ESAC European Interdisciplinary Society for Artificial Intelligence in Cancer Research

Sistema Socio Sanitario
Fondazione IRCCS Istituto Nazionale dei Tumori
Regione Lombardia

I3LUNG

Apollo 11
STRENGTH IS UNITY

AI-ON-Lab

Artificial Intelligence and OECI priorities

Sistema Socio Sanitario
Fondazione IRCCS Istituto Nazionale dei Tumori
Regione Lombardia

neare
POLITECNICO MILANO 1863

Arsela Prelaj, MD, PhD,
Fondazione IRCCS Istituto Nazionale Tumori di Milano, Italy
arsela.prelaj@istitutotumori.mi.it

WHERE WE ARE WITH AI?

Different way to build our AI biomarkers

PREDICTIVE AI TOOLS

FOUNDATION MODELS

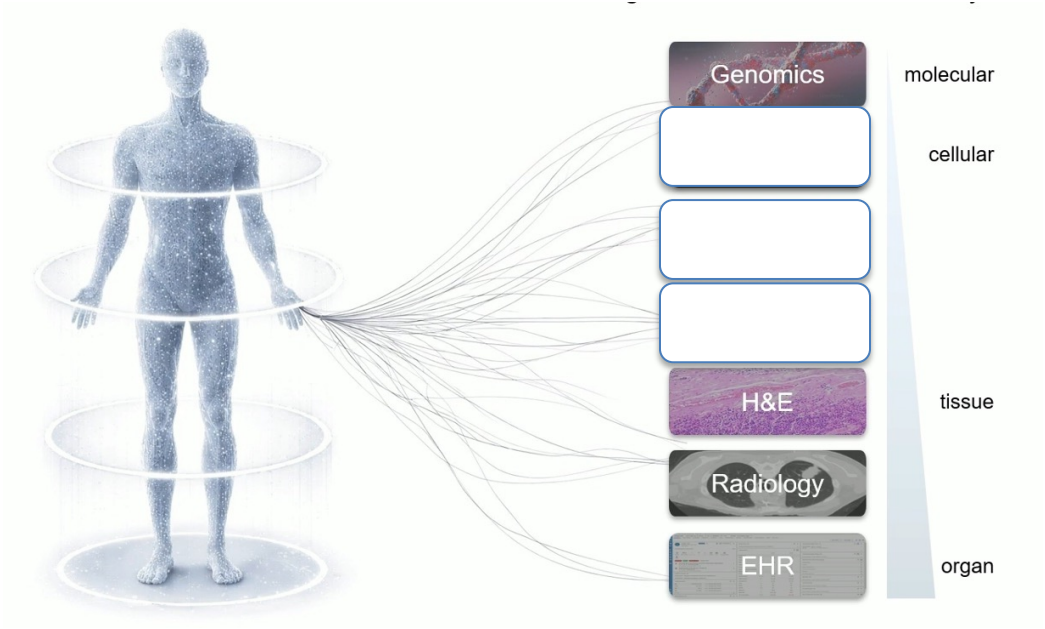
LLM and MULTIMODAL FOUNDATION MODELS

AGENTIC AI (All in ONE MODELS in a Dynamic Fashion)

STATIC MODELS → DYNAMIC MODELS

INTEGRATED CLINICAL BIOMARKERS

THE VISIBLE LAYER



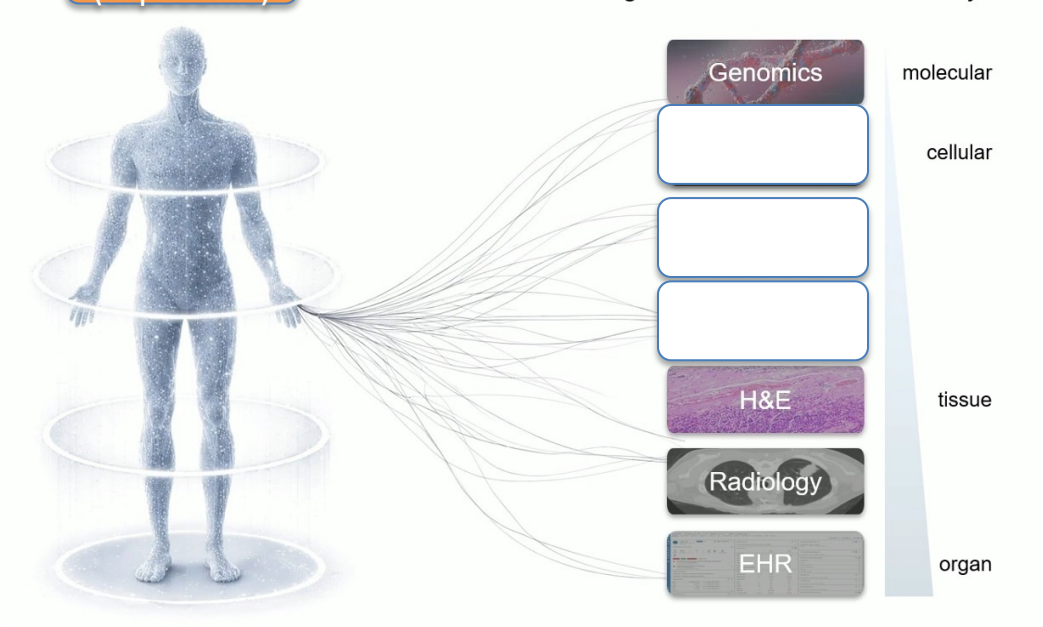
AI Biomarkers as a result of integration, leveraging data from Clinical Practice data

With this *clinical phenotype biomarker* we are able to predict around 80% of the patients

INTEGRATED CLINICAL + EXPOSOME BIOMARKERS

Digitalomics (Exposome)

THE NEW LAYER

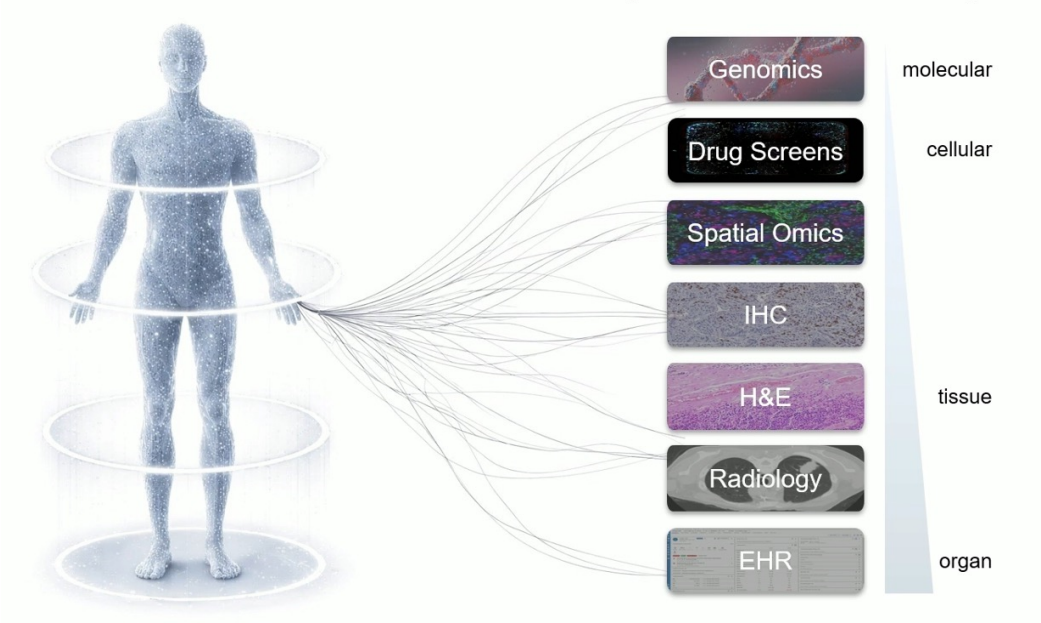


New Biomarkers collected through devices (via wearable or via app), generally ignored before using AI in healthcare

With this *expotype* we are able to further increase the prediction (no estimation is possible at the moment)

DEEP-OMIC BIOMARKERS

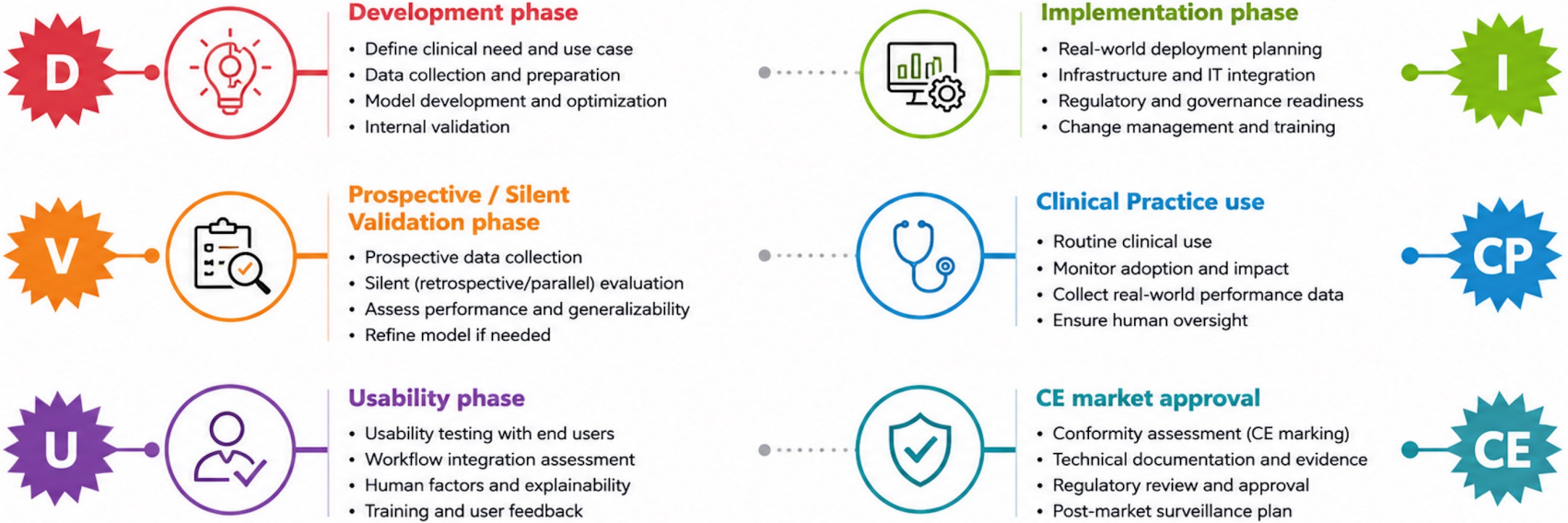
THE HIDDEN LAYER



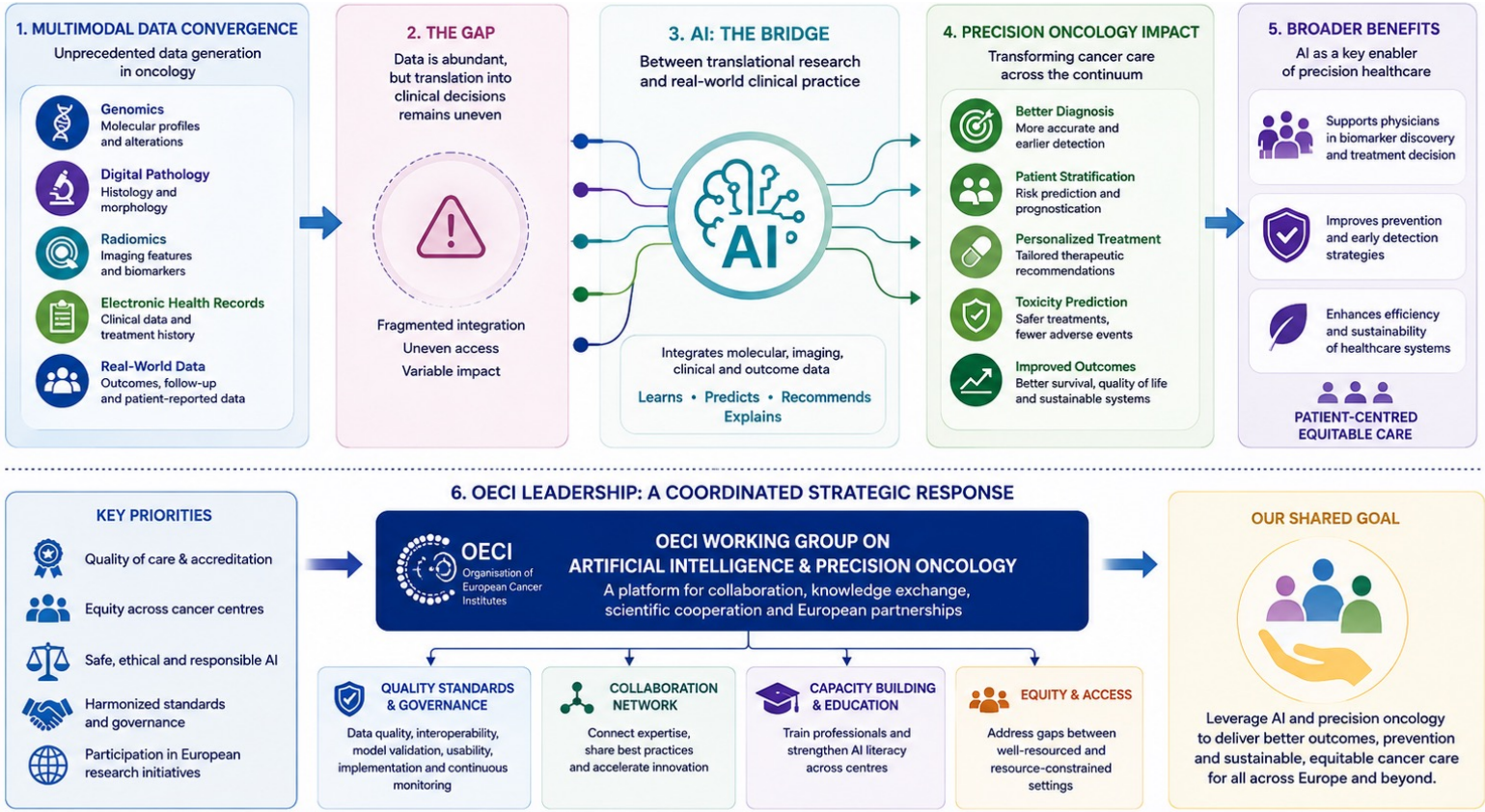
AI Biomarkers as a result of interconnection of different tissue, blood and ever-biosample omics (Not collected per clinical practise)

With this *genotype biomarker* we are able to further increase the prediction (no estimation is possible at the moment)

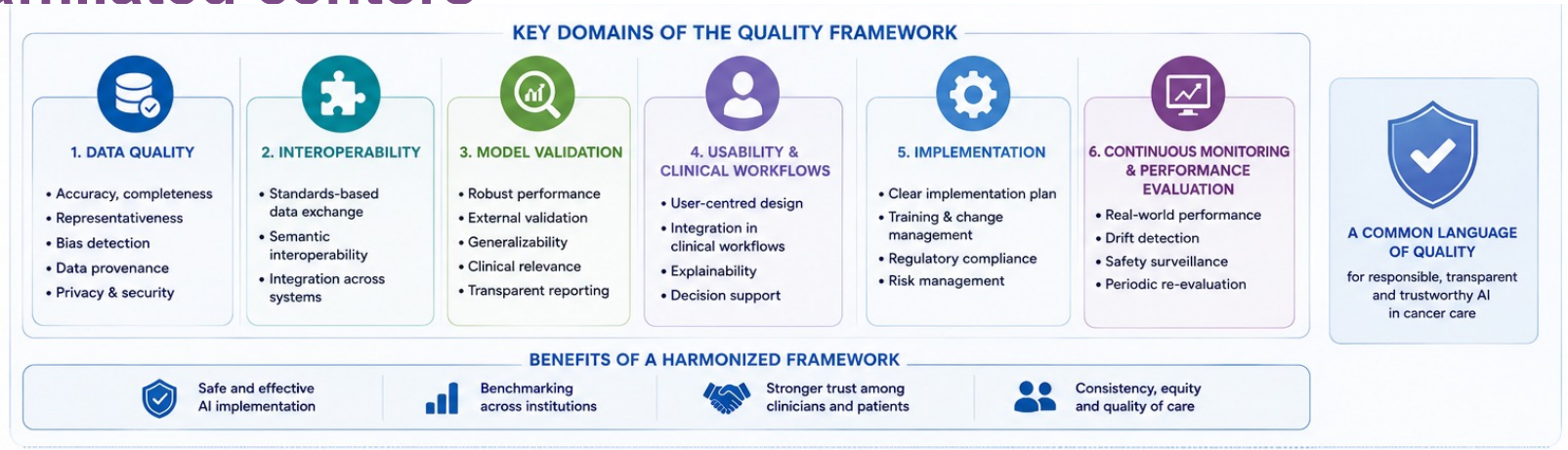
Life cycle of an AI biomarkers: what studies we need?

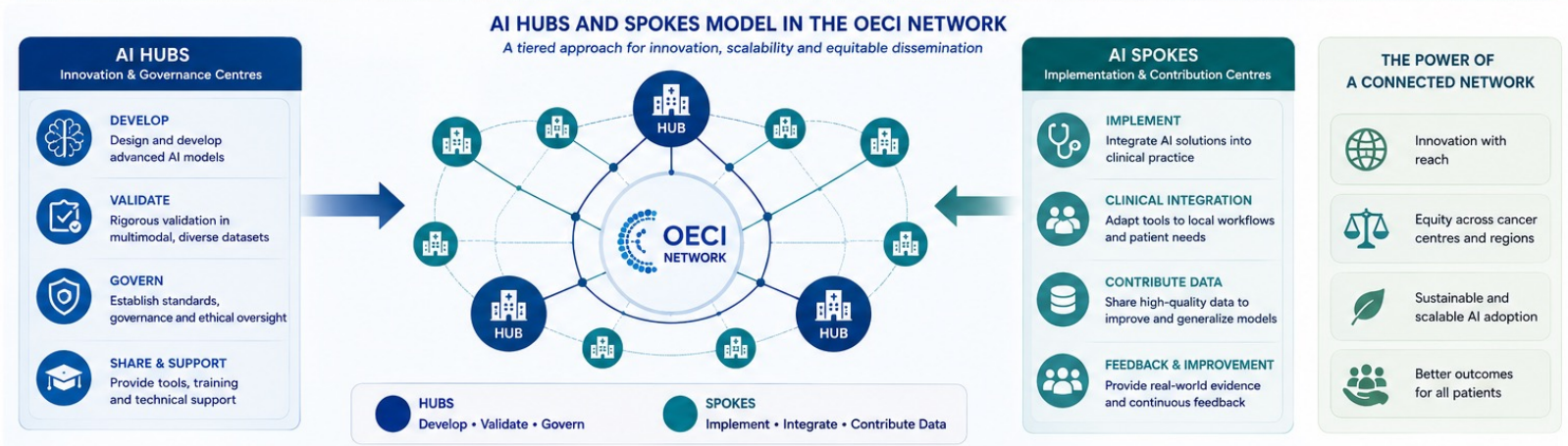


Where can OECI stand within the AI area?



Harmonize Quality standards for AI usage in OECI affiliated centers

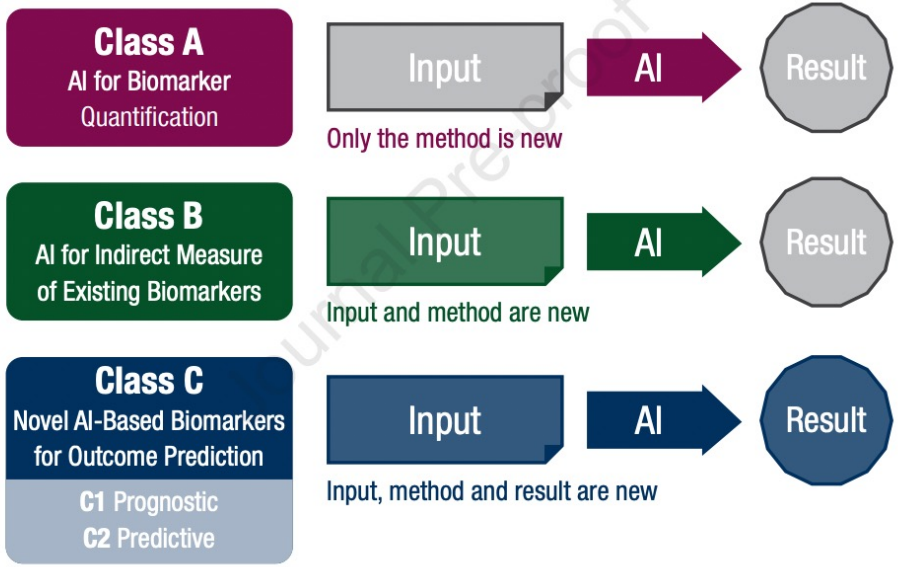




ESMO ANNALS OF ONCOLOGY DRIVING INNOVATION IN ONCOLOGY

ESMO Journals Articles Publish Topics Multimedia About Contact Subscribe

A



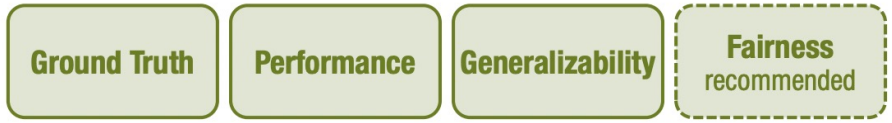
ORIGINAL ARTICLE · Articles in Press, November 17, 2025 · *Open Access* Download Full Issue

ESMO Basic Requirements for AI-based Biomarkers In Oncology (EBAI)

M. Aldea [1,2,3] · M. Salto-Tellez [4,5] · A. Marra [6,7] · R. Umeton [8,9,10,11] · A. Stenzinger [12] · M. Koopman [13] · A. Prelaj [14,15] · K.L. Kehl [3] · S. Gilbert [16] · M-E Leßmann [16] · J. Lipkova [17,18,19] · L. Provenzano [20,21] · F. Meric-Bernstam [22] · S. Halabi [23] · J. Wu [24] · A. Pellat [25] · K.P.M. Suijkerbuijk [26] · B. Besse [1,2] · B. Ryll [27] · C. Marchió [28,29] · M. Crispin-Ortuzar [30] · R. Fehrmann [31] · J. Vibert [32] · D. Ferber [33] · C. Pauli [34] · A. Valachis [35] · Federica Corso [14] · T.J. Brinker [36] · J. Mateo [37] · N. Harbeck [38] · E.C. Winkler [39] · F. Lopez-Rios [40] · R. Perez-Lopez [41] · G. Pentheroudakis [42] · S. Delaloge [1] · C. Benedikt Westphalen [43,44] · J.N. Kather [16,33,45] Show less

Affiliations & Notes · Article Info

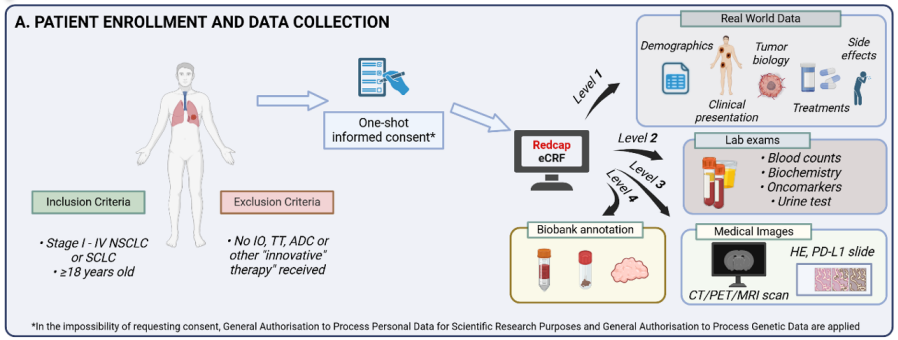
B

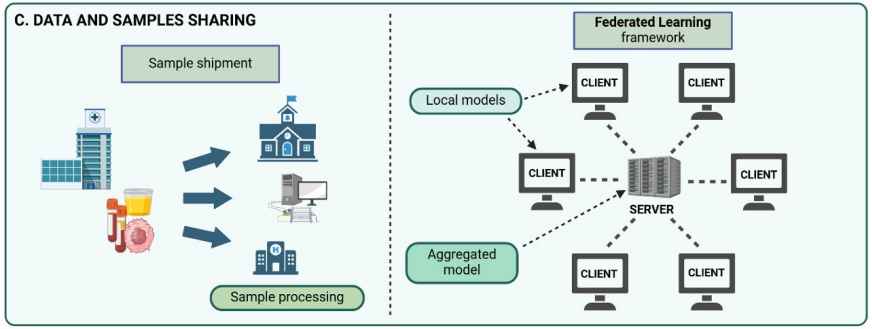


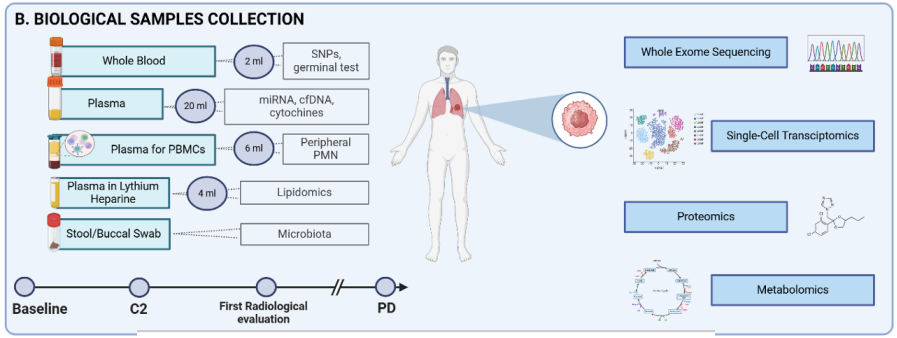
VALIDATION is key

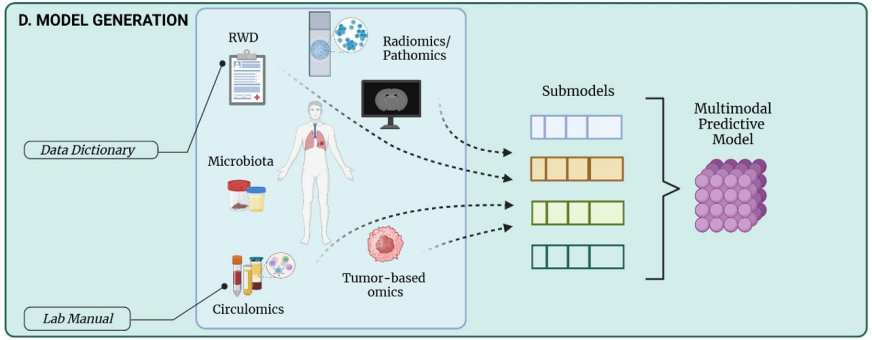
APOLLO11, BIO DATA DRIVEN MODEL

Apollo 11
STRENGTH IS UNITY









npj | precision oncology

Explore content · About the journal · Publish with us

nature › npj precision oncology › articles › article

Article | Open access | Published: 29 January 2026

APOLLO11: a bio-data-driven model for clinical and translational research in lung cancer

Arsela Prelaj, Leonardo Provenzano, Vanja Miskovic, Monica Ganzinelli, Laura Mazzeo, Maria Gemelli,

DATA remain key

Synthetic data for clinical trials

nature reviews cancer

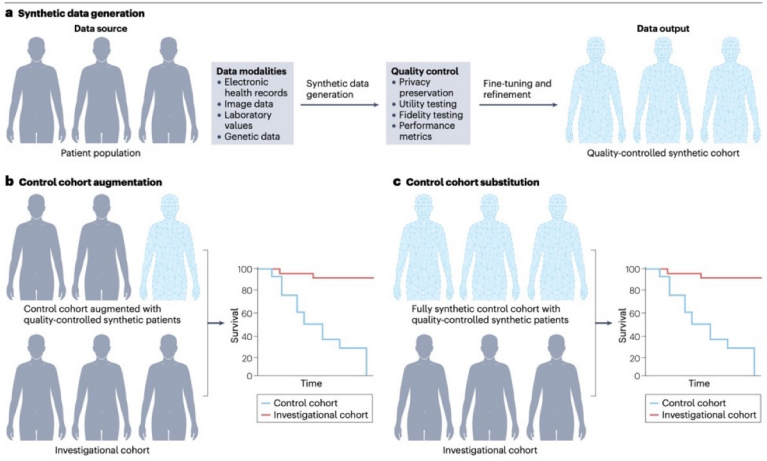
https://doi.org/10.1038/s41568-026-00912-4

Review article

Check for updates

Artificial intelligence-generated synthetic data for cancer research and clinical trials

Jan-Niklas Eckardt[1,2], Waldemar Hahn[3,4], Arsela Prelaj[5], Martin Bornhäuser[1,6,7], Jan Moritz Middeke[1,2,9] & Jakob Nikolas Kather[1,2,8,9]



Evaluation must benchmark fidelity and utility against real-data references, with strict filtering for clinical plausibility.
Privacy auditing is mandatory. Transparent workflows enable reproducibility, trust, and safe downstream use.

Table 1 | Common evaluation metrics for synthetic tabular and image data

Category	Data type	Key metrics	Purpose
Fidelity	Tabular	Statistical tests (e.g., Kolmogorov–Smirnov, X^2) Quantification of correlation difference (e.g., Jensen–Shannon or Wasserstein distance) Propensity score classifier or distinguishability (area under the curve)	Compares real and synthetic data sets regarding their distributional and correlation similarity
	Image	FID KID Precision and recall for generative models Inception score	Measures perceptual quality, alignment and diversity of generated images
Utility	Tabular	TSTR using task-specific metrics (e.g., AUROC, F1, R^2, RMSE) Ability to reproduce published results or statistical associations	Evaluates whether synthetic data can substitute for real data in predictive modelling or hypothesis testing
	Image	TSTR on intended tasks (e.g., classification, segmentation, detection) Human expert evaluation in medical contexts	Tests whether synthetic images support real-world tasks
Privacy	Tabular	Membership-inference attack Privacy budget and/or differential privacy DCR	Quantifies risk of patient re-identification or leakage of sensitive attributes
	Image	Membership-inference attack Nearest-neighbour analysis Duplicate detection Model inversion attack (emerging)	Identifies memorization or leakage of real images into the synthetic set

Patients engagement and Usability

AI Scribe

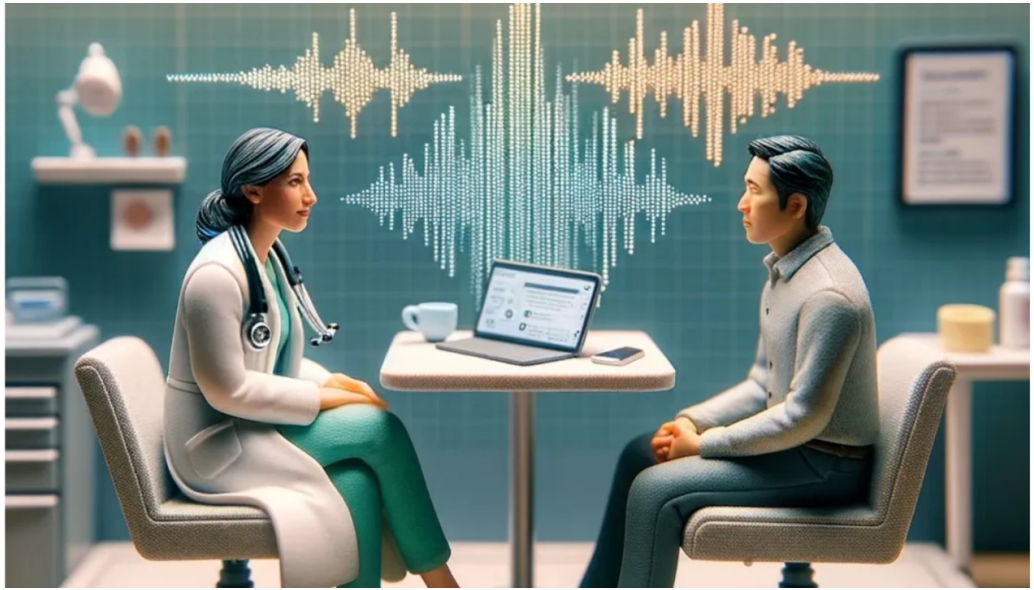
EndPoints

- User trust, transparency, and perceived explainability of AI outputs.
- **Impact of the AI tool on user experience, workflow integration, and decision-making.**
- Fairness, privacy protection, and bias mitigation across different user groups.

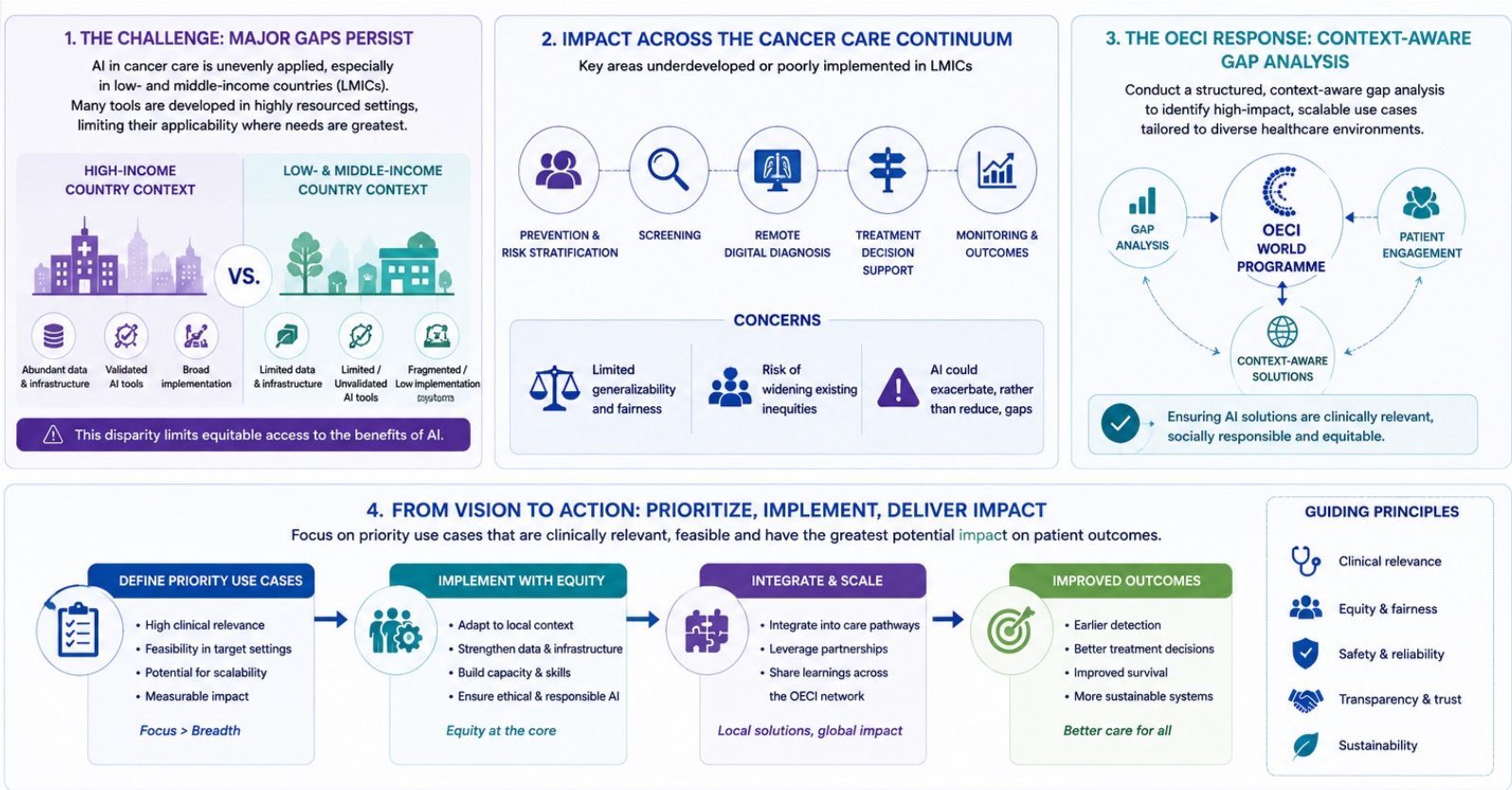
Ambient AI Scribes in Clinical Practice: A Randomized Trial

Authors: Paul J. Lukac, M.D., M.B.A., M.S. , William Turner, B.S. , Sitaram Vangala, M.S. , Aaron T. Chin, M.D. , Joshua Khalili, M.D. , Ya-Chen Tina Shih, Ph.D. , Catherine Sarkisian, M.D., M.S.H.S. , Eric M. Cheng, M.D., M.S. , and John N. Mafi, M.D., M.P.H. Author Info & Affiliations

Published November 26, 2025 | NEJM AI 2025;2(12) | DOI: 10.1056/AIoa2501000 | VOL. 2 NO. 12



How OECI can help in addressing “AI Gaps” or Gaps with AI



IMPLEMENTAION GAP

nature health

Article

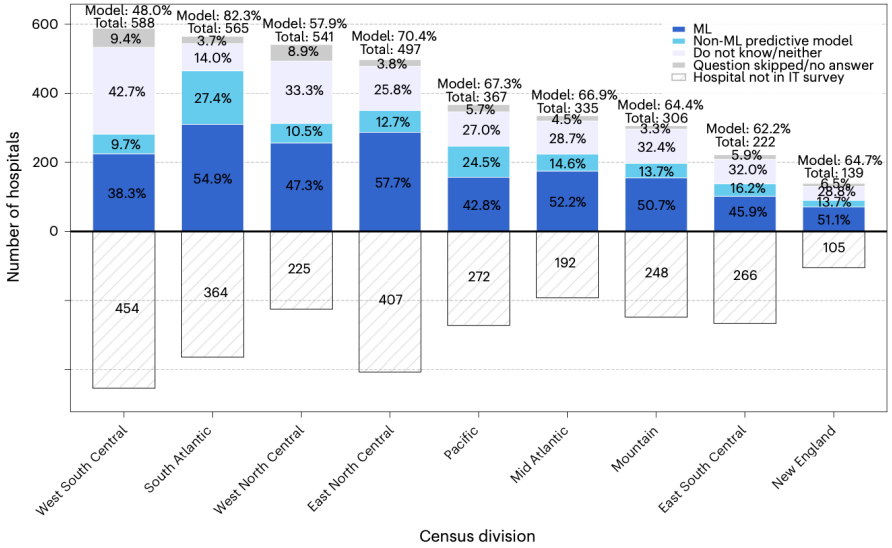
https://doi.org/10.1038/s44360-025-00016-7

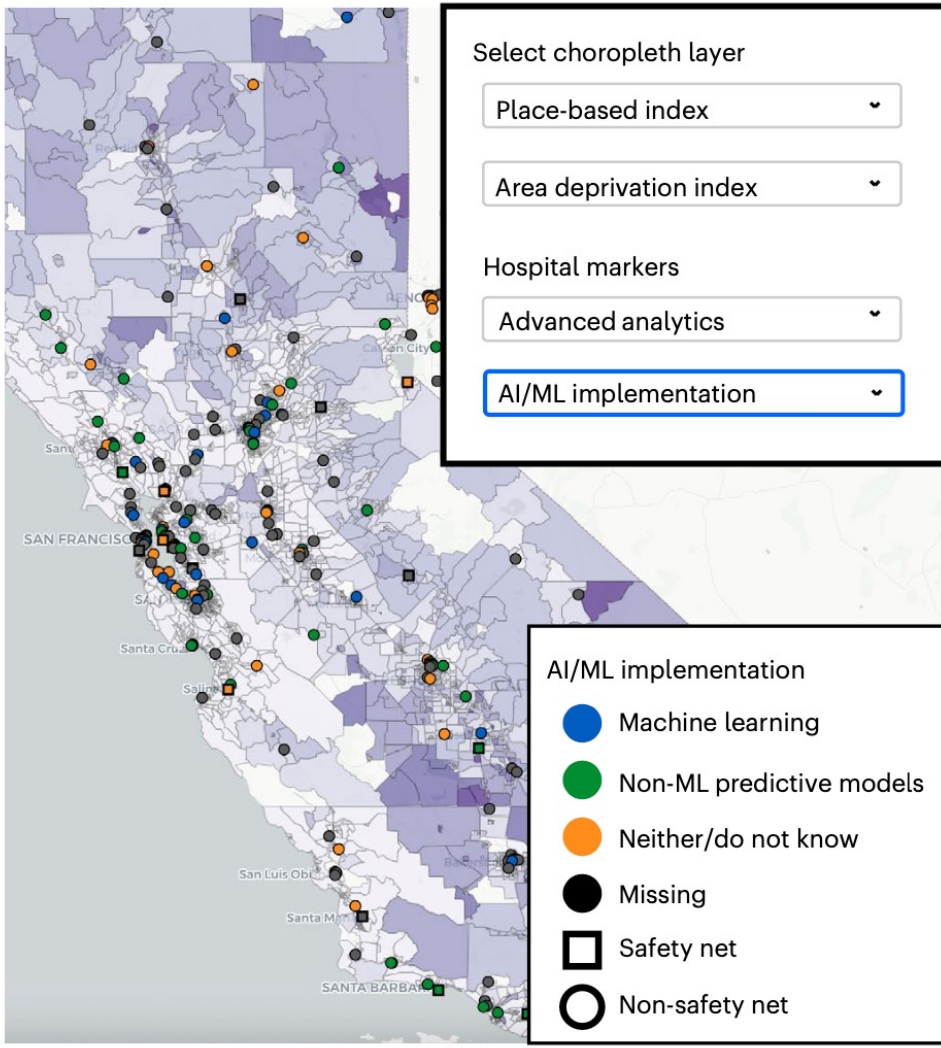
The landscape of AI implementation in US hospitals

Received: 23 August 2025

Accepted: 22 October 2025

Yeon-Mi Hwang[1], Madelena Y. Ng[1,2], Malvika Pillai[1,3], Michelle P. Sahai[1] & Tina Hernandez-Boussard[1,4]



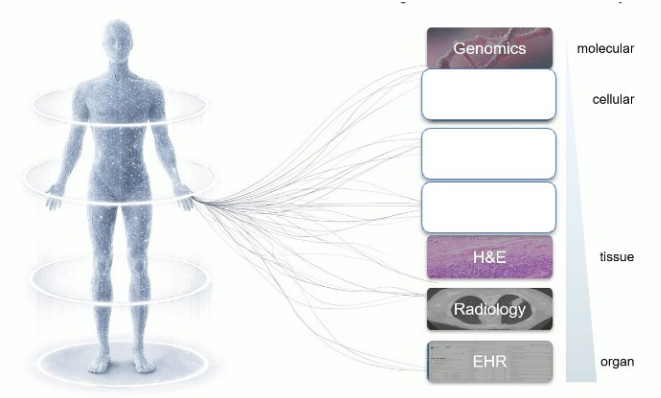


PREDICTIVE
AI TOOLS

FOUNDATION
MODELS

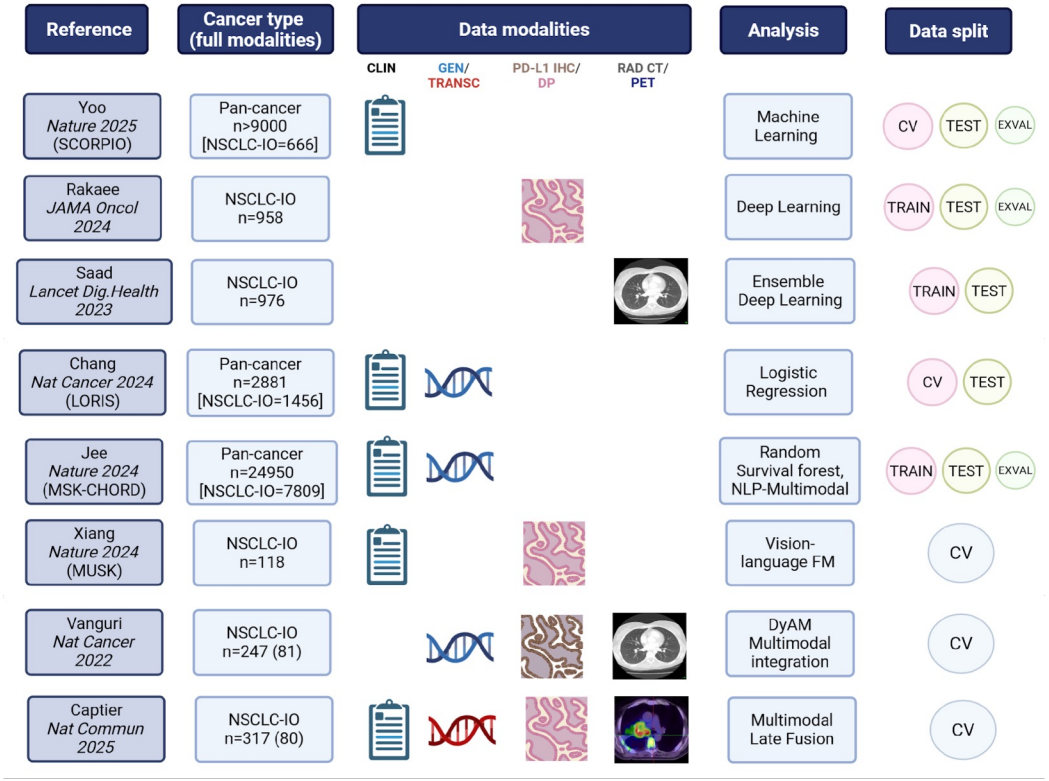
INTEGRATED CLINICAL BIOMARKERS

THE VISIBLE LAYER

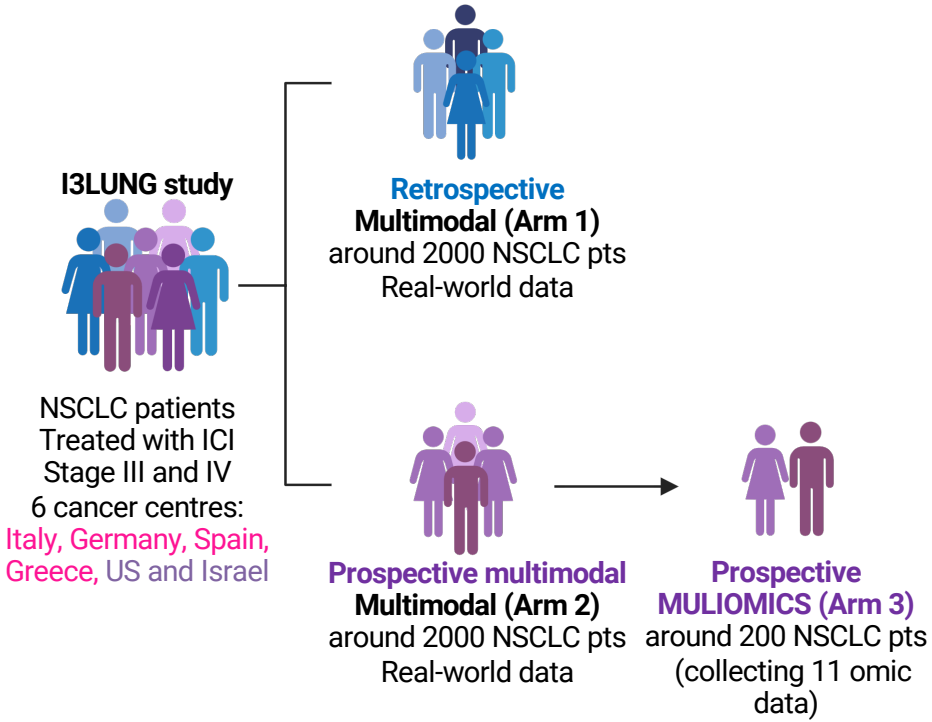


Immune prediction: the best use case for AI (more patients compared to target)

IMMUNOTHERAPY PREDICTION OVER THE LAST 7 YEARS WITH AI



I3LUNG Project arms



Objective: Developing AI-based tools by using multimodal data for clinical decision making in NSCLC patients treated with immunotherapy

A project funded by the European Union

I3LUNG

Prelaj et al. Under Revision in Nature Medicine

Prelaj et al. Clinical Lung Cancer 2023.
medRxiv 2026.01.16.25342913;
doi:https://doi.org/10.64898/2026.01.16.25342913

Prelaj A, Miskovic V. ….. Pedrocchi A, Garassino M. ***Nature Medicine, Accepted, 7TH MAY 2026***

Retrospective Arm and Cohorts=2396 patients

Cohort collection

Study population

2396 patients, III-IV stage NSCLC ,
Immune checkpoint inhibitors
6 clinical cancer centers:
INT, GHD, SZMC, VHIO; UOC, MH

Collected baseline data

2252 clinical data demographic; treatment; lab; molecular

1723 Genomics

936 H&E slides

1705 CT scans

Cohort creation

Cohort 1
Curative setting

Cohort 2
First metastatic line

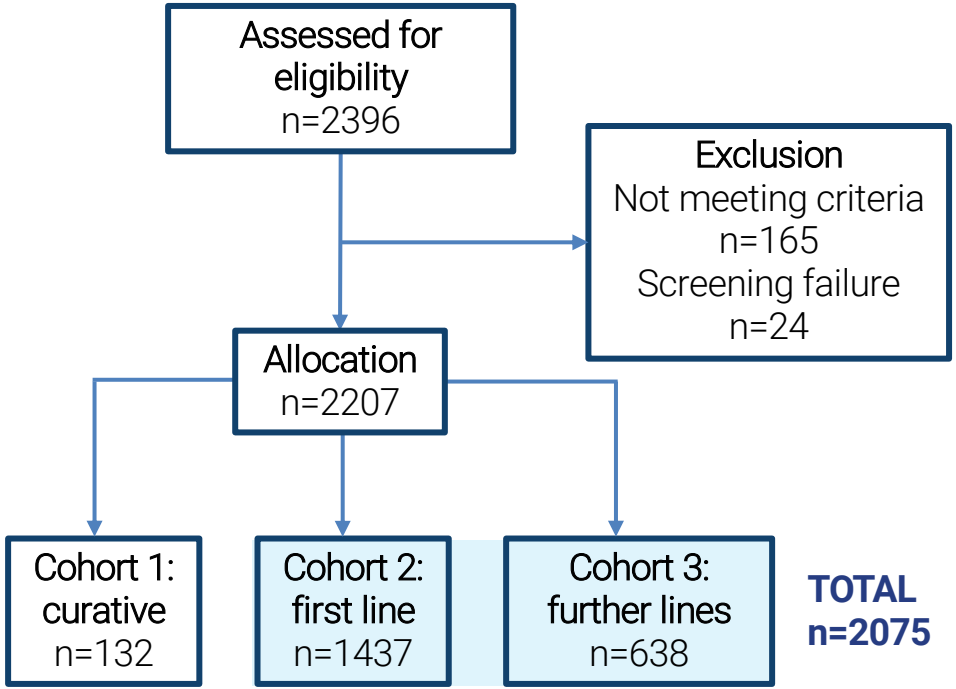
Cohort 3
Later metastatic lines

used in analysis (n=2075)

Train set
(85%, n=1550)

Test set
(15%, n=274)

Ex. val. set (UOC)
(n=251)



Prelaj A, Miskovic V. ….. Pedrocchi A, Garassino M.
Nature Medicine, Accepted, 7TH MAY 2026

I3LUNG: Clinical Usability study (Expert vs Non-Expert)

PATIENTS COHORT

100 NSCLC pts treated with ICI, 6 I3LUNG centers

I³LUNG

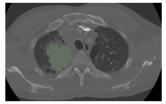
PHYSICIANS

10 experts in lung cancer

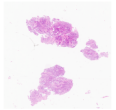
10 non-experts in lung cancer

PHASE 1 - NO XAI support

Patients' data: Clinical, demographic, lab, Genomics, molecular



CT scan



H&E slide

Questions:

Will this patient achieve DCR with IO-based treatment?

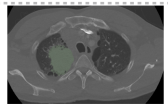
YES NO

What is the most likely survival range for this patient?

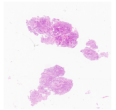
<6m, 6-12m, 12-18m, ≥24m

PHASE 2 - WITH XAI support

Patients' data: Clinical, demographic, lab, Genomics, molecular

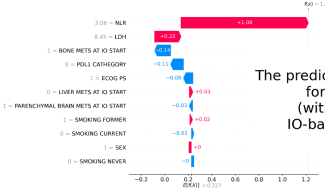


CT scan



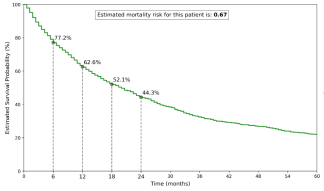
H&E slide

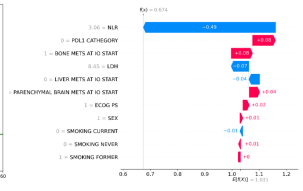
XAI DCR prediction



The predicted response to IO-based therapy for this patient is: **Responder** (with likelihood of responding to IO-based therapy equal to **77.22%**)

XAI OS prediction





Questions:

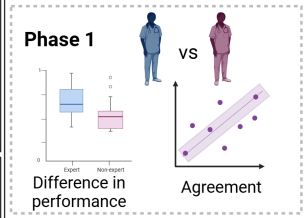
After seeing the XAI DCR prediction Will this patient achieve DCR with IO-based treatment?

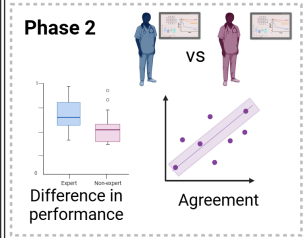
YES NO

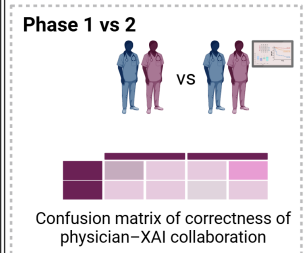
After seeing the XAI OS prediction what is the most likely survival range for this patient?

<6m, 6-12m, 12-18m, ≥24m

ANALYSIS







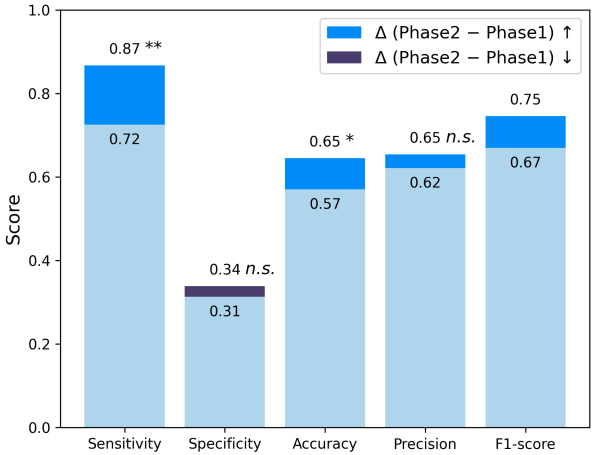
AI-ON-Lab

ESAC

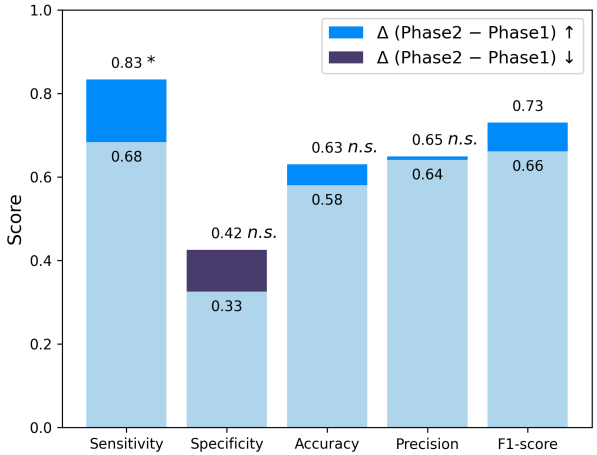
Prelaj A, Miskovic V. ….. Pedrocchi A, Garassino M. ***Nature Medicine, Accepted, 7TH MAY 2026***

Clinical usability: IO efficacy prediction and user feedbacks

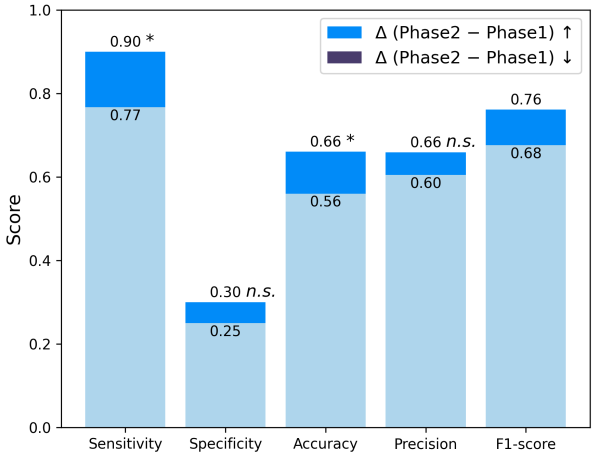
All physicians



Lung Cancer Experts



General Oncologist
Non-experts

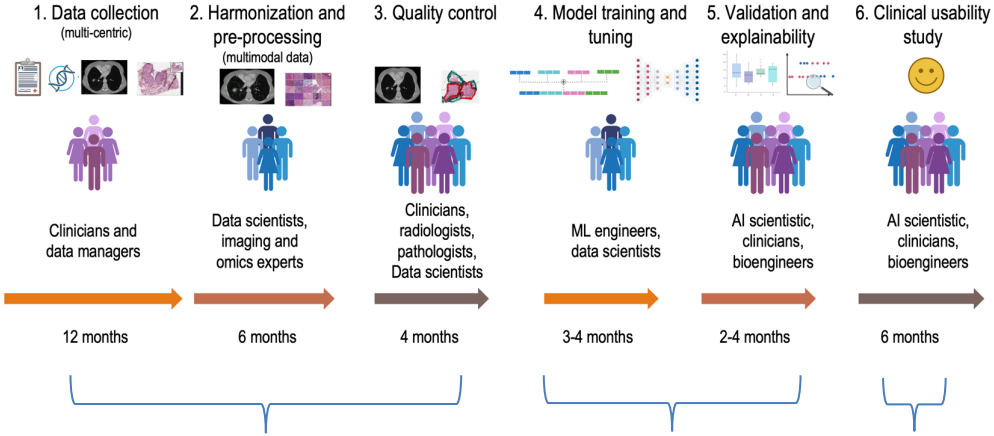


XAI increases the probability of predicting the true DCR by 37% and true OS by 36% for all physicians (**DCR: 23%** for experts vs **53%** non-experts) (**OS: 14%** for experts vs **61%** non-experts)

Prelaj A, Miskovic V. ….. Pedrocchi A, Garassino M. ***Nature Medicine, Accepted, 7TH MAY 2026***

From Data To Implementation

Our experience on human and time resources: I3LUNG use case



(DATA)
22 months

(MODEL)
8 months

(UTILITY)
6 months

D

V

U

Finished

Finished

7. Implementation and data interoperability

(IMPLEMENTATION and Prospective Validation)
24 months

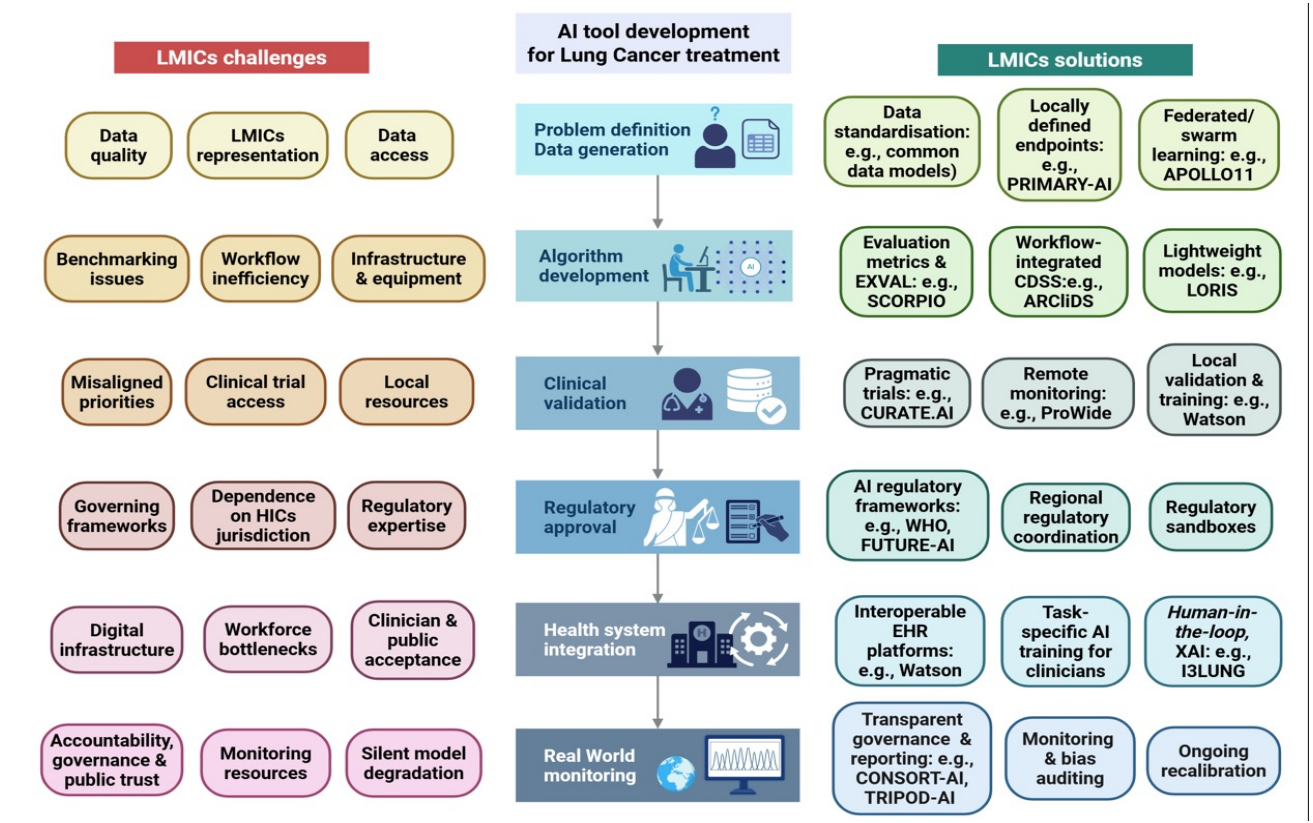
I

PV

Ongoing

ORIENTING LOW AND MIDDLE INCOME COUNTRY PRIORITIES

Addressing LMICs Challenges for Equitable AI:

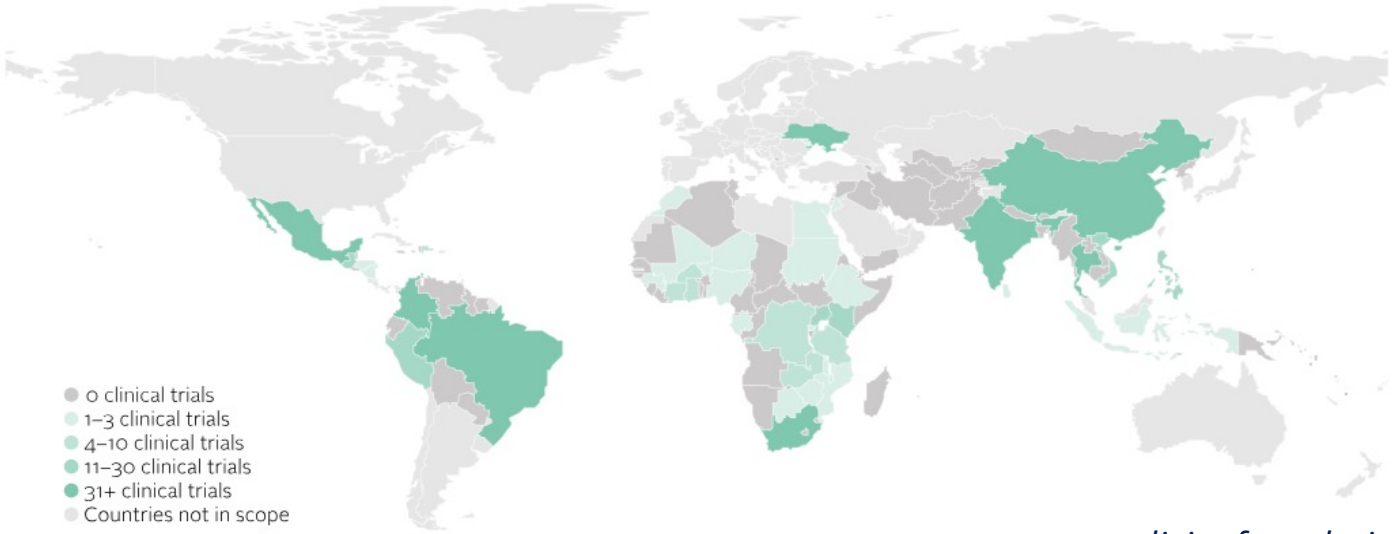


Romanò R... Prelaj A; under review The Lancet Oncol

LMICs Underrepresentation in Clinical Trials: A Global Challenge

Over half (70/113) the countries in scope have no active clinical trials

Clinical trials carried out in countries covered by the Index, are concentrated in a small cohort of countries, such as China, Brazil, Mexico, South Africa and Thailand. While these countries are all LMICs covered in the Index analysis, all five are classified as upper-middle-income countries, with few companies' conducting clinical trials in low-income countries.



accesstomedicinefoundation.org

AI-Based Augmentation of Oncology Clinical Trials
Invited Article: Nature Reviews Clinical Oncology

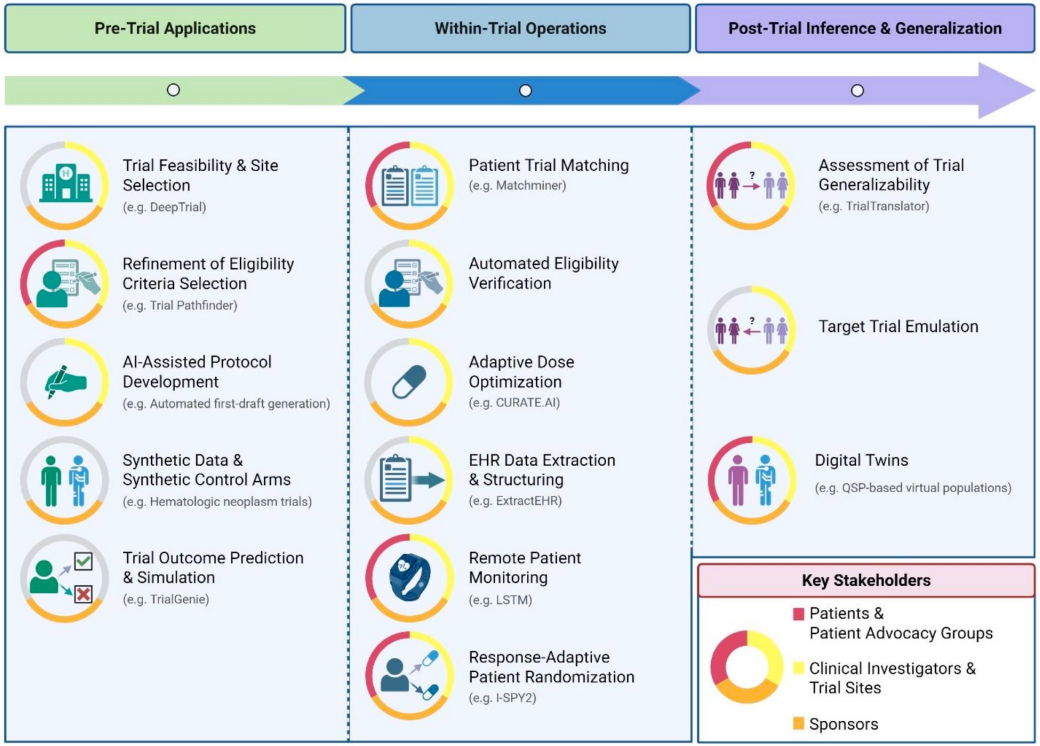
Authors: Andrea Villa[1*], Ashley L. Eadie[2*], David Synnott[3,4], Rebecca Romanò[1], Max Piffoux[5], Evelyn Yi Ting Wong[6], Naomi Schneinerman[7], Daniel Tao Shao Weng[6], Filippo Guglielmo Maria de Braud[1], Miriam Koopman[8], Jarushka Naidoo[3,9], Madhusmita Behera[10], Selen Bozkurt[10], Susan Halabi[11], Rodrigo Dienstmann[12,13], Loic Verlingue[5,14], Arsela Prelaj[1**], Ravi B. Parikh[2,10**]

AI-driven site selection can identify underserved regions with high cancer burden and low trial access.

Automated patient screening, eligibility matching, and EHR data extraction can reduce workload and resource requirements.

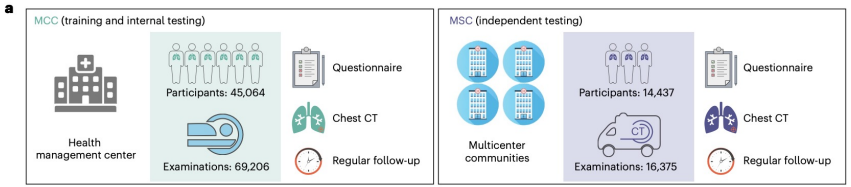
Support capacity building across the OECI network

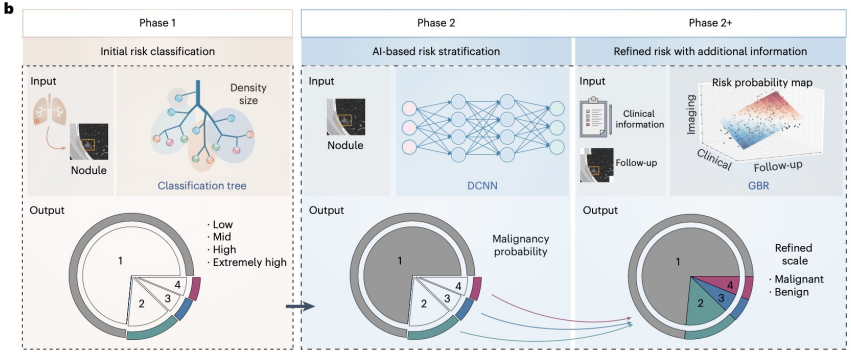
AI can help emerging cancer centres participate in clinical trials and translational research.



FPOCUS ON OPEN CHALLENGES OF PERSONALIZED HEALTH CARE SYSTEM

Lung-RADS: lung cancer screening





c

Category	1: low risk	2: mid risk	3: high risk	4: extremely high risk
Findings	Refined four-category risk scale			
Management	Continue annual screening with LDCT in 12 months	6-month CT	3-month CT	Immediate clinical assessment

nature medicine

Developed using **45.064 participants** from a medical check-up cohort and validated on **14,437 participants** from mobile CT units across Western China

Internal dataset: AUC 0.918 vs 0.881 for single-dimension models; Independent dataset: Sensitivity 87.1%

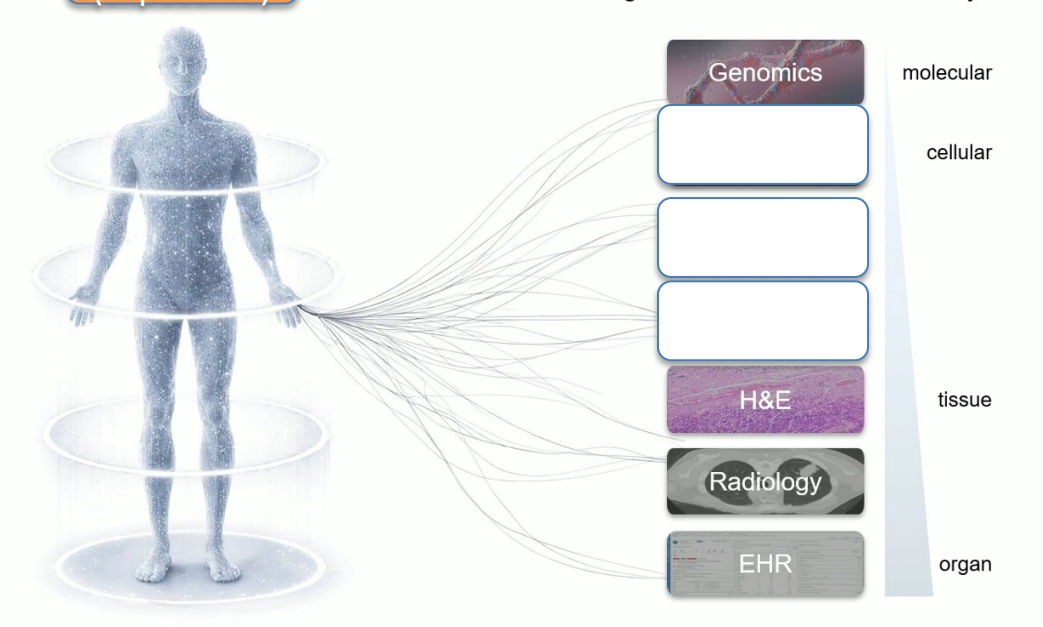
Negative predictive value: 99%, reducing risk of missed cancers

Chengdi Wang et al. Nature Medicine 2024

INTEGRATED CLINICAL + EXPOSOME BIOMARKERS

Digitalomics (Exposome)

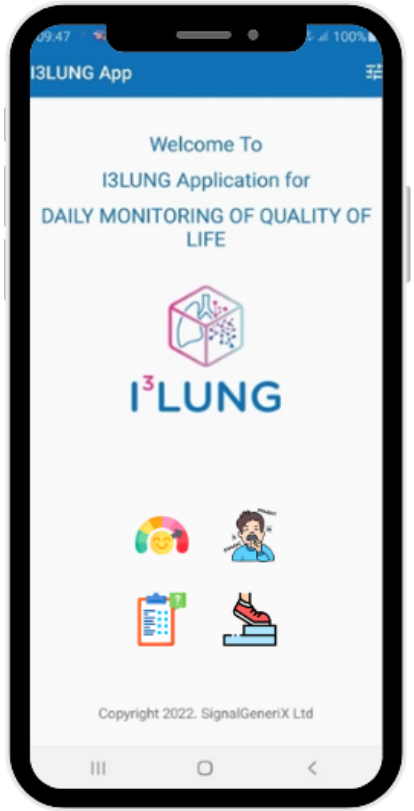
THE NEW LAYER



New Biomarkers collected through devices (via wearable or via app), generally ignored before using AI in healthcare

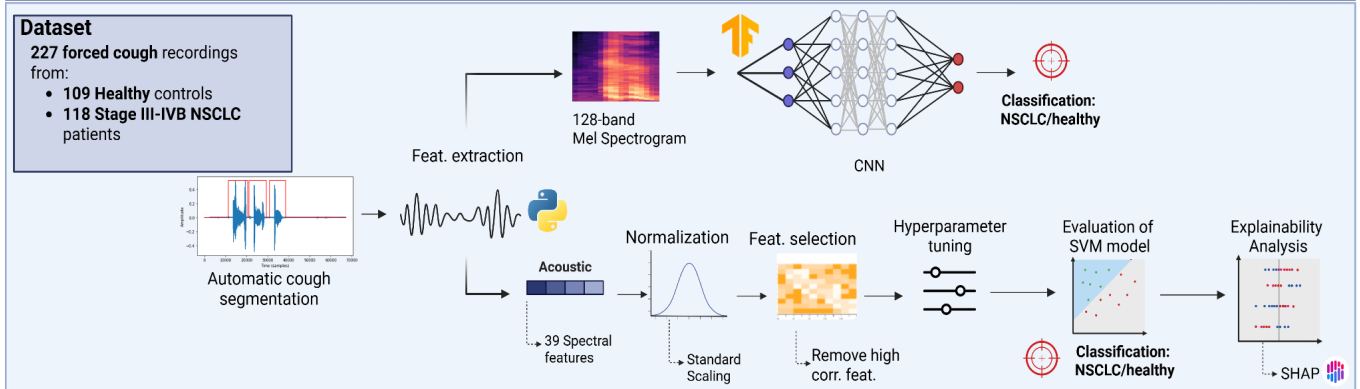
With this *expotype* we are able to further increase the prediction (no estimation is possible at the moment)

Cough Analysis: Prescreening Tools

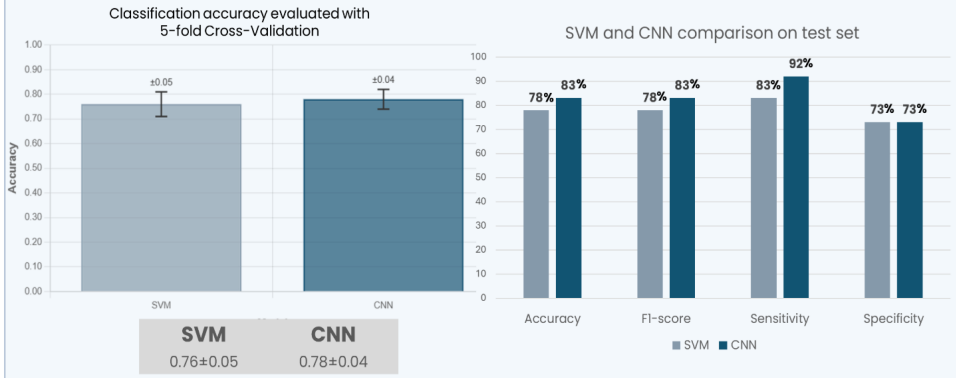


METHODS

Dataset
227 forced cough recordings from:
- **109 Healthy** controls
- **118 Stage III-IVB NSCLC** patients



RESULTS



SVM	CNN
0.76±0.05	0.78±0.04

Comparison with literature

Method	Sensitivity	Specificity
Cough analysis	**0.92**	**0.73**
VOCs in exhaled breath	0.86	0.69
E-nose technology	0.71-0.96	0.33-1.00
LDCT	0.94-0.98	0.82-0.91

References
1. X. Chen, et al. "A study of the volatile organic compounds exhaled by lung cancer cells in vitro for breath diagnosis," 2007.
2. Horvath, et al., "Exhaled biomarkers in lung cancer", 2009
3. Guo, L., et al., Accuracy of baseline low-dose computed tomography lung cancer screening: a systematic review and meta-analysis. 2023

Cough analysis in High Risk for Lung Cancer population

STUDY COHORTS

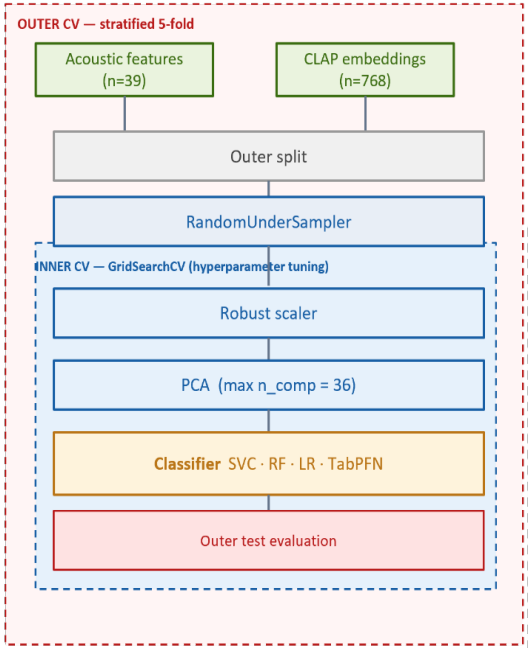
Healthy control
Healthy subjects
Median age: 45 (25-68), non-smokers mainly
109 cough recordings

At-risk group
Age ≥55
Current smokers ≥15 pack-years
Former smokers (last 20 yrs)
234 cough recordings

Surgical ONGOING
Patients with pulmonary nodules undergoing pre-surgical staging
15 cough recordings so far

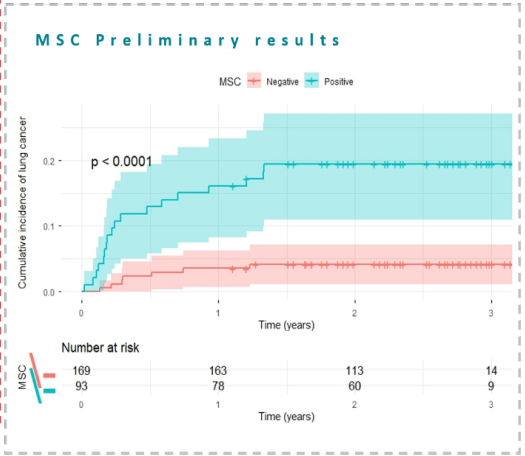
NSCLC patients
Stage III–IVB
Undergoing immunotherapy
150 cough recordings

ANALYSIS PIPELINE EXAMPLE



Cough Analysis

Daily coughs are recorded and a-posteriori analyzed to relate them with trend of therapy



Method (Preliminary results)	Sens.	Spec.
Cough analysis (at-risk vs NSCLC)	0.82	0.70
Cough analysis (healthy vs NSCLC)	0.92	0.73

Method (From Literature)	Sens.	Spec.
VOCs in exhaled breath	0.86	0.69
E-nose technology	0.71–0.96	0.33–1.00
LDCT screening	0.94–0.98	0.82–0.91

HOW WE CAN USE ETHICALLY GENERATIVE AI FOR CLINICAL DECISION MAKING GLOBALLY

nature health

Article

https://doi.org/10.1038/s44360-026-00082-5

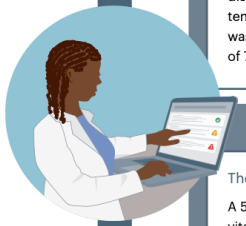
Safety of a large language model-based clinical decision support system in African primary healthcare

An EMR-integrated LLM clinical decision support system was assessed across **16 primary care clinics in Kenya**

Hallucinations were uncommon (**3.4%**) and recommendations were highly concordant with local guidelines (**99%**).

Implementation challenge: Despite guideline alignment, clinicians adopted only a minority of AI suggestions, while some potentially harmful recommendations (**7.8%**) were incorporated into final documentation.

The system corrected or mitigated pre-existing clinical risks in approximately **8% of encounters**, demonstrating potential to improve care quality.



An Unsafe LLM Response Overridden

The case:

A 28-year-old woman presented, following treatment for candidiasis/UTI five days prior, on this occasion with right lower quadrant (RLQ) abdominal pain and an elevated heart rate (102). She had tenderness in the RLQ, but no distention, guarding or rebound tenderness. A complete blood count was unremarkable except for an MCV of 78.

The LLM response:

The AI Consult suggested further evaluation of possible abdominal and gynaecological pathologies, including appendicitis, ovarian torsion and ectopic pregnancy. However, the tool argued that appendicitis should be the primary differential, and suggested an ultrasound.

The outcome:

The clinician, independent of the AI Consult output, ordered a pregnancy test, which turned out to be positive. The clinician made a provisional diagnosis of ectopic pregnancy, and the patient was urgently referred to specialized care at an inpatient hospital.

A harmful LLM response enacted

The case:

A 5-year-old child, with normal vital signs, presents with a mild non-productive cough, runny nose, fevers and abdominal pain with four episodes of vomiting on the morning of the presentation. A stool sample showed mucous and moderate pus cells.

The LLM response:

The AI Consult's interpretation of the symptoms and stool results was that they suggested a gastrointestinal infection. It recommended prescribing oral rehydration solution (ORS), as well as an antibiotic such as metronidazole or amoxicillin.

The outcome:

The clinician took heed of the advice to prescribe amoxicillin, even though the presentation did not warrant it. Moreover, amoxicillin is not the preferred antibiotic for a GI infection and is likely to cause additional GI symptoms.

Overfitting to the prompt

The case:

A 21-month-old child presented with typical cold symptoms. There were no concerning findings in the history or examination. The clinician planned to prescribe OTC symptomatic treatments.

The LLM response:

In addition to confirming the clinician's initial plan, the LLM recommended, "over-the-counter throat lozenges if the child is old enough."

Within the AI Consult prompt is a one-shot example that involves viral pharyngitis. The prompt encourages the model to also consider symptomatic treatments that may not be part of standard guidelines, such as [explicitly mentioned] throat lozenges for cases of viral pharyngitis. In our post-deployment testing, we found that unprompted models do not make the mistake of recommending throat lozenges to a 1-year-old, who is clearly not old enough for that treatment to make sense. This is a good example of how detailed prompt instructions may have unintended consequences in model outputs.

LLMs for CDM (AMIE)

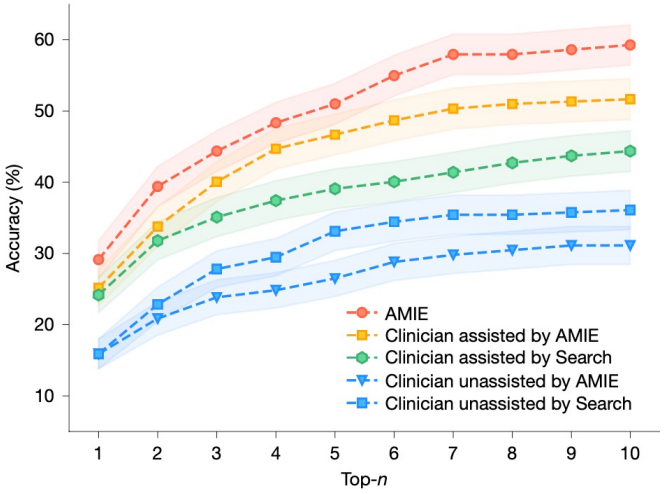
nature

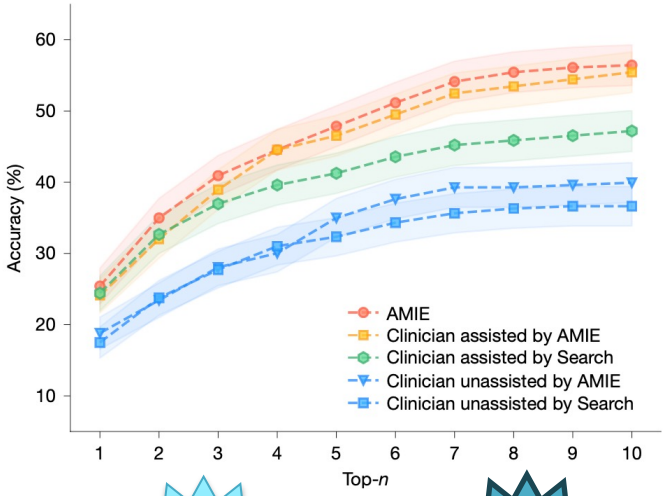
Explore content · About the journal · Publish with us

nature › articles › article

Article | Open access | Published: 09 April 2025

Towards accurate differential diagnosis with large language models





I

CP ? CE ?

nature medicine

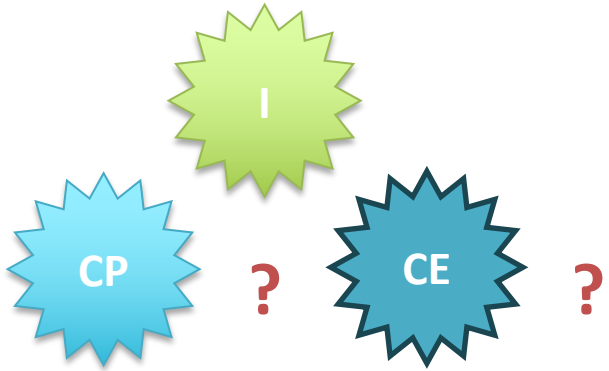
Article

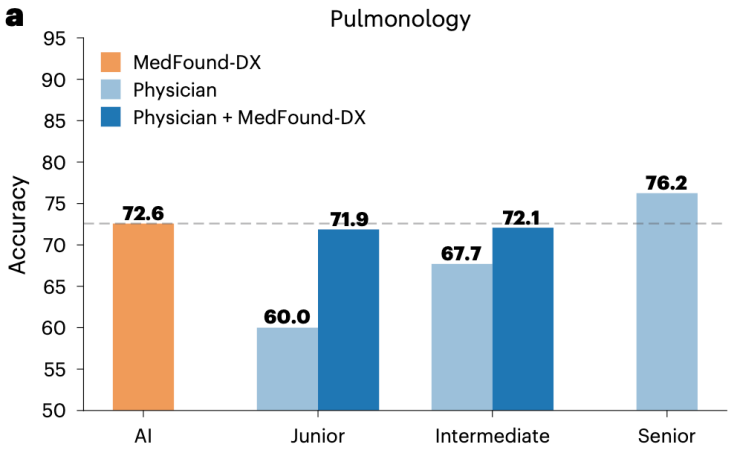
https://doi.org/10.1038/s41591-024-03416-6

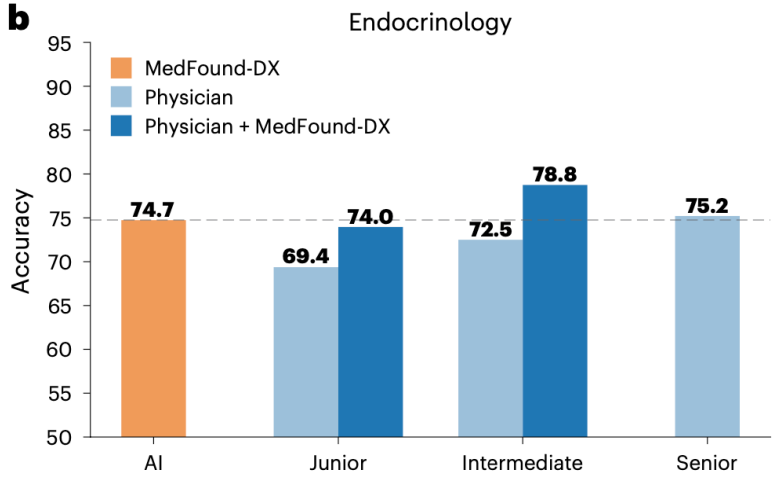
A generalist medical language model for disease diagnosis assistance

These results demonstrate that our LLM-based diagnostic generalist outperformed junior and intermediate physicians in both specialties and was similar to senior physicians.

MedFound-DX-PA: a BIOLLMs







Limitation: *they did not compared MedFound+Senior*

Liu et al. A generalist medical language model for disease diagnosis assistance. Nat Med 2024.

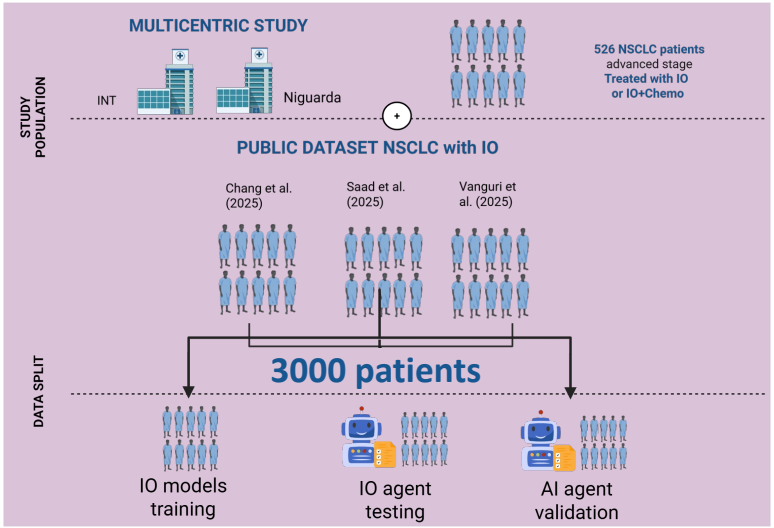
AGENTIC AI
(All in ONE MODELS in a Dynamic Fashion)

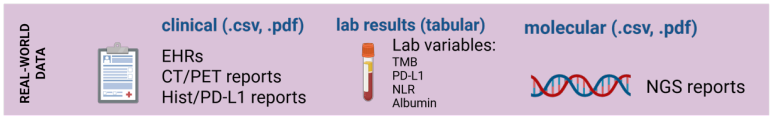
AI Agents

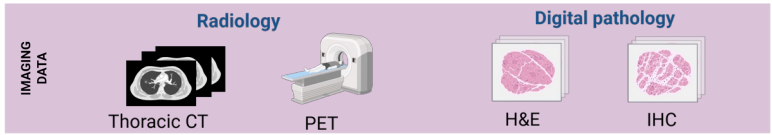
Treatment-based AI Agents

POLITECNICO MILANO 1863 · Apollo 11 STRENGTH IS UNITY · AI-ON-Lab · Fondazione IRCCS Istituto Nazionale dei Tumori · Sistema Socio Sanitario Regione Lombardia

Collaborative Human-Agent for Therapeutic Decision-Making In Cancer Immunotherapy Prediction

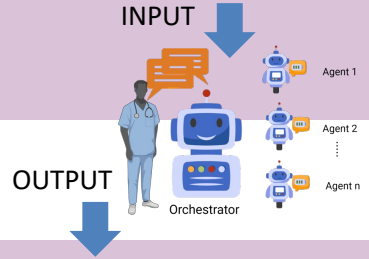






You are the Lead Oncologist.
Construct a final, detailed treatment plan based on the reports from your team.
Synthesize all the above information into a final response. Output a structured, professional medical report. Each statement you make should be supported by the input data. Cite the input source of each statment (i.e. "Patient has IV stage NSCLC)
Follow EXACTLY this format and these headings:

****Key Findings****
****Guideline-based Evidence****
****Predictive Tools****
****Regimen:****
****Contraddictions****



****Key Findings****
- The patient is a 63-year-old female with stage IV, adenocarcinoma of the lung (T1bN0M1b) and multiple metastatic sites (liver, bone, adrenal, pleural) with no brain metastases (Key Findings).
- PD-L1 tumor proportion score (TPS) is 10 % (PD-L1 = 10 % → category 1-49 %) (Key Findings).
- ECOG performance status is 1 (Key Findings).

****Guideline-based Evidence**** : NCCN 2024 and ESMO 2024 guidelines recommend
****Predictive Tools****
- ASTEP predicts a higher benefit for ICI-chemo over chemo-alone (CATE = -0.072, confidence = moderate) (Predictive Tools).
- LORIS data are unavailable; however, the ASTEP result aligns with guideline recommendations and supports the chosen regimen.

****Regimen:** Pembrolizumab + Carboplatin + paclitaxel (triple-combination).**

V

Unpublished material

ESAC LinkedIn

ESAC Website

Become a Member!
Shape the Future!

ESAC

EDUCATION AND PATIENT ENGAGMENT IN AI

THE EUROPEAN INTERDISCIPLINARY SOCIETY OF ARTIFICIAL INTELLIGENCE FOR CANCER RESEARCH

Acknowledgments

https://Ai-onlab.com

I³LUNG

A project funded by the European Union

https://i3lung.eu

Fondazione IRCCS Istituto Nazionale dei Tumori

Sistema Socio Sanitario Regione Lombardia

Arsela.prelaj@istitutotumori.mi.it

@arselaprelaj.bsky.social

ESAC

POLITECNICO MILANO 1863