

Velkommen til workshop om validering og testing av KI- løsninger for klinisk bruk i radiologi

Tirsdag 29. april 2025



Velkommen til workshop om validering og testing av KI- løsninger for klinisk bruk i radiologi

Tirsdag 29. april 2025

Hilde Myhren
Ulrich Spreng



Mål

Økt bruk

Trygg bruk

Bedre eller lik
kvalitet

Effektivt og
frigjøre tid

Felles KI-plan for trygg og effektiv bruk av KI i helse og omsorgstjenesten 2024- 2025



Innsatsområder

1. Sektorsamarbeid om KI
2. Tverretatlig informasjonsside
3. Tverretatlig regulatorisk veiledning
4. Rammer for bruk av KI i helse- og omsorgstjenesten
5. Bruk av store språkmodeller
6. Kompetanse om KI

Kunstig intelligens

Disse temasideene samler ressurser som skal hjelpe og veilede helse- og omsorgstjenesten og forsknings- og innovasjonsmiljøer i offentlig og privat sektor, slik at de kan lykkes med å utvikle og ta i bruk kunstig intelligens på en trygg måte.



Rammer for kvalitetssikring

"Rapport om kvalitetssikring: Bruk av kunstig intelligens i helse- og omsorgstjenesten" er utarbeidet for å gi støtte til virksomheter som vurderer å anskaffe et KI-system i helse- og omsorgstjenesten. Rapporten omhandler særlig KI-relaterte aspekter som virksomheter må være oppmerksom på for å kunne vurdere om et KI-system er tillitsverdig og kan brukes trygt og effektivt.



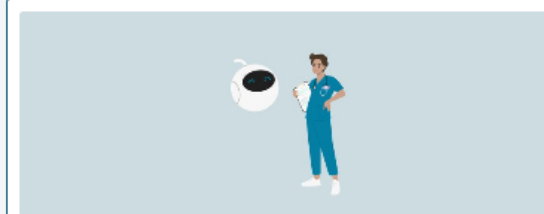
Felles KI-plan for helse og omsorgstjenesten

Den felles KI-planen for helse- og omsorgstjenesten skal bidra til å øke bruken av KI-løsninger som er trygge og bærekraftige.



En-til-en-veiledning i bruk av kunstig intelligens

Du får hjelp av jurister fra Helsedirektoratet, Direktoratet for medisinske produkter og Helsetilsynet, og andre eksperter ved behov.



Store språkmodellar i helse- og omsorgstjenesten

Rapporten gjør eit statusbilde av modenskapen til språkmodellar i helse- og omsorgssektoren. Den skal også peike framover på moglege bruksområde og risikoar, i tillegg til overordna vegval og tiltak.



Arrangementer

Rundskriv, veiledningsmateriell og faktaark om KI

Slik jobber vi

Rapporter



Program



- **10.00 – 10.10 Velkommen og ramme for dagen**
Hilde Myhren (Helsedirektoratet) og
Ulrich Spreng (Helse Sør-Øst RHF)
- **10.10 – 11.30 Presentasjoner fra ulike land (English)**
 - Nederland: *Strategies for validation of AI in clinical settings*, Stefan Romeijn
 - Danmark: *Validation of AI in radiology in the North Jutland Region*, Søren Pihlkjær Hjortshøj, Chief Medical Officer (CMO) Aalborg University hospital:
- **Diskusjon**
- **11.30 – 12.00 Lunsj**
- **12.00 – 14.00 Validering Norge rundt**
 - **Vestre Viken:** Jon Haakon Malmer-Høvik, avdelingssjef bildediagnostikk
 - **Universitetssykehuset i Nord-Norge:** Karl Øyvind Mikaelson, leder SPKI og/eller Camilla Andersen, KI-radiograf
 - **Haraldsplass diakonale sykehus:** Anagha P. Parkar, fagansvarlig radiolog
 - Helse Møre og Romsdal HF: Endre Grøvik, medisinsk fysiker og leder for forskningsgruppen «AI in Medical Imaging» og/ eller Ida Jørgensen, rådgiver Klinikk for diagnostikk
- **Diskusjon og refleksjon** rundt hvordan vi best mulig kan dele erfaringer og eventuelt samarbeide fremover.



Health Directorate

AI implementation and validation

Stephan Romeijn

Romion Health | Health AI Register

Who am I?

Some background.

- Technical Physician, @University of Twente
- Innovation manager, clinical implementation AI, @Leiden University Medical Center



Who am I?

Some background.

- Technical Physician, @University of Twente
- Innovation manager, clinical implementation AI, @Leiden University Medical Center
- Health AI Register
- Romion Health



RADIOLOGY.
Health AI Register



Why are we even talking about AI?



1:12

70s



zorg



andere beroepen

1:6

Now



zorg



andere beroepen

1:3

2060



zorg



andere beroepen

Why are we even talking about AI?



Kernpunten Technologische ontwikkelingen

- Technologische en (bio)medische ontwikkelingen in de zorg hebben geleid tot grote gezondheidswinst, maar hebben ook een aanagend effect op zorguitgaven, zowel via de prijs als via de vraag naar zorg.
- In potentie biedt een toenemende inzet van faciliterende technologieën als robotica, domotica, thuismonitoring en eHealth kansen op het gebied van personele en financiële houdbaarheid.
- Deze verwachtingen zijn echter in Nederland en andere landen tot op heden niet waargemaakt. Dit hangt mede samen met een aantal belemmerende factoren – zoals de ICT-infrastructuur, de financiering gericht op behandelvolume en het draagvlak onder patiënten en zorgmedewerkers.
- Het netto-effect van technologische innovatie in de zorg is vermoedelijke uitgavenverhogend.

The increase of facilitating technologies [...] potentially offer opportunities for the sustainability of financial and human capital.

Why are we even talking about AI?



Kernpunten Technologische ontwikkelingen

- Technologische en (bio)medische ontwikkelingen in de zorg hebben geleid tot grote gezondheidswinst, maar hebben ook een aanagend effect op zorguitgaven, zowel via de prijs als via de vraag naar zorg.
- In potentie biedt een toenemende inzet van faciliterende technologieën als robotica, domotica, thuismonitoring en eHealth kansen op het gebied van personele en financiële houdbaarheid.
- Deze verwachtingen zijn echter in Nederland en andere landen tot op heden niet waargemaakt. Dit hangt mede samen met een aantal belemmerende factoren – zoals de ICT-infrastructuur, de financiering gericht op behandelvolume en het draagvlak onder patiënten en zorgmedewerkers.
- Het netto-effect van technologische innovatie in de zorg is vermoedelijke uitgavenverhogend.

It is expected that in the Netherlands and in other countries, this potential is currently not being delivered.

Why are we even talking about AI?



Kernpunten Technologische ontwikkelingen

- Technologische en (bio)medische ontwikkelingen in de zorg hebben geleid tot grote gezondheidswinst, maar hebben ook een aanagend effect op zorguitgaven, zowel via de prijs als via de vraag naar zorg.
- In potentie biedt een toenemende inzet van faciliterende technologieën als robotica, domotica, thuismonitoring en eHealth kansen op het gebied van personele en financiële houdbaarheid.
- Deze verwachtingen zijn echter in Nederland en andere landen tot op heden niet waargemaakt. Dit hangt mede samen met een aantal belemmerende factoren – zoals de ICT-infrastructuur, de financiering gericht op behandelvolume en het draagvlak onder patiënten en zorgmedewerkers.
- Het netto-effect van technologische innovatie in de zorg is vermoedelijke uitgavenverhogend.

The net effect of technological innovation in healthcare is presumably cost increasing.

Options to get an AI solution



In house development

Developing and maintaining custom AI solutions based on data from own population



Co-create

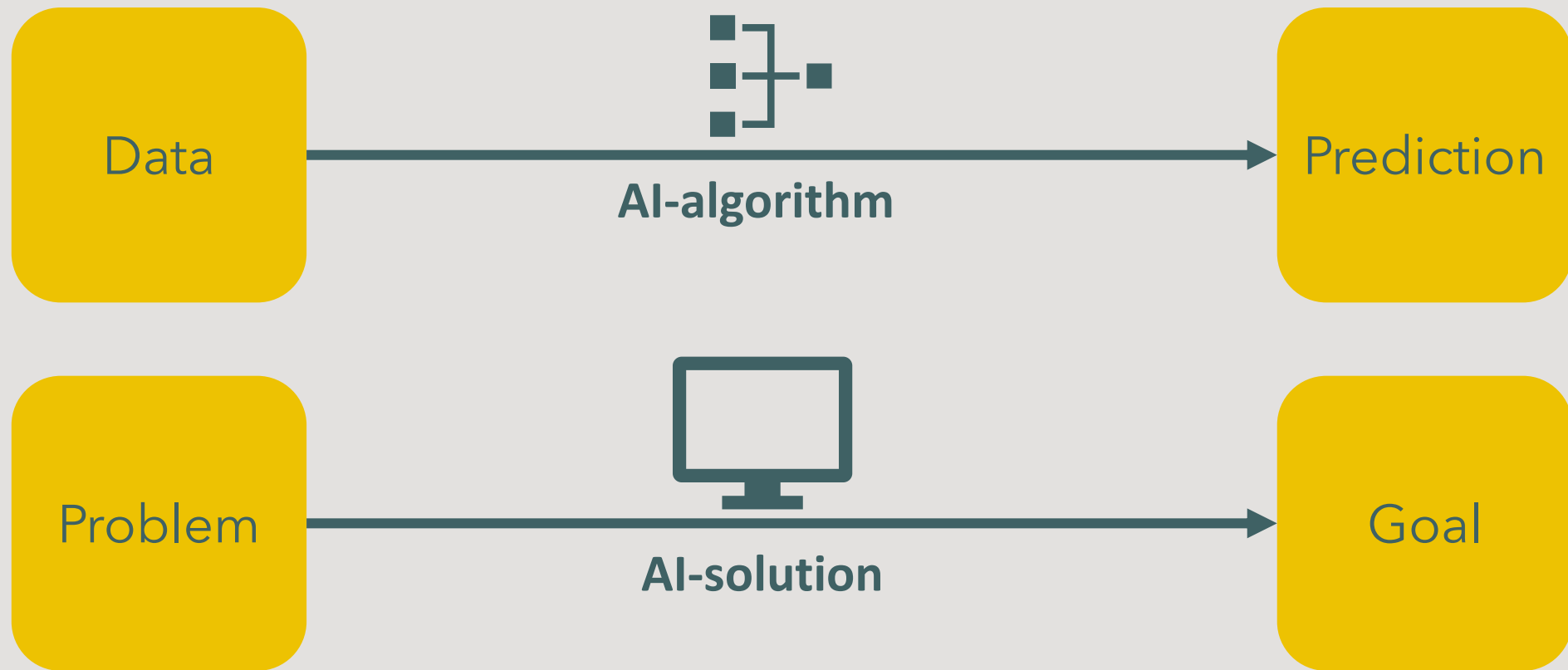
Leveraging experience from vendor and access to data and clinical knowledge from center



Buy

Buying and implementing regulated, commercially available AI solutions

AI algorithm vs. AI solution



“For every \$1 you spend on developing an algorithm, you must spend \$100 to deploy and support it”

Data Collection

Annotations

Modelling

Clinical Validation

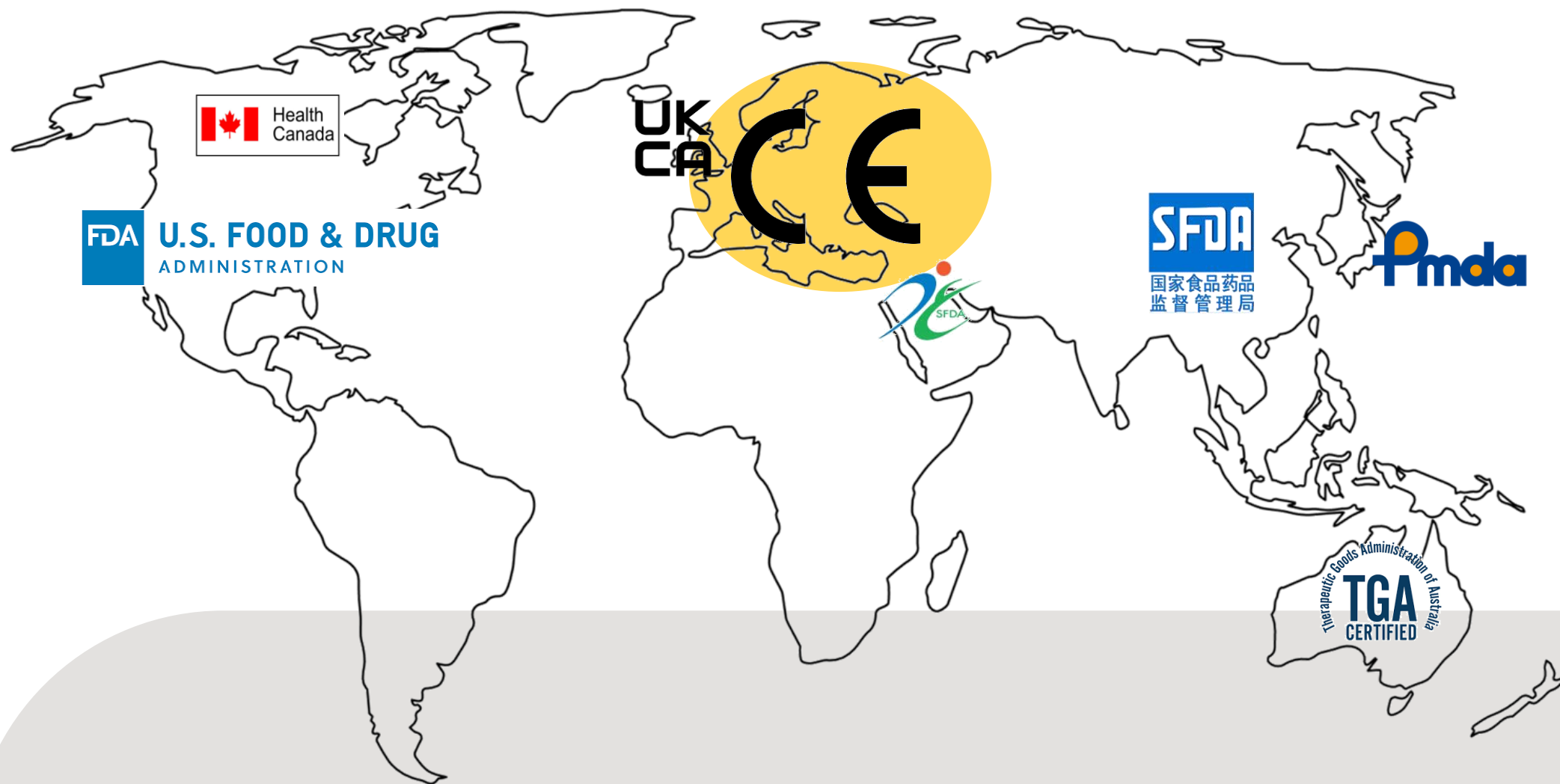
Product Integration

Certification

Deployment & service

- Four years of product development;
- A team of 60+ with an emphasis on software and delivery;
- €12.5m in funding to validate the solution at scale.

Regulatory compliance





**CE-marking tells me that
in my clinical setting the
AI solution is:**

- A. Safe
- B. Accurate
- C. Effective
- D. All of the above
- E. None of the above



**CE-marking tells me that
in my clinical setting the
AI solution is:**

- A. Safe**
- B. Accurate
- C. Effective
- D. All of the above
- E. None of the above

Commercial AI software in radiology

radiology.healthairegister.com

Status as of Apr, 2025
radiology.healthairegister.com

 **RADIOLOGY.**
Health AI Register

ProductsCompaniesNewsAbout ▾Contact

Radiology

Find the artificial intelligence based software for radiology that you are looking for.
All products listed are available for the European market (CE marked).

Subspeciality:Modality:CE ⓘ:CE class ⓘ:FDA class ⓘ:Sort by:

All ▾All ▾All ▾All ▾All ▾last modified ▾

Search...  Search

232/232 results



Carebot
AI DRIVEN HEALTHCARE

Carebot

Carebot AI CXR

Radiologic abnormal finding detection, textual report generation, worklist prioritisation

Carebot AI CXR is a deep learning-based decision support software that assists radiologists in interpreting chest X-ray images in AP/PA projection. It serves as a secondary reader of seven key ...

CE: Class IIa - MDR 

FDA: 

Information sources: Vendor

Clinical use of commercial AI in the Netherlands

NEURO CT			NEURO MR		MSK XR		MR			
StrokeViewer (5) Nicolab		Intracranial Hemorrhage (2) Aidoc	Rapid CTA (2) RapidAI		Quantib ND (2) Quantib	Rapid MRI RapidAI	BoneXpert (10) Visiana		IB Lab LAMA ImageBiopsy Lab	
		Rapid ASPECTS (3) RapidAI			NeuroQuant cortechs.ai	SyMRI Neuro Synthetic MR			BoneView Gleamer	
		Rapid ICH (2) RapidAI			cNeuro MRI Combinostics					
Rapid CTP (3) RapidAI										
CHEST CT			CHEST XR		MSK CT		BREAST MAMMO		ABDOMEN MR	
Veye Lung Nodules (4) Aidence		AI-Rad Companion Chest CT (2) Siemens Healthineers	Veolity MeVis Medical Solutions	ChestEye CAD Oxipit		C-Spine (4) Aidoc		Quantib Prostate (2) Quantib		
		CAD4COVID-CT Thirona	Lunit INSIGHT CXR3 Lunit							
Pulmonary Embolism (3) Aidoc		contextflow SEARCH Lung CT contextflow	ClearRead CT - Vessel Suppress Riverain Technologies	AI-Rad Companion Chest X-ray Siemens Healthineers		Transpara (3) ScreenPoint Medical		cvi42 for Cardiac MR (2) Circle Cardiovascular Imaging		
		AI-Rad Companion Chest X-ray Siemens Healthineers	ClearRead CT – Compare Riverain Technologies	ClearRead Xray - Bone Suppress Riverain Technologies						
		Pulmonary nodules QC Aidoc	ClearRead CT – Detect Riverain Technologies							

Leeuwen, K. G. van, et al. (2023). Clinical use of artificial intelligence products for radiology in the Netherlands between 2020 and 2022. European Radiology.

Tunnel vision

"I saw this conference talk on an AI solution with a great sensitivity - this is what we need!"

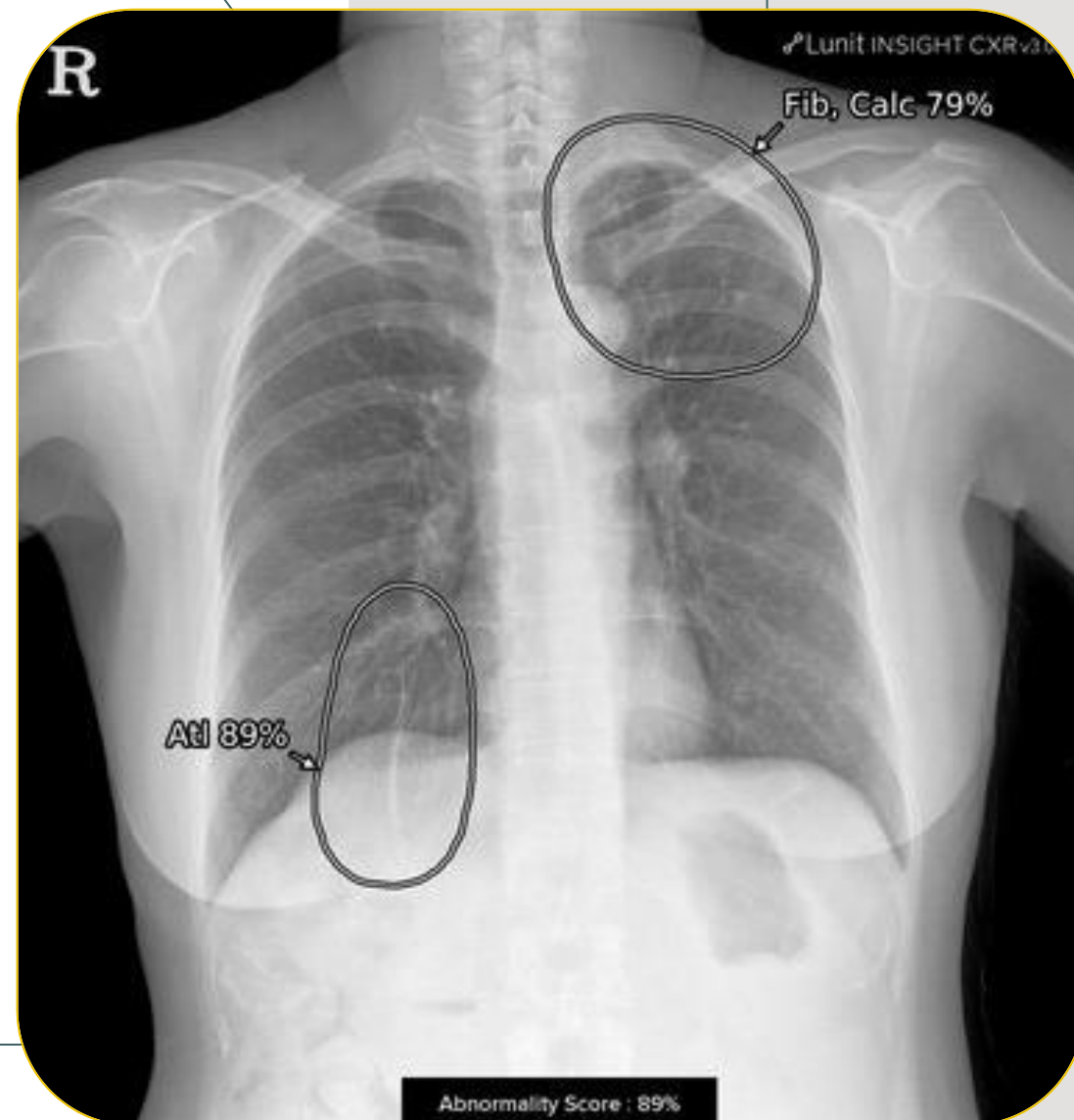
"We got contacted by this company, their product looked like something we could use."

"We have invested already so much in this, let's now move ahead with it."

Goal

Examples where it did not work out

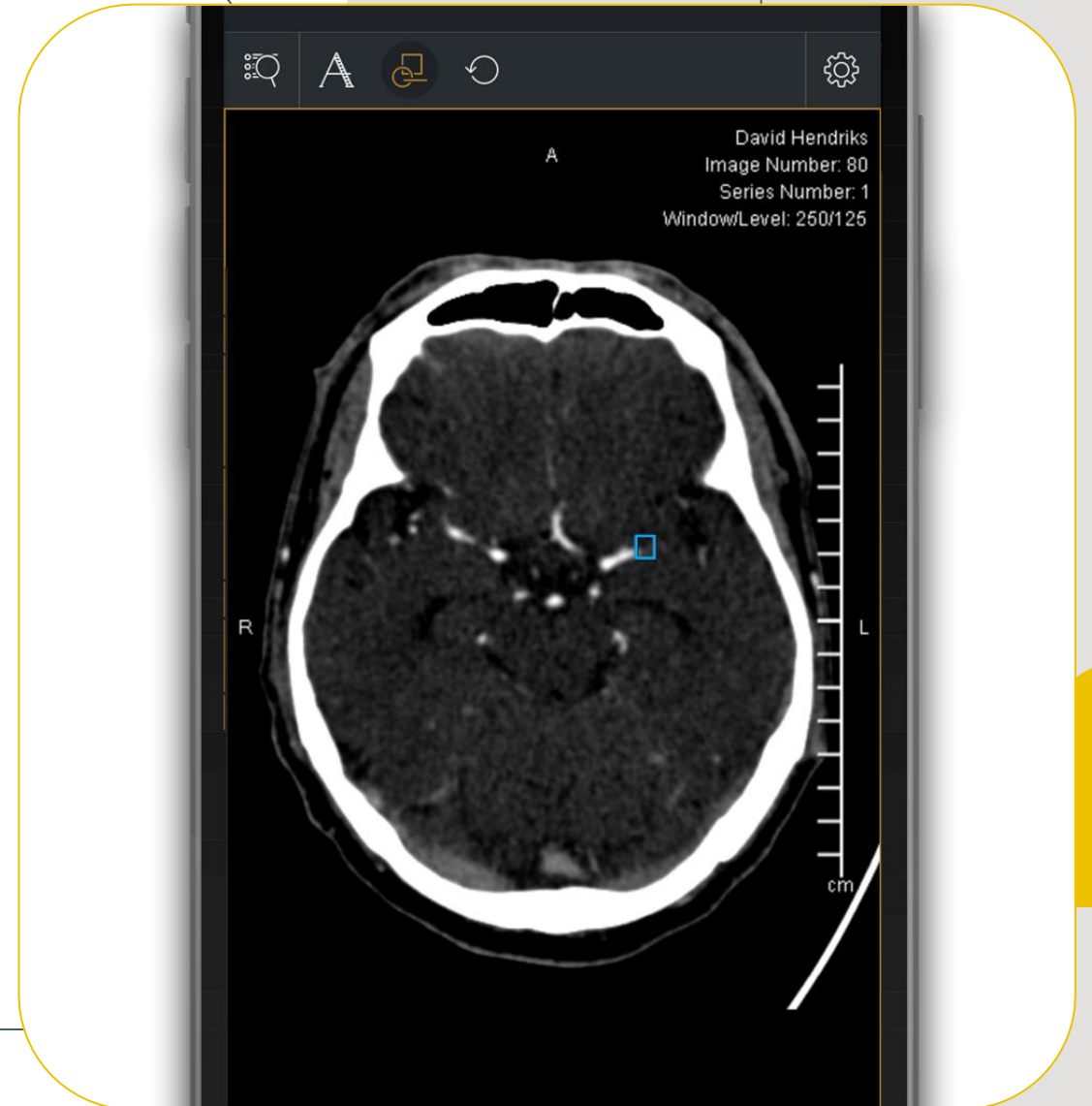
The AI solution was not reaching the actual goal of increased efficiency.



Performance

Examples where it did not work out

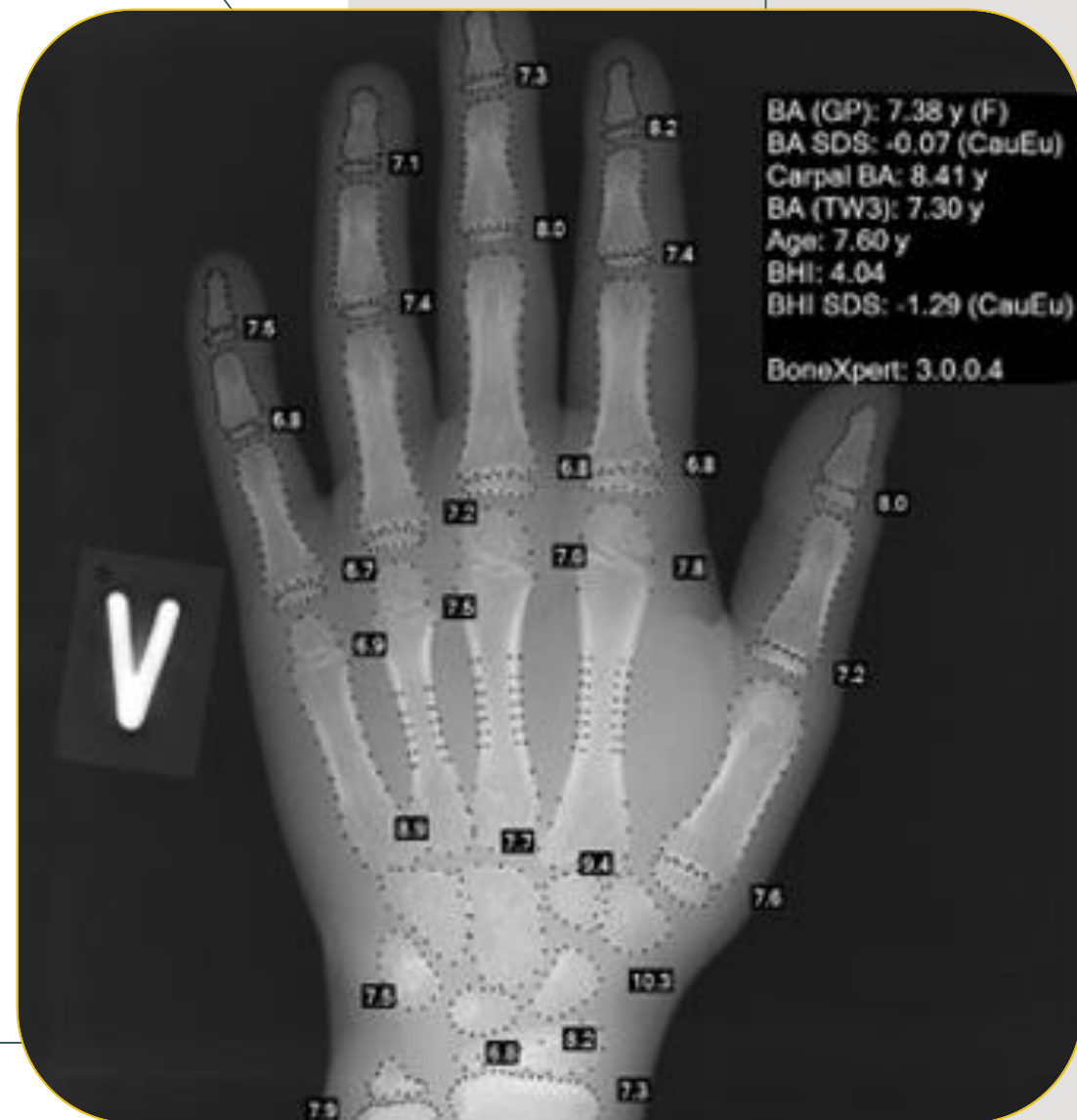
The AI solution was struggling with the cases the radiologists were looking support for.



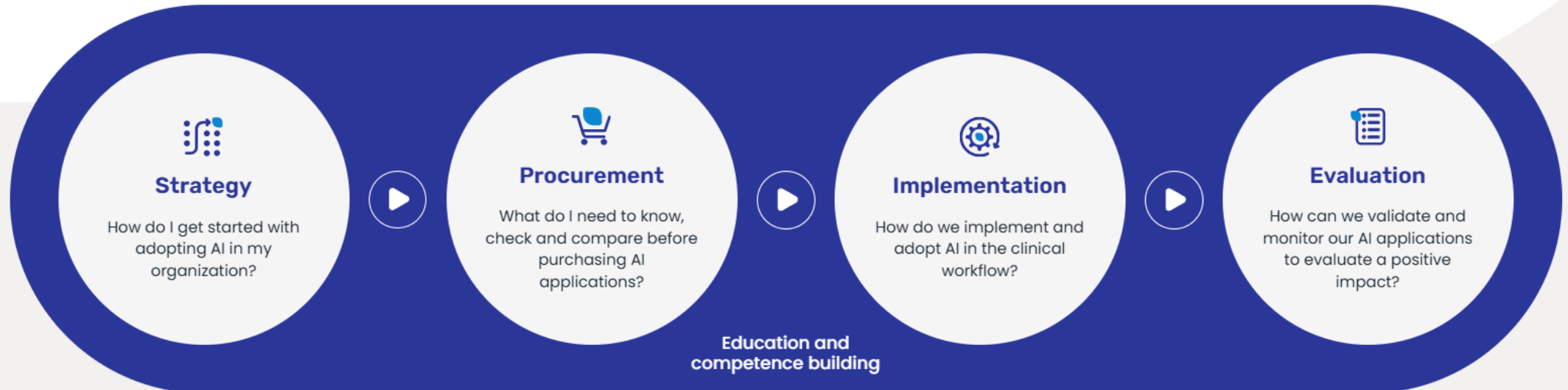
Updates

Examples where it did not work out

After an update on the acquisition device the AI solution was no longer able to process the images.



Responsible AI adoption



Ensuring quality of AI

- What evidence is available?
- Should we locally validate it in every hospital? Or as a region?
- Is evidence from similar centers enough?
- How many studies should we validate on?
- Should we run it in shadow-mode first?
- How to monitor once in use?
- ...

Scientific evidence

INSIGHT CXR3 (Lunit)

<https://www.healthairegister.com/product/lunit-insight-cxr3/>

Evidence

Peer reviewed papers on performance

Impact of AI-assisted CXR analysis in detecting incidental lung nodules and lung cancers in non-respiratory outpatient clinics ([read](#))

Independent evaluation of the accuracy of 5 artificial intelligence software for detecting lung nodules on chest X-rays ([read](#))

Computer-aided detection of tuberculosis from chest radiographs in a tuberculosis prevalence survey in South Africa: external validation and modelled impacts of commercially available artificial intelligence software ([read](#))

Comparison of Commercial AI Software Performance for Radiograph Lung Nodule Detection and Bone Age Prediction ([read](#))

Effects of Implementing Artificial Intelligence-Based Computer-Aided Detection for Chest Radiographs in Daily Practice on the Rate of Referral to Chest Computed Tomography in Pulmonology Outpatient Clinic ([read](#))

The impact of artificial intelligence on the reading times of radiologists for chest radiographs ([read](#))

Incidentally found resectable lung cancer with the usage of artificial intelligence on chest radiographs ([read](#))

AI Improves Nodule Detection on Chest Radiographs in a Health Screening Population : A Randomized Controlled Trial ([read](#))

Multicentre external validation of a commercial artificial intelligence software to analyse chest radiographs in health screening environments with low disease prevalence ([read](#))

Validation study of machine-learning chest radiograph software in primary and emergency medicine ([read](#))

Association of Artificial Intelligence–Aided Chest Radiograph Interpretation With Reader Performance and Efficiency ([read](#))

Successful Implementation of an Artificial Intelligence-Based Computer-Aided Detection System for Chest Radiography in Daily Clinical Practice ([read](#))

Diagnostic performance of artificial intelligence approved for adults for the interpretation of pediatric chest radiographs ([read](#))

Deep Learning for Detecting Pneumothorax on Chest Radiographs after Needle Biopsy: Clinical

ChestView (GLEAMER)

<https://www.healthairegister.com/product/gleamer-ai-chestview/>

Evidence

Peer reviewed papers on performance

Using AI to Improve Radiologist Performance in Detection of Abnormalities on Chest Radiographs ([read](#))

Learning from the machine: AI assistance is not an effective learning tool for resident education in chest x-ray interpretation ([read](#))

Pneumothorax (Aidoc)

<https://www.healthairegister.com/product/aidoc-pneumothorax/>

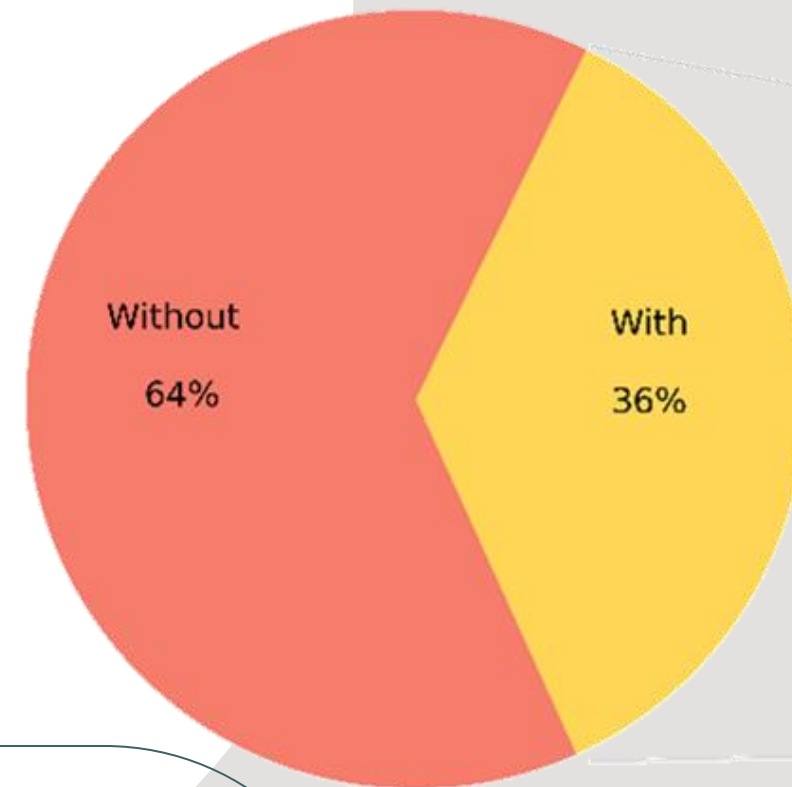
Evidence

Peer reviewed papers on performance

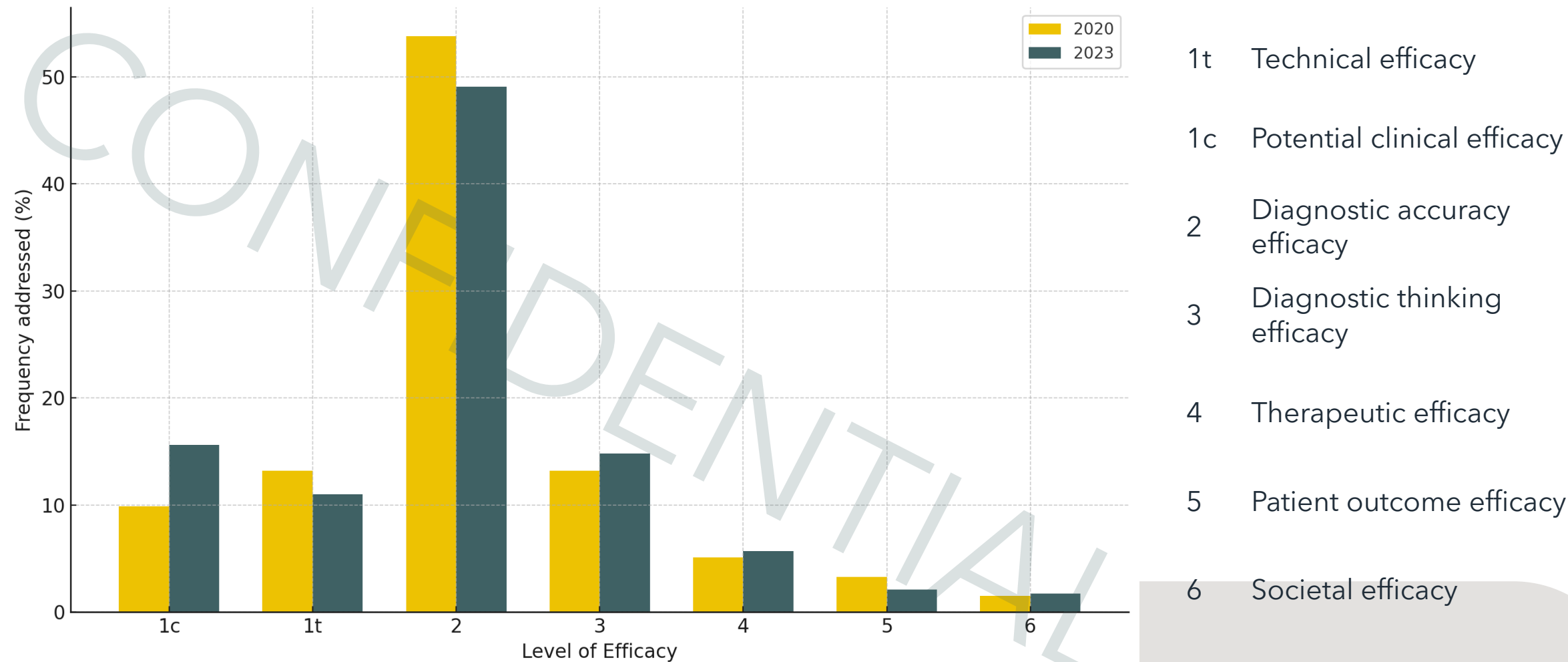
Available evidence

- Publicly available scientific evidence first 100 AI products, CE marked.

Products with peer-reviewed publications



Evidence of 100 AI products (2020) vs 189 products (2023)





How to generate (scientific) evidence?

Validation of AI

- Local validation in every hospital
- Regional/national validation
- Benchmarking studies

Validation types



Retrospective analysis

How well does the solution do on my data?



Acceptance test

Does the implemented solution work end-to-end in my clinical setting?



Pilot

If the solution would have been live, would it have functioned as expected?



Post deployment monitoring

Is the solution used in clinical practice functioning as expected?

Analysis of the clinical evidence

Yes

Is there sufficient clinical evidence available on relevant population data?

No

What is the clinical risk?

What is the clinical risk?

Low

High

Low

High



Retrospective analysis



Acceptance test



Shadow-mode



Post deployment monitoring



Retrospective analysis



Acceptance test



Shadow-mode



Post deployment monitoring



Retrospective analysis



Acceptance test



Shadow-mode



Post deployment monitoring



Retrospective analysis



Acceptance test



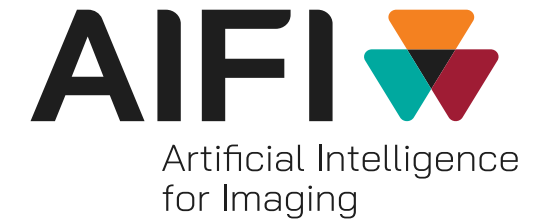
Shadow-mode



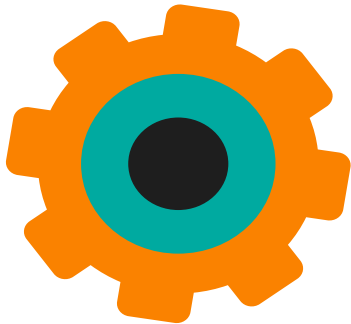
Post deployment monitoring

**Some examples of how we have approached
the quality assessment practice**

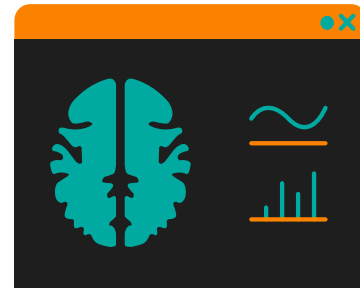
National infrastructure for AI adoption in radiology



Technical



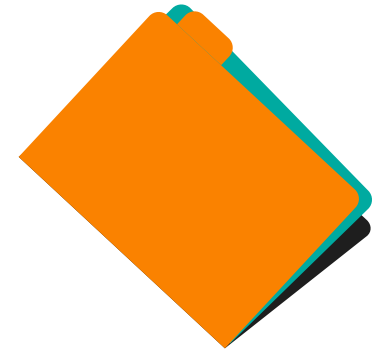
Clinical

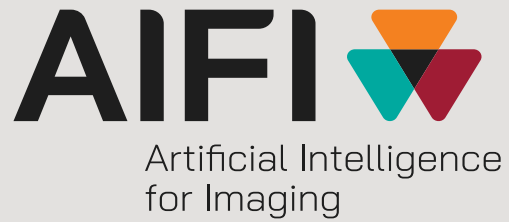


Business



Legal





Bone age, fracture detection, iPE

During procurement:

- Retrospective analysis (5 per center)

During implementation:

- Acceptance test
- (Retrospective analysis, shadow-mode)
- Post deployment monitoring



Teleradiology

Large heterogeneity in data

During procurement:

- Retrospective analysis (5-10 studies)

During implementation:

- Acceptance test
- Retrospective analysis (10-20 studies per country)
- Post deployment monitoring

National breast cancer screening program

- Retrospective analysis
 - Part of tender
- Acceptance test
- Shadow-mode
- Post deployment monitoring

Wrap up

- A critical mindset is needed for the responsible and succesful adoption of AI.

Wrap up

- A critical mindset is needed for the responsible and succesful adoption of AI.
- The quality assurance strategy depends on:
 - The available relevant evidence
 - The risk of the use of the AI

Wrap up

- A critical mindset is needed for the responsible and succesful adoption of AI.
- The quality assurance strategy depends on:
 - The available relevant evidence
 - The risk of the use of the AI
- The infrastructure of AI platforms could support validation and monitoring over time

Why are we even talking about AI?



Kernpunten Technologische ontwikkelingen

- Technologische en (bio)medische ontwikkelingen in de zorg hebben geleid tot grote gezondheidswinst, maar hebben ook een aanagend effect op zorguitgaven, zowel via de prijs als via de vraag naar zorg.
- In potentie biedt een toenemende inzet van faciliterende technologieën als robotica, domotica, thuismonitoring en eHealth kansen op het gebied van personele en financiële houdbaarheid.
- Deze verwachtingen zijn echter in Nederland en andere landen tot op heden niet waargemaakt. Dit hangt mede samen met een aantal belemmerende factoren – zoals de ICT-infrastructuur, de financiering gericht op behandelvolume en het draagvlak onder patiënten en zorgmedewerkers.
- Het netto-effect van technologische innovatie in de zorg is vermoedelijke uitgavenverhogend.

The net effect of technological innovation in healthcare is presumably cost ~~increasing~~. reducing.



Newsletter sign up!

Stephan.romeijn@romionhealth.com

<https://www.linkedin.com/in/stephanromeijn/>

Independent, objective & transparant advice.

► **Contact.**
info@romionhealth.com

► **Find us online!**
www.romionhealth.com

► **Our services.**
www.healthairegister.com

Program etter lunsj



Spørsmål vi ønsker å adressere i presentasjonene

- Hvorfor og hvem gjør kliniske validering?
- Hva kreves av validering og testing når helsetjenesten tar i bruk KI-løsninger?
- Hva kan vi gjenbruke eller dele – og hvilke vurderinger må vi uansett gjøre lokalt?
- Har det betydning om man er først ute eller følger etter?
- Hvordan har dere tenkt å gjøre løpende kvalitetssikring, for eksempel løpende validering, oppdage modell-drift, menneskelig overblikk?

• 12.00 – 14.00 Validering Norge rundt

- **Vestre Viken:** Jon Haakon Malmer-Høvik, avdelingssjef bildediagnostikk
- **Universitetssykehuset i Nord-Norge:** Karl Øyvind Mikaelson, leder SPKI og/eller Camilla Andersen, KI-radiograf
- **Haraldsplass diakonale sykehus:** Anagha P. Parkar, fagansvarlig radiolog
- **Helse Møre og Romsdal HF:** Endre Grøvik, medisinsk fysiker og leder for forskningsgruppen «AI in Medical Imaging» og/ eller Ida Jørgensen, rådgiver Klinikk for diagnostikk
- **Diskusjon og refleksjon** rundt hvordan vi best mulig kan dele erfaringer og eventuelt samarbeide fremover.

Validering av statiske CE-merkede KI-applikasjoner



Hvert 15. minutt
døgnet rundt blir
en pasient i Vestre
Viken diagnostisert
med støtte fra KI

Jon Haakon Malmer-Høvik avdelingssjef Bildediagnostikk Vestre Viken

Målsetning for innføring av KI i Bildediagnostikk i VV



- Frigjøre tid for helsepersonell for å bidra til kapasitet til pasientene som må prioriteres

Formål med validering

Sikre akseptabel risiko ved bruk av KI-applikasjon i ny arbeidsflyt

- Vurdere om diagnostisk resultat er som forventet
 - Forskning
 - Andre sykehus
- Kunnskap mtp å tilpasse bruk av KI-applikasjon og arbeidsflyt
 - Styrker
 - Svakheter
- Endringsverktøy – tillit

Leder må gjøre en helhetlig vurdering av risiko



Med KI

Klinisk risiko
IKT-sikkerhet
Personvern
Implementeringsressurser



Uten KI

Dagens risiko
Fremtidig risiko
Rekruttering
Kompetanseutvikling
Eldrebølgen

Eksempel på resultat fra validering

Icometrix Icobrain – kontroll av multippel sklerose på MR

KI-Resultat	Ratio	CI-Lav/høy	Radiolog	
Nøyaktighet	0,73	[0,65 ; 0,80]	0,97	[0,92 ; 0,99]
Sensitivitet	0,83	[0,55 ; 0,95]	0,75	[0,51 ; 0,90]
Spesifisitet	0,72	[0,63 ; 0,79]	1,00	[0,97 ; 1,00]
PPV	0,23	[0,13 ; 0,38]	1,00	[0,76 ; 1,00]
NPV	0,98	[0,92 ; 0,99]	0,97	[0,92 ; 0,99]

(n=130), 0,02 ml terskelverdi

Ca 88% av undersøkelsene er negative, dvs uten endring

Eksempel på resultat fra validering og faseinndelt implementering

Gleamer Boneview - frakturer på røntgen

	Sensitivitet	Spesifisitet	Positiv prediktiv verdi	Negativ prediktiv verdi
Radiologer	92,6%	97,1%	95,6%	95,0%
KI	90,7%	91,7%	88,3%	93,8%

(n=634)

Fase 1:

- A. Bærum Sykehus, retrospektiv n=634, alle ledd/ben, normale og patologi
-> Pilot ny pasientflyt: Pasient kan reise hjem ved negativt funn, beskrives første virkedag
- B. Bærum Sykehus, retrospektiv n=462, fortløpende vanlig daglig drift
- C. Bærum Sykehus, prospektiv n=502, fortløpende vanlig daglig drift
- D. Bredding sykehus VV, prospektiv n>100, sammenholdt med A, B, C før endelig implementering

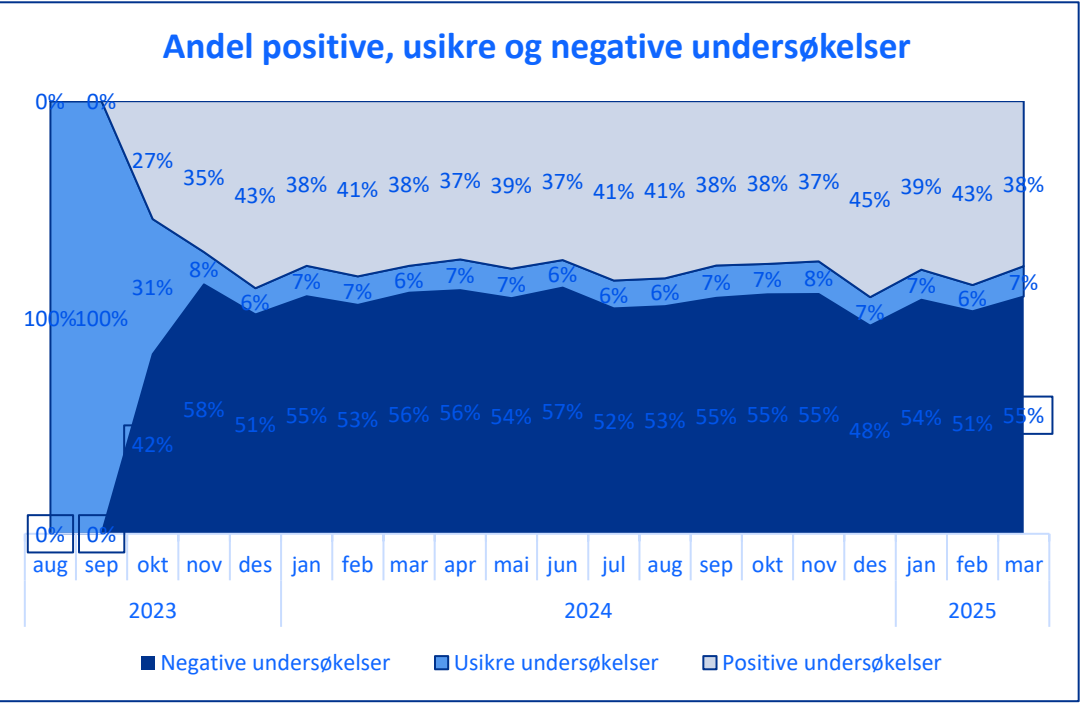
Fase 2: Ortoped støttet av KI håndterer positive uten radiologisk beskrivelse (kvalitetsstudie beslutningsgrunnlag)

Fase 3: Utvalg av negative beskrives ikke av radiolog (publiserer studie som beslutningsgrunnlag)

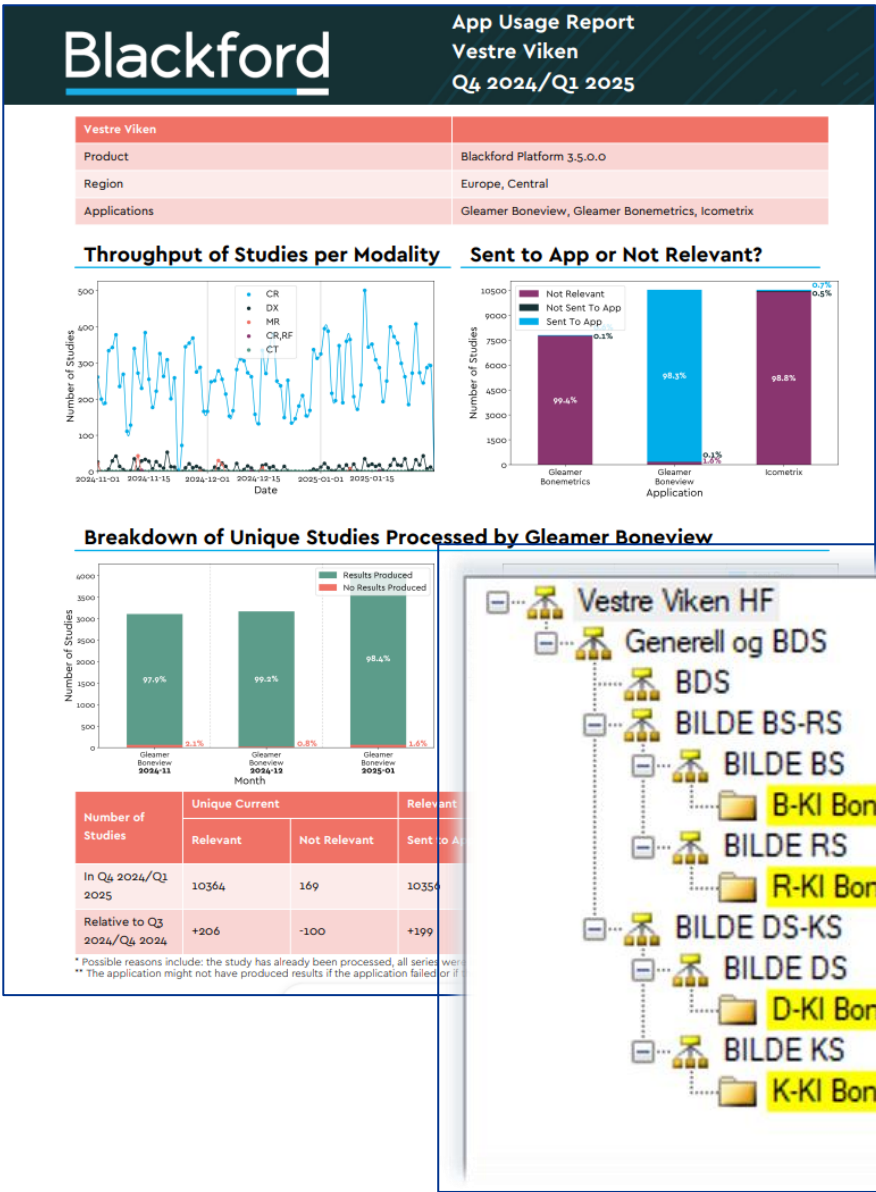
Validering

- Involvering av fagfolk fra de spesialiteter som blir berørt av KI-applikasjonen
 - Radiolog med spisskompetanse på feltet (MSK, thorax, abdomen, barn, etc)
 - Kliniker, f.eks ortoped/nevrolog/etc
 - I tillegg bidrar KI-leger, LIS og medarbeidere med forskningskompetanse
- Risiko ved bifunn vurderes i forbindelse med valideringen
- Validering på pilot-sykehus mer omfattende
 - Mindre valideringer på øvrige sykehus i HF/RHF/nasjonalt

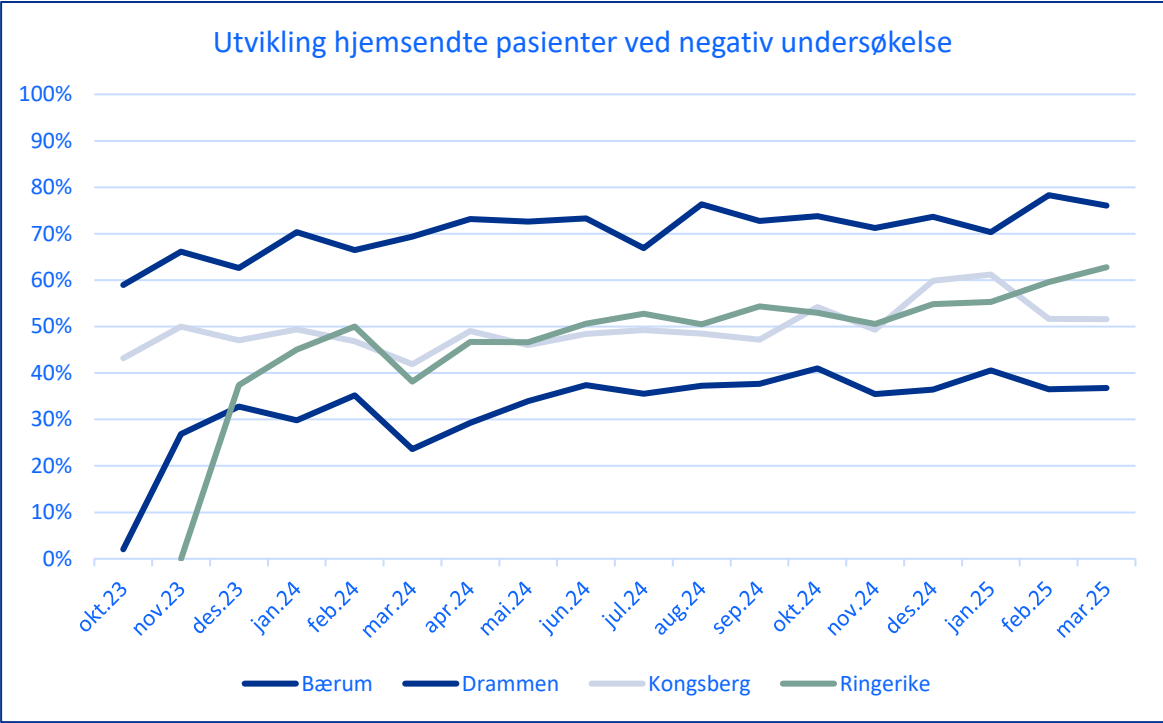
Monitorere KI-applikasjon



- Utveksling av valideringsresultater i HSØ KI nettverk
- Monitorere for evt økning av annen diagnostikk
- Nivå på monitorering avhenger av nivå på risiko



Monitorere gevinster



Spart ventetid for pasienter					
	Bærum	Drammen	Kongsberg	Ringerike	Totalt
Spart ventetid (min) dag	37 860	52 800	7 504	21 425	119 589
Spart ventetid (min) Val	185 553	46 090	22 605	123 417	377 665
Totalt	223 413	98 890	30 109	144 842	497 254
Sparte døgn	155	69	21	101	345

Sparte konsultasjoner / konsultasjonstid					
	Bærum	Drammen	Kongsberg	Ringerike	Totalt
Antall konsultasjoner	6 502	838	411	653	8 404
Spart konsultasjonstid (timer)	1 409	182	89	141	1 821

Samarbeid - HSØ KI nettverk

- Samarbeid om KI-anskaffelser
- Diskusjoner og deling av kunnskap
- Validering deles



KI Startkit

-  Oppstartsmøte
-  Møte med prosjektgruppen
-  Orientering om validering med KI lege
-  Deltagelse i workshop

Vi i Vestre Viken kan tilby løpende dialog gjennom oppstart og implementering



- HSØ samarbeid med øvrige RHF

Leder tar beslutning om å driftsette en KI-applikasjon

- Hva skal KI brukes til?
 - Risiko må gjenspeiles i omfang av validering, risikoanalyse og dokumentasjon, inkl evt studie, samt på hvilket nivå beslutning om bruk tas.
 - Er det et støtteverktøy eller vil det ikke lenger være behov for radiolog, evt faser
 - Hvilken diagnostikk brukes den til
- Valideringen er et grunnlag for beslutning.
- Støtte fra medisinskfaglig rådgiver der leder ikke selv er lege.
 - Ved behov løftes beslutning til HF
- Pilotere ett sykehus, deretter kunne bredde raskt i HF/RHF/nasjonalt via nettverk





Validering og testing av KI løsninger for klinisk bruk i radiologi i Helse Nord

Karl Øyvind Mikalsen, Avdelingsleder SPKI
Camilla H. Andersen, KI radiograf og Implementeringsleder



Anbefalinger for validering av KI-løsninger innen radiologi som skal tas i bruk i Helse Nord:

- ✓ Krev dokumentasjon fra produsenten om trening, testing og validering.
- ✓ Gjennomgå fagfelle-vurderte og andre tilgjengelige studier, og valideringer gjennomført av andre (særlig i Norge)
- ✓ Gjennomfør retrospektive valideringer på norske pasientdata før/under anskaffelse.
- ✓ Tilpass valideringskrav etter risiko
 - ✓ Eksempelvis knytter det seg større risiko ved å innføre algoritmer innen områder som hjerneslag og kreftdiagnostikk enn for løsninger innen brudddeteksjon; slike forskjeller i risiko bør også få konsekvenser for kravene til validering
- ✓ Krev livsløps-evaluering og kontinuerlig kvalitetssikring etter implementering.
- ✓ Praktiske utfordringer å stille krav til prospektiv validering.
 - ✓ Arbeidsgruppens anbefaling om validering, begrenser seg derfor til å gjøre retrospektive valideringer basert på egne pasientdata.



En illustrasjon laget av den generative KI-tjenesten DALL-E

7.1 Anbefalinger om validering av KI-løsninger som tas i bruk i Helse Nord

Siden det i Helse Nord tas sikte på å bare anskaffe CE-merkede KI-løsninger som er godkjent etter MDR, vil disse på forhånd - i regi av produsenten - ha gjennomgått en omfattende intern validering som er godkjent av et autorisert sertifiseringsorgan (Notified Body i EU).

Arbeidsgruppen anbefaler at Helse Nord:

- Før signering av eventuell kontrakt for anskaffelse, krever tilgang til produsentens dokumentasjon av hvordan og med hvilke data KI-løsningen er trent, testet og validert.
- Gjennomgår publiserte fagfelle-vurderte studier, der KI-løsninger som planlegges anskaffet er eksternt og uavhengig validert.
- Gjennomgår andre relevante og tilgjengelige valideringer, som ikke-fagfelle-vurderte whitepapers, og valideringer utført av andre helsevirksomheter (eksempelvis Vestre Viken HF mv).

Ingen av de CE-merkede KI-løsninger innen radiologi er utviklet i Norge. Disse løsningene er derfor trent på andre populasjoner, og foreløpig er bare et fåtall av disse implementert (BoneView fra Gleamer) eller validert (Transpara og Lunit innen mammografi) mot norske data. For løsninger som tidligere ikke er utprøvd i Norge, bør det derfor gjennomføres valideringer basert på retrospektive norske data før en avtale om anskaffelse signeres.

Med utgangspunkt i annoterte data fra UNN HF, vil SPKI gjennom et pilotprosjekt, i løpet av de nærmeste måneder gjennomføre en retrospektiv validering av KI-løsninger på tre funksjonelle områder (Rtg brudddeteksjon, Rtg thorax og CT thorax) som er aktuelle for anskaffelse.

Kravene til validering bør tilpasses risiko ved bruk av algoritmen, vurdert som konsekvenser for pasientene (ved lav sensitivitet), eventuelt konsekvenser for ressursbruk (ved lav spesifisitet). Eksempelvis knytter det seg større risiko ved å innføre algoritmer innen områder som hjerneslag og kreftdiagnostikk enn for løsninger innen brudddeteksjon; slike forskjeller i risiko bør også få konsekvenser for kravene til validering. Ovennevnte rapport om kvalitetssikring av KI-løsninger i helse- og omsorgstjenesten, i regi av Helsedirektoratet, gir et godt utgangspunkt for å utarbeide nasjonale retningslinjer for validering av løsninger. Arbeidsgruppen ser det som ønskelig at slike retningslinjer etableres med det formål å sikre at de lokale valideringene i størst mulig grad kan gjennomføres etter en standardisert mal. Dermed legges det bedre til rette for kumulativ kvalitetssikring, erfaringsutveksling og læring, basert på større pasientpopulasjoner.

Det samarbeidet mellom RHFene som er etablert om anskaffelse, vil øke sannsynligheten for at helseforetakene i Norge vil komme til å gjøre avrop på mange av de samme KI-løsningene. Dette vil muliggjøre samarbeid mellom RHFer og helseforetak også om validering av KI-løsninger som vurderes anskaffet. Dette kan enten skje ved samarbeid om validering av de samme algoritmene, eller ved en arbeidsdeling der man validerer ulike algoritmer, basert på en forutsetning om at pasientene er rimelig like på tvers av helseforetak. Det bemerkes imidlertid at det er variasjoner i maskinparken mellom de ulike helseforetakene, noe som innebærer at det uansett vil være behov for mindre lokale valideringer.



Bruk av KI på stråleterapi for inntegning av organ, målvolum og regionale lymfeknuder

Sykehus	Programvare	Leverandør	Versjon	Bruksområder	Status	Vurdering/kommentar
UNN	CT go.OpenPro DirectORGANS Contours Advanced	Siemens	VA40	utvalgte OAR	Brukt siden 2022	OK på noen OAR
UNN	Raystation	Raysearch	?	proton (og noe mer?)	Venter på installering	
NLSH	-					
NLSH	-					
SOH	CT go.OpenPro DirectORGANS Contours Advanced	Siemens	VA40	noen få OAR for et fåtall diagnoser	brukt siden 2022/2023	strukturer er evaluert på ca 10 caser før bruk
SOH	Raystation	Raysearch	2023B	utvalgte OAR for mange diagnoser, målvolum for bryst og tilhørende lymfeknuder	brukt siden desember 2021	strukturer er evaluert på minst 10 case av minst 1 onkolog før rutinemessig bruk
AAS	RayStation	RaySearch	2024B	Alt av OAR + MV bryst og prostata	I klinisk bruk siden des 2021	Bruker KI på alle pasienter - meget gode erfaringer med KI segmentering
HUS	ART-plan	Therapanacea	02.03.2001	OAR/lymfekaner til ØNH, prostata og bryst. CTV til bryst	brukt siden jan 2023	fungerer bra
HUS	RayStation	RaySearch	2024B	lunge	under evaluering	ser bra ut enn så lenge
SUS	CT go.OpenPro DirectORGANS Contours Advanced	Siemens	VA40		kun opsjon for leveranse - ikke midler i NyeSUS...	
SUS	ART-plan	Therapanacea	02.03.2001		mulighet for påkobling av samme kontrakt som SUS under vurdering	
SSK	Raystation	RaySearch	18B	Utvalgte OAR. Legene var ikke fornøyd med målvolum for bryst. Venter for nye modeller med neste oppgradering	brukt siden 2023	OAR er tilfredstillende hvis litt begrenset i antall. Flere tilgjengelig i neste versjon. Legene var ikke fornøyd med Bryst MV. Skal re-evaluere neste versjon.
SSK						
OUS						
OUS						
SIG	MIM	GE health care		Det meste	Installeres ila april.	
SIG						
DS	AI Rad companion organs RT	Siemens	VA60B		Under installering	
DS	* Dette er det samme som CT go.OpenPro DirectORGANS Contours Advanced, men kalles noe annet da det er kjøpt via TPS anskaffelse.					

Stråleterapi:

- Holder på å forberede anskaffelse
- usikker på hvor vi skal legge listen for validering
- Mange har tatt i bruk KI, har derfor kartlagt hva som er gjort
- Ikke kapasitet til å gjøre like grundig jobb som Haukeland gjorde



Behov for nasjonale retningslinjer

- Det bør rigges opp et nasjonalt system for kvalitetssikring og validering av KI-løsninger:
 - Nasjonale retningslinjer med formål å sikre at de lokale valideringene i størst mulig grad kan gjennomføres etter en standardisert mal.

Samarbeid

- Samarbeidet mellom RHFene om anskaffelse, vil øke sannsynligheten for at helseforetakene i Norge vil komme til å gjøre avrop på mange av de samme KI-løsningene.
- Dette vil muliggjøre samarbeid om validering av KI-løsninger.
 1. Samarbeid om validering av de samme algoritmene
 2. Arbeidsdeling der man validerer ulike algoritmer.
 - ✓ Det bemerkes imidlertid at det er variasjoner i maskinpark og protokoller mellom de ulike helseforetakene, noe som innebærer at det uansett vil være behov for mindre lokale valideringer/testing
 - ✓ Her må de større HF trekke i trådene – de små har ikke ressurser eller kompetanse

Tillegg: KI-forordningen (AI Act)

Regulatoriske sandkasser (art. 57)

- Medlemslandene skal etablere minst én regulatorisk sandkasse
- Fasilitere kontrollert miljø for utvikling, testing og validering av KI-systemer
- Tilrettelegge for «real world»-testing
- Formålet er å fremme innovasjon og konkurranseevne, og særlig støtte SME's og start-ups
- Kan det bli innen helse ?

Innhold

Forord	→
Sammendrag	→
Innledning	→
Fase 1: Beskrive utfordringen og vurdere om KI er løsningen	→
Beskrive hvilket behov KI-systemet skal ivareta	→
Overordnet risikovurdering ved bruk av et KI-system	→
Overordnet gevinstvurdering	→
Involvere rett fagkompetanse	→
Relevante kilder til fase 1	→
Fase 2: Kartlegge KI-systemer på markedet	→
Identifisere aktuelle KI-systemer	→
Regelverk for bruk av KI-systemer	→
Vurdering av kvaliteten til et KI-system	→
Relevante kilder til fase 2	→
Fase 3: Vurdere egen virksomhets evne til å kunne ta i bruk et KI-system	→
Menneskelige faktorer	→
Tekniske faktorer	→
Planlegge anskaffelse	→
Erfaringsdeling og forskning	→
Relevante kilder til fase 3	→
Fase 4: Anskaffe et KI-system	→
Vurdere samarbeid om anskaffelse av KI-systemer	→
Regelverk for anskaffelser	→
Anskaffelsesprosedyre	→
Prismodeller	→
Avtaler	→
Kravspesifikasjon	→
Vurdering av leverandørene	→
Vurdering av tilbudte KI-systemer	→
Relevante kilder til fase 4	→
Fase 5: Innføre og kvalitetssikre et KI-system	→
Ny risikovurdering	→
Iverksette tekniske og organisatoriske tiltak	→
Testing av KI-systemet	→
Overlevering med plan for drift og forvaltning	→
Relevante kilder til fase 5	→

Hva kan vi gjenbruke eller dele?



Fra teori til praksis – ny ekspertgruppe skal få fart på kunstig intelligens i arbeidslivet

Pressemelding | Dato: 17.03.2025

Statsråd Karianne Tung setter ned en ekspertgruppe som får i oppgave å lage en praktisk veileder for å øke tempoet i bruken av kunstig intelligens (KI) på norske arbeidsplasser. Tung vil ha konkrete oppskrifter på hvordan virksomheter kan ta i bruk assistenter som baserer seg på KI.

En bredt sammensatt ekspertgruppe fra akademien, næringslivet og offentlig sektor skal utarbeide veilederen. Målet er å gi virksomheter tydelige råd for ansvarlig utvikling og bruk av KI-assistenter. Veilederen skal være praktisk rettet, tilpasset norske forhold, og relevant for sentrale næringer. Gruppen besitter ekspertise på KI, både på strategisk og operativt nivå. Partene i arbeidslivet, NHO og LO, er også representert. Ekspertgruppen ledes av Jon Atle Gulla, direktør for NorwAI og professor ved NTNU.

Deltakerne i ekspertgruppen:

- Jon Atle Gulla (leder), direktør for NorwAI & Professor ved Institutt for datateknologi og informatikk, NTNU, Trondheim
- Rebekka Borsch, avdelingsdirektør for kompetanse, innovasjon og digitalisering, NHO, Oslo
- Kjetil Stålesen, spesialrådgiver, LO, Oslo
- Astrid Undheim, konserndirektør - Teknologi og utvikling, SpareBank 1 SMN, Trondheim
- Mette Rønning Raabel, Product Manager – Advance, DNV Maritim, Bærum
- Malin Tønseth, partner, head of compliance & Risk, Simonsen Vogt Wiig, Oslo
- Torbjørn Folgerø, Direktør for IT og digitalisering, Equinor, Stavanger
- Ann Merethe Lysø Sommerseth, chief commercial officer, Secure Practice, Trondheim
- Arne Ingebrigtsen, kommunedirektør, Kristiansund kommune
- Hilde Margrethe Lovett, seniorrådgiver og prosjektleder for samordnet KI-plan i helse- og omsorgssektoren, Helsedirektoratet, Oslo
- Øyvind Husby, administrerende direktør, IKT Norge, Oslo
- Trond Trosterud, professor i samisk lingvistik, Norges arktiske universitet, Giellatekno, Tromsø
- Kari Anna Fiskvik, teknologidirektør og VP Technology, Strawberry, Trondheim
- Per Kristian Vareide, kommunedirektør, Stavanger kommune

Sekretariatet ledes av Digital Norway med deltakelse fra Digitaliseringsdirektoratet



Etablere et regionalt system for kvalitetssikring:

- Implementeringen av KI innen radiologi vil stille andre krav til forvaltning enn man er vant til fra tidligere.
 - ✓ behov for tettere samhandling mellom fagpersoner med domene- og KI-kompetanse i foretakene og personell med IKT kompetanse i HN IKT
- I Helse Nord anbefales det å etablere et regionalt system for kvalitetskontroll og evaluering av KI-løsningene, i samarbeid mellom HNIKT og fagmiljøene
 - ✓ Periodisk overvåkning av KI-løsningenes ytelse for å sikre at løsningene over tid gir samme kvalitetsmessige resultat - behov for ferdigutviklet programvare som kan benyttes til formålet
 - ✓ Det anbefales å involvere radiologer eller andre kliniske ressurser med kompetanse til å ivareta den medisinskfaglige vurderingen i kvalitetssikringsarbeidet.
 - ✓ Det anbefales å benytte Kompetansesenter for diagnostisk fysikk (KDF) i arbeidet med foretakenes optimalisering av undersøkelsesprotokoller og kvalitetssikring av bildediagnostikken.
 - ✓ Det anbefales opprettelse av lokale superbrukere som skal være bindeledd mellom klinikk og IKT.
- Vi har ikke begynt å jobbe med dette enda...



Fra en radiologs perspektiv

Det er viktig at valideringen er robust nok og adresserer aktuelle forskningsspørsmål for å skape tillit:

- a. Verktøyets sensitivitet og spesifisitet
- b. Valideringen må inkludere et stort nok antall av ulike patologier
- c. Scoringen må være blindet
- d. Effekt på diagnostikk og behandling av pasienten

- ☐ Hvis valideringen er robust nok vil det potensielt være høyre aksept for gjenbruk
 - ✓ vil være ressurs- og tidsbesparende



Fordeler med lokal validering

- ✓ Validering er mer enn å undersøke KI-verktøyets treffsikkerhet.
 - ✓ Hands-on erfaring for brukerne
 - ✓ Teste muligheter innenfor brukergrensesnittet og se eksempler - Tillit
- ✓ Kunnskapsgrunnlag ved etablering av ny arbeidsflyt og prosedyrer
 - ✓ Detaljer om hva KI-verktøyet er god/mindre god på
 - ✓ Hvordan vi kan implementere verktøyet i arbeidsflyten på en god måte – gevinster
 - ✓ Nyttig ved opplæring

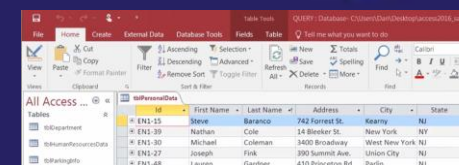


Hvorfor:

- ✓ Erfaringer med brukervennlighet og integrasjon i PACS
- ✓ Kunnskaper om KI-verktøyets treffsikkerhet, styrker og svakheter
- ✓ Behov for grundigere validering av barnespesifikke frakturer
- ✓ Tillit

Hvem gjør klinisk validering:

- KI-radiograf har utført det praktiske forberedelsesarbeidet, og har konferert med radiolog ved behov
 - ✓ planlagt valideringsmetodikk
 - ✓ hentet ut datasett
 - ✓ laget koblingsnøkkel
 - ✓ anonymisert caser i prod-PACS
 - ✓ laget strukturert svarrapport i REDcap
 - ✓ fører inn KI-resultater i strukturert svarrapport
- Radiolog satt Ground truth
- LIS tolker us. uten og med KI – svar i strukturert rapport
- Analyse av resultater – KI-radiograf, radiolog og ressurser fra SPKI

First Name	Last Name	Address	City	State
Nathan	Cole	14 Bleecker St.	New York	NY
Michael	Coleman	3420 Broadway	West New York	NJ
Joseph	Fink	390 Summit Ave.	Union City	NJ
Lauren	Gardner	430 Princeton Rd.	Warren	NJ





Datasett Barn

- ✓ 600 pasienter - totalt 656 us.
- ✓ Øyeblikkelig hjelp med spørsmål om skjelettskade

Poweranalyse

- ✓ 600 us. 3 aldersgrupper, 5 regioner med 120 i hver region gir en ok Power.
- ✓ Skal publisere resultatene av barne-valideringen

Boneview Barn 2-18 år	2-5 år	6-10 år	11-15 år	Antall
Overekstremitet				
Hånd, håndledd og underarm	40	40	40	120
Albue	40	40	40	120
Clavicula, Skulder, overarm og scapula	40	40	40	120
Underekstremitet				
Bekken, hofte, lår, kne og legg	40	40	40	120
Ankel og fot	40	40	40	120
Antall pasienter	200	200	200	600

Datasett Voksne

- ✓ 300 pasienter - totalt 370 us.
- ✓ Øyeblikkelig hjelp med spørsmål om skjelettskade
- ✓ 50% 18-60 år
- ✓ 50% > 61 år
- ✓ Balansert datasett 50 pos/neg
- ✓ Inkludert et bredt spekter av ulike frakturer, både åpenbare og subtile, vanlige og uvanlige frakturer.

Ground truth:

Radiolog registrerte funn i strukturer svarrapport i REDcap

- Hadde tilgang til klinisk informasjon og bilder til sammenligning

- ✓ Funn (Fraktur/Usikker/Ingen Fraktur)
- ✓ Lokalisasjon (knokkel)
- ✓ Segment (Proksimal/skaft/distal)
- ✓ Øvrige funn (Luksasjon/hydrops/lesjon)
- ✓ Behov for supplerende bilder
- ✓ Videre tiltak (7 alternativer)

Validering voksne

	
Radiolog registrerer funn (ground truth/fasit)	Vi registrerer funn gjort av KI-algoritme 1 (stand-alone)

- ✓ Treffsikkerheten for KI-verktøyene alene
- ✓ Treffsikkerheten for radiolog både med og uten KI som beslutningsstøtte

Validering barn LIS får se redigert henvisningstekst – ingen bilder til sml.

					
Radiolog registrerer funn (ground truth/fasit)	Vi registrerer funn gjort av KI-algoritme (stand-alone)	LIS A registrerer funn uten KI	LIS A registrerer funn med KI	LIS B registrerer funn uten KI	LIS B registrerer funn med KI

Validering Norge rundt Presentasjon fra Helse Vest

Oslo, 28.04.25

Anagha P. Parkar

Fagansvarlig radiolog, PhD.

Haraldsplass Diakonale Sykehus

Dagens «oppdrag»

- Hvorfor og hvem gjør kliniske validering?
- Hva kreves av validering og testing når helsetjenesten tar i bruk KI-løsninger?
- Hva kan vi gjenbruke eller dele – og hvilke vurderinger må vi uansett gjøre lokalt?
- Har det betydning om man er først ute eller følger etter?
- Hvordan har dere tenkt å gjøre løpende kvalitetssikring, for eksempel løpende validering, oppdage modell-drift, menneskelig overblikk?

Hvorfor AI i radiologi?

- Før undersøkelsen: automatisk prioritering av henvisninger ?
- Maskiner: bedre bildekvalitet/ postprosessering
- Worklist prioritisation
- Deteksjon av patologi (røntgen, CT og MR)
- Volummålinger av lesjoner, mest utbredt lunge cancer screening
- Differensiering og radiologisk diagnose (røntgen, CT, MR)

Hvorfor AI i radiologi?

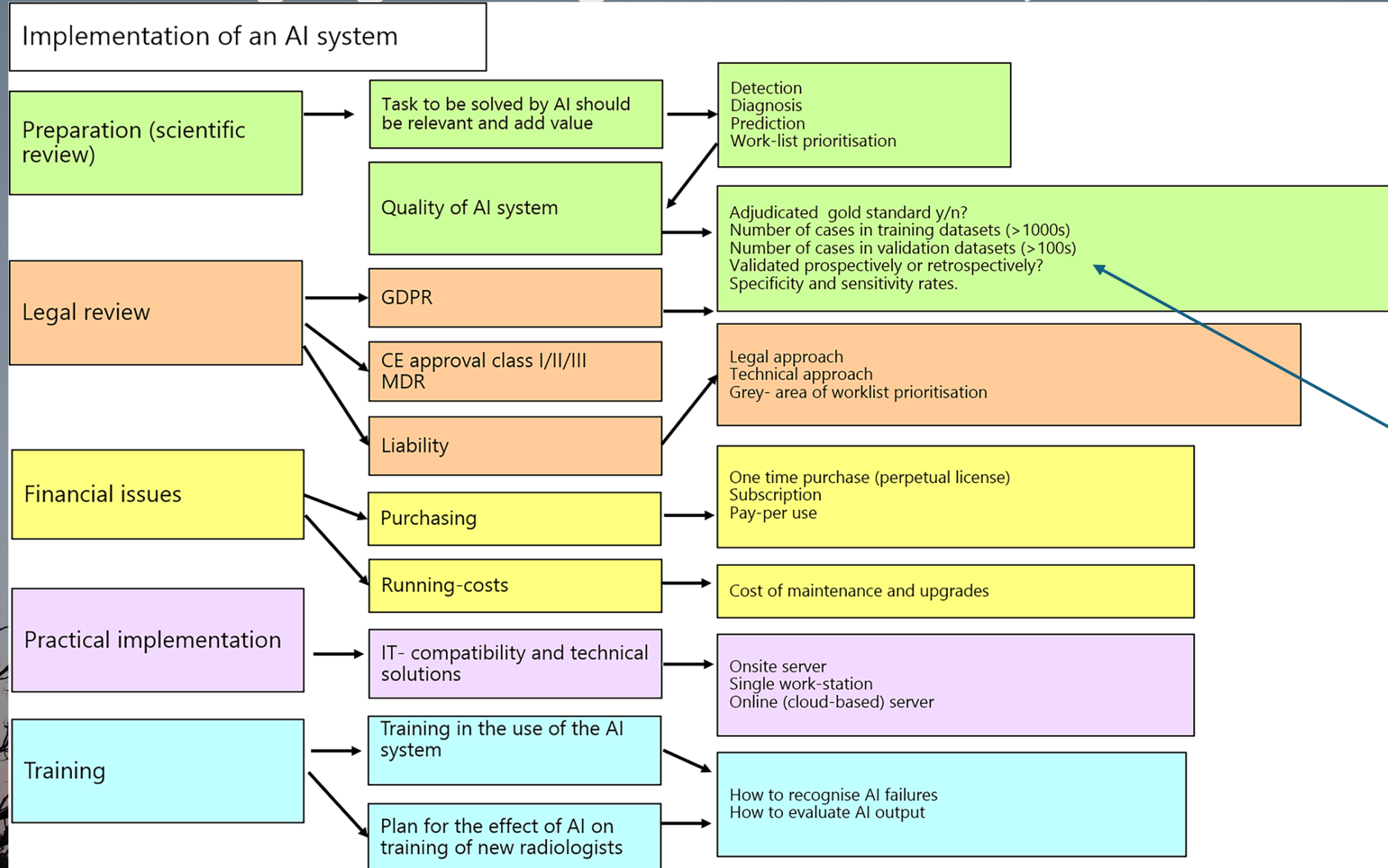
- Før undersøkelsen: automatisk prioritering av henvisninger ?
- Maskiner: bedre bildekvalitet/ postprosessering ✓
- Worklist prioritisation
- Deteksjon av patologi (røntgen, CT og MR)
- Volummålinger av lesjoner, mest utbredt lunge cancer screening
- Differensiering og radiologisk diagnose (røntgen, CT, MR)

Hvorfor og hvem gjør kliniske validering?

- Hvorfor:

- Radiologisk avdeling består av kun overleger i hjemmevakt på natt.
- Røntgen for brudd: ortopeder veldig flinke til å se på selv
- Røntgen thorax ses på av kliniker på natt, beskrives tidlig neste dag av radiolog.
- Tenkte at AI kunne bli en god støtte for klinikere på natt.
- Vi ga uttrykk for ønsker om testing høsten 2022.
- Kom i gang i 12. juni 2024.

Validering og testing av KI i helsetjenesten



Using AI to Improve Radiologist Performance in Detection of Abnormalities on Chest Radiographs

Soubail Bennani, MD • Nor-Eddine Regnard, MD • Jeanne Ventre, PhD • Louis Lassalle, MD • Toan Nguyen, MD • Alexis Ducarouge, MSc • Lucas Dargent, MD • Enora Guillo, MD • Elodie Goubier, MD • Sophie-Hélène Zaimi, MD • Emma Canniff, MD • Cécile Malandrin, MD • Philippe Khafagy, MD • Hasmik Koulakian, MD • Marie-Pierre Revel, MD, PhD • Guillaume Chassagnon, MD, PhD

From the Department of Thoracic Imaging, Cochin Hospital, AP-HP; 27 Rue du Faubourg Saint-Jacques, Paris 75014, France (S.B., L.D., E. Guillo, E. Goubier, S.H.Z., E.C., M.P.R., G.C.); Gleamer, Paris, France (S.B., N.E.R., J.V., L.L., T.N., A.D.); Réseau d'Imagerie Sud Francilien, Lieusaint, France (N.E.R., L.L., C.M.); Department of Pediatric Radiology, Armand Trousseau Hospital, AP-HP, Paris, France (T.N.); HFR Fribourg, Fribourg, Switzerland (P.K.); and Centre d'Imagerie Médicale de l'Ouest Parisien, Paris, France (H.K.). Received April 20, 2023; revision requested June 26; revision received October 9; accepted October 23. Address correspondence to M.P.P. (email: marie-pierre.revel@aphp.fr).

Supported by Gleamer.

Conflicts of interest are listed at the end of this article.

Radiology 2023; 309(3):e230860 • <https://doi.org/10.1148/radiol.230860> • Content codes: CH AI

Chestview (Gleamer)

- Kun for stående front og side
- Ser på 5 diagnose:
 - Alveolar syndrome (fortetning)
 - Nodule
 - Mediastinal mass
 - Pleural effusion
 - Pneumothorax
- Fullt integretert i Sectra, resultater kommer opp i eget “vindu” som et bilde



NEGATIV

1

ANALYSERT



1

MOTTATT

MOTTOK DU ALLE
BILDENE?

Vær oppmerksom på at
resultatene kan være
unøyaktige dersom Gleamer
ikke mottok alle bildene fra
studien



CHESTVIEW

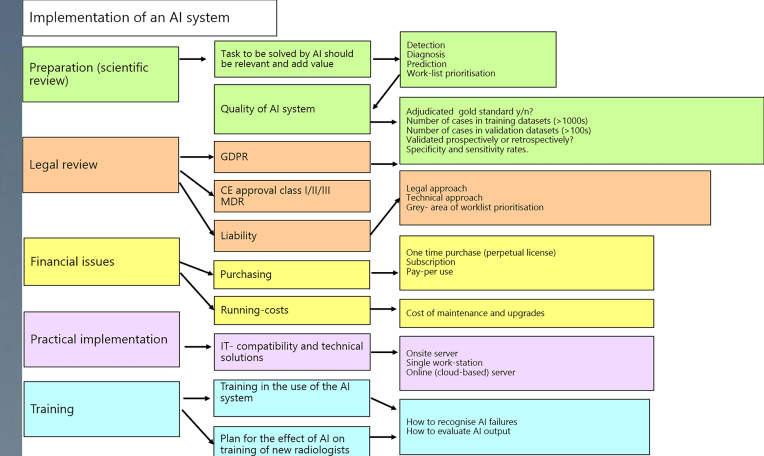
ALVEOLART SYNDROM
NODUL
OPPFYLNINGER I MEDIASTINUM

PLEURAL EFFUSJON
PNEUMOTORAKS

NEI

Resultater

- Valgte en prospektiv validering
- Vanskelig å fylle ut skjema og gjøre jobben ila sommeren
- Ble ikke gjort riktig utfylt for korrekt utregning av spesifisitet/sensitivitet
- Kun stående front/ front og side:
 - 71/98 samsvar (ekte positive/negative)
 - 27/98 ikke samsvar (falsk positive 8stk, falsk negative 13 stk).
 - Og 24 stk hadde diagnoser som ikke var en del av de 5.



Resultater

- Valgte en prospektiv validering
- Vanskelig å fylle ut skjema og gjøre jobben ila sommeren
- Ble ikke gjort riktig utfylt for korrekt utregning av spesifisitet/sensitivitet
- På stående front/ front og side og liggende front:
 - 56/96 samsvar (ekte positive/negative)
- Og vi hadde ca 30% liggende bilder

*
*

Hva kan vi gjenbruke eller dele – og hvilke vurderinger må vi uansett gjøre lokalt?

- Vi trenger flere diagnoser ved neste test, sammenligning med gamle undersøkelser og forliggende bilder også (finnes ikke per dags dato).
- Skal gjøre et endelig arbeid og fullstendig validering etter hvert – deler da gjerne sensitivitet/spesifisitet etc.
- Lokalt- alle må se på behovene for sine egne avdelinger

Har det betydning om man er først ute eller følger etter?

- Ikke nødvendigvis -fordeler og ulemper.
- Fikk erfare «automation bias» på kroppen!
- DET VAR SKREMMENDE!

*
*

Har det betydning om man er først ute eller følger etter?

- Ikke nødvendigvis -fordeler og ulemper.
- Fikk erfare «automation bias» på kroppen!

European Radiology (2023) 33:8263–8269
<https://doi.org/10.1007/s00330-023-09747-1>

European R
ESIR EL
O

CHEST

Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography

Michael H. Bernstein^{1,2,3} · Michael K. Atalay^{1,3} · Elizabeth H. Dibble¹ · Aaron W. P. Maxwell^{1,3} · Adib R. K. Saurabh Agarwal¹ · Robert C. Ward¹ · Terrance T. Healey¹ · Grayson L. Baird^{1,2,3}

Clinical relevance statement When AI is wrong, radiologists make more errors than they would have without AI. Based on human factors psychology, our manuscript provides evidence for two AI implementation strategies that reduce the deleterious effects of incorrect AI.

Key Points

- *When AI provided incorrect results, false negative and false positive rates among the radiologists increased.*
- *False positives decreased when AI results were deleted, versus kept, in the patient's record.*
- *False negatives and false positives decreased when AI visually outlined the region of suspicion.*



ELSEVIER

Contents lists available at ScienceDirect

Journal of Safety Science and Resilience

journal homepage: www.keaipublishing.com/en/journals/journal-of-safety-science-and-resilience/

Review

Exploring the risks of automation bias in healthcare artificial intelligence applications: A Bowtie analysis

Moustafa Abdelwanis, Hamdan Khalaf Alarafati, Maram Muhanad Saleh Tammam, Mecit Can Emre Simsekler *

Department of Management Science and Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

M. Abdelwanis et al.

Journal of Safety Science and Resilience 5 (2024) 460–469

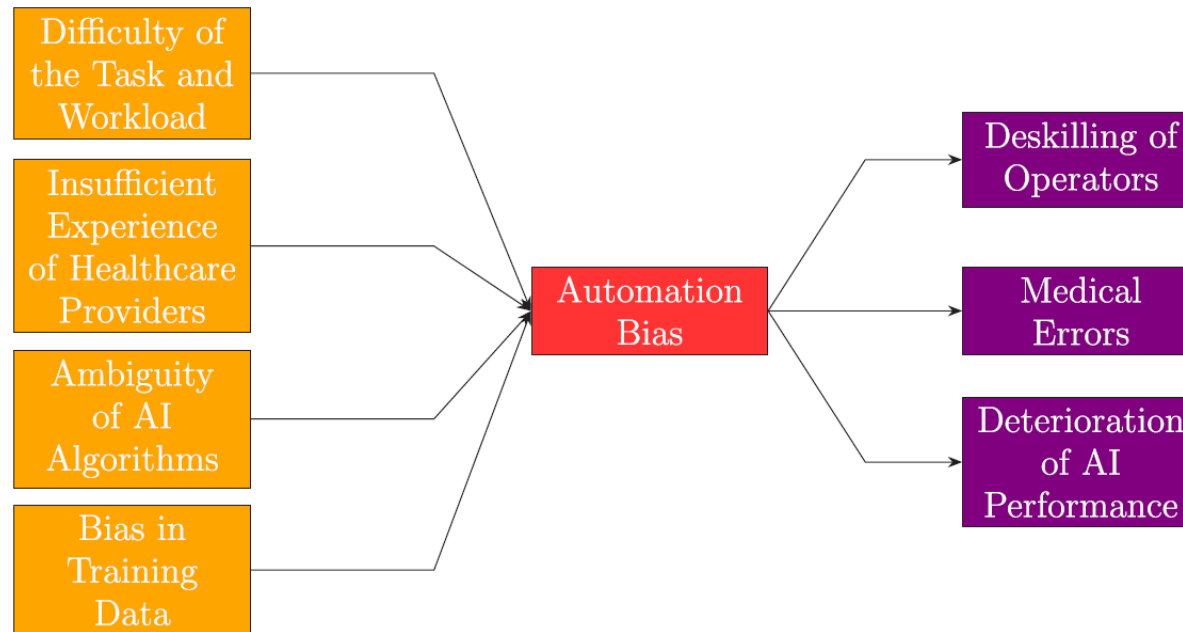
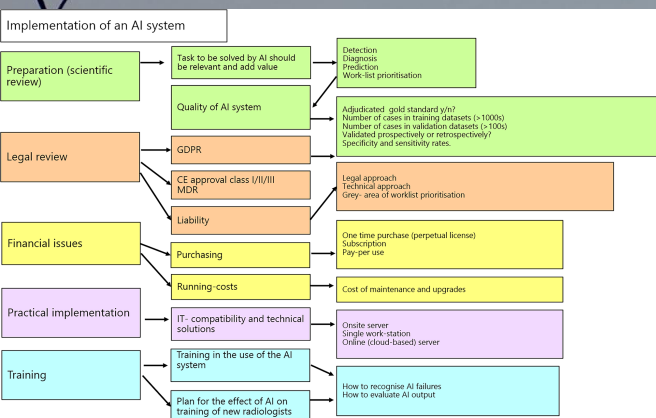


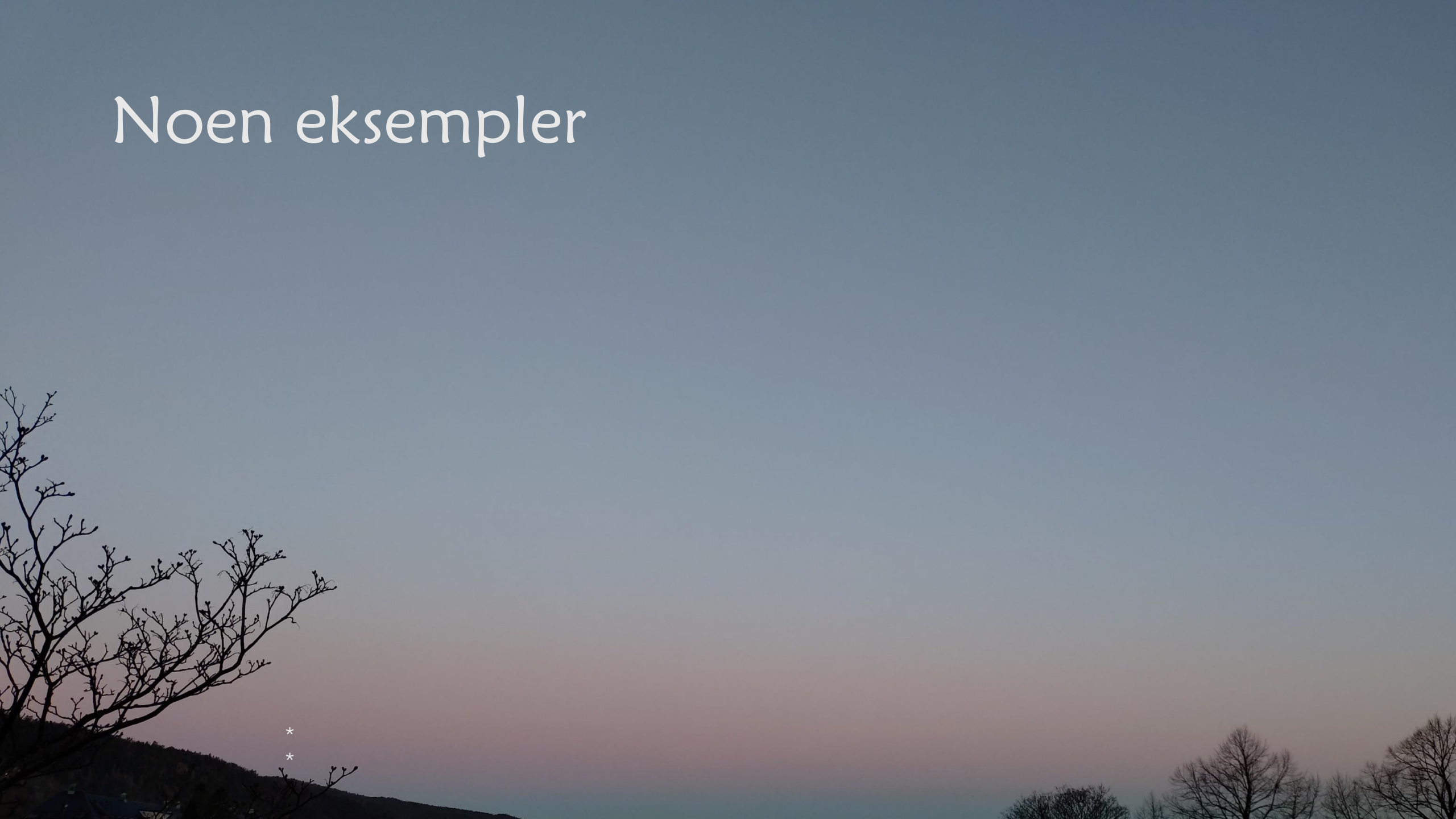
Fig. 1. Causes (orange boxes) and consequences (purple boxes) of Automation Bias.

Hvordan har dere tenkt å gjøre løpende kvalitetssikring, for eksempel løpende validering, oppdage modell-drift, menneskelig overblikk?

- Vi må gjøre årlige revisjoner av AI'en, og sjekke at sensitivitet/spesifisitet ikke faller.
- Sette av tid for radiologer å sjekke/permisjoner
- Alle kliniske bilder skal beskrives av radiologer uansett
- MEN- hvordan skal vi sikre at yngre radiologer blir og nok til å skjønne når AI feiler?
- Hvor er den (nasjonale) planen??



Noen eksempler



*
*





POSITIV

1

ANALYSERT

1

MOTTATT

MOTTOK DU ALLE
BILDENE?

Vær oppmerksom på at
resultatene kan være
unøyaktige dersom Gleamer
ikke mottok alle bildene fra
studien



CHESTVIEW

PNEUMOTORAKS

JA

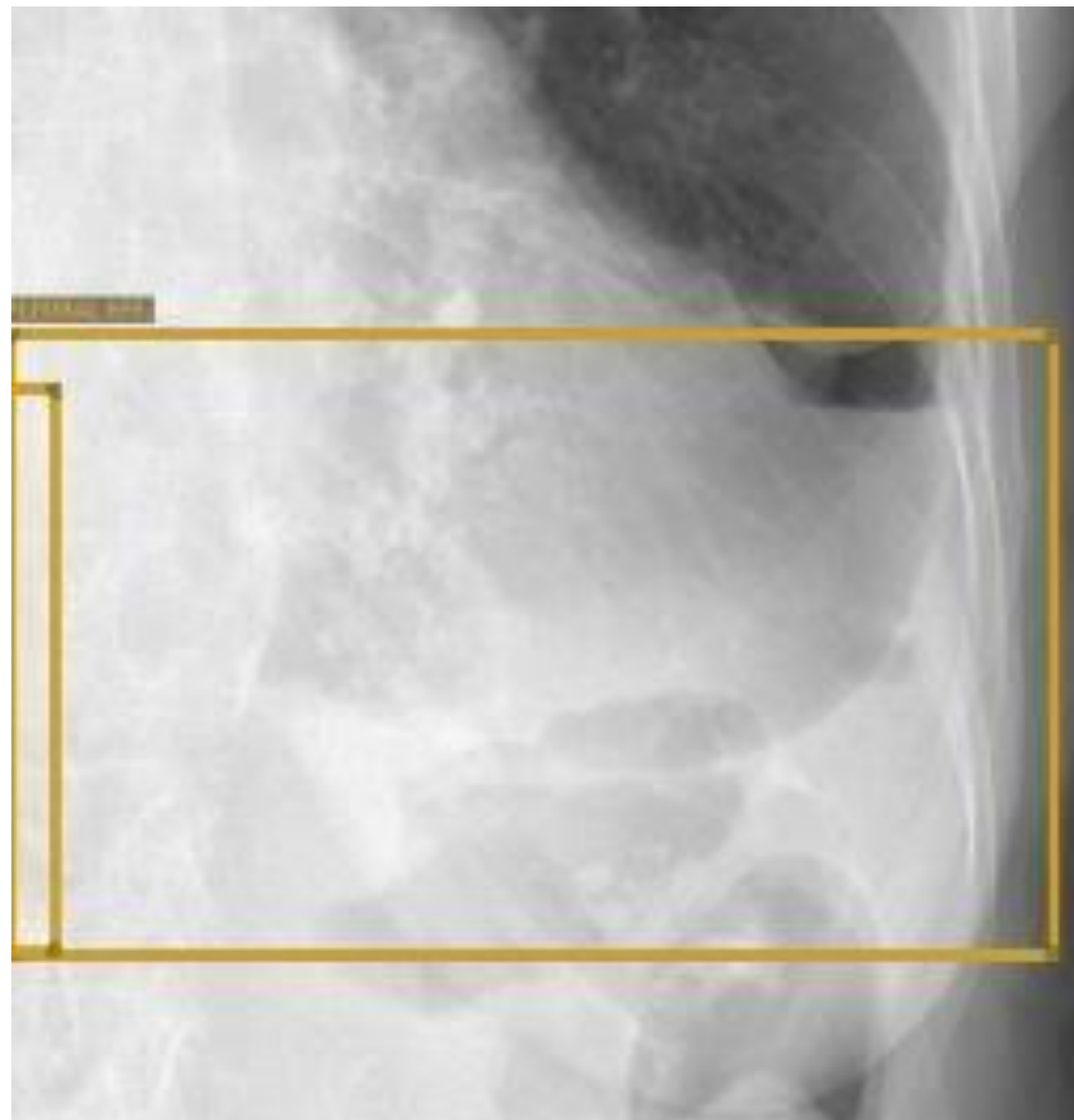
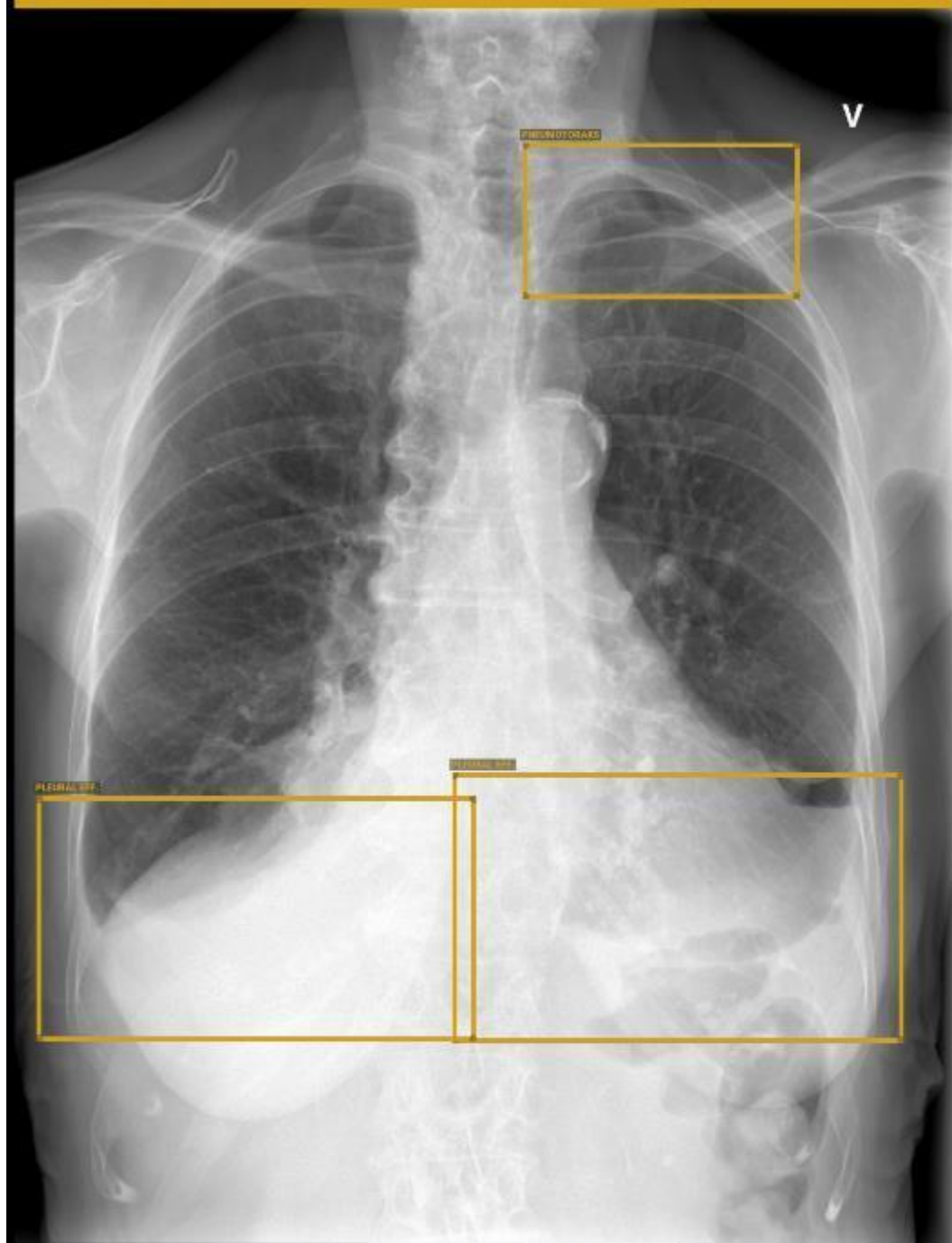
PLEURAL EFFUSJON

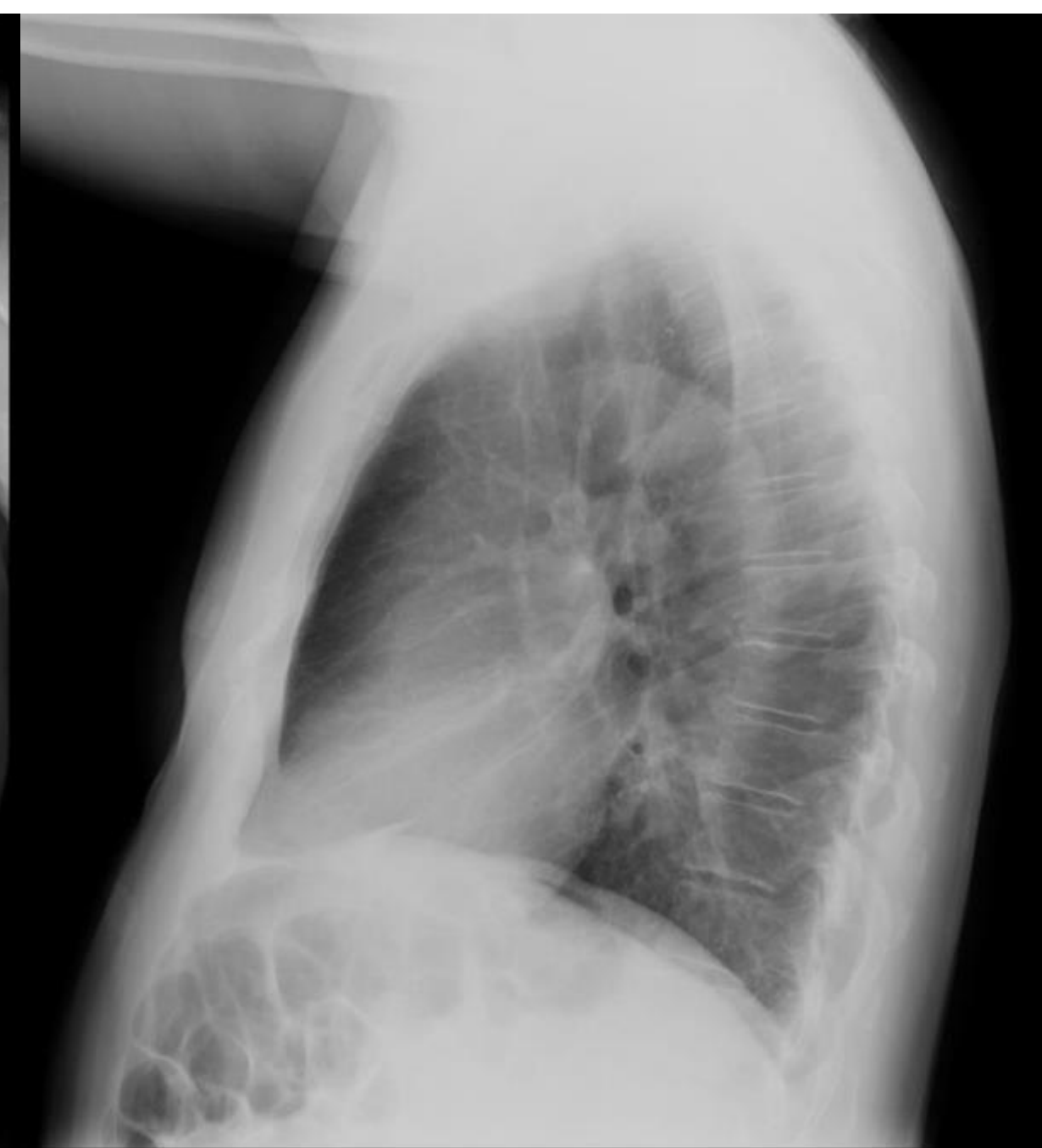
JA

ALVEOLART SYNDROM
NODUL

OPPFYLNINGER I MEDIASTINUM

NEI







CHESTVIEW

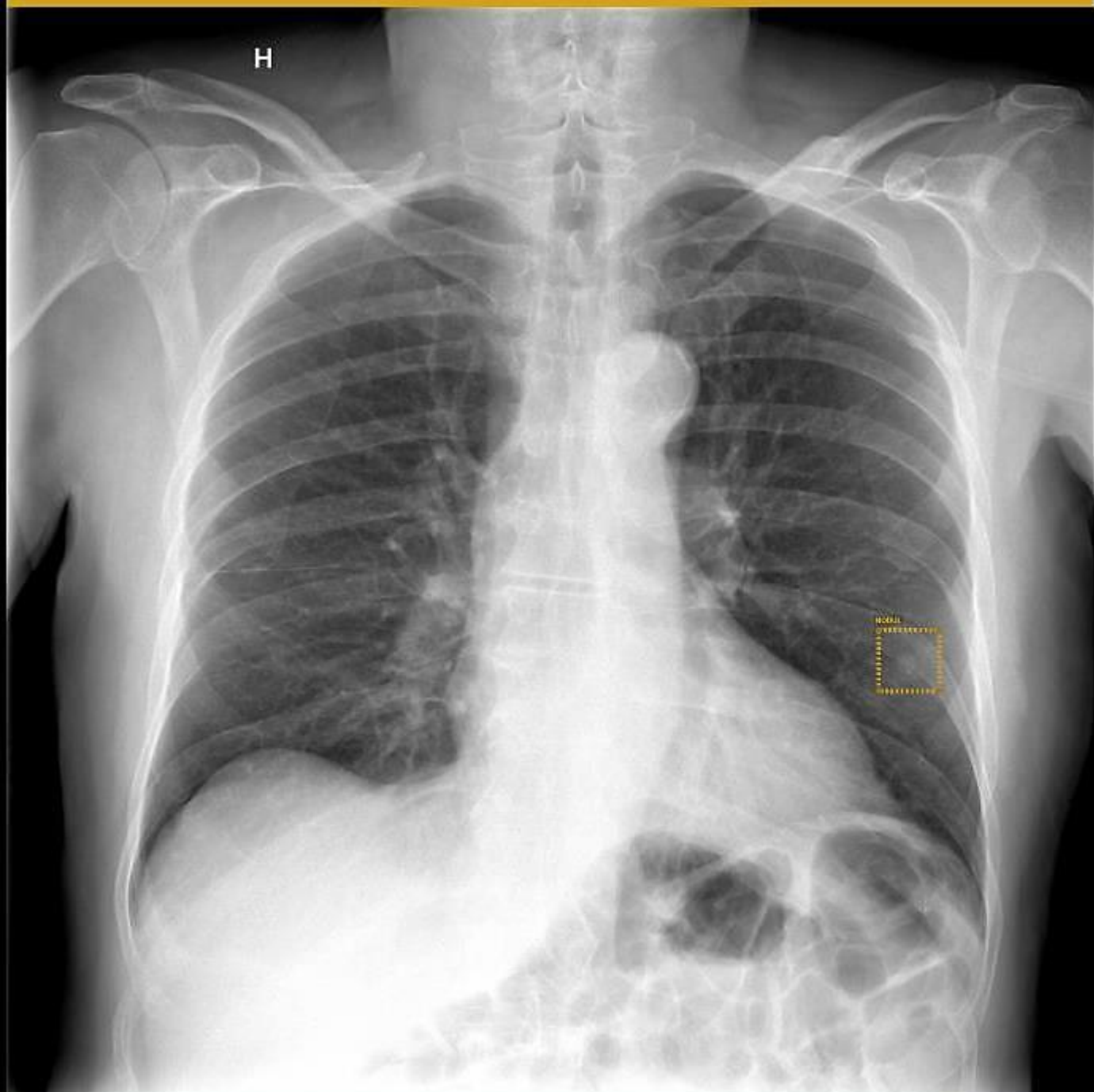
NODUL

TVIL

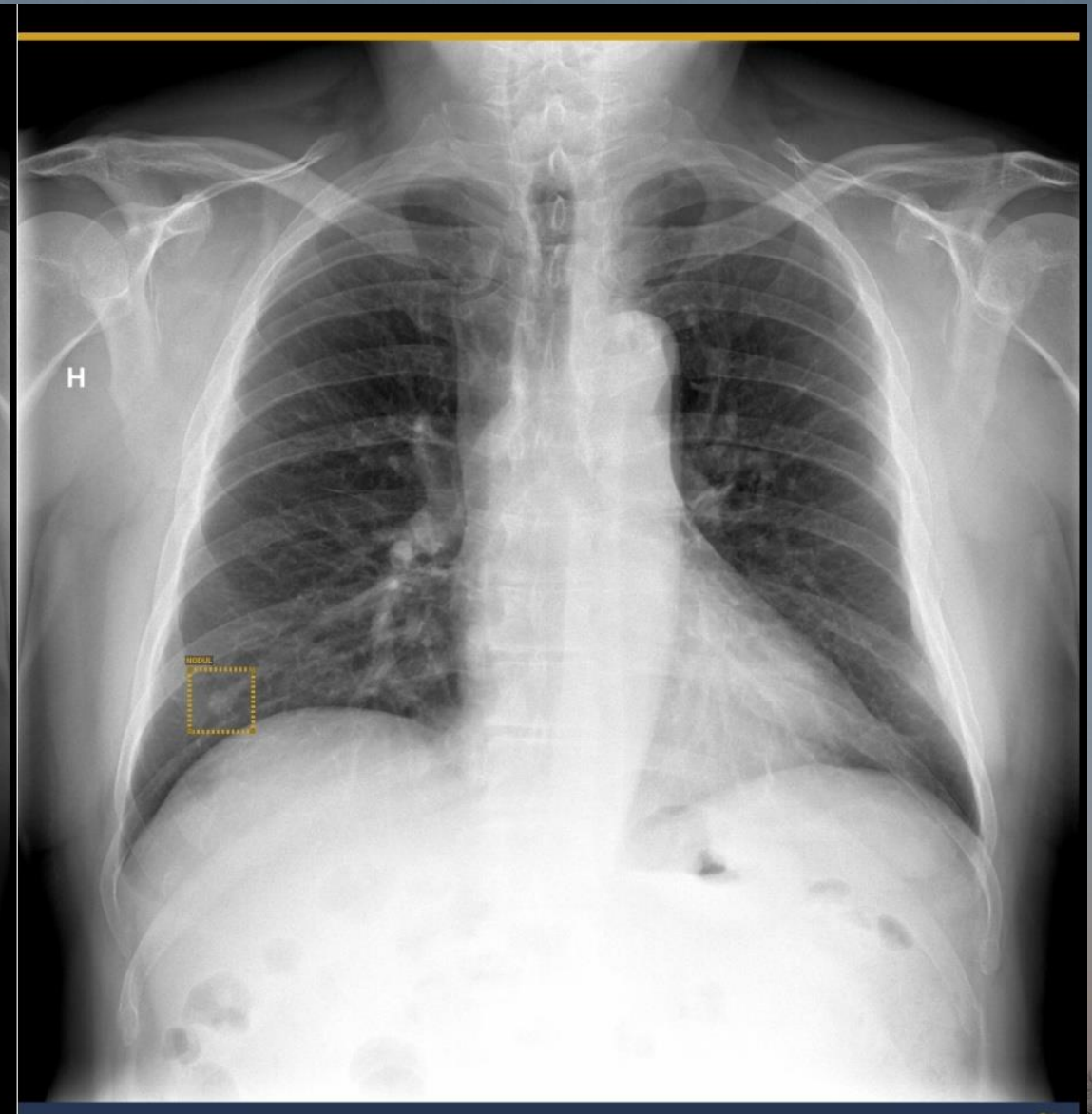
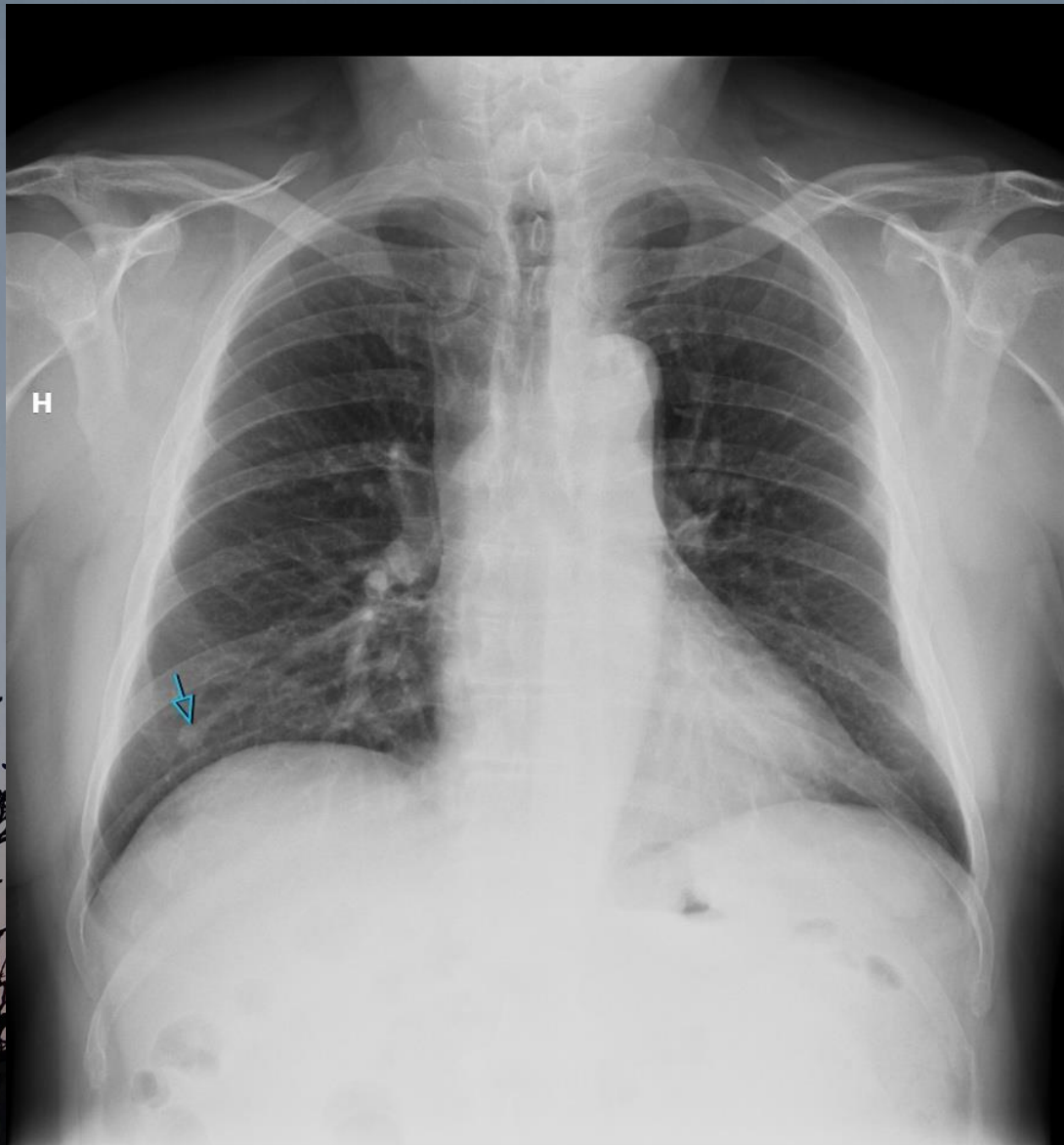
ALVEOLART SYNDROM
OPPFYLNINGER I MEDIASTINUM

PLEURAL EFFUSJON
PNEUMOTORAKS

NEI

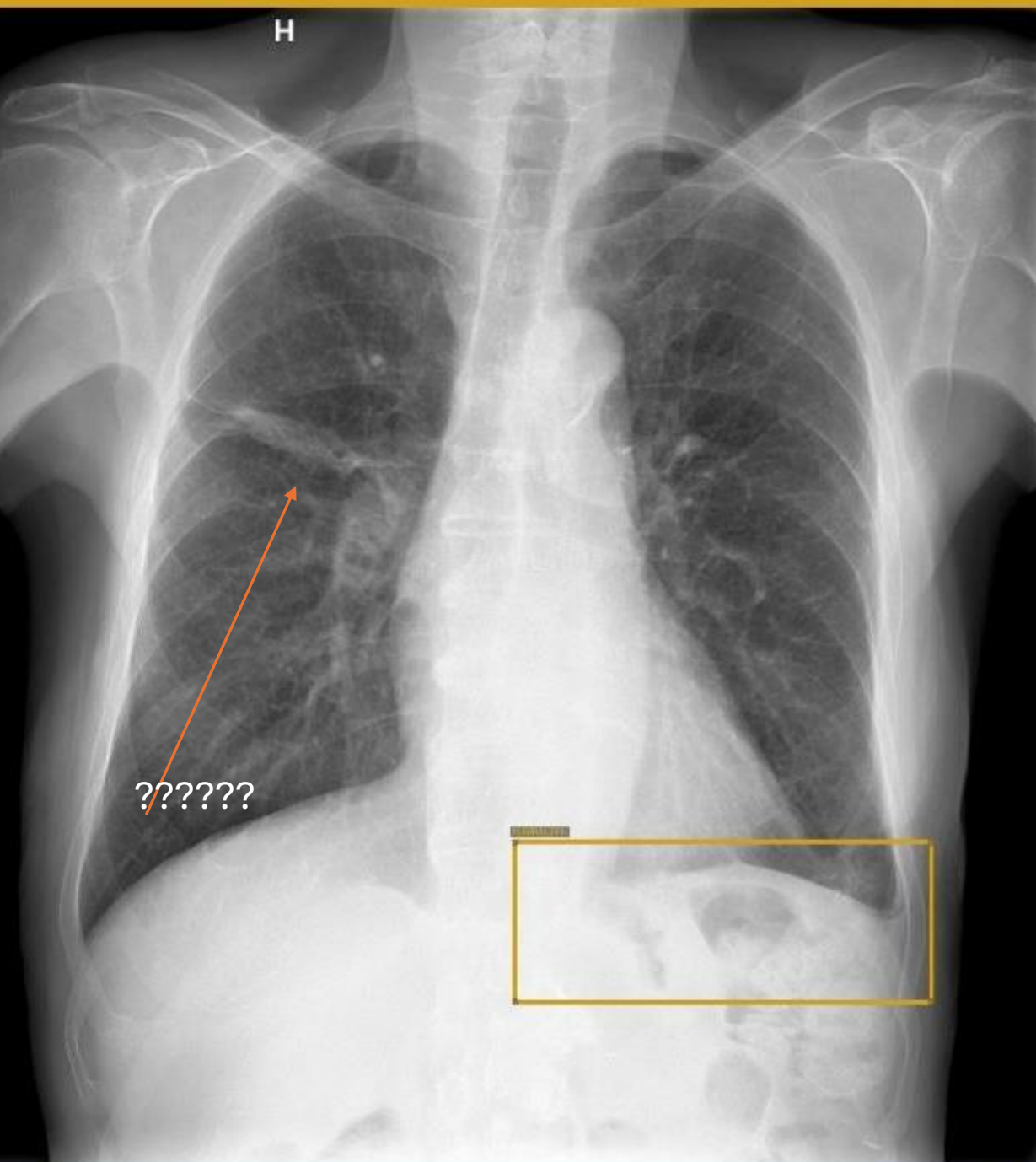


Automation bias



H





 **POSITIV**

2
ANALYSERT

2
MOTTATT

MOTTOK DU ALLE
BILDENE?

Vær oppmerksom på at
resultatene kan være
uhøykålige dersom Gleam
ikke mottok alle bildene
studen



CHESTVIEW

PLEURAL EFFUSJON

JA

ALVEOLART SYNDROM
NODUL

OPPFYLNINGER I MEDIASTINUM
PNEUMOTORAKS

NEI



POSITIV

2
ANALYSER

2
NOTIS

NOTIS DEL ALC
BLISS

Vor dem Einsatz der
Bliss-App kann es
zu Verzögerungen bei
der Analyse kommen.



CHESTVIEW

PLEURAL EFFUSION

JA

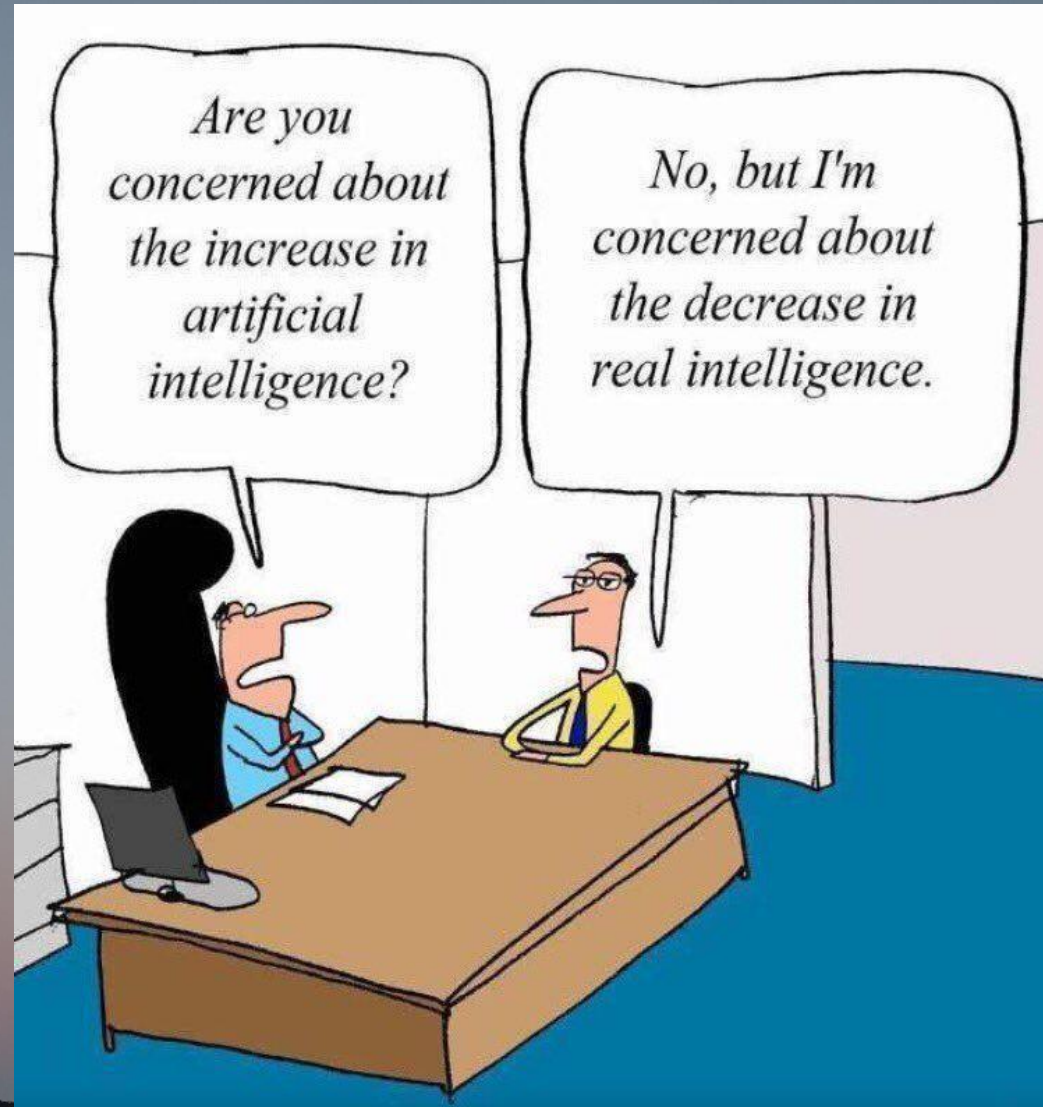
ALVEOLÄR SYNDROM
NODUL

DEPLETIONSBILDE IN MEDIASTINUM
PNEUMOTORAKS

NEI



Takk for at vi ble invitert!



Validering KI-verktøy bildediagnostikk Helse Midt-Norge

Workshop Helsedirektoratet 29.4.2025

Ida Fallmyr Jørgensen

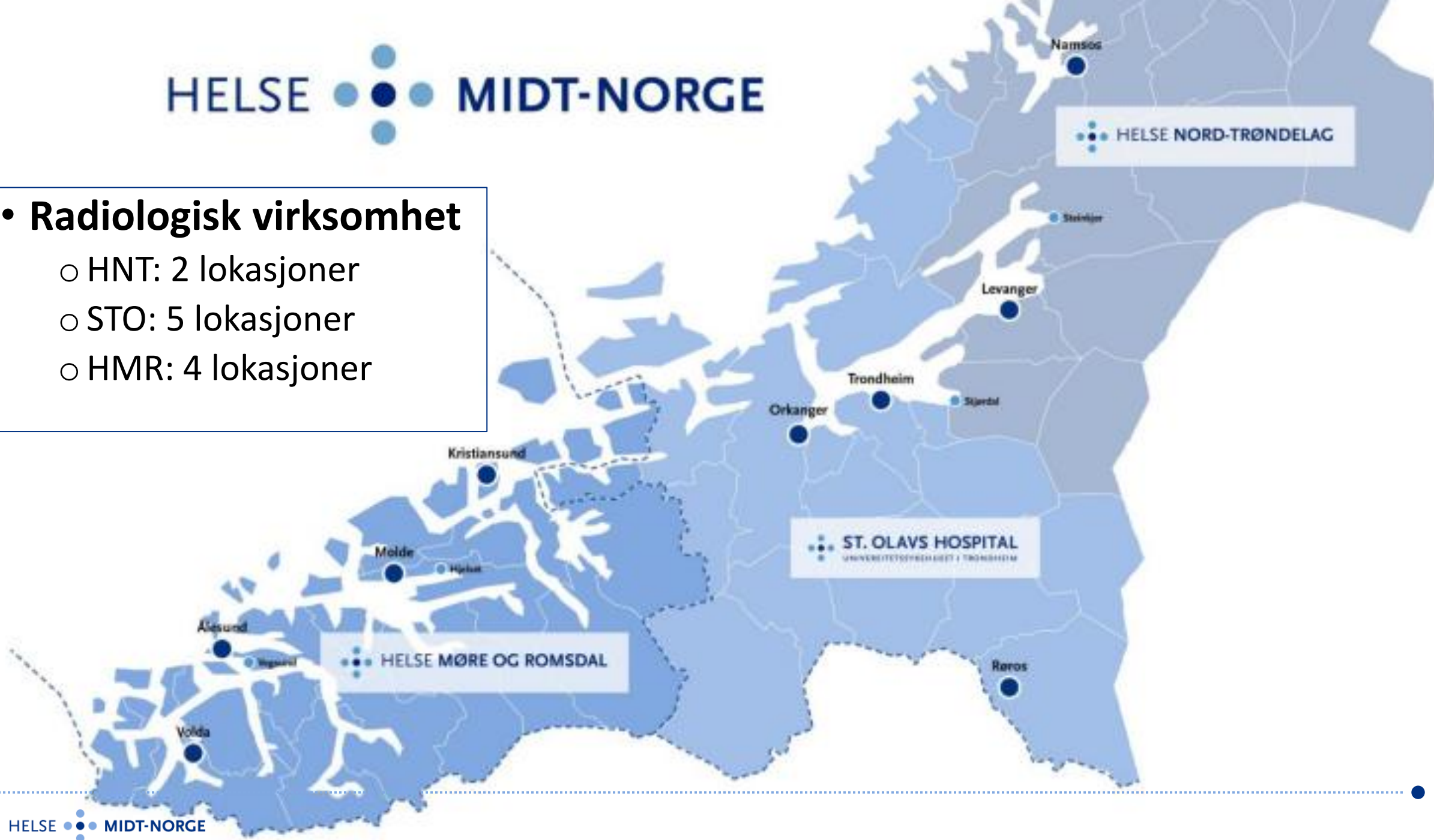
Leder regionalt KI – team bildediagnostikk - Helse Midt Norge

Rådgiver Klinikk for diagnostikk - Helse Møre og Romsdal

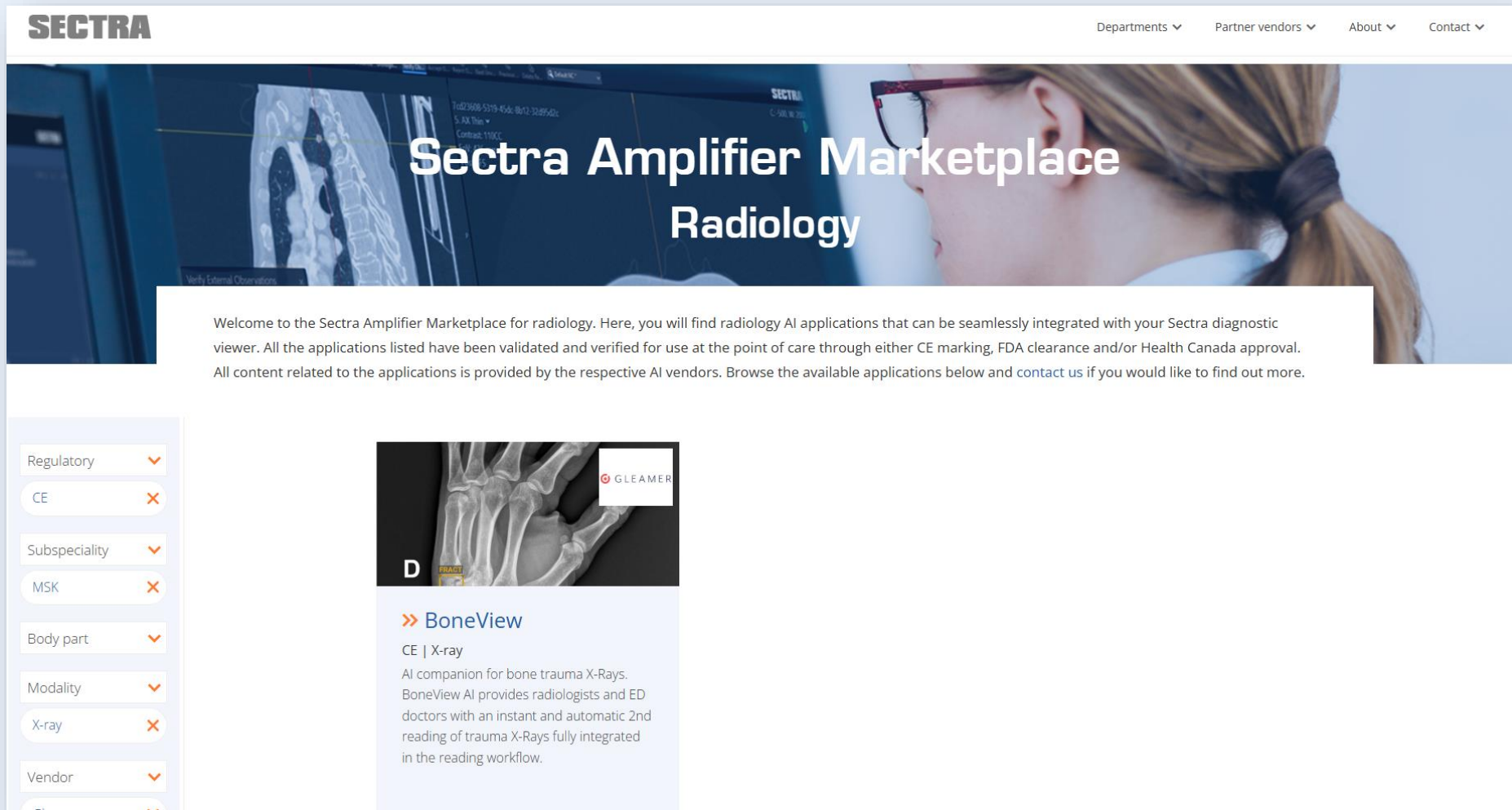
Ida.Fallmyr.Jorgensen@helse-mr.no

- **Radiologisk virksomhet**

- HNT: 2 lokasjoner
- STO: 5 lokasjoner
- HMR: 4 lokasjoner



Vår erfaring er basert på følgende:



SECTRA Departments ▾ Partner vendors ▾ About ▾ Contact ▾

Sectra Amplifier Marketplace Radiology

Welcome to the Sectra Amplifier Marketplace for radiology. Here, you will find radiology AI applications that can be seamlessly integrated with your Sectra diagnostic viewer. All the applications listed have been validated and verified for use at the point of care through either CE marking, FDA clearance and/or Health Canada approval. All content related to the applications is provided by the respective AI vendors. Browse the available applications below and [contact us](#) if you would like to find out more.

Regulatory ▾
CE ✕

Subspecialty ▾
MSK ✕

Body part ▾

Modality ▾
X-ray ✕

Vendor ▾

D **GLEAMER**

>> BoneView

CE | X-ray

AI companion for bone trauma X-Rays. BoneView AI provides radiologists and ED doctors with an instant and automatic 2nd reading of trauma X-Rays fully integrated in the reading workflow.

Involverte i valideringen

- Tverrfaglig sammensetning
- Regional sammensetning



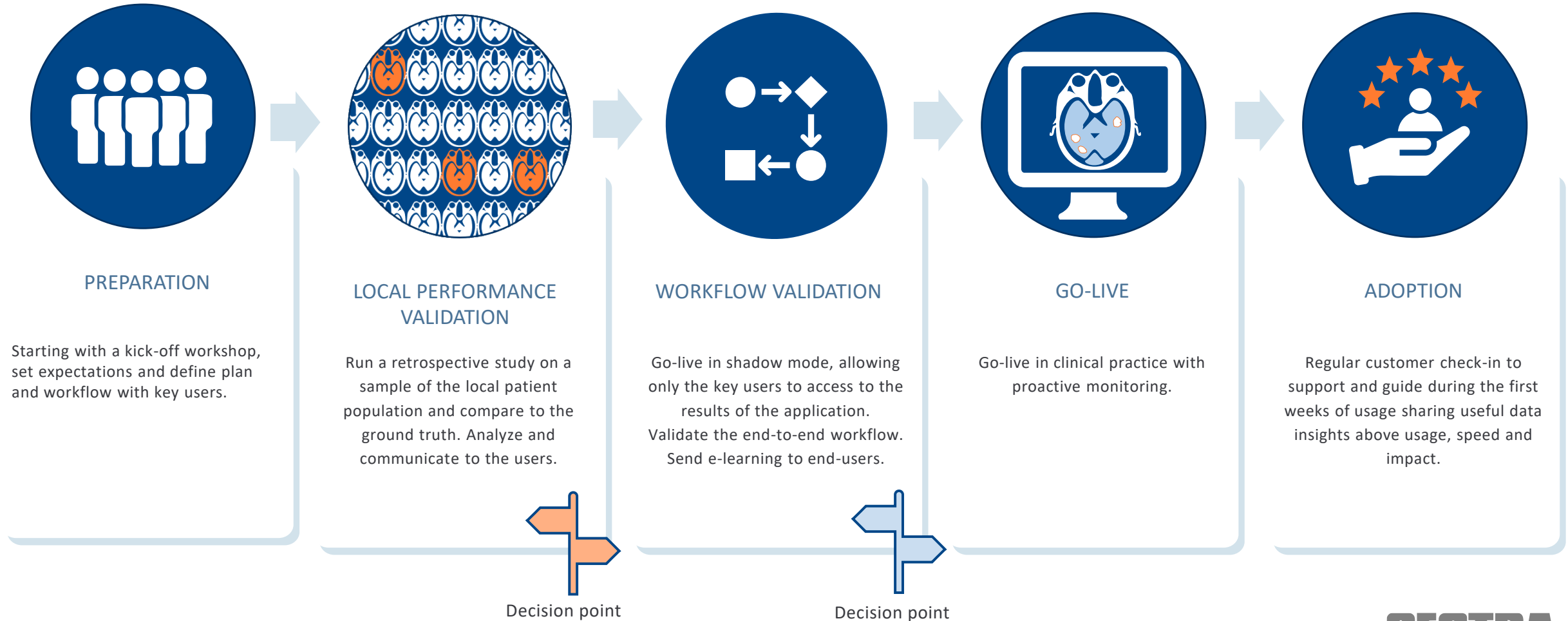
Radiograf

Radiolog

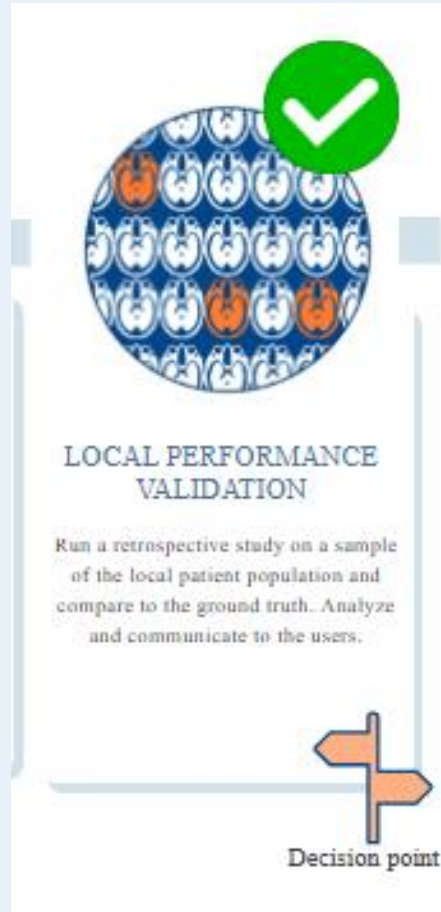
Fysiker

PACS/RIS radiograf

Deployment process BoneView with Amplifier Services

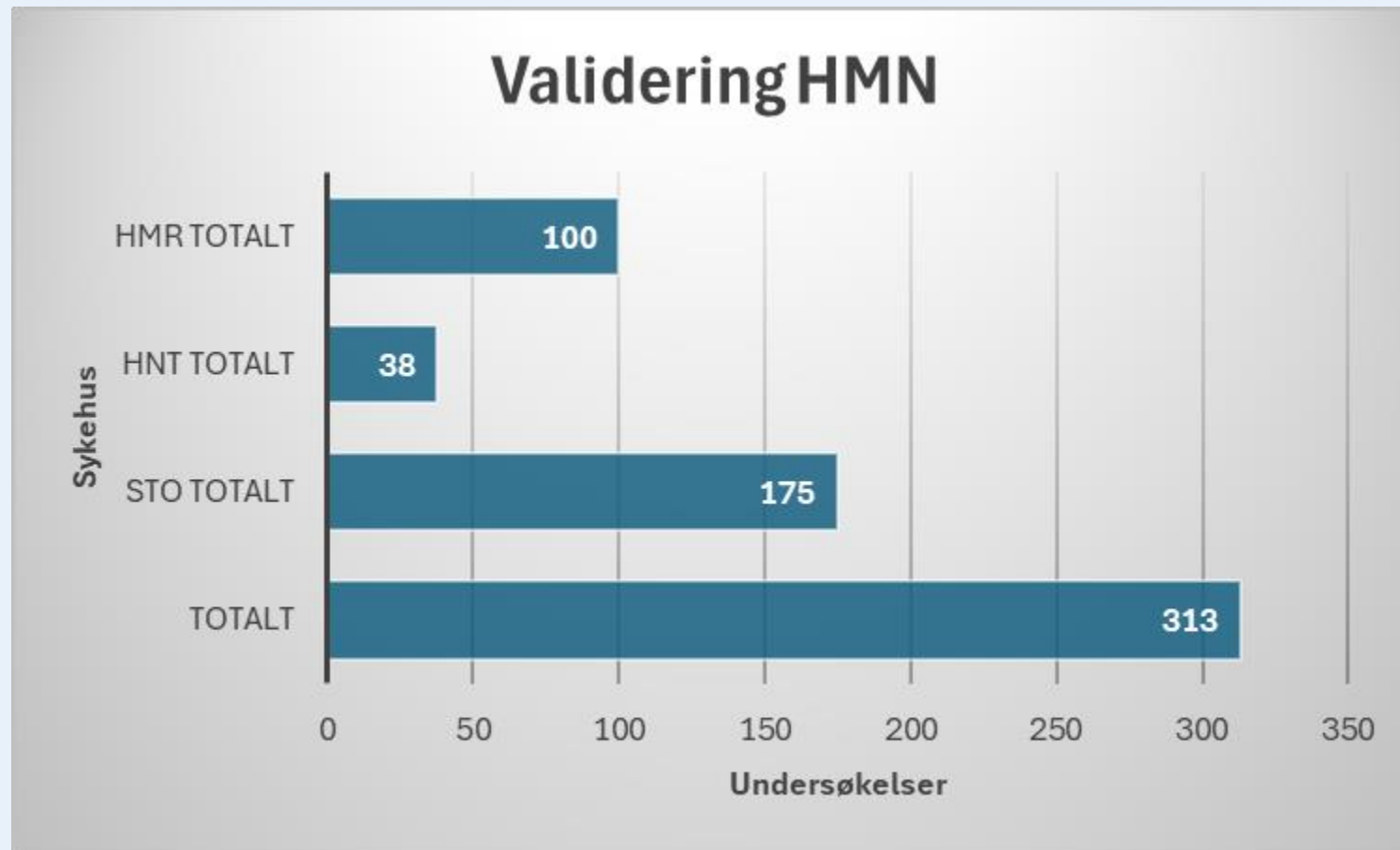


Gjenbruk av validering?



- Diskusjonspunkt - gjennomføre eller ikke?
 - Læringsprosess - ga nyttig kunnskap
 - Gjennomgang av FN/FP - får trygghet
 - Skape tillit
- Det at andre har gjennomført bidro til en raskere og mindre omfattende validering

Utvalg til validering HMN



Conflicts

38 out of 285

Discrepancy between the ground truth and BoneView

Negative Predictive Value

89,86%

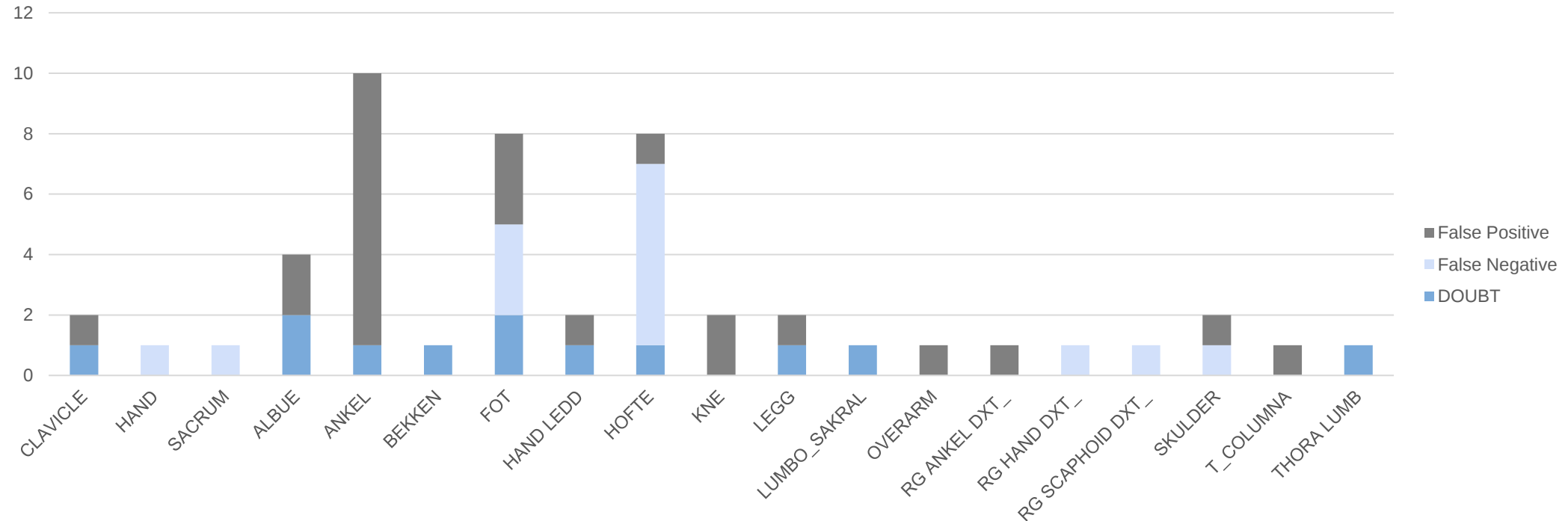
If BoneView says there is no fracture, it is right 89 times out of 100.

Doubts

4,21%

Of all cases that are read by the AI, you might be confronted for 4,21% of them.

Analysis per Body Part



88,8% Sensitivity

*If there are 100 positive fractures,
BoneView detects this % of them.*

83,8% Specificity

*If there are 100 negative fractures,
BoneView detects this % of them.*



Gjennomgang av konfliktcaser

- Radiologer gikk gjennom liste med FP/FN vs GT
- Korrigerte/trakk ut caser som ikke var aktuell

Conflicts

25 out of 297

Discrepancy between the ground truth and BoneView

Negative Predictive Value

93,01%

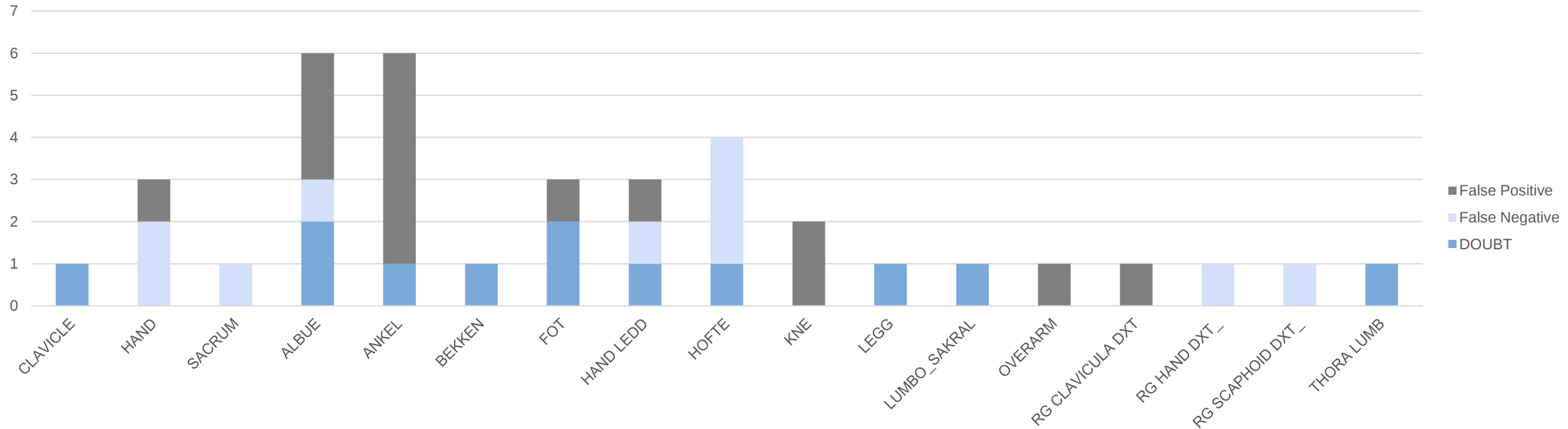
If BoneView says there is no fracture, it is right 93,01% times out of 100.

Doubts

4,04%

Of all cases that are read by the AI, you might be confronted for 4,04% of them.

Analysis per Body Part



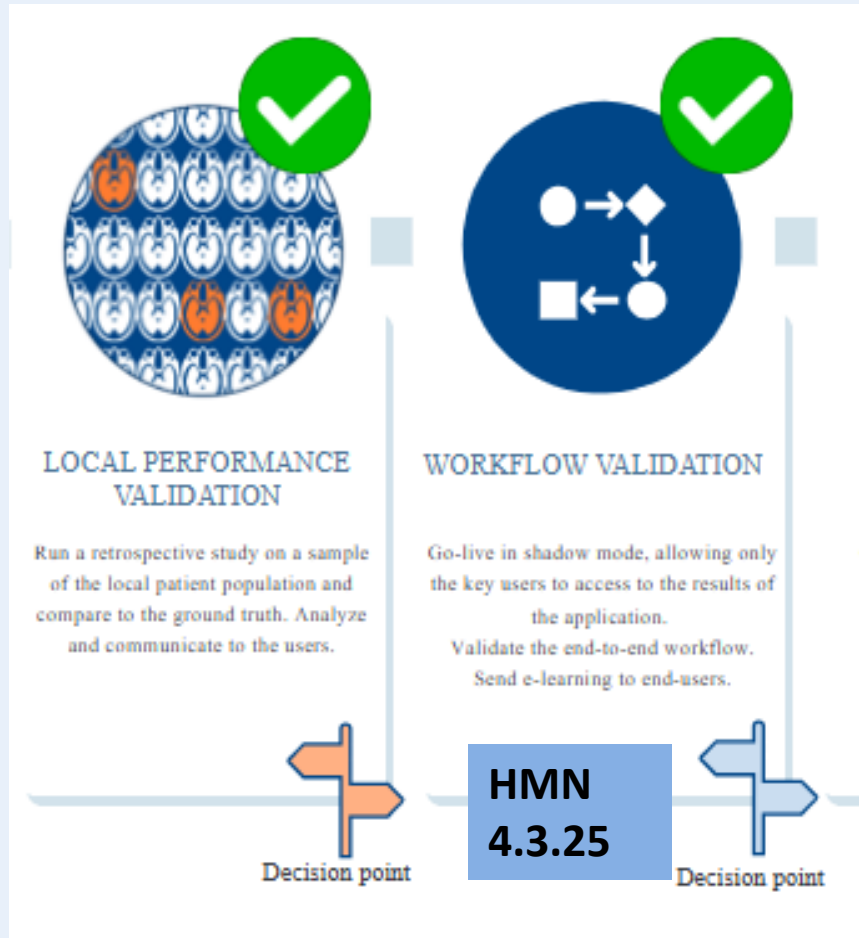
92,7% Sensitivity

If there are 100 positive fractures,
BoneView detects this % of them.

89,9% Specificity

If there are 100 negative fractures,
BoneView detects this % of them.

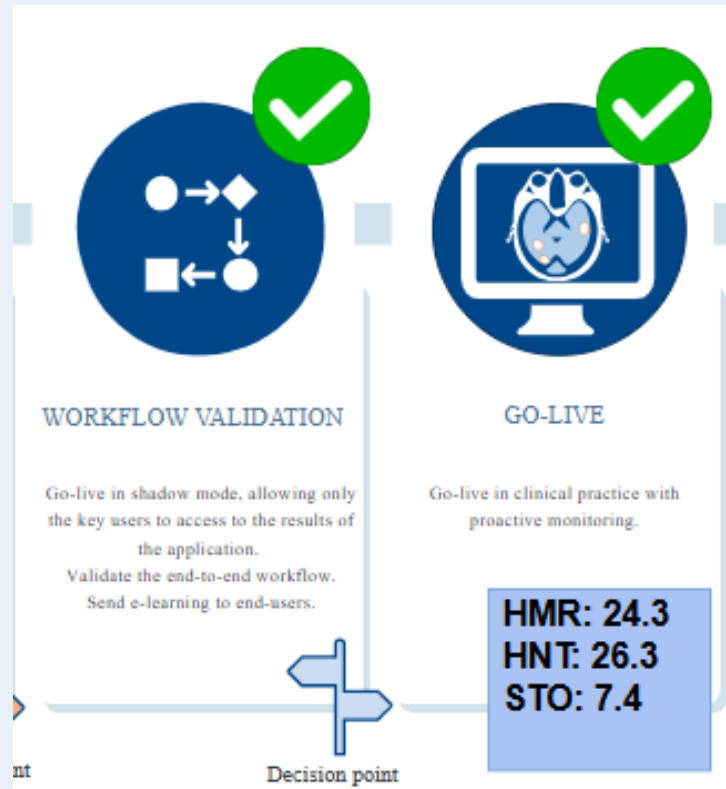
Teknisk validering



- Neste steg: teknisk validering med testing av KI verktøy i klinisk miljø
- Utvalgte brukere fikk tilgang
- Jevnlig korrigering i perioden i samarbeid med leverandør
- **Viktig punkt for å lære mer om funksjon og "hvordan det fungerer hos oss"!**



Klinisk implementering



- Beslutning om klinisk implementering
- Tett dialog med SECTRA
 - Fortløpende feilretting
- Rask klinisk implementering for alle sykehus
 - Fordel med å "følge etter"



Kvalitetssikring

Nåværende prosjekt: Bone View

- Logging av FP og FN på egne lister i PACS til oppfølging
- Korrigerer filter ved behov
- Utarbeide statistikk til forbedring
- Barn - se på nøyaktighet i ulike aldre
- Uttak av potensiale for å forsvare videre bruk av KI-verktøyet

Punkter til videre diskusjon

- Nivå og omfang av kvalitetssikring avhenger av hva KI-verktøyet skal brukes til
- Nasjonalt samarbeid - viktig med erfarings- og kunnskapsdeling på tvers
- Oppdateringer av KI-verktøy - behov for ny validering/testing? Ansvar?

Takk for meg!



Photo: Arild Solberg