

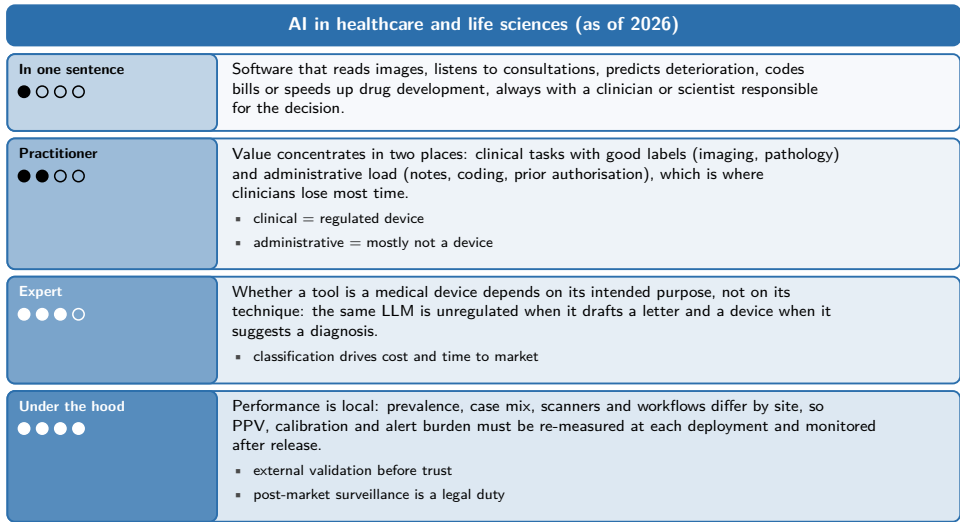


AI in Healthcare and Life Sciences

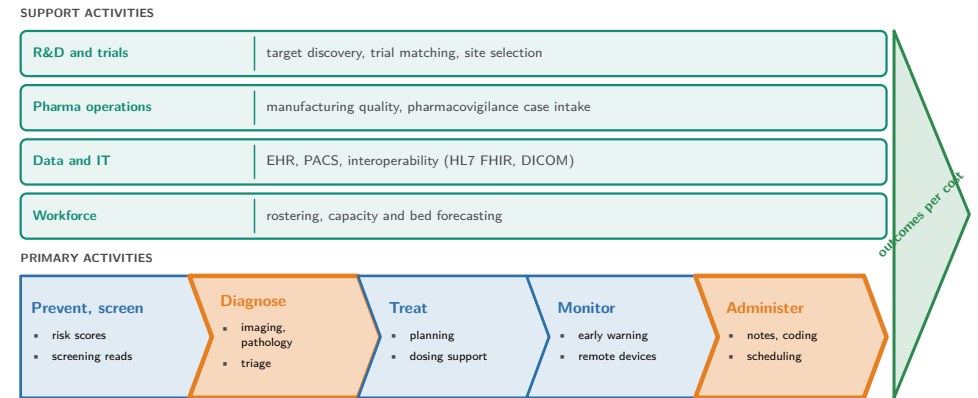
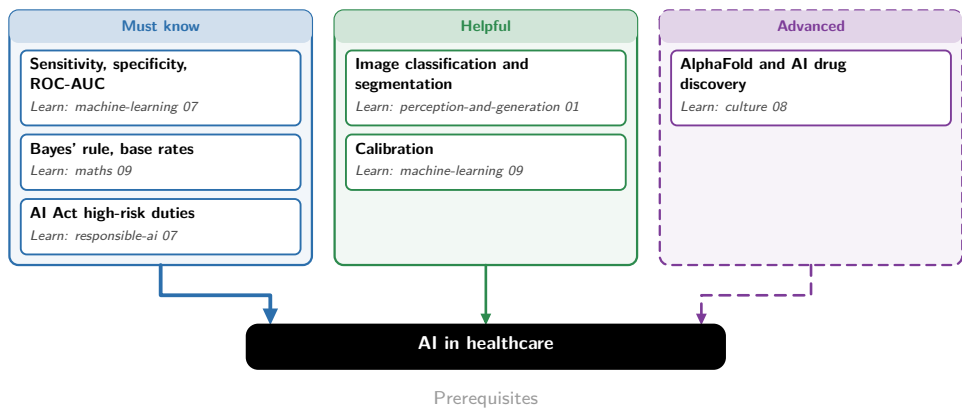
AI by industry | Care and pharma value chains, use cases with their data, models, clinical KPIs, medical-device rules (EU MDR Rule 11, FDA pathways, PCCP) and the AI Act, build or buy, failure stories, PPV at local prevalence; as of 2026

#Healthcare #MedicalDevice #MDR #FDA #LifeSciences

Where AI fits in care and pharma



AI in healthcare at four depths



The care value chain (primary) over its support functions. Highlighted: where AI is most deployed in 2026, imaging-led diagnosis and administrative load

Top use cases and their data

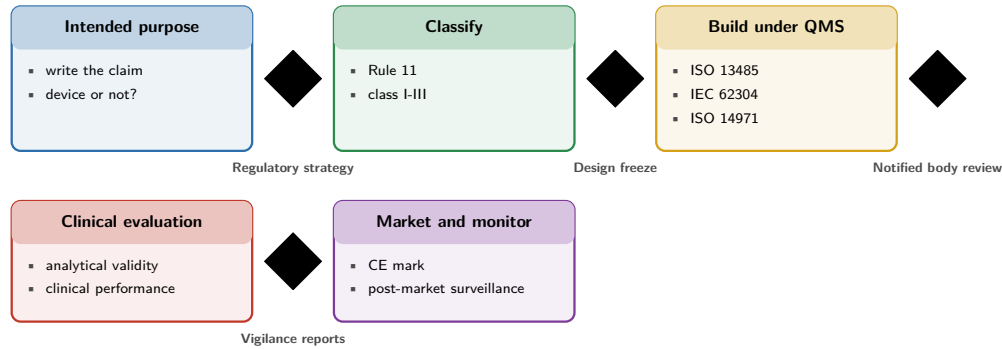
Use case	Data needed	Model family	KPI	Device?
Imaging detection and worklist triage	DICOM images with labels confirmed by follow-up or pathology	CNN, ViT, detection and segmentation	sensitivity at fixed specificity, time to report	yes
Digital pathology	whole-slide images, slide-level diagnoses	multiple-instance learning, pathology foundation models	agreement with panel, review time	yes
Deterioration and sepsis early warning	EHR vitals, labs, medications, time stamps	gradient boosting, survival, time-series	PPV, lead time, alerts per patient-day	often
Ambient clinical documentation	consultation audio, note templates, EHR context	speech recognition + LLM	note time, after-hours EHR time, edit rate	usually no
Clinical coding and billing	notes, discharge summaries, code sets (ICD-10)	LLM or text classifiers with code retrieval	coder throughput, denial rate	no
Capacity, bed and no-show forecasting	admissions, appointments, calendars	gradient boosting, time-series	MAPE, empty slots filled	no
Trial matching and site selection	eligibility criteria text, EHR cohorts	LLM extraction + rules	screen-to-enrol ratio, recruitment time	no
Pharmacovigilance case intake	adverse-event reports, literature, call notes	NER, LLM extraction	cases per FTE, missed serious cases	no (GxP)

Use case → data it needs → the KPI that proves it. "Device?" = likely a medical device under its usual intended purpose

⚠ Labels decide the ceiling

A label from one reader copies that reader's errors. Use reference standards confirmed by pathology, follow-up or expert panels, and record disease prevalence and case mix in the test set: they change PPV and the alert burden at every new site.

● Regulation and certification



EU route to market for AI medical software (MDR 2017/745; IVDR 2017/746 for in-vitro diagnostics). The notified-body gate applies from class IIa upwards

Rule	What it says	Consequence
EU MDR Rule 11	decision support for diagnosis or therapy: IIa; IIb if a wrong decision can cause serious deterioration or surgery; III if death or irreversible deterioration; monitoring of physiological processes IIa (IIb for vital parameters in immediate danger)	almost every clinical AI tool needs a notified body
US FDA	510(k) (predicate), De Novo (new low-moderate risk type), PMA (class III); final PCCP guidance for AI device software (Dec. 2024)	most AI devices so far cleared by 510(k)
IMDRF SaMD framework	risk category from the state of the condition (critical, serious, non-serious) × the role of the output (treat or diagnose, drive, inform management)	shared vocabulary across regulators
Data law	GDPR Art. 9 (health data, EU); HIPAA (US: de-identify by Safe Harbor's 18 identifiers or expert determination); European Health Data Space Regulation (EU) 2025/327 for secondary use, phased in from the late 2020s	data access often takes longer than modelling
Medicines lifecycle	EMA reflection paper on AI in the medicinal-product lifecycle (2024); FDA draft guidance on AI supporting drug regulatory decisions (Jan. 2025); EU GMP Annex 22 on AI (draft 2025)	GxP validation for AI used in trials and manufacturing

Classification and pathways for AI software (sector rules first)

📄 Where the AI Act adds duties

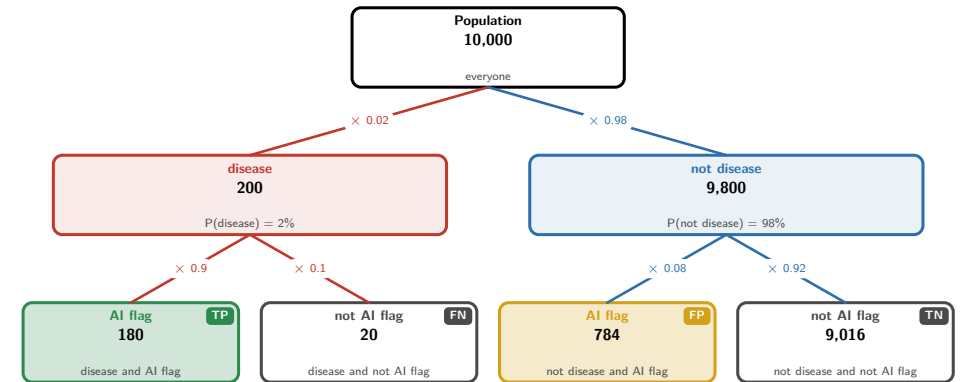
AI that is a medical device needing a notified body is high-risk through Annex I; those duties apply from 2028-08-02 (was 2027-08-02), assessed together with the MDR. Annex III adds emergency triage of patients and AI for access to public healthcare services (from 2027-12-02 (was 2026-08-02)). Mechanics of the Act: responsible-ai 07.

Evidence pack an AI device buyer should ask for

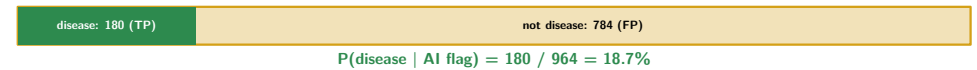
✓ baseline / mandatory □ recommended – verify in your context

- ✓ Intended purpose, class and certificate (CE number or FDA decision)
- ✓ External validation on a population like yours, with prevalence stated
- ✓ Performance by subgroup (sex, age, ethnicity, device vendor)
- **MUST** PCCP or change policy: what happens when the model is retrained
- **SHOULD** Post-market monitoring data you will receive
- **COULD** Reporting guideline used for studies (TRIPOD+AI, CONSORT-AI, DECIDE-AI)

📊 KPIs and a worked example



Evidence "AI flag": TP + FP = 180 + 784 = 964



Natural frequencies: $P(H|E) = TP / (TP + FP) = 180 / (180 + 784) = 0.187$ (18.7%)

Bayes' rule: $P(H|E) = P(E|H) \cdot P(H) / [P(E|H) \cdot P(H) + P(E|\neg H) \cdot P(\neg H)] = 0.9 \cdot 0.02 / (0.9 \cdot 0.02 + 0.08 \cdot 0.98) = 0.187$

Prior 2% → posterior 18.7%: the evidence multiplies the odds by the likelihood ratio $LR = P(E|H) / P(E|\neg H) = 0.9 / 0.08 = 11.2$

Without the evidence: $P(H|\neg E) = FN / (FN + TN) = 20 / (20 + 9,016) = 0.22\%$

Where: H: disease (the hypothesis); E: AI flag (the evidence); $\neg$: not; P(H) = prior, the base rate before seeing E; P(E|H) = sensitivity (true-positive rate); P(E| $\neg$ H) = false-positive rate; P(H|E) = posterior; TP, FN, FP, TN: counts of true / false positives / negatives among N = 10,000 people (or items)

Why PPV, not sensitivity, decides the clinician's workload. Illustrative assumptions: prevalence 2 % in the screened population, sensitivity 90 %, false-positive rate 8 %

KPI	Formula	Read it as
PPV	$\frac{TP}{TP + FP}$	share of flags that are real; 180 / 964 ≈ 19 % above
NNE	$\frac{1}{\frac{PPV}{A}}$	≈ 5.4 work-ups per true case above
Alert burden	$\frac{A}{P \times D}$	alerts per patient-day; set a ceiling with clinicians

Lead time	$t_{event} - t_{alert}$	hours gained; must exceed the time needed to act
-----------	-------------------------	--

where TP, FP = true and false positive flags (patients); A = alerts fired in the period (alerts); P = patients monitored (patients); D = days in the period (days); t_{event} , t_{alert} = time of the clinical event and of the first alert (hours)

Clinical KPIs every deployment should report per site

📄

Worked ROI: ambient scribe for 20 clinicians (every number is an assumption to replace)

Assume C = 20 clinicians, documentation h = 2 h per day, share saved f = 20 % (measure it in a pilot), d = 220 working days, loaded cost w = 100 EUR per hour, licence L = 3,000 EUR per clinician per year. Time value B = C × h × f × d × w = 20 × 2 × 0.2 × 220 × 100 = 176,000 EUR per year; cost = 20 × 3,000 = 60,000 EUR; ROI = (176,000 - 60,000) / 60,000 ≈ 1.9. Caveat: the saving becomes cash only if it buys extra appointments or cuts paid over-time; otherwise report it as after-hours time returned, a burnout KPI.

⚠️

Build or buy, and failure stories

	Recommended			
	Buy cleared device	EHR-vendor module	Build on own data	Academic partnership
Regulatory burden on you	vendor's	vendor's	you are the manufacturer	shared
Fit to local data	validate locally	 partial	best	good
Time to value	months	fast	1-3 years	slow
In-house skills needed	low	low	high + QMS	medium
Control of updates	vendor PCCP	vendor	full	contract

Four routes for a hospital or pharma team. Recommended default for clinical use: buy a cleared device and validate it locally

	Year	Case	Lesson
Clinical advice	2017-2018	IBM Watson for Oncology: an MD Anderson audit (2017) found the related project had cost about USD 62 M without patient use; internal documents reported by STAT (2018) cited unsafe treatment advice trained on synthetic cases	train on real outcomes; independent clinical validation
Imaging	2018	Zech et al. (PLOS Medicine): pneumonia CNNs partly learned which hospital an X-ray came from; performance fell on external sites	test on another hospital before trust
Population health	2019	Obermeyer et al. (Science): a care-management algorithm used past cost as a proxy for need, so Black patients were sicker at the same score; fixing the label raised their share flagged from 17.7 % to 46.5 %	the target variable carries the bias
Imaging	2020	Beede et al. (CHI): a diabetic-retinopathy screening deployment in Thai clinics met poor lighting and slow uploads; many images were rejected as ungradable	accuracy in the lab is not usability in the clinic
Payers	2023-	US class actions allege insurers used post-acute care prediction tools (e.g. naviHealth nH Predict) to deny coverage; allegations, under litigation as of 2026	a prediction must not replace individual medical review

Documented failures, each with its lesson (cross-sector cases such as the Epic sepsis model are in culture 09)

🚩

Myths and recap

MYTH

✖️

A CE-marked tool works at our hospital

FACT

✔️

CE marking shows conformity, not local performance

Prevalence, scanners and workflow change PPV; validate on your own retrospective data before go-live.

✖️

High sensitivity means few wasted work-ups

✔️

Prevalence decides the work-ups

At 2 % prevalence a 90 % sensitive model with 8 % false positives gives about 4 false flags per true one.

✖️

An LLM that drafts notes is a medical device

✔️

Only if its intended purpose is medical

Drafting documentation for a clinician to sign is usually not a device; suggesting diagnoses or treatment is.

✖️

Retraining an approved model needs no approval

✔️

Changes need a PCCP or a new submission

FDA PCCPs and MDR significant-change rules govern updates.

Four beliefs that cost hospitals money

Tick a box only when you could explain the point to someone else



Intended purpose decides device status
then Rule 11 decides the class



Most clinical AI needs a notified body
IIa or above; AI Act Annex I on top



Report PPV and NNE per site
not only sensitivity and AUC



Admin AI has the fastest ROI
scribes, coding, scheduling; not devices



Check the label and the proxy
cost $\neq$ need (Obermeyer 2019)



Validate externally
shortcut learning hides in single-site data

Q1 Under EU MDR Rule 11, what is the minimum class for software informing diagnosis or therapy decisions?

A Class IIa; IIb for possible serious deterioration or surgery, III for possible death or irreversible deterioration.

Q2 What does a PCCP allow?

A Pre-authorized model changes (e.g. retraining) with a stated validation protocol, without a new FDA submission for each change.

Q3 PPV with prevalence 2 %, sensitivity 0.9, false-positive rate 0.08?

A $0.018 / (0.018 + 0.0784) \approx 0.19$, so NNE ≈ 5.4 .

Q4 Which three standards underpin medical AI software development?

A ISO 13485 (QMS), IEC 62304 (software life cycle), ISO 14971 (risk management).

Q5 What did Obermeyer et al. (2019) find?

A Using cost as a proxy for health need under-flagged Black patients; changing the label corrected much of it.

Q6 When does an AI medical device become high-risk under the AI Act?

A When it needs third-party conformity assessment under the MDR or IVDR (Annex I); from 2028-08-02 (was 2027-08-02).

Q7 Why can an early-warning model with good AUC fail on the ward?

A Low PPV and high alert burden cause alert fatigue, and lead time may be too short to act.

Q8 What is GMLP?

A Ten Good Machine Learning Practice principles from the FDA, Health Canada and MHRA (2021).

Q9 Which HIPAA routes de-identify data?

A Safe Harbor (remove 18 identifier types) or expert determination.