AI regulation in simple words.
- Classify pharmacy AI systems into **unacceptable, high, limited, and minimal risk** tiers, with one pharmacy example for each tier.
- Outline, at concept level, the European Union's AI Act (risk-based framework), the US FDA's framework for AI/ML-based medical devices, and the Indian (CDSCO) context — using cautious, "confirm the current edition" language.
- Explain what a risk classification means in practice for a pharmacy AI product: **documentation, human oversight, and post-market monitoring**.
- Apply the framework to judge the likely risk tier of a new pharmacy AI tool.

2. Definitions

Every term is defined here **before** it is used later in the chapter.

- **Regulation** — rules made by an authority that people and companies must follow. Traffic rules are regulation for roads; drug regulation is rules for medicines.
- **Regulator** — the government body that makes the rules and checks that they are followed. Examples: CDSCO in India, FDA in the United States.
- **Regulatory framework** — the full set of laws, rules, and official guidance documents that together govern one area (here: AI in pharmacy and healthcare).
- **Guidance / guideline** — official advice published by a regulator explaining *how* to follow the rules. Simple rule of thumb: the law commands, guidance explains.
- **Artificial Intelligence (AI) system** — software that performs tasks which normally need human intelligence, such as learning from data, making predictions, or supporting decisions.
- **Medical device** — any instrument, machine, or software intended for diagnosis, treatment, monitoring, or prevention of disease. Software alone can be a medical device.
- **Software as a Medical Device (SaMD)** — software that performs a medical function by itself, without being part of a physical instrument. Example: a phone app that analyses a patient's symptoms and suggests a possible diagnosis.
- **Risk** — the chance that something causes harm, combined with how serious that harm could be. Small chance × serious harm can still be high risk.

- **Risk-based approach** — the idea that regulation should be stricter where the risk of harm is higher, and lighter where the risk is low. A toy and a heart pacemaker are not regulated the same way.
- **Unacceptable risk** — AI uses considered so harmful that they are banned or prohibited.
- **High risk** — AI systems that can seriously affect health, safety, or fundamental rights. They are allowed, but only with strict safeguards.
- **Limited risk** — AI systems where the main duty is **transparency**: users must be told they are interacting with AI.
- **Minimal risk** — low-risk AI systems with no special AI obligations beyond ordinary law.
- **Human oversight** — a qualified human being reviews, supervises, or can override the AI's output. The AI advises; the human decides.
- **Post-market monitoring (post-market surveillance)** — watching a product's safety and performance *after* it is sold and in regular use, and acting on problems found.
- **Transparency obligation** — the duty to openly inform users about the AI system: what it is, what it does, and what its limits are.
- **Conformity assessment** — the process of checking that a product meets the regulatory requirements before it is placed on the market.
- **Premarket review** — evaluation of a product by (or for) the regulator *before* it is allowed to be sold.
- **CDSCO** — Central Drugs Standard Control Organisation: India's national regulator for drugs and medical devices.
- **FDA** — the United States Food and Drug Administration: the US regulator for drugs, medical devices, food safety, and related products.

3. Terminology Table

Term	Plain-English meaning	Quick example
Regulation	Must-follow rules from an authority	Prescription-only rule for antibiotics
Regulator	The body that makes and enforces rules	CDSCO, FDA
Regulatory framework	All the laws + rules + guidance for one area, taken together	The set of rules governing AI medical devices
Guidance	Official "how to comply" advice from a regulator	A document explaining how to document an AI model
AI system	Software that learns from data and supports decisions	A tool predicting which patients may miss doses
Medical device	Product intended for diagnosis/treatment/prevention	A diagnostic app (software can qualify)
SaMD	Software that is itself the medical device	An app that interprets ECG data
Risk	Chance of harm × seriousness of harm	Wrong-dose AI advice = high risk
Risk-based approach	Stricter rules for higher risk, lighter rules for lower risk	Pacemaker vs. pedometer app
Unacceptable risk	Banned AI uses	AI that secretly manipulates patients

Term	Plain-English meaning	Quick example
High risk	Allowed only with strict safeguards	AI that adjusts medicine doses by itself
Limited risk	Main duty is transparency	A chatbot that must say "I am an AI"
Minimal risk	No special AI obligations	A simple pill-reminder alarm
Human oversight	A qualified human checks or can override the AI	Pharmacist reviews AI dose suggestion
Post-market monitoring	Safety watch after the product is sold	Tracking AI errors reported by hospitals
Transparency obligation	Duty to tell users about the AI and its limits	"This advice was generated by AI"
Conformity assessment	Pre-sale check that requirements are met	Testing + paperwork review before launch
CDSCO	India's drug and medical-device regulator	Approves medical devices in India
FDA	The US regulator for drugs and devices	Reviews AI/ML-based devices in the US

4. Explanation

4.1 What is regulation? (Starting from zero)

Technical concept: A regulatory framework is the system of laws, rules, and guidance that controls how a product or activity is developed, tested, sold, and monitored.

Simple explanation: Think of traffic rules. Nobody is "against" cars — but without speed limits, signals, and licences, roads would be deadly. Regulation does the same job for medicines and medical technology: it does not stop innovation, it makes innovation safe enough to trust.

Example: A pharmacy cannot sell antibiotics without a prescription. That is regulation. It slows down sales a little, but it protects patients from misuse and resistance.

> **Key takeaway:** Regulation = the safety rules of the game. AI in pharmacy is entering the game, so it needs its rules too.

4.2 Why does AI in pharmacy need regulation?

Technical concept: AI systems in healthcare are "high-stakes" systems: their errors can injure people, their biases can discriminate, and their opacity can hide mistakes.

Simple explanation: A wrong answer from a movie-recommendation AI is a minor annoyance. A wrong answer from a pharmacy AI can mean the wrong dose, a missed interaction, or a delayed treatment. Three specific dangers make regulation necessary:

1. **Patient harm from errors.** An AI that miscalculates a dose or misses a drug interaction can directly injure a patient.
2. **Bias and unfairness.** An AI trained mostly on data from one population may work poorly for others — for example, a dosing model trained on adult data applied to elderly patients.
3. **Opacity and over-trust.** If nobody understands how the AI reached its answer (the "black box" problem — see Chapter 7), errors go unnoticed and users trust the machine too much.

Example: Imagine an AI tool that suggests antibiotic doses. If it was trained only on data from one hospital and is used in another hospital with different resistance patterns, its suggestions could be wrong — and a busy pharmacist might accept them without checking. Regulation requires testing across settings, clear labelling of limits, and a human in charge.

> **Key takeaway:** Pharmacy AI needs regulation because its mistakes can harm patients, its biases can be unfair, and its inner workings are often invisible.

4.3 The risk-based approach — the big idea behind modern AI regulation

Technical concept: A risk-based approach classifies AI systems by the level of risk they pose and applies proportionate obligations: bans for unacceptable risk, strict safeguards for high risk, transparency duties for limited risk, and no special duties for minimal risk.

Simple explanation: Not all AI is equally dangerous, so it should not be equally regulated. A strict rule for everything would crush small, harmless innovations; no rules at all would let dangerous systems loose. The risk-based approach is the middle path: **match the strictness of the rule to the size of the risk.**

Example: Compare three pharmacy tools:

- An AI that **autonomously changes a patient's warfarin dose** → could cause bleeding or clotting → **high risk** → strict rules.
- An AI **chatbot that answers "where is the cough syrup shelf?"** → worst case is a wrong shelf direction → **limited risk** → just tell users it is a bot.
- An AI that **sorts pharmacy emails into folders** → negligible harm → **minimal risk** → no special AI rules.

Diagram: See the risk pyramid in Section 7.

> **Key takeaway:** Risk-based regulation = heavy helmet for the dangerous ride, light cap for the safe walk.

4.4 The EU AI Act — a risk-based framework (concept overview)

Technical concept: The European Union's Artificial Intelligence Act is a comprehensive, risk-based legal framework for AI. It sorts AI systems into risk tiers and attaches obligations to each tier. (Describe only the concept here — for the exact legal text, articles, dates, and timelines, always consult the current official edition.)

Simple explanation: The EU looked at all possible AI uses and asked one question: "How much harm could this cause?" Then it built four boxes:

1. **Unacceptable risk → banned.** AI uses that clearly threaten safety or rights are prohibited.
2. **High risk → allowed with strict safeguards.** Systems that influence important decisions (health, safety, employment, education) must meet requirements on data quality, documentation, human oversight, accuracy, and post-market monitoring.
3. **Limited risk → transparency duties.** If you deploy a chatbot or AI-generated content, you must tell people they are dealing with AI.

4 **Minimal risk → free to use.** Ordinary low-risk AI (spam filters, video-game AI) faces no extra obligations.

Pharmacy example for each box:

- *Unacceptable:* an AI marketing tool that secretly manipulates vulnerable patients into buying medicines they do not need.
- *High risk:* an AI system that independently adjusts insulin doses.
- *Limited risk:* a pharmacy website chatbot that answers stock queries — it must disclose it is AI.
- *Minimal risk:* an AI tool that organises the pharmacy's internal duty roster.

Cautious note: This is a concept-level description. The Act's exact scope, articles, dates, and transition timelines change — **always confirm the current official edition** before quoting specifics.

Key takeaway: The EU AI Act = sort AI by risk into four boxes; the box decides the rules.

4.5 The US FDA framework for AI/ML-based medical devices (concept overview)

Technical concept: In the United States, the FDA regulates AI/ML-based software that qualifies as a medical device (including SaMD). The framework works across the **total product lifecycle**: careful premarket review before sale, plus post-market monitoring after sale. Because AI/ML models can learn and change after deployment, the framework pays special attention to managing updates and real-world performance.

Simple explanation: The FDA asks: "Does this AI software diagnose, treat, or guide treatment?" If yes, it is treated like a medical device:

- **Before sale (premarket):** the maker must show evidence the product is safe and works as claimed — testing data, validation, and clear labelling of what the AI does and does not do.
- **After sale (post-market):** the maker must keep watching real-world performance, report problems, and manage changes to the model carefully, because a learning system can drift.

Pharmacy example: An AI app that reads a photo of a prescription and flags possible prescribing errors is doing a medical job (error detection affecting treatment). Under this concept it would face premarket review and ongoing monitoring. A simple pill-reminder alarm that only rings at set times is not making medical judgments, so it would not be treated the same way.

Cautious note: FDA thinking in this area evolves through guidance documents. Product teams should **confirm the current guidance edition** rather than relying on any single summary.

Key takeaway: FDA concept = if the AI does a medical job, prove it is safe before sale and keep watching it after sale — across the product's whole life.

4.6 CDSCO and the Indian context (cautious wording)

Technical concept: In India, CDSCO is the national regulator for drugs and medical devices. AI-based software that qualifies as a medical device falls within the medical-device regulatory system.

Simple explanation: India does not have one single famous "AI law" of the EU type described above. Instead, the practical question for a pharmacy AI product in India is: **"Is this software a medical device?"** If an AI tool diagnoses, monitors, or guides treatment, it is likely to be treated as a medical device and must follow the medical-device rules — registration/licensing, quality requirements, labelling, and adverse-event reporting, as applicable.

What this means for students and pharmacists:

- An AI inventory chatbot for a pharmacy → unlikely to be a medical device → general consumer/IT law applies.
- An AI tool that interprets lab values and recommends dose changes → likely a medical device → CDSCO pathway applies.
- When in doubt, the product's **intended use** (what the maker claims it does) decides the category — not the technology inside.

Cautious note: Regulatory guidance in this area is developing. **As per the latest guidance, confirm the current edition** of CDSCO medical-device rules and any AI-specific advisories before making compliance decisions. Do not quote version numbers, dates, or fees from memory or secondary summaries.

Key takeaway: In India, the key question is "is this AI a medical device?" — if yes, the CDSCO medical-device framework applies; always verify the current edition.

4.7 The four risk tiers, with pharmacy examples

Tier 1 — Unacceptable risk (banned/prohibited). *Concept:* AI uses so harmful to safety or rights that they should not exist. *Pharmacy example:* An AI sales tool that deliberately hides a drug's side effects from patients to increase sales, or that manipulates addicted patients into buying more. Such deceptive, harmful manipulation is the kind of use regulators prohibit. *Student note:* You will rarely meet this tier in legitimate pharmacy practice — it exists to draw a hard line.

Tier 2 — High risk (allowed with strict safeguards). *Concept:* AI that can significantly affect health, safety, or rights. Permitted, but the maker must meet strict requirements. *Pharmacy examples:*

- AI that **autonomously adjusts doses** (e.g., insulin, warfarin) without a clinician confirming each change.
- AI **clinical decision support** that recommends treatments or flags critical interactions in a hospital.
- AI that **triages patients** (decides who needs urgent pharmacist/doctor review).

What is required (concept level): good training data, testing and validation, clear documentation, human oversight, accuracy and robustness, and post-market monitoring.

Tier 3 — Limited risk (transparency duties). *Concept:* The AI is not dangerous enough for strict safeguards, but users have a right to know they are dealing with AI. *Pharmacy examples:*

- A pharmacy **inventory chatbot** ("Is paracetamol 650 mg in stock?") — it must disclose it is an AI assistant.
- An AI tool that **drafts patient information leaflets** — the output should be labelled as AI-assisted so a pharmacist verifies it before use.

Tier 4 — Minimal risk (no special AI obligations). *Concept:* Ordinary, low-stakes AI. General law still applies (e.g., consumer protection), but no extra AI-specific duties. *Pharmacy examples:* AI that sorts supplier emails, a smart duty-roster planner, an AI grammar checker for internal memos.

4.8 What classification means for a pharmacy AI product

Technical concept: A product's risk tier decides its compliance burden: what must be documented, how much human oversight is required, and how closely it must be monitored after launch.

Simple explanation: The tier is not just a label — it is a **to-do list**. Here is what each tier demands from the product team (and what the pharmacist user should check before buying or trusting the tool):

1. **Documentation.** Higher tiers need written proof: what data trained the model, how it was tested, what its limits are, and how it should (and should not) be used. *Pharmacist's check:* "Show me the validation report and the intended-use statement."
2. **Human oversight.** High-risk tools must keep a qualified human in control — the AI suggests, the pharmacist decides and can override. *Pharmacist's check:* "Can I see, understand, and reject the AI's suggestion?"
3. **Post-market monitoring.** Higher tiers require active watching after launch: collecting error reports, tracking performance drift, and updating or withdrawing the product if problems appear. *Pharmacist's check:* "How do I report it when the AI is wrong, and what happens then?"

Example: A hospital considers buying an AI that flags drug interactions. Classified as high risk, the vendor must provide test evidence, build in pharmacist review of every critical alert, and run a system for collecting missed-interaction reports after deployment. A chatbot for stock queries (limited risk) mainly needs a clear "I am an AI assistant" disclosure.

Key takeaway: Risk tier → to-do list: document it, keep a human in charge, and keep watching it after launch.

5. Classifications

5.1 Classification of AI systems by risk tier (summary)

Tier	Core idea	Regulatory consequence	Pharmacy example
Unacceptable	Too harmful to permit	Prohibited/banned	AI that deceives patients to push sales
High	Can seriously affect health/safety/rights	Strict safeguards: data quality, documentation, human oversight, monitoring	Autonomous dose-adjustment AI
Limited	Low danger but users must know it is AI	Transparency/disclosure duties	Inventory chatbot ("I am an AI assistant")
Minimal	Negligible risk	No special AI obligations	Duty-roster planner

5.2 Classification of common pharmacy AI products

Pharmacy AI product	Likely tier	Reasoning
AI that independently changes insulin doses	High	Direct effect on patient safety; error can kill

Pharmacy AI product	Likely tier	Reasoning
AI clinical decision support recommending antibiotics	High	Influences treatment decisions
AI triage: which prescriptions need urgent review	High	Delay can harm patients
Drug-interaction alert system in hospital software	High	Safety-critical advice
Pharmacy stock chatbot	Limited	Must disclose AI identity; low harm
AI-drafted patient leaflets (pharmacist-reviewed)	Limited	Transparency + human verification needed
AI demand forecasting for inventory	Limited/Minimal	Business tool; harm is financial, but if it causes stock-outs of critical drugs, treat with care
Email sorter / roster planner	Minimal	No health impact

Note: "likely tier" is illustrative — the real classification depends on the product's exact intended use and the applicable regulation's current text.

6. Mechanisms

Regulation works through mechanisms **before** sale and **after** sale. Think of it as two gates.

6.1 Premarket mechanisms (before the product reaches users)

1. **Classification of the product.** First decide: is this AI a medical device (or high-risk AI)? The intended use decides.
2. **Conformity assessment / premarket review.** The maker checks (and the regulator may check) that requirements are met: suitable training data, validated performance, risk analysis, clear labelling and instructions.
3. **Documentation (technical file).** A written record: what the AI is for, how it was built and tested, its limits, and how users should operate it safely.
4. **Registration/listing.** The product is registered with the regulator so it can be traced.

Simple picture: like getting a driving licence — learn, get tested, carry the licence — before driving alone.

6.2 Post-market mechanisms (after the product is in use)

1. **Post-market monitoring/surveillance.** The maker actively collects reports of errors, near-misses, and unexpected behaviour.
2. **Incident reporting.** Serious problems (e.g., the AI contributed to a patient injury) must be reported to the regulator.
3. **Corrective action.** If a pattern of harm appears: warn users, update the model, restrict use, or withdraw the product.
4. **Managing AI updates.** Because AI/ML models can learn and change, updates that alter performance must be controlled and re-validated — a model is not "approved once, safe forever."

6.3 How the pieces fit: the product lifecycle



Key takeaway: Regulation is not a one-time exam; it is a licence plus lifelong supervision.

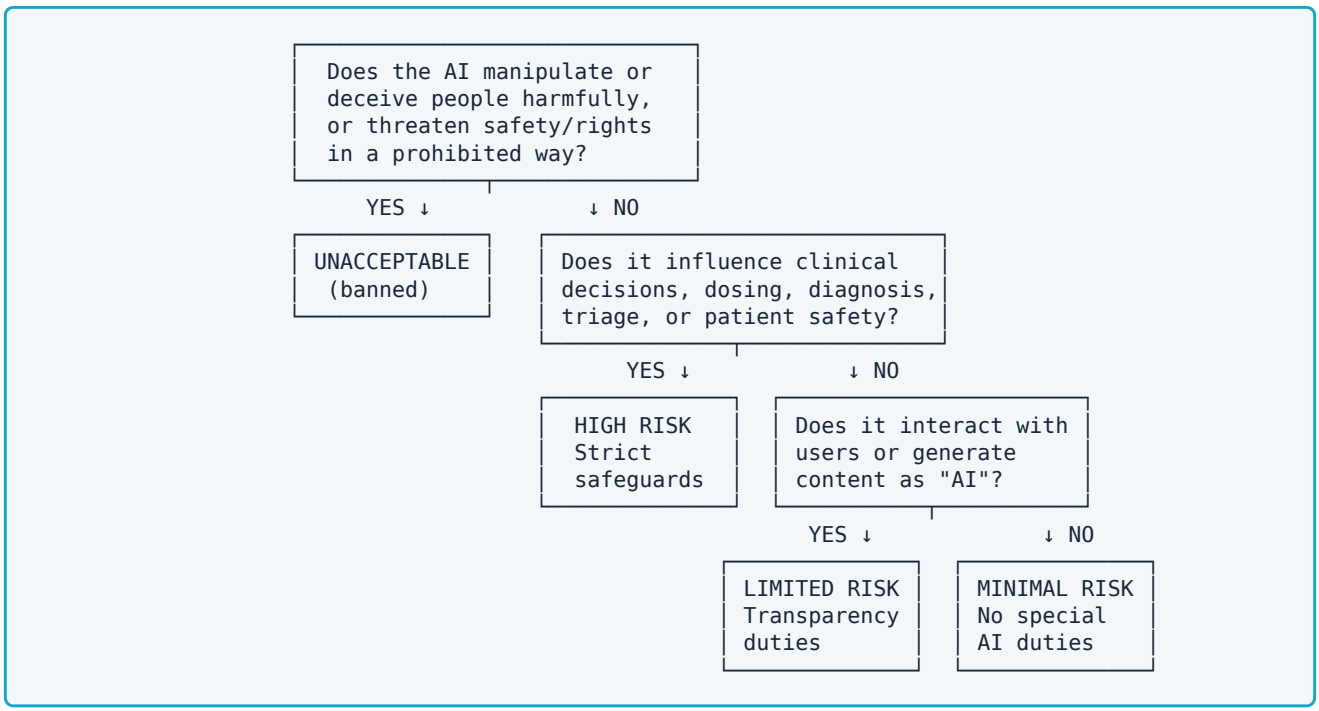
7. Diagrams / Flowcharts

7.1 The risk pyramid

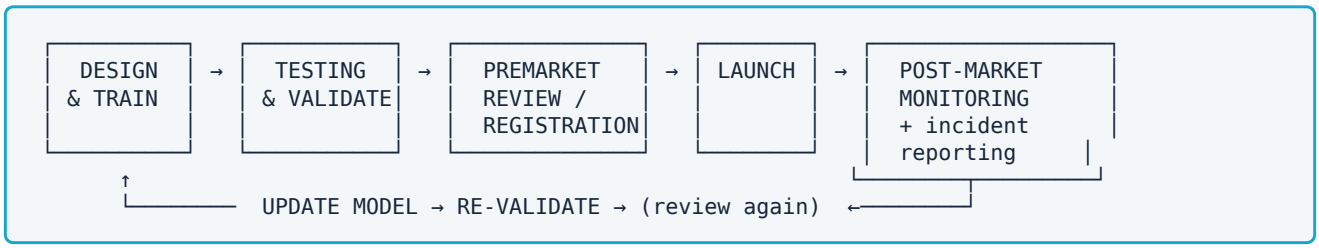


Reading the pyramid: the higher the tier, the fewer systems it contains but the stricter the rules.

7.2 "Which risk tier is my pharmacy AI product?" — decision flowchart



7.3 AI product regulatory lifecycle



8. Infographic Concepts

(Descriptions a designer could use to draw visual summaries.)

1 **"The Four Boxes" poster.** Four large boxes stacked as a pyramid, coloured red (unacceptable — with a prohibition sign), orange (high — with a shield and checklist), yellow (limited — with a speech bubble saying "I am AI"), and green (minimal — with a relaxed tick). Each box holds one pharmacy icon: a banned manipulative ad, a dose-adjustment robot with a pharmacist supervising, a chatbot, and a roster calendar. Caption: "Same technology, different risk — different rules."

2 **"Before and after the sale" timeline.** A horizontal timeline split into two halves: left half "BEFORE SALE" (design → testing → review → registration, with document icons) and right half "AFTER SALE" (monitoring → incident reports → updates, with magnifying-glass icons). A loop arrow shows updates returning to testing. Caption: "Regulation is a licence plus lifelong supervision."

3 **"The pharmacist's three questions" card.** Three big question marks, each with a checklist: ① "Show me the proof" (validation report, intended use); ② "Can I override it?" (human oversight); ③ "How do I report errors?" (post-market monitoring). Designed as a pocket card for pharmacy practice.

9. Tables

Table 6.1 — Risk tiers at a glance

Tier	Meaning	Regulatory consequence	Pharmacy example
Unacceptable	Too harmful to permit	Prohibited	AI that deceives patients to push unnecessary sales
High	Can seriously affect health/safety	Strict safeguards	Autonomous dose-adjustment AI
Limited	Low danger; users must know it is AI	Transparency duties	Inventory chatbot
Minimal	Negligible risk	No special AI duties	Duty-roster planner

Table 6.2 — Regulator comparison (concept level)

Region	Body	Approach	Key idea
European Union	EU AI Act framework	Risk-based, four tiers	Box decides the rules
United States	FDA	Medical-device framework for AI/ML devices; total product lifecycle	Prove safety before sale; monitor after sale
India	CDSCO	Medical-device regulation; AI software that is a medical device falls under it	"Is it a medical device?" decides the pathway; confirm current edition

Table 6.3 — What each tier demands from a product team

Obligation	Unacceptable	High	Limited	Minimal
Ban	Yes — must not be built/deployed	No	No	No
Documentation & validation	—	Extensive	Basic (intended use, limits)	None special
Human oversight	—	Required	Recommended where advice is given	Not required
Post-market monitoring	—	Required	Good practice	Not required
Transparency ("this is AI")	—	Yes, plus more	Yes — the core duty	Not required

10. Worked Toy Examples

All examples below are illustrative teaching scenarios, not real products or real classifications.

Example A — "DosePilot": AI that autonomously adjusts warfarin doses. Step 1: Does it manipulate/deceive? No. Step 2: Does it influence clinical decisions and patient safety? Yes — directly changes doses of a high-risk drug. → **High risk.** Consequences: needs validation data, pharmacist/doctor oversight built in, full documentation, post-market error tracking.

Example B — "StockBuddy": chatbot answering "is this medicine in stock?" Step 1: Manipulative? No. Step 2: Clinical influence? No — it reports stock levels. Step 3: Interacts with users as AI? Yes. → **Limited risk.** Consequence: must disclose "I am an AI assistant"; answers about stock should still be verifiable.

Example C — "RotaSmart": AI that plans pharmacist duty rosters. Step 1–3: No to all. → **Minimal risk.** No special AI duties; ordinary employment/IT rules apply.

Example D — "AdPush": AI that hides side-effect information to boost sales of a costly brand. Step 1: Deceptive and harmful to patients? Yes. → **Unacceptable risk.** This must not be built or deployed.

Example E — "InteractGuard": AI that flags dangerous drug interactions in hospital prescriptions. Step 1: No. Step 2: Influences treatment safety? Yes — a missed alert can kill; a false alert can cause alert fatigue. → **High risk.** Needs clinical validation, pharmacist review workflow, documentation, and post-market monitoring of missed/false alerts.

Key takeaway from the examples: Always classify by **intended use and potential harm**, never by how "advanced" the technology sounds.

11. Analogies

1 **Traffic rules.** Regulation does not ban cars; it sets speed limits, signals, and licences. AI regulation does not ban AI; it sets risk tiers, safeguards, and oversight.

2 **Driving licence + traffic police.** Premarket review is the driving test and licence; post-market monitoring is the traffic police watching the roads afterwards.

3 **Food labelling.** Just as packaged food must list ingredients and expiry dates, AI products must disclose what they are, what data they used, and their limits. The label does not make the food tasty — it makes it trustworthy.

4 **Helmet sizes.** A cyclist and a motorbike racer both wear helmets, but the racer's protection is far stricter. Same idea: minimal-risk AI gets light duties; high-risk AI gets the full protective gear.

12. Pharmacy Relevance

Why must a **pharmacist** — not just an engineer or lawyer — understand AI regulation?

1 **You will buy and use these tools.** Hospitals and retail chains already purchase AI dispensing aids, interaction checkers, and inventory predictors. The pharmacist on the purchase committee must ask: "What risk tier is this? Where is the validation report? Can I override it?"

2 **You are the human in "human oversight."** For high-risk tools, regulation assumes a qualified professional supervises the AI. In pharmacy, that professional is usually the pharmacist. If you blindly accept every AI suggestion, the safeguard fails.

3 **You counsel patients about AI apps.** Patients will ask, "Doctor Google says… / this app says…" You need to explain which AI health tools are regulated medical devices and which are unregulated wellness toys.

4 **You report AI errors.** Post-market monitoring depends on frontline reports. A pharmacist who notices an AI interaction checker repeatedly missing a serious interaction is a vital sensor in the safety system.

5 **Your licence is on the line.** If an AI error harms a patient, "the computer told me to" is not a defence. Understanding your oversight duty protects patients and yourself.

13. India Context

- **CDSCO is the anchor.** In India, AI software that qualifies as a medical device is handled within the medical-device regulatory system overseen by CDSCO. There is no need to memorise a separate "Indian AI Act" — the exam-relevant point is the **"is it a medical device?"** test described in Section 4.6.
- **Cautious wording is mandatory.** Guidance in this area is evolving. In exams and in practice, write and say: *"as per the latest guidance; confirm the current edition"* — never quote article numbers, version numbers, dates, or fees from memory.
- **Jan Aushadhi and public supply chains.** Government drug-supply programmes handle huge volumes at low margins. AI tools for demand forecasting and stock monitoring here are typically limited/minimal risk — but a forecasting failure that causes a stock-out of an essential medicine has real public-health impact, so even "low-risk" tools deserve validation and human review.
- **E-pharmacy growth.** Online pharmacies use AI for order processing, chat support, and logistics. Pharmacists should know which of these are transparency-duty chatbots (limited risk) and which cross into clinical advice (high risk — e.g., an AI that recommends medicines based on uploaded symptoms).
- **Practical tip for students:** in viva, if asked about Indian AI regulation, the safe, correct answer is: "AI-based medical device software falls under the CDSCO medical-device framework; the exact requirements depend on the product's intended use and risk class — I would confirm the current edition of the guidance."

14. Key Points

- Regulation = safety rules; it enables trust in AI rather than stopping innovation.
- Modern AI regulation is **risk-based**: stricter rules for higher risk.
- Four tiers: **unacceptable (banned) → high (strict safeguards) → limited (transparency) → minimal (no special duties)**.
- EU AI Act (concept): four risk boxes; the box decides the rules.
- US FDA (concept): AI/ML medical-device software needs premarket evidence and post-market monitoring across the total product lifecycle.
- India/CDSCO (concept): ask "is this AI a medical device?"; confirm the current edition of guidance.
- A product's tier decides three practical duties: **documentation, human oversight, post-market monitoring**.
- Classify by **intended use and potential harm**, not by how advanced the technology sounds.
- The pharmacist is usually the "human" in human oversight — blind acceptance breaks the safeguard.
- Post-market monitoring depends on frontline error reporting by pharmacists.

15. Common Confusions

Confusion	Clarification
"Regulation and guidance are the same."	The law/regulation commands; guidance explains how to comply. Both matter, but they are different documents.
"High risk means banned."	No — high risk means **allowed with strict safeguards**. Only the unacceptable tier is banned.
"If the EU passed it, it applies in India."	No. The EU AI Act applies in the EU. It may *influence* global practice, but Indian products follow Indian (CDSCO) requirements.
"FDA-cleared/approved once = safe forever."	No. AI/ML models can learn and drift; the framework requires ongoing monitoring and controlled updates.
"A chatbot is too simple to be regulated."	Limited-risk systems still carry **transparency duties** — users must know it is AI.
"The vendor handles compliance; the pharmacist just uses the tool."	The pharmacist provides human oversight and error reporting — compliance is shared.

16. Common Mistakes

1. **Quoting article numbers, dates, or version numbers** of the EU AI Act or FDA guidance from memory — these change; the rule is "confirm the current edition."
2. **Classifying by technology instead of intended use** — e.g., calling every chatbot "minimal risk" even when it gives clinical advice.
3. **Forgetting post-market monitoring** — writing only about premarket approval in exam answers.
4. **Saying "the AI is responsible"** — accountability always sits with humans (see Chapter 7).
5. **Ignoring the Indian context** — in Indian exams, always add the CDSCO line.

6 **Confusing transparency with safety** — disclosing "this is AI" does not make a high-risk system safe; each tier has its own duties.

17. Mnemonics

- **Risk tiers: "U–H–L–M" → "Usually Harmless? Look More!"** — Unacceptable, High, Limited, Minimal. Read top-down: the higher you go, the harder you must look.
- **The three product duties: "DOP" → "Document, Oversee, Post-watch."** Documentation, human Oversight, Post-market monitoring.
- **Premarket vs post-market: "Licence, then police."** Licence (premarket review) first, traffic police (post-market monitoring) afterwards.
- **India question: "Is it a medical device?"** — one question that decides the CDSCO pathway.

18. Exam Summary

AI regulation in pharmacy follows a **risk-based approach**: obligations grow with potential harm. The four tiers are **unacceptable risk** (banned — e.g., deceptive patient-manipulation AI), **high risk** (allowed with strict safeguards — e.g., autonomous dose adjustment, clinical decision support, triage AI), **limited risk** (transparency duties — e.g., inventory chatbots that must disclose they are AI), and **minimal risk** (no special duties — e.g., roster planners). Conceptually, the **EU AI Act** sorts AI into these risk boxes; the **US FDA** treats AI/ML medical-device software (including SaMD) under a total-product-lifecycle framework with premarket review and post-market monitoring, including control of model updates; in **India**, AI software that is a medical device falls under the **CDSCO** medical-device framework, decided by intended use — always with the caution "confirm the current edition." A product's tier creates three practical duties: **documentation** (validation, intended use, limits), **human oversight** (a qualified professional — usually the pharmacist — can review and override), and **post-market monitoring** (error collection, incident reporting, corrective action). Regulation works through **premarket mechanisms** (classification, conformity assessment/review, registration) and **post-market mechanisms** (surveillance, incident reporting, update control).

19. One-Minute Revision

- Risk-based regulation: strict rules for high risk, light rules for low risk.
- Four tiers: **Unacceptable** (banned) → **High** (safeguards) → **Limited** (transparency) → **Minimal** (nothing special).
- Autonomous dose AI = high risk; inventory chatbot = limited risk; roster planner = minimal risk.
- EU AI Act = four risk boxes (concept; confirm current edition).
- FDA = premarket proof + post-market watch, whole product lifecycle.
- India = CDSCO; key question: "is it a medical device?"
- Tier → to-do list: **Document, Oversee, Post-watch (DOP).**
- Pharmacist = the human in human oversight; report AI errors.
- Never quote article numbers/dates/versions from memory.

20. MCQs

Q1. The core idea of a risk-based approach to AI regulation is: a) Banning all AI in healthcare b) Applying stricter rules where risk of harm is higher, lighter rules where it is lower c) Letting companies self-certify everything d) Regulating only foreign AI products **Answer: b**

Q2. An AI system that autonomously adjusts a patient's insulin dose would most likely be classified as: a) Minimal risk b) Limited risk c) High risk d) No risk **Answer: c**

Q3. A pharmacy inventory chatbot ("Is this medicine in stock?") most likely falls under: a) Unacceptable risk b) High risk c) Limited risk, with transparency duties d) Complete ban **Answer: c**

Q4. The main obligation for limited-risk AI systems is: a) A complete ban on deployment b) Transparency — users must know they are interacting with AI c) No obligation at all d) Daily reporting to the regulator **Answer: b**

Q5. "Post-market monitoring" means: a) Advertising the product after launch b) Watching the product's safety and performance after it is sold and acting on problems c) Monitoring competitors' products d) Counting sales after launch **Answer: b**

Q6. In the US, AI/ML-based software that qualifies as a medical device is regulated under the framework of the: a) CDSCO b) FDA, across the total product lifecycle c) Local municipal corporation d) Pharmacy council only **Answer: b**

Q7. In India, the key question that decides whether a pharmacy AI product falls under medical-device regulation is: a) How expensive the software is b) Whether the software is a medical device, based on its intended use c) Whether it was made in India d) How many users it has **Answer: b**

Q8. Which of the following is a premarket regulatory mechanism? a) Incident reporting by hospitals b) Conformity assessment / premarket review before sale c) Post-launch advertising d) Customer satisfaction surveys **Answer: b**

Q9. Why is controlling AI model updates part of the regulatory framework? a) Updates are always illegal b) A learning model can change its behaviour after deployment, so updates need re-validation c) Regulators dislike software updates d) Updates increase the price **Answer: b**

Answer key: 1-b, 2-c, 3-c, 4-b, 5-b, 6-b, 7-b, 8-b, 9-b

21. Viva / Application Questions

1 Define "risk-based approach" and give one pharmacy example each for high, limited, and minimal risk AI.

2 Your hospital wants to buy an AI tool that flags drug interactions. What three things (DOP) would you check before approving the purchase?

3 Explain the difference between premarket review and post-market monitoring with a non-pharmacy analogy.

4 A vendor says their AI dose calculator is "FDA approved, so it is 100% safe forever." What is wrong with this statement?

5 How would you classify an AI chatbot on a pharmacy website that answers questions about medicine stock? Justify your answer.

6 What is the role of the pharmacist in "human oversight" of a high-risk AI system?

7 In the Indian context, how do you decide whether a pharmacy AI product falls under CDSCO regulation? What cautious wording should you use?

8 Describe the EU AI Act's four risk tiers in your own words, without quoting any article numbers or dates.

22. References

- European Union — Artificial Intelligence Act: risk-based framework for AI systems (four risk tiers: unacceptable, high, limited, minimal). Consult the current official edition for the exact legal text, articles, and timelines.
- United States Food and Drug Administration (FDA) — framework for AI/ML-based medical devices, including Software as a Medical Device (SaMD) concepts and total product lifecycle thinking (premarket review and post-market monitoring). Confirm the current guidance edition.
- Central Drugs Standard Control Organisation (CDSCO), India — national regulator for drugs and medical devices; AI-based software qualifying as a medical device falls within the medical-device regulatory system. As per the latest guidance, confirm the current edition.
- General concept sources used for teaching: standard regulatory-science teaching material on risk-based regulation, conformity assessment, and post-market surveillance (no specific edition cited; verify current versions before professional use).

CHAPTER 7: Explainable AI (XAI) — Transparency, Interpretability, Accountability

1. Learning Objectives

By the end of this chapter, you will be able to:

- Define Explainable AI (XAI) and distinguish **black-box** from **interpretable** models, starting from zero.
- Explain why explainability matters in pharmacy practice — a pharmacist must be able to **justify** a recommendation.
- Describe, at concept level and without mathematical derivations, the explanation methods **feature importance**, **LIME**, and **SHAP**, using simple toy illustrations.
- Distinguish **global** from **local** explanations, and **transparency** from **interpretability** from **accountability** (a classic confusion).
- Discuss **accountability** in automated decision systems: who is responsible when the AI errs, and the role of **human-in-the-loop**.
- Work through a toy example explaining an ADR-risk prediction for one patient (labelled "ILLUSTRATIVE TOY DATASET").

2. Definitions

Every term is defined here **before** it is used later in the chapter.

- **Explainable AI (XAI)** — the set of methods and practices that make an AI system's decisions understandable to humans. If the AI says "high risk," XAI answers the follow-up question: "why?"
- **Black-box model** — an AI model whose internal reasoning cannot be easily understood by humans. You can see the input and the output, but not *why* one led to the other. Example: a large deep neural network.
- **Interpretable (white-box / glass-box) model** — a model whose reasoning a human can follow step by step. Example: a small decision tree or a simple scoring rule.
- **Explanation** — a human-understandable statement of why the AI produced a particular output, e.g., "the risk is high mainly because the patient takes 8 medicines and is 76 years old."
- **Feature** — one input variable the model uses, e.g., age, number of medicines, kidney function.
- **Feature importance** — a ranking of which features influenced the model's prediction the most.
- **LIME** (Local Interpretable Model-agnostic Explanations) — a method that explains **one single prediction** by building a tiny, simple model that mimics the big model *just around that case*. Think: "ask the big model many what-if questions near this patient, then fit a simple rule to its answers."
- **SHAP** (SHapley Additive exPlanations) — a method that **fairly divides credit** for a prediction among the features, like dividing a restaurant bill fairly according to what each person ordered. Each feature gets a contribution score; the scores add up to the final prediction.
- **Global explanation** — an explanation of how the model behaves **in general**, across all cases ("overall, number of medicines is the strongest factor").
- **Local explanation** — an explanation of **one specific prediction** for one specific case ("for *this* patient, age mattered most").
- **Transparency** — openness *about* the system: what it is, what data it was trained on, what it is intended for, and what its limits are.
- **Interpretability** — the ability to understand *why* the model made a particular decision.
- **Accountability** — clarity about **who answers** when the AI system causes harm or makes an error. Accountability always belongs to humans or organisations, never to the software.
- **Human-in-the-loop (HITL)** — a design where a human reviews and approves the AI's output before any action is taken. The AI proposes; the human disposes.
- **Human-on-the-loop** — a weaker form: the human supervises the system and can intervene, but does not approve every single decision.
- **ADR (Adverse Drug Reaction)** — a harmful, unintended response to a medicine taken at a normal dose.

3. Terminology Table

Term	Plain-English meaning	Quick example
XAI	Making AI decisions understandable	"Why did it flag high ADR risk?"
Black-box model	Input/output visible; reasoning hidden	Deep neural network
Interpretable model	Reasoning a human can follow	Small decision tree
Feature	One input variable	Patient's age
Feature importance	Which inputs mattered most	"No. of medicines ranked #1"
LIME	Simple local stand-in model for one prediction	What-if questions around one patient
SHAP	Fair credit-sharing among features	Each feature's contribution adds up to the answer
Global explanation	How the model behaves overall	"Age matters most in general"
Local explanation	Why this one prediction happened	"For Mrs. Sharma, polypharmacy mattered most"
Transparency	Openness about the system	Model card: data, purpose, limits
Interpretability	Understanding the why of a decision	Feature contributions for one case
Accountability	Who answers when it goes wrong	The supervising pharmacist / the deploying hospital
Human-in-the-loop	Human approves each AI output	Pharmacist confirms every AI dose suggestion
ADR	Harmful unintended drug effect	Bleeding from an anticoagulant

4. Explanation

4.1 What is Explainable AI? (Starting from zero)

Technical concept: Explainable AI (XAI) is the field concerned with producing human-understandable reasons for the outputs of AI systems, especially complex models whose internal logic is not directly readable.

Simple explanation: Imagine two students answering an exam question. Student A writes only the final answer. Student B shows every step. If the answer is wrong, you can find Student B's mistake — but with Student A, you are stuck. Most powerful AI models are like Student A: they give answers without showing their work. XAI is the technique of forcing the AI to "show its work."

Example: An AI flags a prescription as "high risk of adverse drug reaction." The pharmacist's next question is obvious: *why?* Is it the patient's age? The number of medicines? A specific interaction? Without an explanation, the pharmacist cannot judge whether to trust the flag.

Key takeaway: XAI = making the AI show its work, so humans can check it.

4.2 Black-box vs interpretable models

Technical concept: Models sit on a spectrum. Interpretable models (small decision trees, simple rules, linear scoring) expose their logic directly. Black-box models (deep neural networks, large ensembles) achieve high accuracy on complex data but their internal reasoning is not human-readable.

Simple explanation: A **glass box** lets you watch the gears turn; a **black box** hides them.

- *Interpretable example — a decision tree for ADR risk:*
- "Does the patient take more than 5 medicines?" → Yes → "Is the patient over 65?" → Yes → **High risk.** You can read this like a flowchart and argue with each step.
- *Black-box example — a deep neural network for the same task:* it takes 40 inputs, passes them through layers of mathematical transformations, and outputs "0.78 risk." Accurate, perhaps — but no human can trace *why*.

The trade-off: often (not always), the most accurate models are the least interpretable, and the most interpretable are less accurate. In pharmacy, accuracy without understanding is dangerous — so we either use interpretable models or add XAI tools on top of black boxes.

> **Key takeaway:** Glass box = you can follow the reasoning. Black box = you see only input and output. Pharmacy needs the reasoning.

4.3 Why explainability matters in pharmacy

Technical concept: Pharmacy is a high-stakes, professional-accountability domain: pharmacists must justify clinical recommendations to patients, doctors, and regulators, detect model errors, and satisfy ethical and legal duties.

Simple explanation: A pharmacist is not a button-presser. When you recommend a medicine change, you must be able to say *why* — to the patient ("why are you changing my dose?"), to the doctor ("why do you disagree with my prescription?"), and to yourself ("am I sure this is right?"). Five reasons explainability is non-negotiable in pharmacy:

1. **Justification.** "The computer said so" is not a professional answer. You need reasons you can defend.
2. **Error detection.** Explanations reveal mistakes. If the AI flags high ADR risk "because the patient is 25 years old," you immediately smell a bug.
3. **Trust.** Patients trust pharmacists who can explain. Nobody trusts a shrug.
4. **Learning.** Explanations teach: a junior pharmacist learns *which* factors drive risk by reading AI explanations.
5. **Legal and ethical duty.** If harm occurs, investigators will ask what the AI considered. "We don't know" is not acceptable (links to Chapter 6's documentation duties and accountability below).

Example: An AI suggests switching a patient's antibiotic. With an explanation — "suggested because of documented penicillin allergy + local resistance pattern + kidney function" — the pharmacist can agree confidently. Without it, the pharmacist is gambling.

> **Key takeaway:** In pharmacy, an unexplained AI recommendation is an unjustified one — and unjustified recommendations are unprofessional.

4.4 Concept-level XAI methods (no maths beyond toys)

Method 1 — Feature importance

Technical concept: Feature importance methods rank the input features by how much they influenced a prediction (or the model's behaviour overall).

Simple explanation: After a cricket match, analysts argue about what decided the game — the toss? the dropped catch? the last over? Feature importance does this for AI predictions: it produces a ranked list like "1. Number of medicines, 2. Age, 3. Kidney function."

Toy illustration (ILLUSTRATIVE TOY DATASET): An ADR-risk model looked at 1,000 toy patients and reported:

1 Number of medicines — most important

2 Age — second

3 Kidney function — third

4 History of allergy — fourth

This is a **global** explanation: it describes the model in general, not any single patient.

> **Key takeaway:** Feature importance = the ranked list of "what mattered."

Method 2 — LIME (explaining one prediction with a tiny local model)

Technical concept: LIME explains a single prediction by probing the complex model with many slightly varied versions of that case, then fitting a simple, interpretable model to those probe results in the neighbourhood of the case.

Simple explanation: Imagine a wise but silent expert. You cannot ask "how do you think?" — but you *can* ask hundreds of what-if questions: "What if this patient were 5 years younger? What if she took one medicine fewer?" By watching how the expert's answers change, you build a small, simple rule that mimics the expert **just for patients like this one**. That small rule is the explanation.

Toy illustration (ILLUSTRATIVE TOY DATASET): For one patient, the big model says "ADR risk: 78%." LIME asks 500 what-if questions around her profile and fits a tiny rule: "Each extra medicine above 5 adds ~6% risk; age above 65 adds ~10%." The pharmacist reads the tiny rule — not the giant network.

> **Key takeaway:** LIME = a simple stand-in that imitates the big model near one case, so you can read the stand-in.

Method 3 — SHAP (fair credit-sharing)

Technical concept: SHAP assigns each feature a contribution value such that the contributions fairly add up to the difference between the prediction and an average baseline, inspired by a fair-division idea from game theory.

Simple explanation: Four friends share a taxi and split the fare fairly by distance each travelled. SHAP does this for predictions: it asks, "how much of this 78% risk score did each feature *contribute*?" Each feature gets a fair share of credit (or blame), and all the shares add up to the final answer.

Toy illustration (ILLUSTRATIVE TOY DATASET): Baseline average ADR risk = 8%. This patient's predicted risk = 78%. SHAP divides the 70-point difference:

- Number of medicines (8 medicines): **+25 points**
- Age (76 years): **+18 points**

- Kidney function (low): **+15 points**
- History of allergy: **+12 points**
- Total: 8 + 25 + 18 + 15 + 12 = **78%** ✓ (adds up exactly — that is SHAP's signature property)

Key takeaway: SHAP = fair division of credit among features; the pieces sum to the prediction.

4.5 Global vs local explanations

Technical concept: Global explanations describe overall model behaviour; local explanations describe individual predictions. Both are needed: global for validation and policy, local for clinical decisions about a specific patient.

Simple explanation: A **global** explanation is the textbook chapter ("in general, polypharmacy drives ADR risk"). A **local** explanation is the case discussion ("for *this* patient, her kidney function was the deciding factor"). A doctor needs the textbook to trust the model and the case discussion to treat the patient.

Example: Global: "Across all patients, number of medicines is the #1 risk factor." Local: "For Mr. Verma specifically, his history of allergy contributed the most to his 78% score." Both can be true at once.

Key takeaway: Global = the model's habits. Local = this patient's story.

4.6 Transparency vs interpretability vs accountability — three different things

Technical concept: These three terms are often confused but mean different things: transparency is openness *about* the system; interpretability is understanding *why* a decision was made; accountability is clarity about *who answers* for outcomes.

Simple explanation — the restaurant analogy:

- **Transparency** = the restaurant displays its licence, menu prices, and kitchen hygiene certificate. You know *what the place is*.
- **Interpretability** = the waiter explains *why* your dish contains peanuts ("the recipe includes groundnut oil"). You understand *this decision*.
- **Accountability** = if the meal makes you sick, you know *who is responsible* — the restaurant owner, not the recipe book.

Pharmacy mapping:

- Transparency: the AI vendor publishes what the tool does, its training data description, intended use, and limits ("model card").
- Interpretability: for Mrs. Sharma's prescription, the tool shows which factors drove the alert.
- Accountability: the hospital and supervising pharmacist are answerable if the tool's error harms a patient — the software itself cannot be blamed.

Key takeaway: Transparency = open about the system. Interpretability = understand the decision. Accountability = know who answers. Three words, three jobs.

4.7 Accountability in automated decision systems — who is responsible when AI errs?

Technical concept: AI systems have no moral or legal personhood; responsibility must be allocated across the human chain: developers, validators, deploying organisations, and supervising professionals. Human-in-the-loop designs keep a qualified human as the final decision-maker.

Simple explanation: If a self-driving car crashes, we do not arrest the car. We investigate the manufacturer, the software team, the owner, and the driver. Pharmacy AI is the same: **the AI is a tool; humans are accountable.**

The accountability chain in pharmacy AI:

1. **Developers/vendors** — responsible for building and testing the model properly, documenting limits, and warning about misuse.
2. **Deploying organisation (hospital/pharmacy chain)** — responsible for choosing a suitable product, training staff, and setting safe workflows.
3. **Supervising pharmacist** — responsible for reviewing AI outputs before acting (human-in-the-loop), and for reporting errors.
4. **Regulator** — responsible for setting and enforcing the rules (Chapter 6).

Human-in-the-loop in practice: The AI suggests; the pharmacist reviews the suggestion *and its explanation*, then accepts, modifies, or rejects it. The final decision — and its consequences — belong to the human. A weaker setup, **human-on-the-loop** (supervising many AI decisions without approving each), is acceptable only for lower-risk tasks.

Example: An AI recommends doubling a digoxin dose. The pharmacist sees the explanation ("based on low blood levels") but notices the patient just started a interacting drug the AI was not told about. She rejects the suggestion. If she had clicked "accept" blindly and the patient was harmed, accountability would fall on her — the human in the loop.

> **Key takeaway:** AI proposes, human disposes — and the human answers for the outcome.

5. Classifications

5.1 Classification of XAI methods

Family	Scope	Question it answers	Example method
Global, model-level	Whole model	"What does the model rely on in general?"	Overall feature importance
Local, prediction-level	One prediction	"Why this answer for this patient?"	LIME, SHAP
Example-based	One prediction	"Which past cases were similar?"	Nearest-neighbour case display
Rule extraction	Whole model	"Can the model be summarised as rules?"	Decision-rule summary

5.2 Classification of human involvement (oversight levels)

Level	Who decides	Suitable for
Human-in-the-loop	Human approves every AI output	High-risk: dosing, diagnosis support
Human-on-the-loop	Human supervises; intervenes when needed	Medium-risk: triage queues, stock alerts
Human-out-of-the-loop (full automation)	AI acts alone	Only minimal-risk tasks; never for clinical decisions

5.3 The transparency–interpretability–accountability split

Concept	Answers	Provided by	Example artefact
Transparency	"What is this system?"	Vendor / deployer	Model card, intended-use statement
Interpretability	"Why this decision?"	XAI methods	SHAP contributions for one patient
Accountability	"Who answers?"	Law, contracts, professional duty	Named supervising pharmacist; incident reporting chain

6. Mechanisms

6.1 How LIME works (conceptual steps)

1. Take the patient case to be explained.
2. Create many slightly varied "what-if" versions of it (change age a little, drop one medicine, etc.).
3. Ask the big black-box model for its prediction on each version.
4. Fit a tiny, simple model (like a short rule or mini-equation) to those answers — only valid near this case.
5. Read the tiny model as the explanation.

6.2 How SHAP works (conceptual steps)

1. Start from a baseline: the average prediction over all patients.
2. Consider each feature's contribution by comparing predictions with and without that feature (in fair combinations).
3. Assign each feature its fair share of the difference from baseline.
4. The shares add up exactly to the final prediction — check the sum as a sanity test.

6.3 How accountability works (the mechanism, not just the slogan)

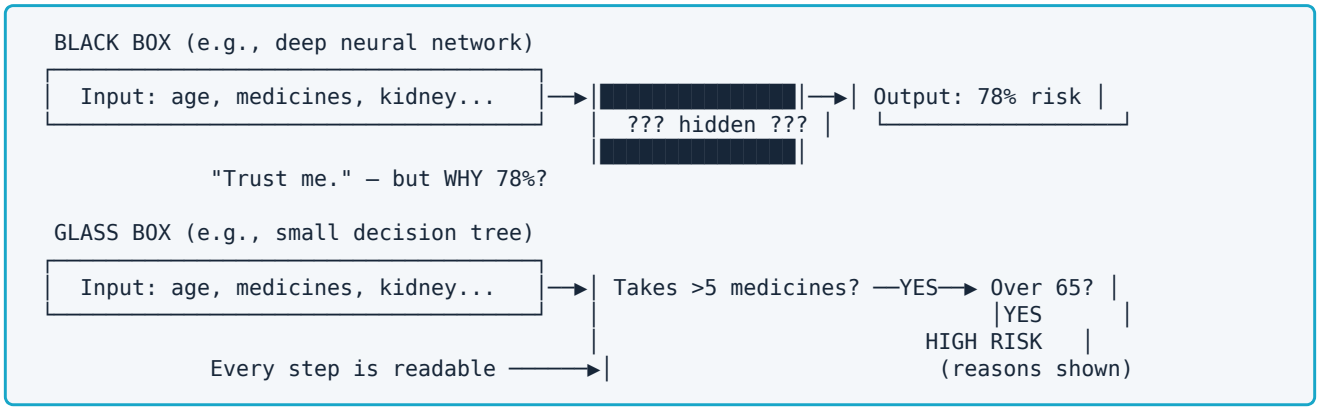
1. **Named responsibility:** every AI deployment names who supervises it (e.g., "chief pharmacist reviews all high-risk alerts").
2. **Logged decisions:** the AI's suggestion, its explanation, and the human's accept/reject decision are recorded.

3 **Error reporting:** wrong AI outputs are reported up the chain (links to Chapter 6 post-market monitoring).

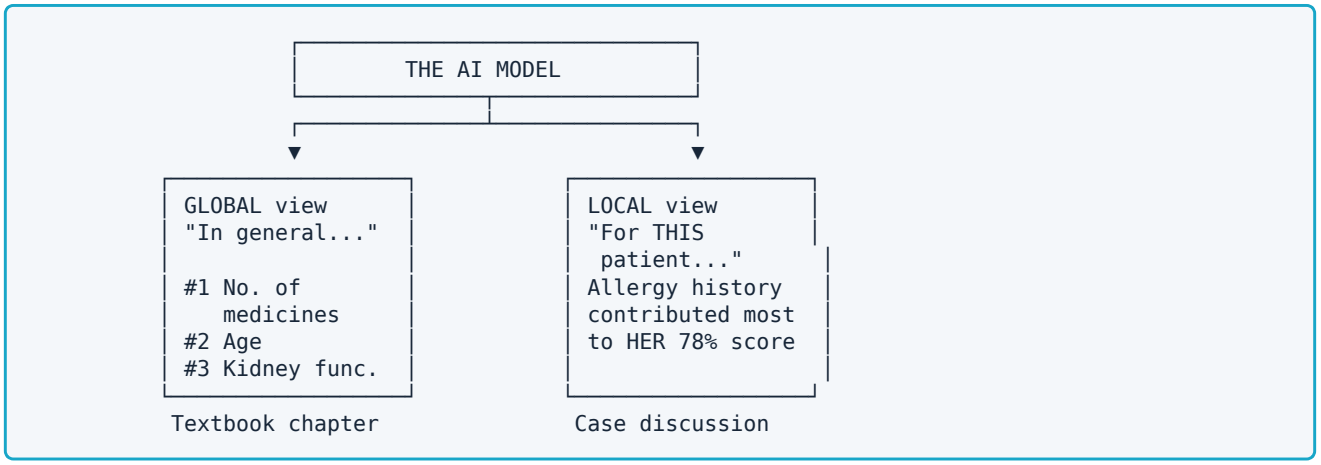
4 **Review and correction:** patterns of errors trigger retraining, workflow change, or product withdrawal.

7. Diagrams / Flowcharts

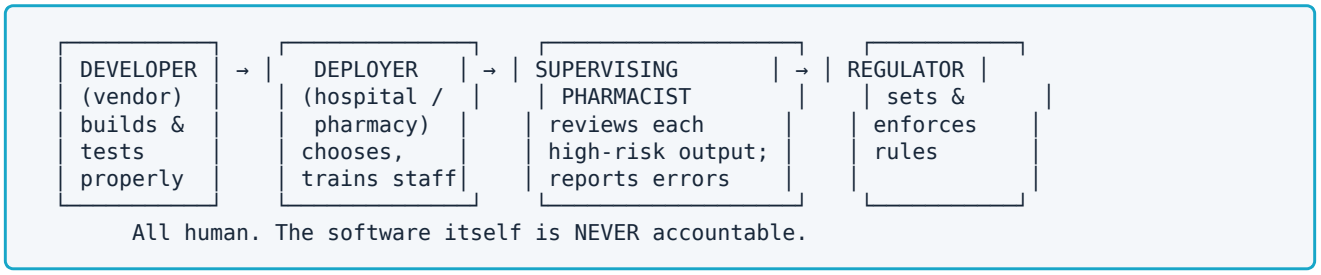
7.1 Black box vs glass box



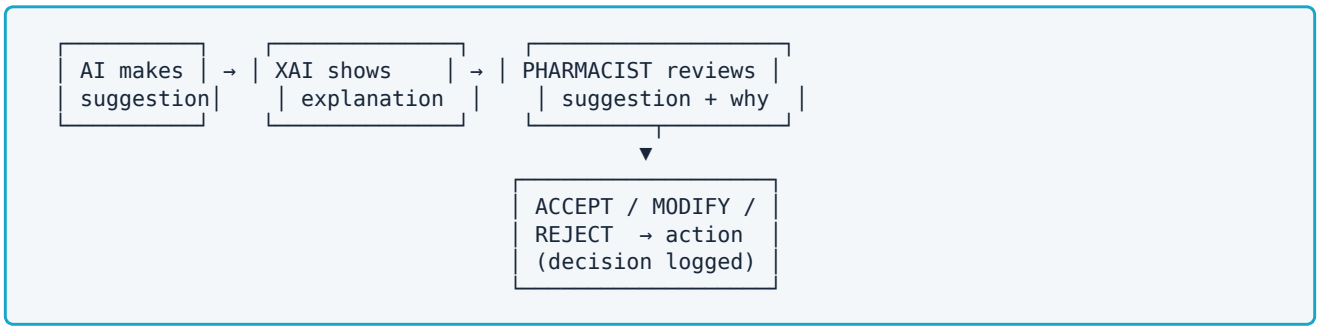
7.2 Global vs local explanation



7.3 The accountability chain



7.4 Human-in-the-loop workflow



8. Infographic Concepts

1 **"Show your work" classroom poster.** Left side: a student handing in only a final answer, stamped "UNCHECKABLE." Right side: a student showing step-by-step working, stamped "VERIFIABLE." Below: an AI doing the same two things. Caption: "In pharmacy, only verifiable answers are professional answers."

2 **"The three words" triangle.** A triangle with three corners labelled TRANSPARENCY ("open about the system"), INTERPRETABILITY ("understand the decision"), ACCOUNTABILITY ("know who answers"), each with its pharmacy icon (document, magnifying glass over a patient case, pharmacist's name badge). Centre text: "Three words. Three different jobs. Don't mix them up."

3 **"Fair shares" pie visual for SHAP.** A pie/plate divided into slices labelled with features (medicines +25, age +18, kidney +15, allergy +12) sitting on a baseline plate of 8%, totalling a 78% "risk plate." Caption: "Every feature pays its fair share — and the shares add up." (Marked: illustrative toy dataset.)

9. Tables

Table 7.1 — Black-box vs interpretable models

Aspect	Black-box (e.g., deep network)	Interpretable (e.g., small decision tree)
Can a human follow its reasoning?	No	Yes, step by step
Typical accuracy on complex data	Often higher	Often lower
Needs XAI add-ons?	Yes, to be usable in pharmacy	Less so — it explains itself
Error detection	Hard — must trust output	Easier — inspect the steps
Pharmacy suitability	Only with explanations + oversight	Preferred where stakes are high

Table 7.2 — XAI methods at a glance (concept level)

Method	Scope	Intuitive idea	Best for
Feature importance	Global (or local)	Ranked list of what mattered	Validating the model overall
LIME	Local	Tiny stand-in model near one case	Explaining one patient's prediction simply
SHAP	Local (can aggregate to global)	Fair credit-sharing; shares sum to the answer	Detailed, checkable per-patient explanations

Table 7.3 — Transparency vs interpretability vs accountability

	Transparency	Interpretability	Accountability
Question	What is this system?	Why this decision?	Who answers for it?
Focus	The system	The decision	The humans
Example	Model card, intended use	SHAP contributions	Named supervising pharmacist

10. Worked Toy Examples

Worked example 1 — Explaining one patient's ADR-risk prediction (ILLUSTRATIVE TOY DATASET)

Setup: A toy ADR-risk model (not a real product) predicts adverse-drug-reaction risk for a patient.

Patient (toy data): Mrs. Rao, age 76, takes 8 medicines, low kidney function, history of drug allergy.

Model output: 78% ADR risk.

Step 1 — Baseline. The average risk across all toy patients is 8%. So we must explain the extra 70 points.

Step 2 — SHAP-style fair shares (toy contributions):

Feature	Patient value	Contribution to risk
Baseline (average patient)	—	8%
Number of medicines	8	+25%
Age	76 years	+18%
Kidney function	Low	+15%
History of allergy	Yes	+12%
Total		**8 + 25 + 18 + 15 + 12 = 78%** ✓

Step 3 — Read the explanation like a pharmacist. The biggest driver is polypharmacy (8 medicines, +25), followed by age (+18). Kidney function and allergy history add the rest. The explanation checks out clinically — no surprise factor like "risk high because patient is female" that would signal bias.

Step 4 — Act. The pharmacist uses the explanation to prioritise: medication review to reduce unnecessary medicines first (biggest slice), then renal dose adjustment.

> **Key takeaway:** The explanation turned a scary number (78%) into an action plan (review medicines → check renal dosing).

Worked example 2 — LIME-style local rule (ILLUSTRATIVE TOY DATASET)

For the same patient, LIME's tiny stand-in rule reads: *"Each medicine above 5 adds about 6% risk; age above 65 adds about 10%; low kidney function adds about 15%."* Check: (8−5)×6 = 18, +10 +15 = 43… plus baseline and allergy terms ≈ 78%. The stand-in is approximate — that is fine; its job is readability, not perfect maths.

Worked example 3 — Spotting a biased model with explanations (ILLUSTRATIVE TOY DATASET)

A different toy model predicts "patient will miss doses." Its feature importance ranking is: 1. Pin code, 2. Number of medicines, 3. Age. The pin code ranking #1 is a red flag — the model may be discriminating by neighbourhood (a proxy for income) rather than clinical factors. **The explanation exposed a problem the accuracy score hid.** This is why XAI matters for fairness.

11. Analogies

1. **Exam with steps vs final answer only.** Student B (steps shown) vs Student A (answer only). XAI forces AI to be Student B.
2. **Glass box vs black box.** Watch the gears turn, or take the output on faith. Pharmacy cannot take outputs on faith.
3. **Restaurant bill split fairly (SHAP).** Each friend pays for what they ordered; each feature "pays" its fair share of the prediction.
4. **Asking the silent expert what-if questions (LIME).** You cannot read the expert's mind, but 500 what-if questions reveal how she thinks near this case.
5. **Textbook chapter vs case discussion (global vs local).** The textbook tells you the disease in general; the case discussion tells you about *this* patient.
6. **Recipe book vs restaurant owner (accountability).** If the food poisons you, you hold the owner responsible — never the recipe book. The AI is the recipe book.

12. Pharmacy Relevance

1. **Justifying recommendations.** Every clinical recommendation a pharmacist makes — dose change, drug switch, counselling point — needs reasons. XAI gives you the AI's reasons so you can adopt, adapt, or reject them professionally.
2. **Catching AI errors before patients do.** Explanations are your early-warning system. A nonsense explanation ("high risk because the patient is 25") tells you the model is broken *before* anyone is harmed.
3. **Patient counselling.** "The computer flagged your prescription because you take 8 medicines and your kidneys are weak, so I want to review them with you" — that is counselling powered by explanation. Patients accept interventions they understand.
4. **Interprofessional communication.** When you phone a doctor to question a prescription, "the AI alert fired" is weak; "the AI alert fired because of X and Y, and here is my assessment" is persuasive.
5. **Medico-legal protection.** If an AI-assisted decision is ever questioned, the logged suggestion + explanation + your documented review is your defence. Unexplained acceptance is indefensible.
6. **Teaching juniors.** Explanations from good models are teaching material — they show trainees which factors matter in real decisions.

13. India Context

- **High patient loads, short counselling time.** Indian pharmacists often serve hundreds of patients daily. XAI that summarises *why* in one line ("main driver: 8 medicines") fits real workflows; a 40-feature SHAP table does not. Demand explanations designed for busy practice.

- **Language and literacy.** Explanations must eventually reach patients in local languages and plain words. "Your kidney is weak, so this dose is risky" counsels better than "renal impairment elevates ADR probability."
- **Low-resource settings.** In rural pharmacies with limited diagnostics, simple interpretable models (scorecards, decision trees) may beat fancy black boxes — they run on paper or basic phones and every step is checkable.
- **Jan Aushadhi and generic substitution.** If an AI recommends switching a brand to a generic, the pharmacist must explain *why* to a sceptical patient ("same salt, same strength, lower price — quality assured"). Explainability supports trust in generics.
- **Regulatory direction.** Indian regulators increasingly expect documented, reviewable AI behaviour for medical devices (see Chapter 6) — XAI is how you produce that evidence. As always: confirm the current guidance edition.

14. Key Points

- XAI = making AI "show its work" so humans can check it.
- **Black box:** input/output visible, reasoning hidden. **Glass box:** reasoning readable.
- In pharmacy, unexplained recommendations are unprofessional — you must justify your actions.
- **Feature importance** = ranked list of what mattered (often global).
- **LIME** = tiny stand-in model mimicking the big model near one case (local).
- **SHAP** = fair credit-sharing among features; shares sum to the prediction (local).
- **Global** = the model's habits; **local** = this patient's story. You need both.
- **Transparency** (open about the system) ≠ **interpretability** (understand the decision) ≠ **accountability** (know who answers).
- **AI proposes, human disposes:** the human-in-the-loop decides and is accountable. Software is never accountable.
- Explanations catch errors and expose bias that accuracy scores hide.

15. Common Confusions

Confusion	Clarification
"Transparency and interpretability are the same."	Transparency = openness *about* the system (model card). Interpretability = understanding *why* a decision was made. Different jobs.
"If the model is accurate, we don't need explanations."	Accuracy hides bias and nonsense reasoning (see the pin-code example). Explanations check the *why*, not just the *what*.
"LIME and SHAP are the same thing."	Both give local explanations, but LIME builds a tiny stand-in model nearby, while SHAP fairly divides credit so shares sum to the prediction.
"Global explanation is enough for patient care."	Global tells you the model's habits; treating *this* patient needs a local explanation.

Confusion	Clarification
"The AI is accountable for its errors."	Never. Accountability sits with developers, deployers, and supervising professionals — humans and organisations.
"Human-on-the-loop is the same as human-in-the-loop."	In-the-loop = human approves *each* output (high-risk). On-the-loop = human supervises and intervenes (lower risk).

16. Common Mistakes

1. **Accepting an AI recommendation without reading its explanation** — the classic automation-bias error.
2. **Mixing up the three words** — writing "transparency" when you mean "interpretability" (or accountability) in exams.
3. **Trusting pretty explanations blindly** — explanations can also be misleading; cross-check against clinical sense.
4. **Using global importance to justify a single patient's decision** — "medicines matter most in general" does not prove they mattered most *here*.
5. **Thinking interpretable = always less accurate** — sometimes simple models perform just as well; the trade-off is "often," not "always."
6. **Forgetting to log** the AI suggestion + explanation + your decision — without the log, there is no accountability trail.

17. Mnemonics

- **The three words: "T–I–A" → "Tell, Interpret, Answer."** Transparency = *Tell* about the system; Interpretability = *Interpret* the decision; Accountability = *Answer* for the outcome.
- **LIME vs SHAP: "LIME is Local-Imitation; SHAP is SHares-Add-up-Precisely."**
- **Global vs local: "Textbook vs Case."** Global = textbook chapter; local = case discussion.
- **Accountability: "AI proposes, human disposes."**
- **Oversight levels: "In > On > Out."** In-the-loop (approves each) is stronger than on-the-loop (supervises), which is stronger than out-of-the-loop (never for clinical decisions).

18. Exam Summary

Explainable AI (XAI) makes AI decisions understandable — forcing the model to "show its work." **Black-box** models (e.g., deep networks) hide their reasoning; **interpretable/glass-box** models (e.g., small decision trees) expose it. Explainability is essential in pharmacy because pharmacists must **justify** recommendations, **detect errors**, build **trust**, and meet **legal duties** — "the computer said so" is never acceptable. Three concept-level methods: **feature importance** (ranked list of what mattered, often global), **LIME** (a tiny stand-in model fitted to what-if probes near one case — a local explanation), and **SHAP** (fair credit-sharing where each feature's contribution sums exactly to the prediction — local). **Global** explanations describe the model's overall habits; **local** explanations explain one patient's prediction. **Transparency** (openness about the system), **interpretability** (understanding a decision), and

accountability (knowing who answers) are three different things. Accountability always belongs to **humans** — developers, deploying organisations, supervising pharmacists — never to software; **human-in-the-loop** design keeps the pharmacist as the final decision-maker, with every AI suggestion, its explanation, and the human's decision logged.

19. One-Minute Revision

- XAI = AI shows its work.
- Black box: reasoning hidden. Glass box: reasoning readable.
- Pharmacy needs the *why* — to justify, to catch errors, to counsel.
- Feature importance = ranked "what mattered" list.
- LIME = tiny local stand-in near one case.
- SHAP = fair shares that sum to the prediction.
- Global = textbook; local = this patient's story.
- **T–I–A: Tell (transparency), Interpret (interpretability), Answer (accountability).**
- AI proposes, human disposes — software is never accountable.
- Log: suggestion + explanation + your decision.

20. MCQs

Q1. Explainable AI (XAI) is best described as: a) AI that runs faster than other AI b) Methods that make an AI system's decisions understandable to humans c) AI that needs no data d) AI that replaces pharmacists **Answer: b**

Q2. A "black-box" model is one where: a) The computer case is painted black b) Inputs and outputs are visible but the internal reasoning is not human-readable c) It only works at night d) It has no inputs **Answer: b**

Q3. Which model is inherently interpretable? a) A large deep neural network b) A small decision tree whose steps can be read like a flowchart c) Any model with more than a million parameters d) A model with no training data **Answer: b**

Q4. Feature importance provides: a) The price of the software b) A ranked list of which input features influenced the prediction most c) The programmer's name d) The computer's serial number **Answer: b**

Q5. LIME explains a prediction by: a) Building a tiny, simple model that mimics the big model just around that one case b) Deleting the training data c) Asking the patient directly d) Repeating the prediction louder **Answer: a**

Q6. The signature property of SHAP is that: a) It works without any data b) Each feature's fair contribution adds up exactly to the final prediction c) It only works on images d) It replaces the pharmacist **Answer: b**

Q7. "For Mrs. Rao specifically, allergy history contributed most to her 78% score" is a: a) Global explanation b) Local explanation c) Transparency statement d) Legal contract **Answer: b**

Q8. Which matches correctly? a) Transparency = understanding why a decision was made b) Interpretability = openness about the system; Accountability = understanding the decision c) Transparency

= openness about the system; Interpretability = understanding the decision; Accountability = knowing who answers d) All three mean the same thing **Answer: c**

Q9. When an AI-assisted decision harms a patient, accountability lies with: a) The software itself b) The humans/organisations — developers, deployers, supervising professionals c) Nobody d) The patient **Answer: b**

Q10. "Human-in-the-loop" means: a) The human watches the AI from far away b) A human reviews and approves the AI's output before any action is taken c) The AI works without any human d) The human is replaced by a loop **Answer: b**

Answer key: 1-b, 2-b, 3-b, 4-b, 5-a, 6-b, 7-b, 8-c, 9-b, 10-b

21. Viva / Application Questions

1 Define XAI. Why is "the computer said so" an unacceptable justification in pharmacy?

2 Distinguish black-box and interpretable models with one example of each.

3 Explain LIME intuitively, as if to a classmate who fears mathematics.

4 Explain SHAP's "fair shares" idea using the restaurant-bill analogy, with the toy ADR numbers.

5 Distinguish global from local explanations. When would you need each?

6 Distinguish transparency, interpretability, and accountability — give one pharmacy example of each.

7 Who is accountable when a pharmacy AI makes an error? Trace the accountability chain.

8 An AI flags a prescription "high risk" but its explanation cites a clinically irrelevant feature. What do you do, and why?

22. References

- Explainable AI (XAI): general teaching literature on making AI decisions human-understandable (black-box vs interpretable models; no specific edition cited).
- LIME — Local Interpretable Model-agnostic Explanations: concept of explaining single predictions with simple local stand-in models (described at concept level; consult original teaching sources for detail).
- SHAP — SHapley Additive exPlanations: concept of fair, additive feature contributions (described at concept level; consult original teaching sources for detail).
- Human-in-the-loop / human-on-the-loop oversight: standard AI-governance teaching concepts for keeping qualified humans as final decision-makers.
- Links to this course: Chapter 6 (regulatory frameworks — documentation and oversight duties); Chapters 8–9 (pharmacy automation and supply-chain AI where explanations support safe use).