Fenice

Gruppo italiano per la ricerca clinica in Medicina d'Urgenza

Fare le giuste domande, avere le giuste risposte



Luca Antiga
Domyn

6 novembre 2026

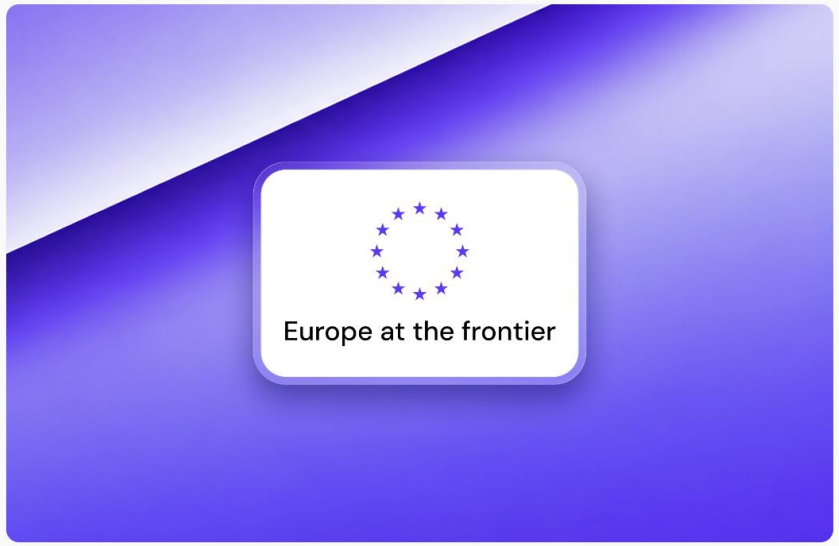

OROBIX
joins
Domyn

Featured

Europe cannot afford to give up on frontier AI

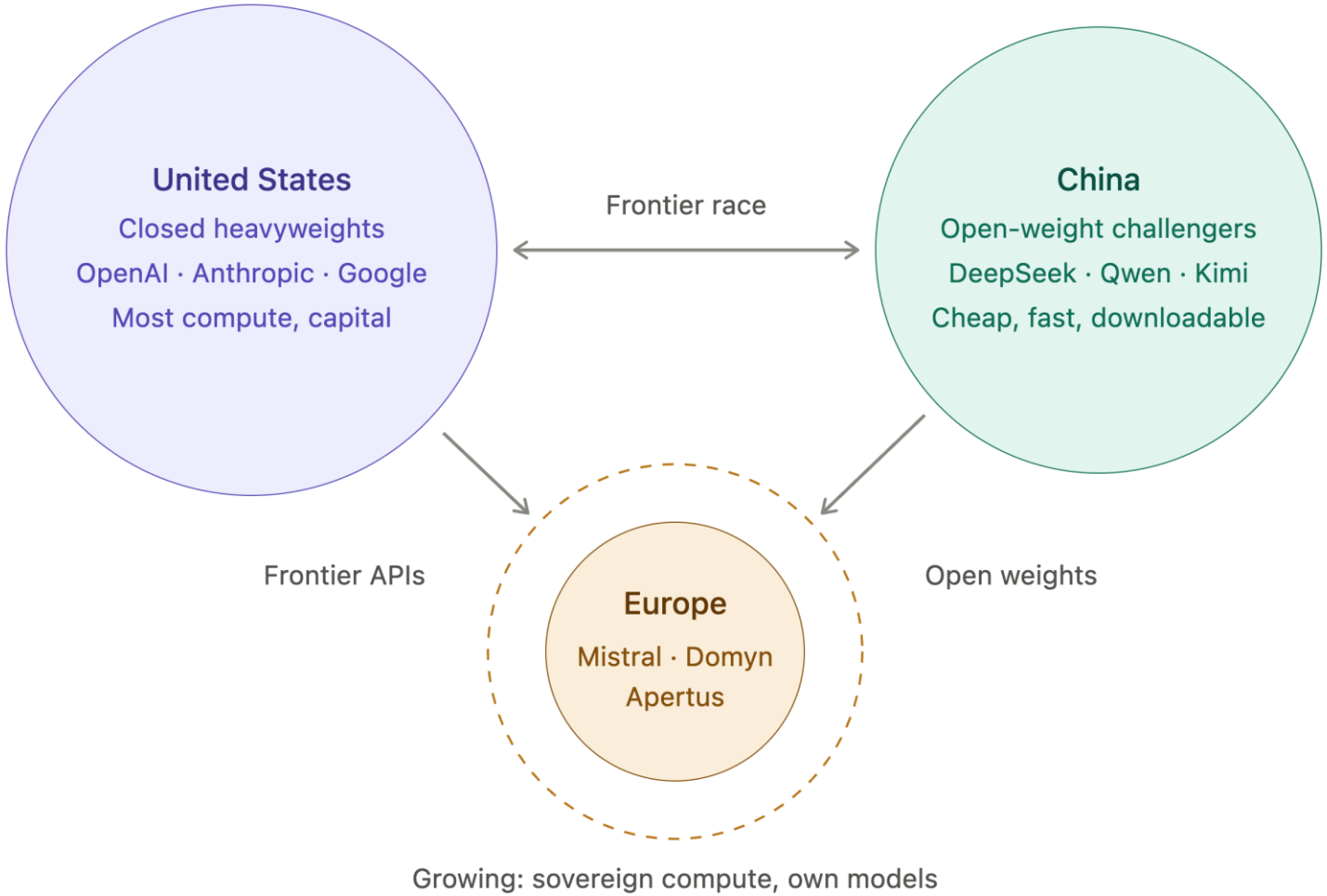
Domyn CEO Uljan Sharka tells the FT why Europe must build its own frontier AI, not rent it, and how sovereign compute and the AI Act can get it there.

Read more

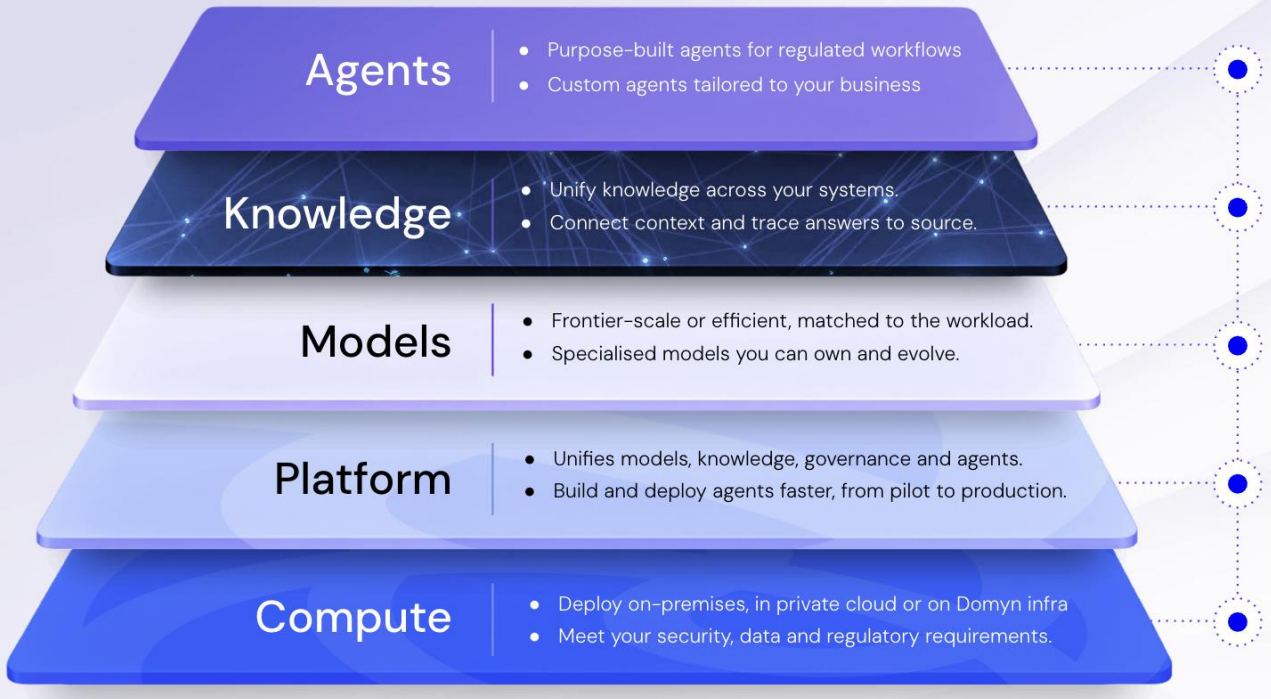


https://www.domyn.com/blog

La partita


United States
Closed heavyweights
OpenAI · Anthropic · Google
Most compute, capital
Frontier race
China
Open-weight challengers
DeepSeek · Qwen · Kimi
Cheap, fast, downloadable
Frontier APIs
Open weights
Europe
Mistral · Domyn
Apertus
Growing: sovereign compute, own models

A vertically integrated AI operating system



Use the whole stack or just the parts you need

Domyn

Governance

Embedded across the Domyn Operating System

Control
Maintain oversight and set guardrails

Explainability
Make every decision understandable

Compliance
Meet regulatory requirements and industry standards

Trust
Build confidence through security and transparency

Domyn

Domyn

What's next: Europa, a frontier scale MoE

An open-source frontier model, 400B+ parameters, built for all 24 EU languages, backed by the European Commission.

Access to three European HPCs starting in September, plus Colosseum HPC in 2027.

Europa LLM

- MoE +400B parameters
- Long-context support
- Agentic capabilities
- 24 European languages
- EU AI Act Compliant
- SOTA general-purpose model

Europa Consortium

Fraunhofer Domyn

Jupiter (DE)

4xGH200

5,938 booster nodes - Exaflop

Mare Nostrum (ES)

4xH100

1,120 nodes – 260 Petaflops

Leonardo (IT)

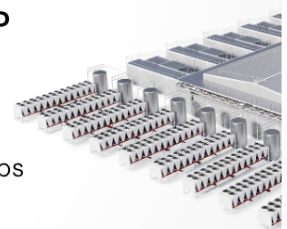
4xA100

3,456 booster nodes – 250 Petaflops

Colosseum HCP

72xGB300 (NVL72)

Several nodes - Exaflops



Confidential

eCREAM
enabling Clinical Research
in Emergency and Acute care Medicine

Project funded by the European Union
www.ecreamproject.eu

eCREAM's coordinating centre is located at the Laboratory of Clinical Epidemiology of Mario Negri Institute
Villa Camozzi via G.B. Camozzi, 3, 24020 Ranica (BG) - Italy

+39 035 4535389

ecream@marionegri.it

CONTESTO

Dalle note cliniche ai dati strutturati

L'AI legge il testo libero, compila una scheda e spiega ogni risposta citando la nota

1 · TESTO LIBERO

Nota clinica

Paziente di 68 anni, operata nel 2021 per adenocarcinoma del colon. Attualmente in chemioterapia. Riferisce stanchezza nelle ultime settimane.

2 · LETTURA

AI
modello linguistico

3 · SCHEDA STRUTTURATA E MOTIVATA

Variabile	Valore	Motivazione
Neoplasia	**Attiva**	*«Attualmente in chemioterapia»*
Livello di coscienza	**Vigile**	*«Riferisce stanchezza»: il paziente comunica*
Età	**68 anni**	*«Paziente di 68 anni»*

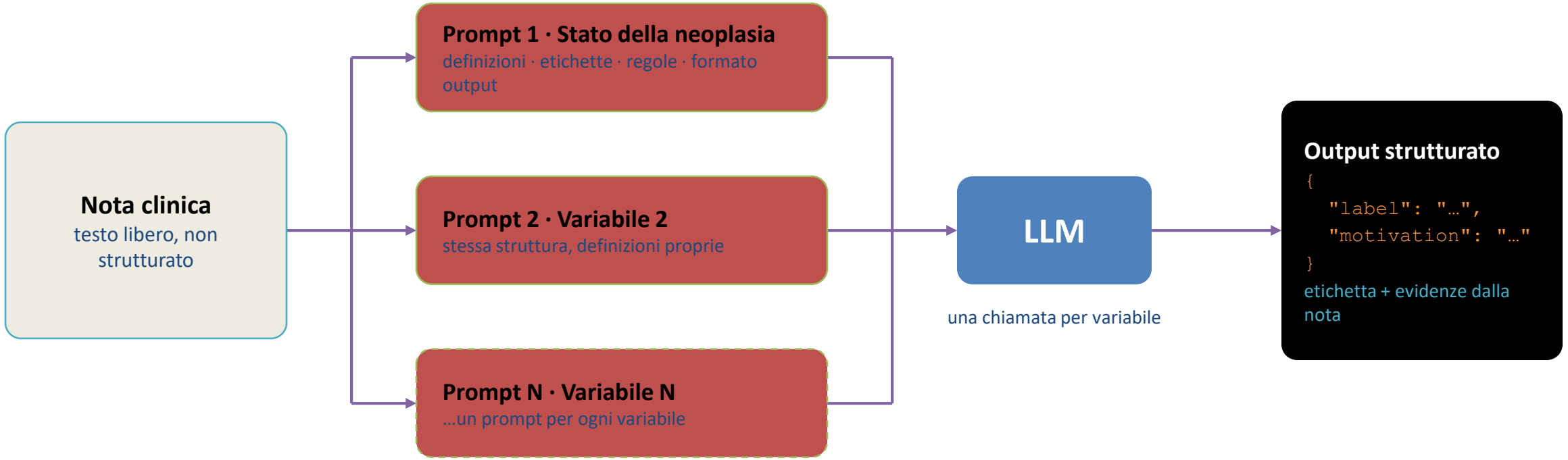
Ogni valore è accompagnato dalla frase della nota che lo giustifica: il risultato si può verificare

Nota clinica

Donna di 82 anni. Non allergie note. Ipertesa, diabetica, dislipidemica, obesa. IRC.
Vascuopatia carotidea (stenosi 60% ICA sx, stenosi 50% ICA dx).
Occlusione succlavia sx con furto.
Pregressa neoplasia vescicale sottoposta a resezione endoscopica nel 2015 (non documentazione).
CAD nota da settembre 2015 esordita con NSTEMI da discrepanza in corso di anemizzazione post intervento ortopedico (protesi ginocchio sx); a ottobre 2015 ripresa dolore toracico, riscontro di TVP e TEP, NSTEMI > coro: malattia plurivasale : IVA con stenosi 50-60% al tratto medio, calcifica, Cx con stenosi critica al tratto medio in vaso di piccolo calibro, coronaria dx con stenosi critica ostiale e subocclusiva al tratto prossimale con circolo omo ed eterocoronarico > PCI+BMS.
FE 58%.
Indicato controllo coronarografico a risoluzione del quadro urologico (non chiaro dalla documentazione se mai eseguito).
Nel 2016 SPECT miocardica minime quote di ischemia.
Ultima visita cardio (9/2020) : sospetti episodi di angina, titolata terapia medica.
TD venitrin TD 15 mg, perindopril/amlodipina, pantorc 20, furosemide 2 cp x 2, cardura 4 mg x 2, allopurinolo 150 mg, ASA, metoprololo 50 mg x 2, cholecomb 5/10 Accesso in DEA il 25/2 ore 9.20 condotta da MSA, allertata per dispnea e astenia dal risveglio > all'ECG eseguito al domicilio RS 67 bpm, T negative in inferiore, scarsa progressione R in anteriore (sostanzialmente invariato rispetto a ECG del 2020).
All'arrivo in DEA PA 140/85 mmHg, FC 80 R, SpO2 80% in aa, apiretica.
Obiettivamente quadro di scompenso cardiaco, all'rx torace sovraccarico del piccolo circolo con iniziale interessamento interstizio-alveolare e addensamento vascolare delle ombre ilari. Agli ematici WBC 5.97, Hb 8.9, PLTs 124, INR 1.13, crea 4.3, urea 210, AST 30, ALT 35, bil tot 0.35, Na 140, k 4.9, PCRneg, TnT 0.08 > 0.08. Somministrato lasix 1 fl ore 20. Durante la notte in DEA allerta il personale per comparsa di dolore toracico oppressivo > eseguito ECG che documenta sopraST 1 mm orizzontale in aVR e sottoST discendente in laterale max 1 mm. EGA A p/f 126, pO2 63, pCO2 29, HCO3 18, k 4.9, lac 0.85, gli 145, Hb 9.6 Al mio arrivo la pz riferisce risoluzione della sintomatologia, eupnoica in MV 50%, PA 120/70 (riferisce episodio di dolore analogo ieri mattina). All'ecoscopia VS ai limiti superiori, FE conservata in presenza di sfumata ipocinesia della parete posteriore, IM moderato-severa da degenerazione dei lembi, degenerazione e ipomobilità delle cuspidi aortiche con iniziale accelerazione della flussimetria anterograda, IT moderata, VDx non dilatato normocontrattile, VCI 22 mm ipocollabente, non versamento pericardico.
CONCLUSIONI Dolore toracico con alterazioni ECG in pz con scompenso cardiaco e cardiopatia ischemica-valvolare, anemica, con IRC avanzata (eGFR 10 ml/min). Prosegue decongestione con lasix ev e avvia nitrato ev da titolare sec PA. Da rivedere con enzimi.

Variable: “neoplasia”
Classificazione: "possibilmente attiva"

Un prompt per variabile



Espliciti e strutturati
definizioni → etichette di classificazione → istruzioni → formato di output JSON

Esaustivi
classi di neoplasie, abbreviazioni (CA, HCC, K…), trattamenti (Sx, RT, CT, IT…)

(Possibilmente) scritti da esperti
definizioni cliniche e criteri decisionali formulati da specialisti del dominio

Classificazione neoplasia (1)

Classifica lo stato della neoplasia.

Dalla nota clinica, si evince che la paziente ha una pregressa neoplasia vescicale, trattata con resezione endoscopica nel 2015. Tuttavia, non è disponibile documentazione dettagliata in merito (es. istologia, stadiazione, follow-up oncologico).

Pertanto, lo stato della neoplasia può essere classificato come:

Neoplasia vescicale pregressa, trattata chirurgicamente (resezione endoscopica) nel 2015, senza evidenza attuale di recidiva documentata. Stato oncologico non valutabile con certezza per mancanza di documentazione.

In termini oncologici, si potrebbe indicare come:
- **Stato: remissione clinica presunta**, ma
- **Stato attuale: non documentabile / non valutabile (ND/NV).**

Classificazione neoplasia (2)

Dimmi se la neoplasia è certamente attiva, possibilmente attiva o certamente non attiva, oppure se non vi sono informazioni sufficienti per determinarlo.

Dalla nota clinica si legge:

"Pregressa neoplasia vescicale sottoposta a resezione endoscopica nel 2015 (non documentazione)."

Non vi sono ulteriori informazioni riguardo a follow-up oncologici, recidive, trattamenti successivi o segni clinici/laboratoristici di malattia attiva.

Pertanto, la risposta corretta è:

Non vi sono informazioni sufficienti per determinarlo.

Classificazione neoplasia (3)

Dimmi se la neoplasia è certamente attiva, possibilmente attiva, certamente non attiva oppure se non vi sono informazioni sufficienti per determinarlo. Dammi solo la classificazione senza ulteriore testo.

Non vi sono informazioni sufficienti per determinarlo.

Classificazione neoplasia (4)

Dimmi se la neoplasia è certamente attiva, possibilmente attiva o certamente non attiva.

Ecco dei criteri di classificazione:
- certamente attiva: diagnosi recente o trattamento in corso
- certamente non attiva: risoluzione recente
- possibilmente attiva: risoluzione dubbia
- non disponibile: non si parla della neoplasia nella nota clinica

Dammi solo la classificazione senza ulteriore testo.

possibilmente attiva

Classificazione neoplasia (5)

Dimmi se la neoplasia è certamente attiva, possibilmente attiva o certamente non attiva.

Ecco dei criteri di classificazione:
- certamente attiva: diagnosi recente o trattamento in corso
- certamente non attiva: risoluzione recente
- possibilmente attiva: risoluzione dubbia
- non disponibile: non si parla della neoplasia nella nota clinica

Dammi solo la classificazione senza ulteriore testo.
Voglio la risposta secondo questo formato JSON:
'stato': 'certamente attiva' | 'possibilmente attiva' | 'certamente non attiva' | 'non disponibile'
'ragionamento': 'testo che spiega il ragionamento alla base della classificazione'

{
"stato": "possibilmente attiva",
"ragionamento": "La paziente ha una storia di neoplasia vescicale trattata con resezione endoscopica nel 2015, ma non è disponibile documentazione che ne confermi la risoluzione o il follow-up. La nota menziona la necessità di risolvere un problema urologico prima di una coronarografia, il che suggerisce una possibile problematica urologica ancora in corso o non risolta. Pertanto, lo stato della neoplasia è incerto e potrebbe essere ancora attiva."
}

CONTESTO

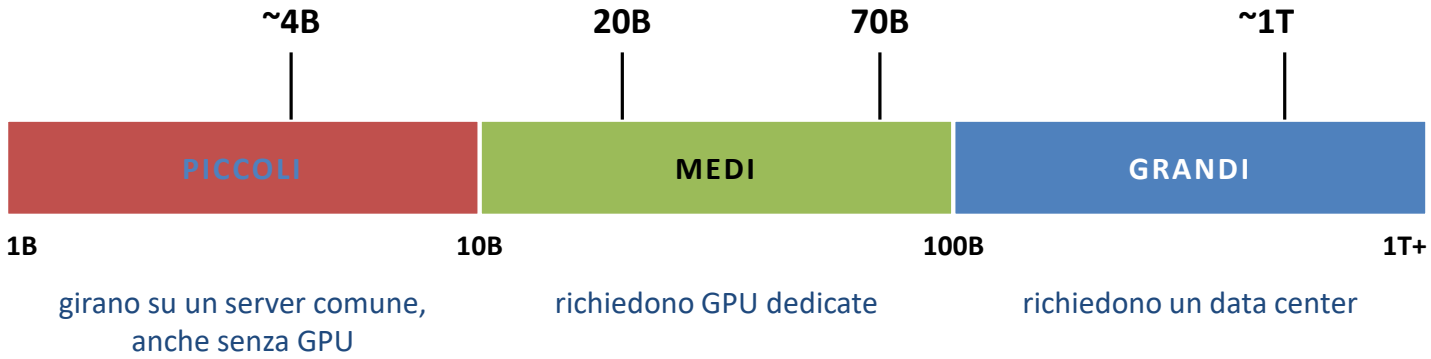
Quanto è grande un modello?

La dimensione si misura in parametri e determina capacità, ma anche risorse

Cosa sono i parametri

Sono i valori che il modello impara durante l'addestramento: più parametri, più conoscenza e capacità di ragionamento.

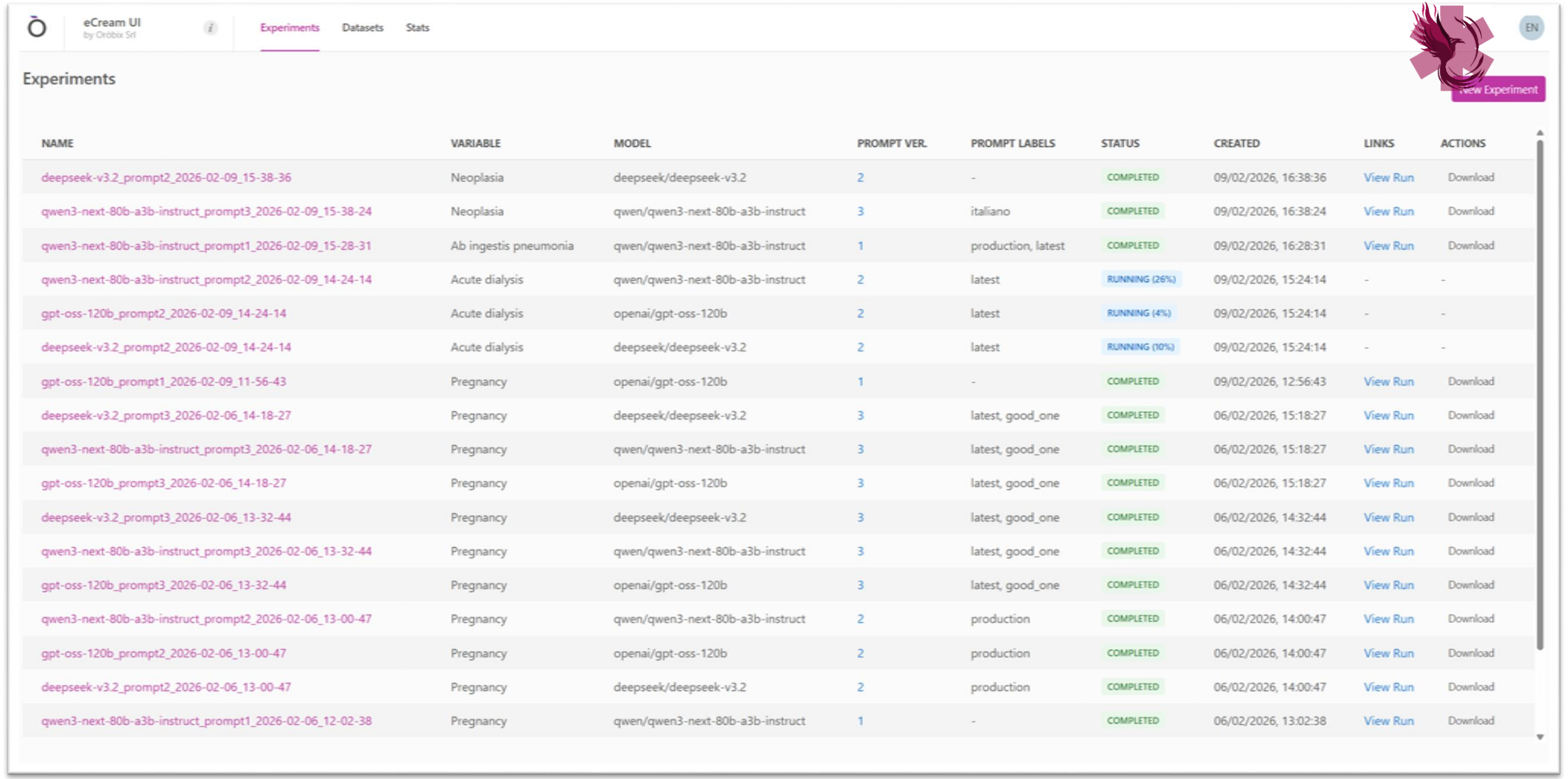
1B = 1 miliardo
1T = 1.000 miliardi



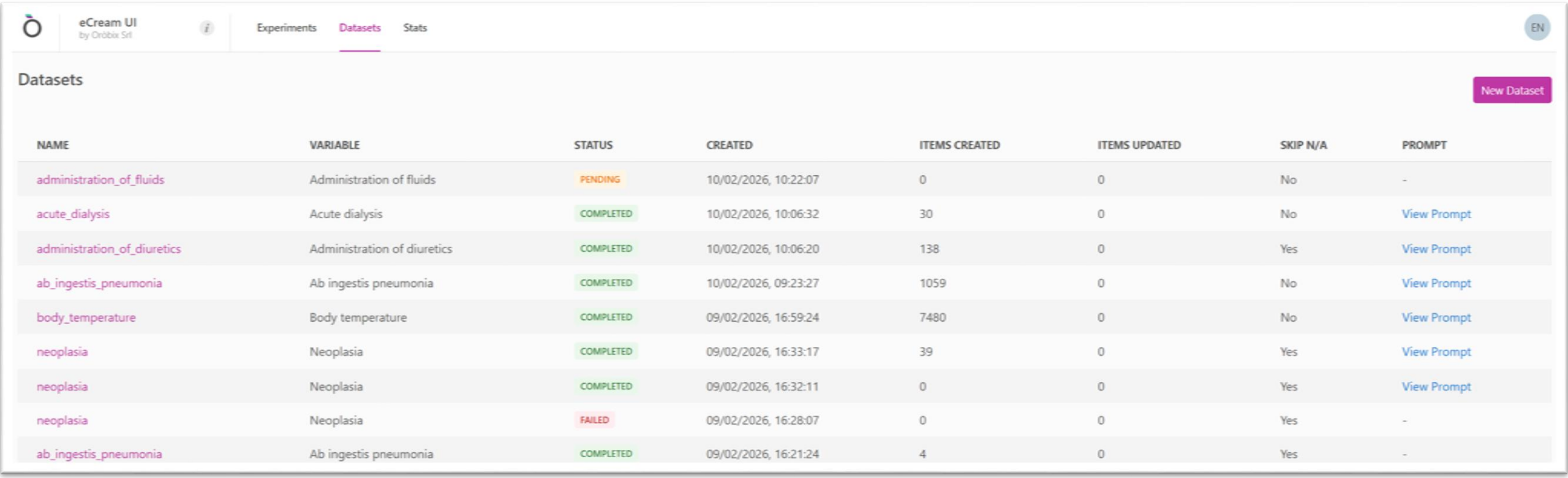
\+ Più parametri: più conoscenza e capacità di ragionamento

− Più parametri: più costi, energia e hardware necessari

La dimensione non è tutto: i modelli più recenti ottengono di più con meno parametri



NAME	VARIABLE	MODEL	PROMPT VER.	PROMPT LABELS	STATUS	CREATED	LINKS	ACTIONS
deepseek-v3.2_prompt2_2026-02-09_15-38-36	Neoplasia	deepseek/deepseek-v3.2	2	-	COMPLETED	09/02/2026, 16:38:36	View Run	Download
qwen3-next-80b-a3b-instruct_prompt3_2026-02-09_15-38-24	Neoplasia	qwen/qwen3-next-80b-a3b-instruct	3	italiano	COMPLETED	09/02/2026, 16:38:24	View Run	Download
qwen3-next-80b-a3b-instruct_prompt1_2026-02-09_15-28-31	Ab ingestis pneumonia	qwen/qwen3-next-80b-a3b-instruct	1	production, latest	COMPLETED	09/02/2026, 16:28:31	View Run	Download
qwen3-next-80b-a3b-instruct_prompt2_2026-02-09_14-24-14	Acute dialysis	qwen/qwen3-next-80b-a3b-instruct	2	latest	RUNNING (26%)	09/02/2026, 15:24:14	-	-
gpt-oss-120b_prompt2_2026-02-09_14-24-14	Acute dialysis	openai/gpt-oss-120b	2	latest	RUNNING (4%)	09/02/2026, 15:24:14	-	-
deepseek-v3.2_prompt2_2026-02-09_14-24-14	Acute dialysis	deepseek/deepseek-v3.2	2	latest	RUNNING (10%)	09/02/2026, 15:24:14	-	-
gpt-oss-120b_prompt1_2026-02-09_11-56-43	Pregnancy	openai/gpt-oss-120b	1	-	COMPLETED	09/02/2026, 12:56:43	View Run	Download
deepseek-v3.2_prompt3_2026-02-06_14-18-27	Pregnancy	deepseek/deepseek-v3.2	3	latest, good_one	COMPLETED	06/02/2026, 15:18:27	View Run	Download
qwen3-next-80b-a3b-instruct_prompt3_2026-02-06_14-18-27	Pregnancy	qwen/qwen3-next-80b-a3b-instruct	3	latest, good_one	COMPLETED	06/02/2026, 15:18:27	View Run	Download
gpt-oss-120b_prompt3_2026-02-06_14-18-27	Pregnancy	openai/gpt-oss-120b	3	latest, good_one	COMPLETED	06/02/2026, 15:18:27	View Run	Download
deepseek-v3.2_prompt3_2026-02-06_13-32-44	Pregnancy	deepseek/deepseek-v3.2	3	latest, good_one	COMPLETED	06/02/2026, 14:32:44	View Run	Download
qwen3-next-80b-a3b-instruct_prompt3_2026-02-06_13-32-44	Pregnancy	qwen/qwen3-next-80b-a3b-instruct	3	latest, good_one	COMPLETED	06/02/2026, 14:32:44	View Run	Download
gpt-oss-120b_prompt3_2026-02-06_13-32-44	Pregnancy	openai/gpt-oss-120b	3	latest, good_one	COMPLETED	06/02/2026, 14:32:44	View Run	Download
qwen3-next-80b-a3b-instruct_prompt2_2026-02-06_13-00-47	Pregnancy	qwen/qwen3-next-80b-a3b-instruct	2	production	COMPLETED	06/02/2026, 14:00:47	View Run	Download
gpt-oss-120b_prompt2_2026-02-06_13-00-47	Pregnancy	openai/gpt-oss-120b	2	production	COMPLETED	06/02/2026, 14:00:47	View Run	Download
deepseek-v3.2_prompt2_2026-02-06_13-00-47	Pregnancy	deepseek/deepseek-v3.2	2	production	COMPLETED	06/02/2026, 14:00:47	View Run	Download
qwen3-next-80b-a3b-instruct_prompt1_2026-02-06_12-02-38	Pregnancy	qwen/qwen3-next-80b-a3b-instruct	1	-	COMPLETED	06/02/2026, 13:02:38	View Run	Download


eCream UI
by Orobix Srl
Experiments
Datasets
Stats
EN
Datasets
New Dataset
NAME
VARIABLE
STATUS
CREATED
ITEMS CREATED
ITEMS UPDATED
SKIP N/A
PROMPT
administration_of_fluids
Administration of fluids
PENDING
10/02/2026, 10:22:07
0
0
No
-
acute_dialysis
Acute dialysis
COMPLETED
10/02/2026, 10:06:32
30
0
No
View Prompt
administration_of_diuretics
Administration of diuretics
COMPLETED
10/02/2026, 10:06:20
138
0
Yes
View Prompt
ab_ingestis_pneumonia
Ab ingestis pneumonia
COMPLETED
10/02/2026, 09:23:27
1059
0
No
View Prompt
body_temperature
Body temperature
COMPLETED
09/02/2026, 16:59:24
7480
0
No
View Prompt
neoplasia
Neoplasia
COMPLETED
09/02/2026, 16:33:17
39
0
Yes
View Prompt
neoplasia
Neoplasia
COMPLETED
09/02/2026, 16:32:11
0
0
Yes
View Prompt
neoplasia
Neoplasia
FAILED
09/02/2026, 16:28:07
0
0
Yes
-
ab_ingestis_pneumonia
Ab ingestis pneumonia
COMPLETED
09/02/2026, 16:21:24
4
0
Yes
-

eCream UI
by Orobix Srl

Experiments Datasets Stats

EN

Stats

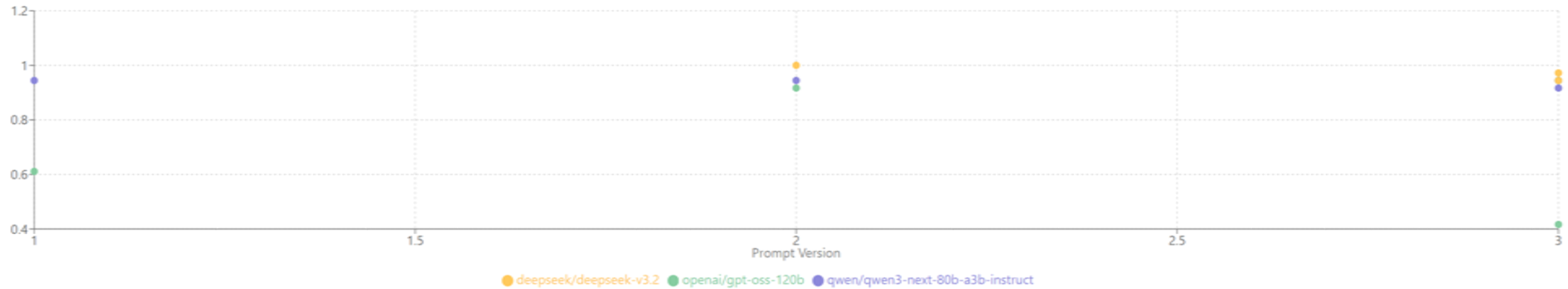
VARIABLE

Pregnancy

PLOTS

Accuracy × Avg class precision ×
Avg class recall × N precision × N recall ×
Runerror × Y precision × Y recall ×

Accuracy



Avg class precision



TASK · ESTRAZIONE DI INFORMAZIONI DA NOTE CLINICHE

Cosa abbiamo imparato

01

VALUTAZIONE

Valutare le prestazioni non è banale e dipende dalla variabile

Prestazioni migliori per le variabili che non richiedono conoscenze di dominio

02

PORTABILITÀ

I prompt sembrano trasferibili tra modelli diversi per il nostro task

Accuratezza indipendente dalla lingua del prompt rispetto a quella della nota

03

CHI SCRIVE I PROMPT

Prompt meno dettagliati portano a un’accuratezza inferiore

I prompt vanno scritti da esperti di dominio; gli esperti tecnici aiutano con la struttura

04

QUALITÀ DEI DATI

Il confronto tra modello ed esperti è utile di per sé

Sulle discordanze a volte vince il modello, a volte l’esperto; più spesso è la categoria a essere ambigua e va rivista

VALUTAZIONE · LIVELLO DI COSCIENZA

Anche il ground truth è ambiguo

L'LLM inferisce informazioni implicite che l'annotatore ha scelto di non annotare

ESEMPIO

Nota clinica: *«Il paziente riferisce che…»*

ANNOTATORE
N.D. · non esplicito nel testo

LLM
A · paziente vigile

57 su 199

casi «non disponibile» per l'annotatore che l'LLM classifica come «vigile»

RIGHE: ANNOTATORE · COLONNE: LLM

	A	V	P	U	N.D.
A	**84**	2	0	0	1
V	0	**2**	0	0	0
P	0	0	**2**	0	0
U	0	0	0	0	0
N.D.	**57**	4	0	1	**137**

Scala AVPU: A = vigile, V = risponde alla voce, P = risponde al dolore, U = non risponde · N.D. = non disponibile

Non sempre sono errori: va deciso a monte se l'inferenza è ammessa, nel prompt e nelle linee guida

TASK · ESTRAZIONE DI INFORMAZIONI DA NOTE CLINICHE

L’LLM come motore di ragionamento

La conoscenza di dominio la forniamo noi nel prompt: l’LLM la applica alla nota clinica

DA EVITARE

LLM come fonte di conoscenza

Prompt generico: «Il paziente ha una neoplasia attiva?»

- Il modello usa definizioni implicite
- Criteri di decisione non controllabili
- Risposte difficili da verificare

vs

IL NOSTRO APPROCCIO

LLM come motore di ragionamento

Prompt con il contesto clinico fornito in dettaglio

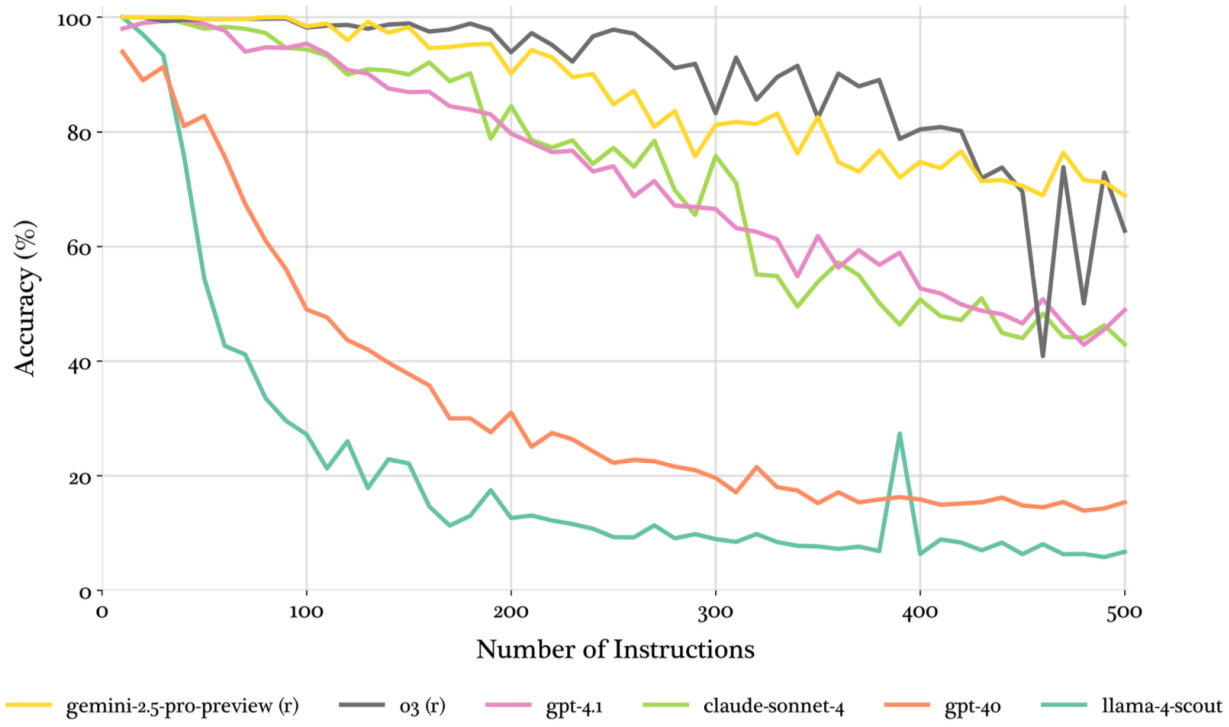
- Definizioni e criteri di classificazione espliciti
- Abbreviazioni e trattamenti del dominio
- Risposta motivata con evidenze dalla nota

Più il contesto è dettagliato, più il ragionamento è affidabile

Large vs Small - Ottobre 2025

Model Size	VRAM (FP16)	VRAM (4-bit)	Local Hardware	Best Use Cases
1–3B	4–6 GB	~2 GB	RTX 3060, laptop GPUs	Basic chat, text classification, autocomplete
7–8B	14–16 GB	~6–8 GB	RTX 4080/4090, A6000	General-purpose assistants, summarization, coding
13–14B	26–28 GB	~12–16 GB	RTX 4090 (quantized only)	Stronger reasoning, better instruction following
70B+	140 GB+	~35–40 GB	Multi-GPU setups (expensive)	SOTA reasoning, enterprise applications

Instruction following - Ottobre 2025



How Many Instructions Can LLMs Follow at Once?

Daniel Jaroslawicz[1] **Brendan Whiting**[1] **Parth Shah**[1] **Karime Maamari**[1]
[1]Distyl AI
{daniel, brendan, parth, karime}@distyl.ai

Instruction following - Ottobre 2025

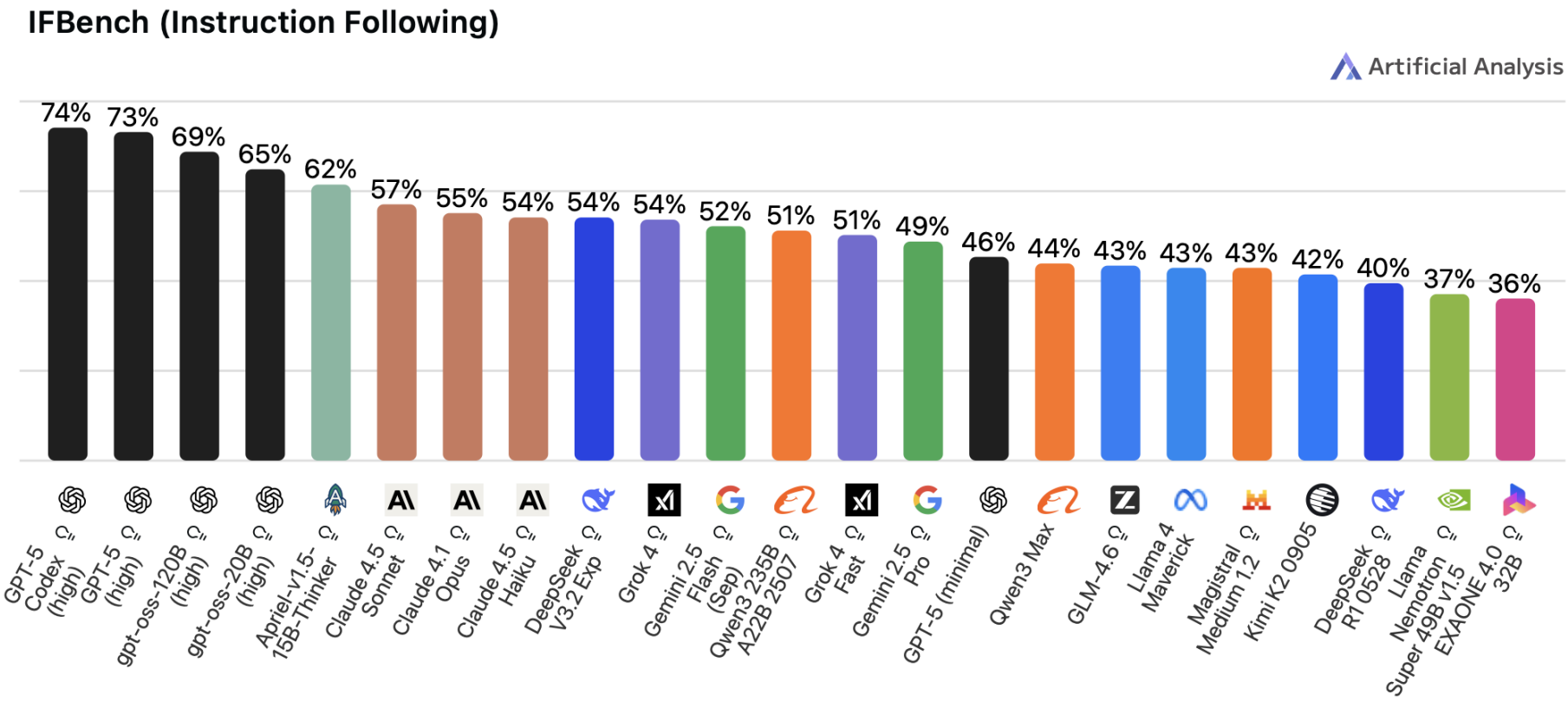
“Modelli di taglia intermedia (20B-70B):

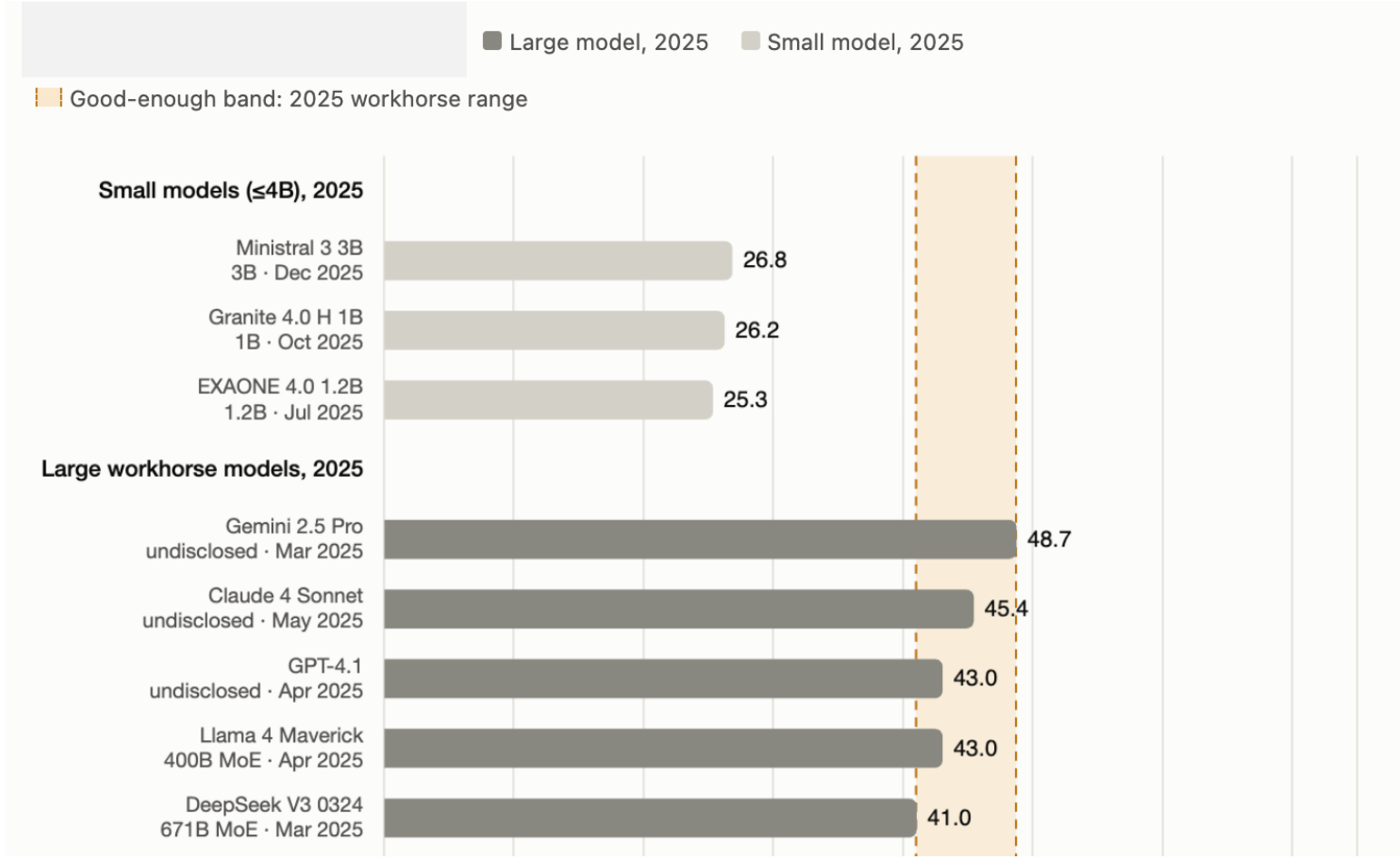
Non sono sempre affidabili rispetto al formato dell’output:
e.g. invalid JSON: 5% con GPT-OSS-20B, 10% llama 3 70B

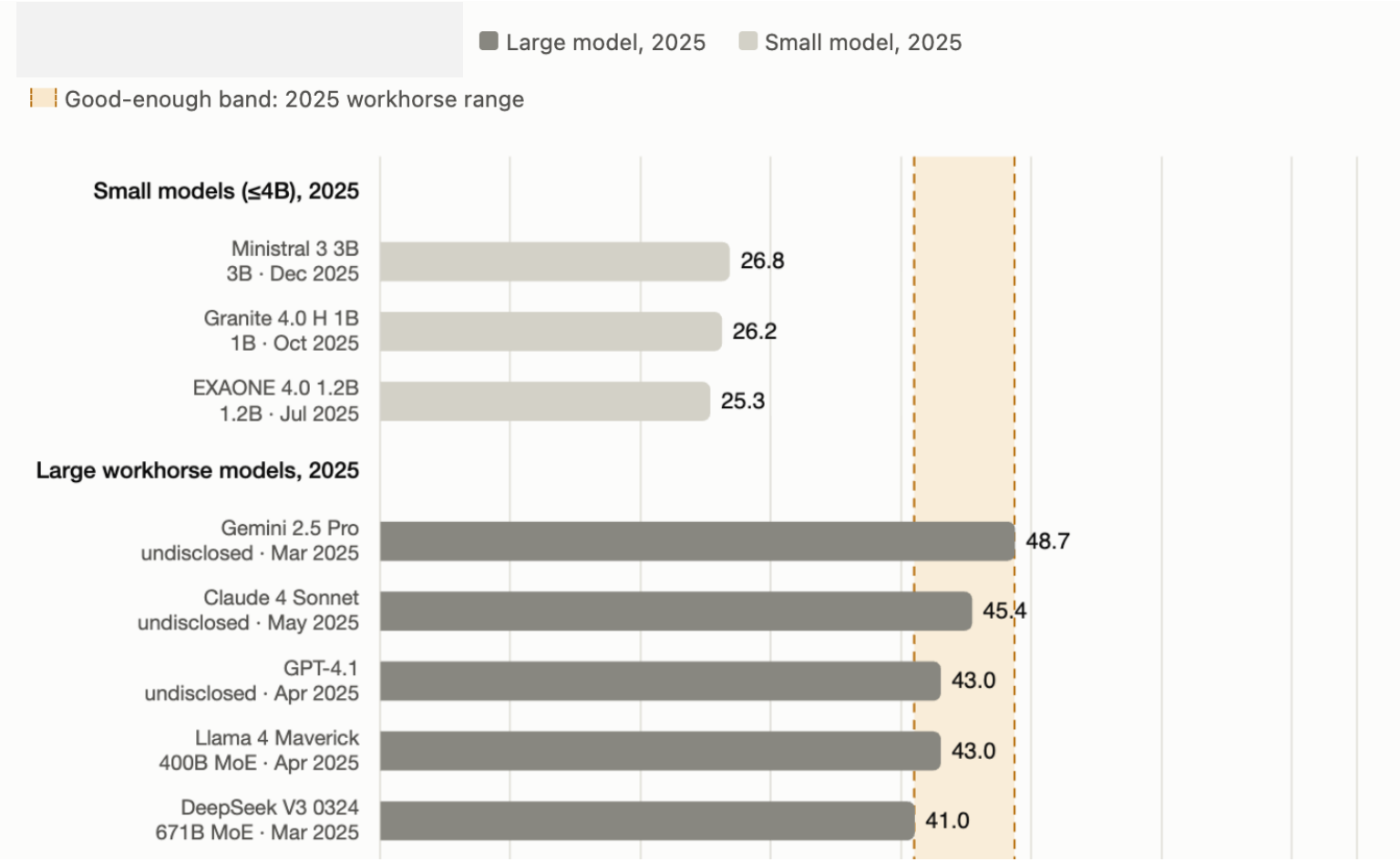
È una possibile indicazione che potrebbero non seguire anche istruzioni concettuali, meno evidenti
Modelli più recenti migliorano l’aderenza alle istruzioni”

IFBench - Ottobre 2025

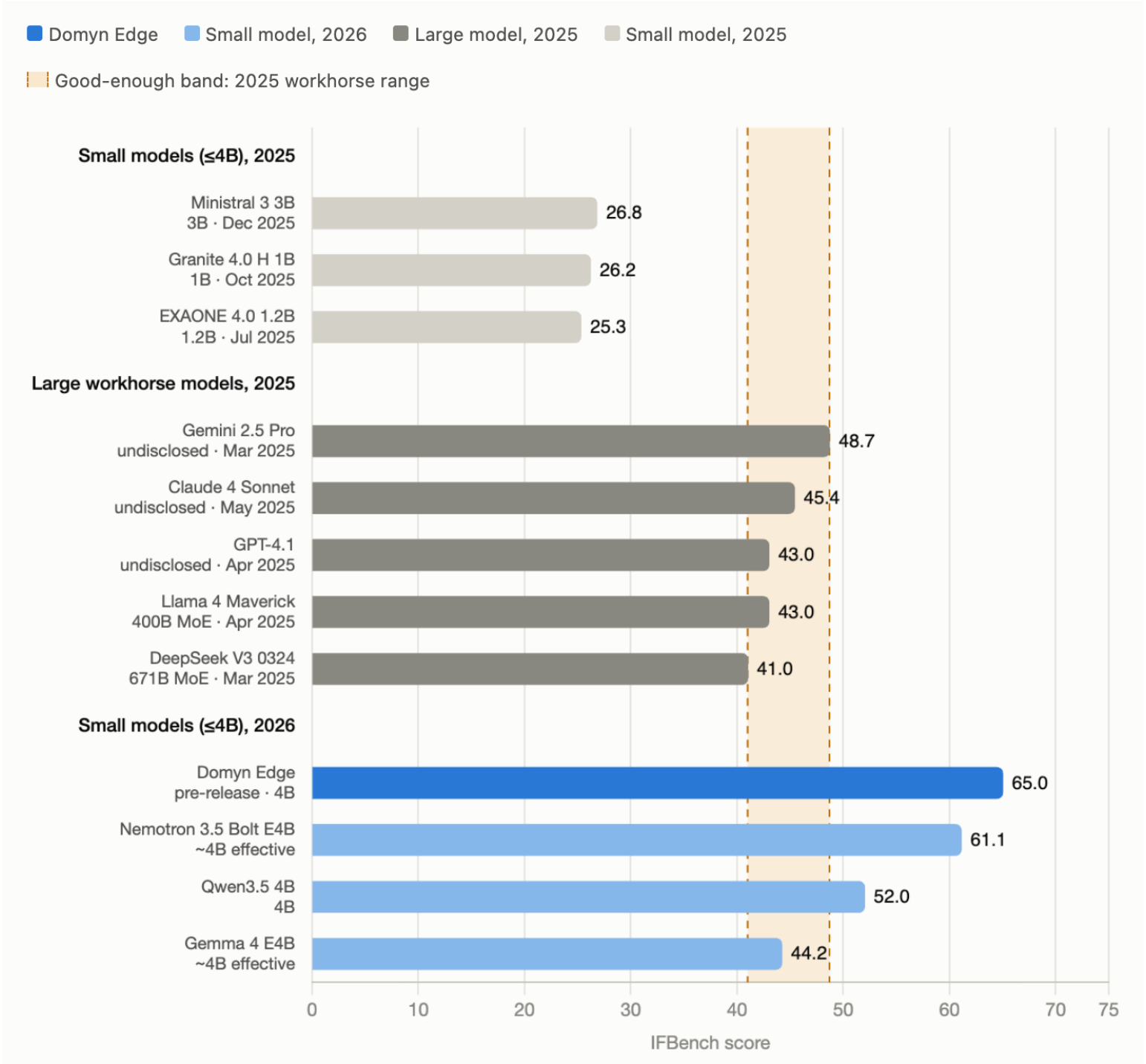
Write an email to Matthew Greene asking him about potential postdoc opportunities. Talk about the work that I worked on and how close it is his and how impactful my contribution will be in his lab. Please personalize it. Matthew Greene works on: interests in the development of mobile sensing (phones, wearables), intervention and AI technology capable of accurately assessing and managing mental illness (e.g., anxiety, depression, schizophrenia) at population-scale. I specialize in building large-scale, deployable sensing systems for smart environments and developing foundation machine learning models for human activity recognition and various health and well-being applications. Include keyword cloud once in your response, keyword rhythm twice in your response, keyword fish three times in your response, keyword puzzle five times in your response, and keyword dawn seven times in your response.

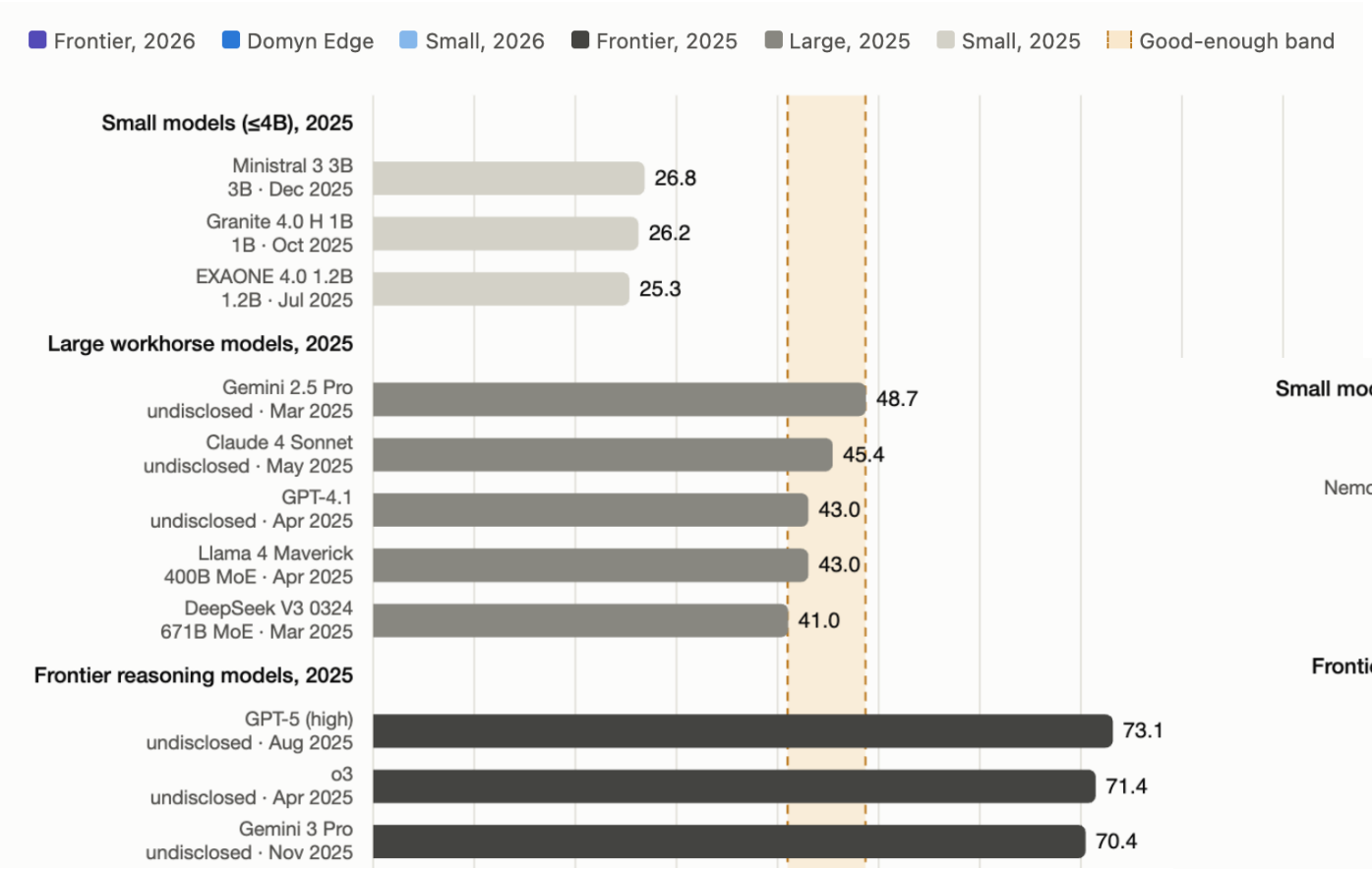


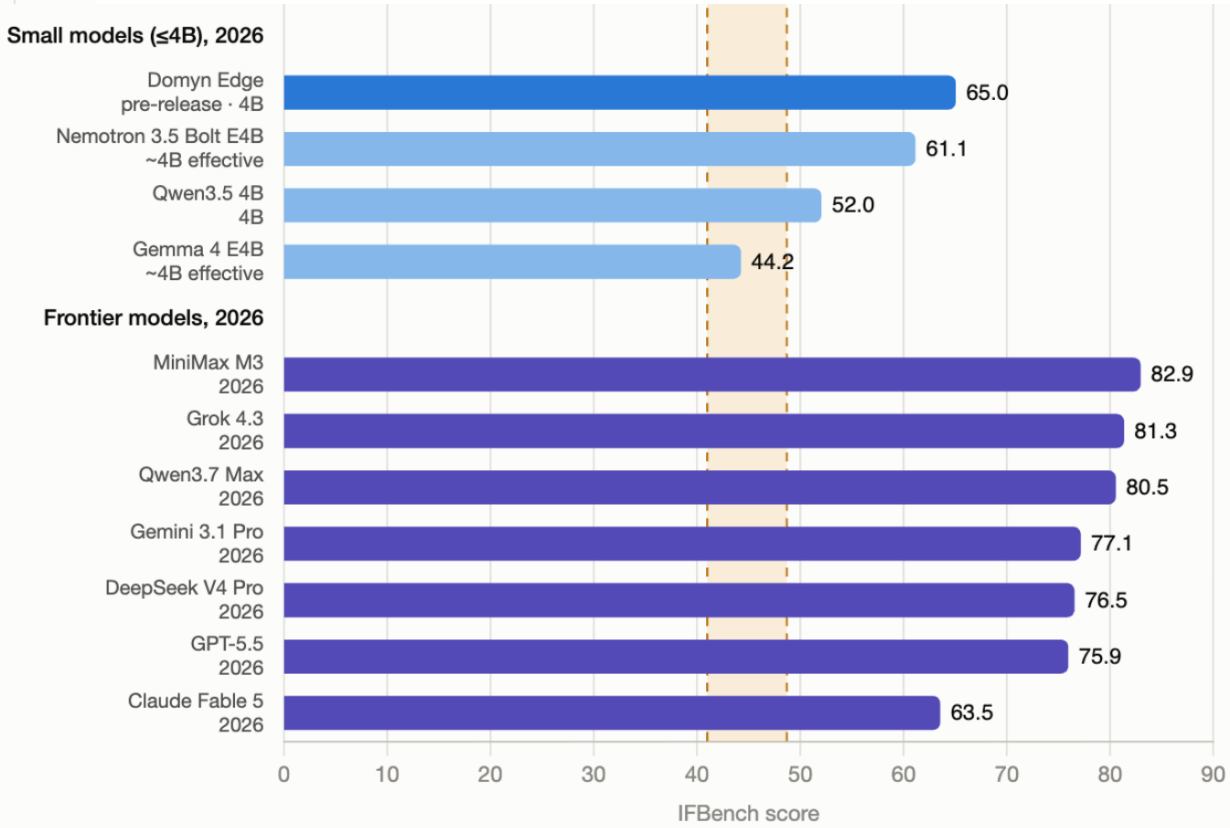

Large model, 2025
Small model, 2025
Good-enough band: 2025 workhorse range
Small models (≤4B), 2025
Ministral 3 3B
3B · Dec 2025
26.8
Granite 4.0 H 1B
1B · Oct 2025
26.2
EXAONE 4.0 1.2B
1.2B · Jul 2025
25.3
Large workhorse models, 2025
Gemini 2.5 Pro
undisclosed · Mar 2025
48.7
Claude 4 Sonnet
undisclosed · May 2025
45.4
GPT-4.1
undisclosed · Apr 2025
43.0
Llama 4 Maverick
400B MoE · Apr 2025
43.0
DeepSeek V3 0324
671B MoE · Mar 2025
41.0


Large model, 2025
Small model, 2025
Good-enough band: 2025 workhorse range
Small models (≤4B), 2025
Ministral 3 3B
3B · Dec 2025
26.8
Granite 4.0 H 1B
1B · Oct 2025
26.2
EXAONE 4.0 1.2B
1.2B · Jul 2025
25.3
Large workhorse models, 2025
Gemini 2.5 Pro
undisclosed · Mar 2025
48.7
Claude 4 Sonnet
undisclosed · May 2025
45.4
GPT-4.1
undisclosed · Apr 2025
43.0
Llama 4 Maverick
400B MoE · Apr 2025
43.0
DeepSeek V3 0324
671B MoE · Mar 2025
41.0

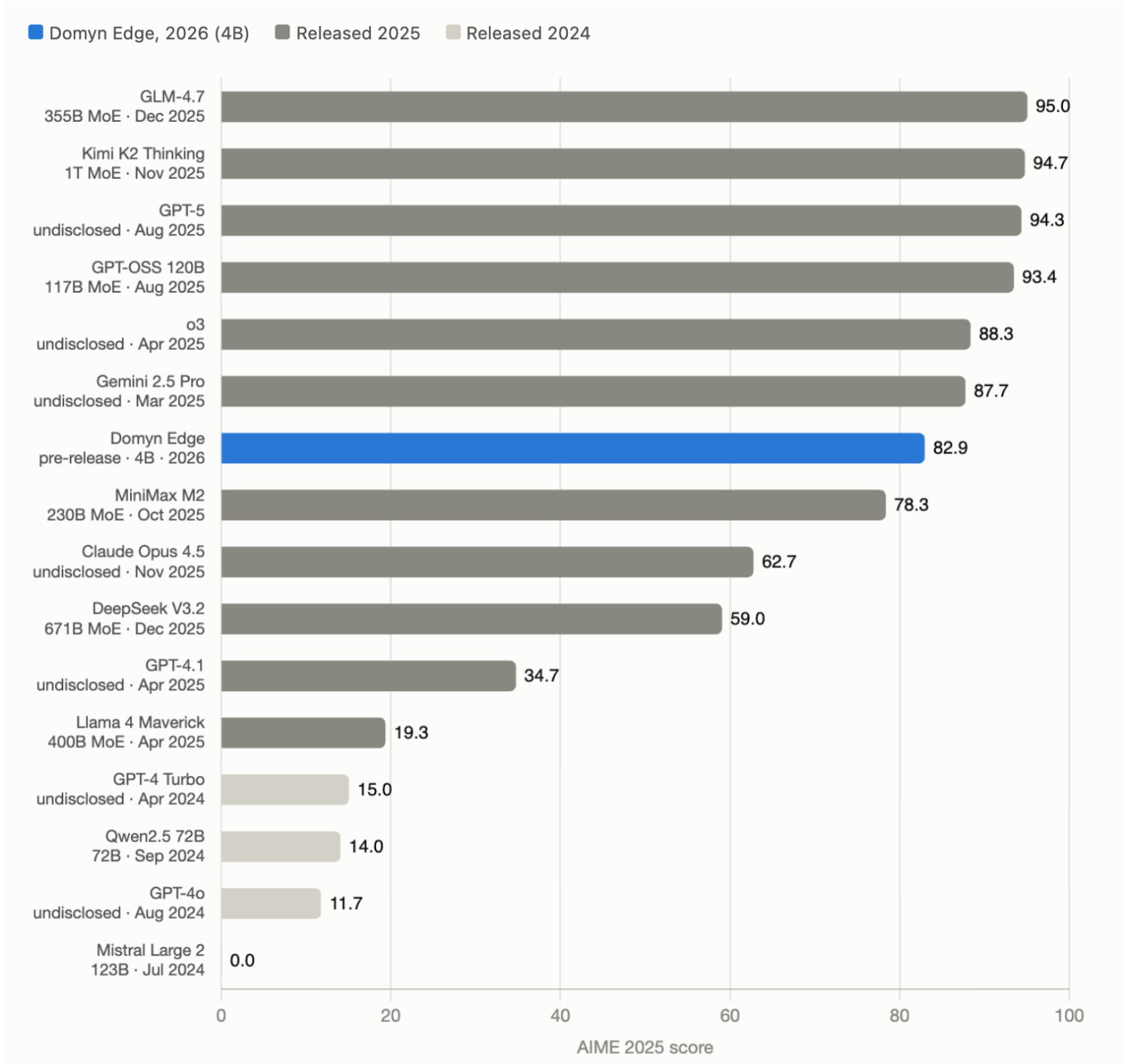
Un anno dopo…

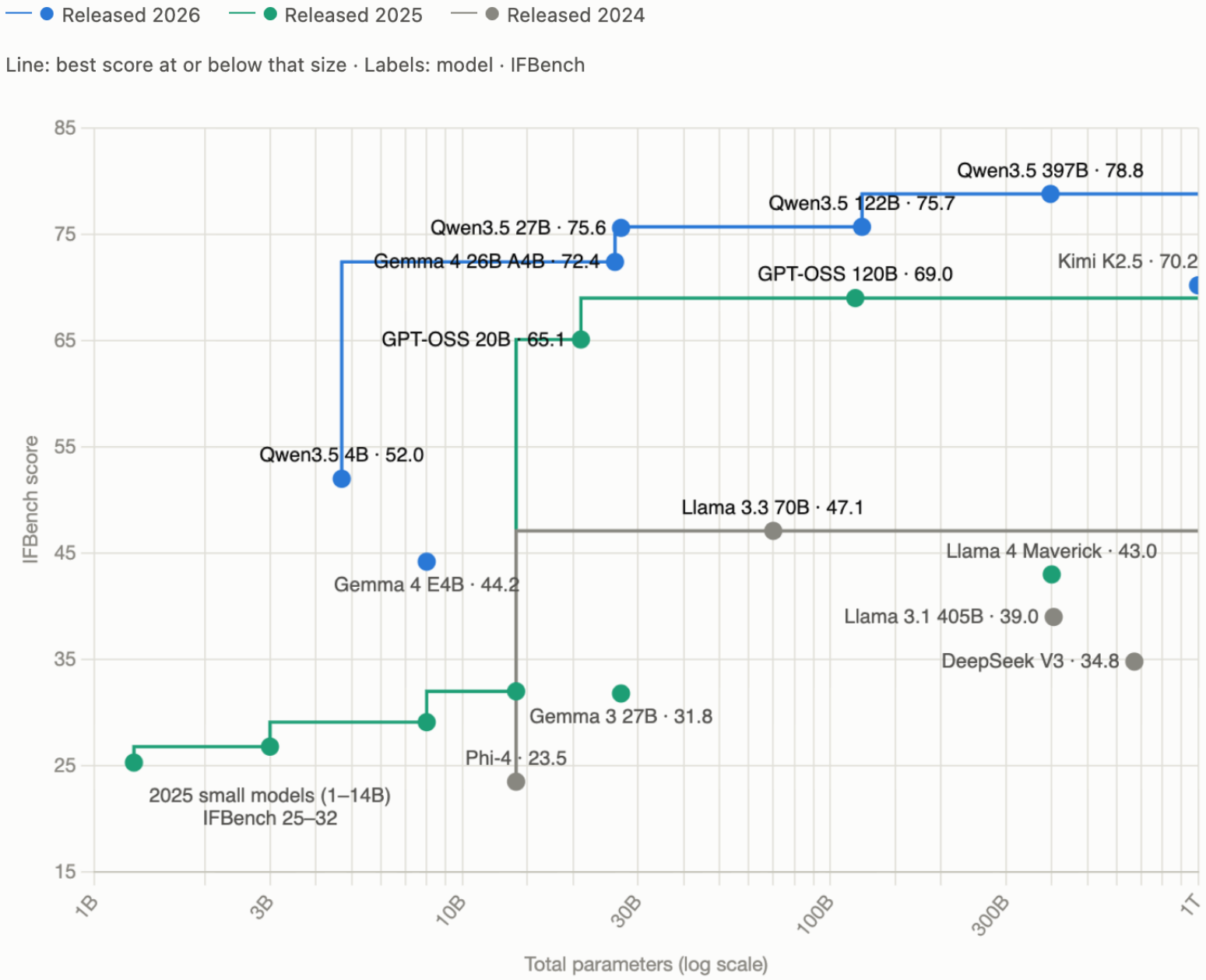

Domyn Edge
Small model, 2026
Large model, 2025
Small model, 2025
Good-enough band: 2025 workhorse range
Small models (≤4B), 2025
Ministral 3 3B
3B · Dec 2025
26.8
Granite 4.0 H 1B
1B · Oct 2025
26.2
EXAONE 4.0 1.2B
1.2B · Jul 2025
25.3
Large workhorse models, 2025
Gemini 2.5 Pro
undisclosed · Mar 2025
48.7
Claude 4 Sonnet
undisclosed · May 2025
45.4
GPT-4.1
undisclosed · Apr 2025
43.0
Llama 4 Maverick
400B MoE · Apr 2025
43.0
DeepSeek V3 0324
671B MoE · Mar 2025
41.0
Small models (≤4B), 2026
Domyn Edge
pre-release · 4B
65.0
Nemotron 3.5 Bolt E4B
~4B effective
61.1
Qwen3.5 4B
4B
52.0
Gemma 4 E4B
~4B effective
44.2
0
10
20
30
40
50
60
70
75
IFBench score





IFBench. All scores: independent runs by Artificial Analysis (via BenchLM, Nov 2026), except Domyn Edge: Domyn internal evaluation of an unreleased checkpoint. Good-enough band: range of 2025 workhorse large models.


Domyn Edge, 2026 (4B)
Released 2025
Released 2024
GLM-4.7
355B MoE · Dec 2025
95.0
Kimi K2 Thinking
1T MoE · Nov 2025
94.7
GPT-5
undisclosed · Aug 2025
94.3
GPT-OSS 120B
117B MoE · Aug 2025
93.4
o3
undisclosed · Apr 2025
88.3
Gemini 2.5 Pro
undisclosed · Mar 2025
87.7
Domyn Edge
pre-release · 4B · 2026
82.9
MiniMax M2
230B MoE · Oct 2025
78.3
Claude Opus 4.5
undisclosed · Nov 2025
62.7
DeepSeek V3.2
671B MoE · Dec 2025
59.0
GPT-4.1
undisclosed · Apr 2025
34.7
Llama 4 Maverick
400B MoE · Apr 2025
19.3
GPT-4 Turbo
undisclosed · Apr 2024
15.0
Qwen2.5 72B
72B · Sep 2024
14.0
GPT-4o
undisclosed · Aug 2024
11.7
Mistral Large 2
123B · Jul 2024
0.0
0
20
40
60
80
100
AIME 2025 score



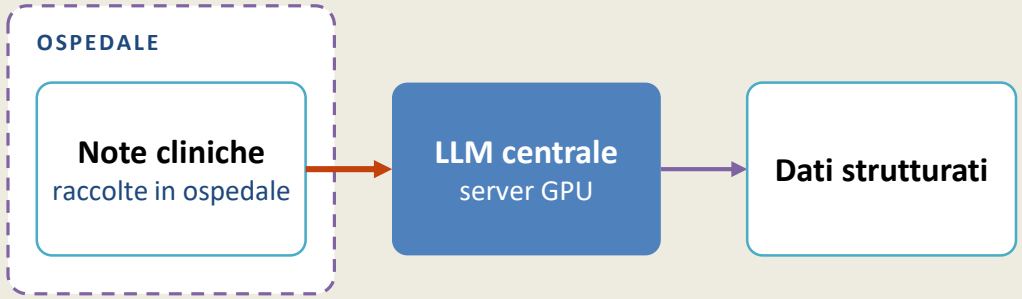
Open-weight models only; total parameters (GPT-OSS 20B: 3.6B active; Qwen3.5 4B: 4B language model + vision encoder; Gemma 4 E4B: 8B total). Llama 3.3 70B released Dec 2024. IFBench: Artificial Analysis (via BenchLM, Nov 2026).

PROSPETTIVE PER IL FUTURO

Scenari architetturali

STATO ATTUALE

Modello centralizzato



Escono dall'ospedale: le note cliniche

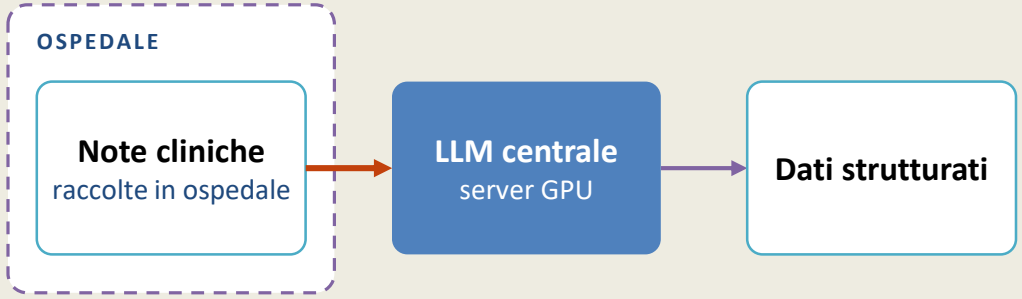
Un unico modello grande, su infrastruttura centrale, elabora le note di tutti gli ospedali

PROSPETTIVE PER IL FUTURO

Scenari architetturali

STATO ATTUALE

Modello centralizzato

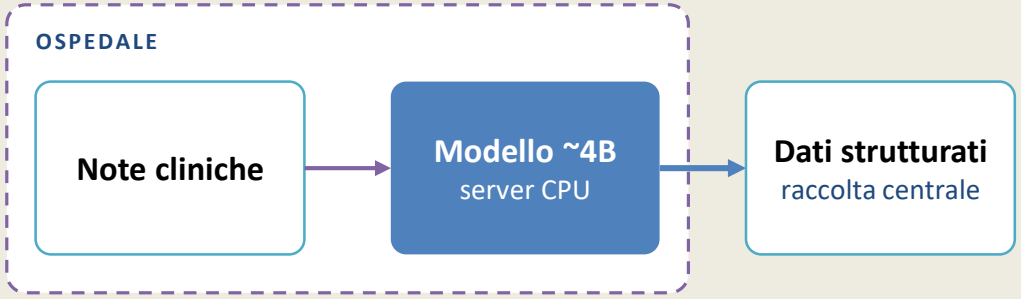


Escono dall'ospedale: le note cliniche

Un unico modello grande, su infrastruttura centrale, elabora le note di tutti gli ospedali

SCENARIO POSSIBILE

Elaborazione in ospedale



Escono dall'ospedale: solo i dati strutturati

Un modello piccolo gira localmente in ogni ospedale, anche senza GPU

PROSPETTIVE PER IL FUTURO

Nuove possibilità



Domyn Edge: Introducing a Compact Model Built in the Open

Domyn Edge is a 4B-parameter open model trained from scratch on 13 trillion tokens, with weights, datasets, checkpoints and training code released under open licenses.



Specializing NVIDIA Nemotron 3.5 Super VL for better financial knowledge graph extraction

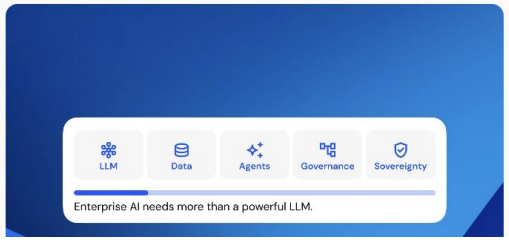
Fine-tuned NVIDIA Nemotron 3.5 Super VL doubles triple recall in one-pass 10-K knowledge graph extraction and halves rule violations on unseen companies.



Accelerating Sovereign AI: Domyn's Journey with NVIDIA

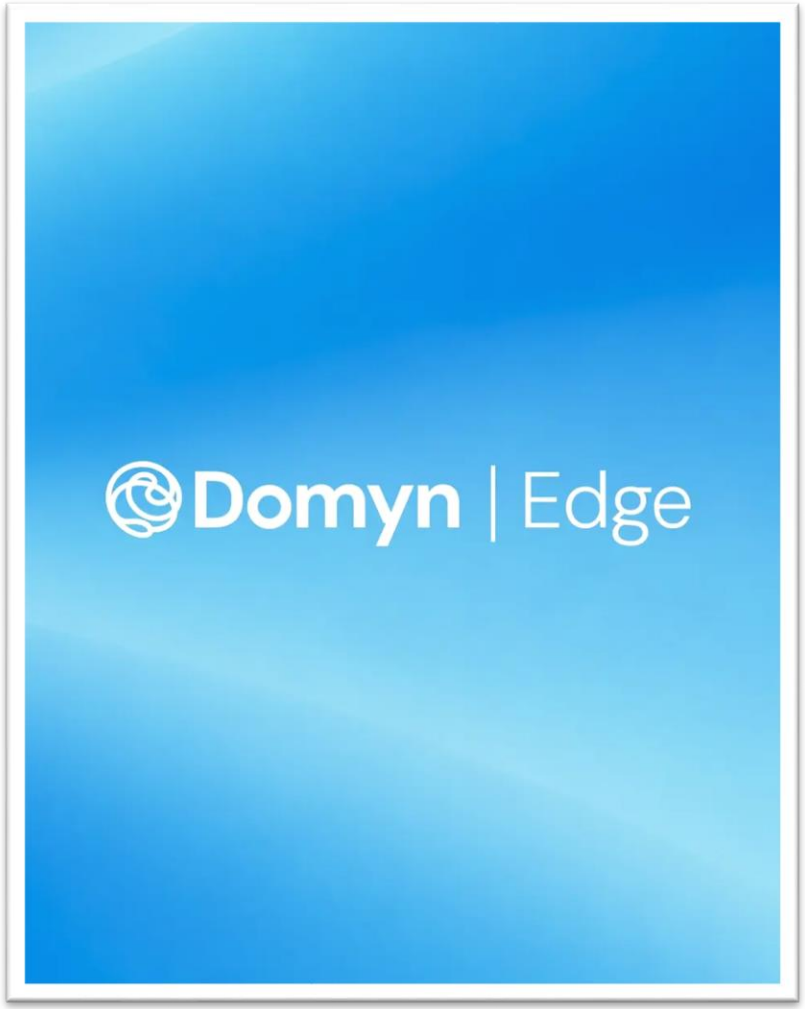
How Domyn built sovereign AI models (Italia, Colosseum, Domyn Large/Small) using NVIDIA NeMo, Megatron-Core & Leonardo supercomputer. Technical insights on pruning, distillation, RL & EU AI Act compliance.







Domyn Edge



4B parametri, addestrato da zero su 13T token

Rilascio **completamente aperto**:

- **pesi del modello** a diverse fasi dell'addestramento
- **codice** di addestramento (GitHub)
- **dati** di addestramento (HuggingFace)

Addestrato su cluster pubblico (Leonardo HPC)



Domyn Edge

Domyn Edge: six pre-training stages over 13 trillion tokens

The bar is to scale. Stage 1 is 65% of the token budget; the two context extensions together are 0.6T tokens, under 5% of the run. Colour marks the context regime the model is training at, not the stage.



1

Pre-training
0.0T to 8.5T
8.5T tokens
8k sequences

Web, code, STEM, multilingual and synthetic data

2

Specialized
8.5T to 10.1T
1.6T tokens
16k sequences

Reasoning, maths and code upweighted

3

Context extension
10.1T to 10.3T
0.2T tokens
to 64k

First step toward long context

4

Mid-training
10.3T to 12.3T
2.0T tokens
32k sequences

Agentic tool-calling and coding trajectories

5

Context extension
12.3T to 12.7T
0.4T tokens
to 1M

Second step, reaching the 1M training context

6

Annealing
12.7T to 13.0T
0.3T tokens
to 512k sequences

Final learning-rate decay

8k context · 16k context · 32k context · 64k context · Long context, up to 1M · Released checkpoint

Domyn Edge

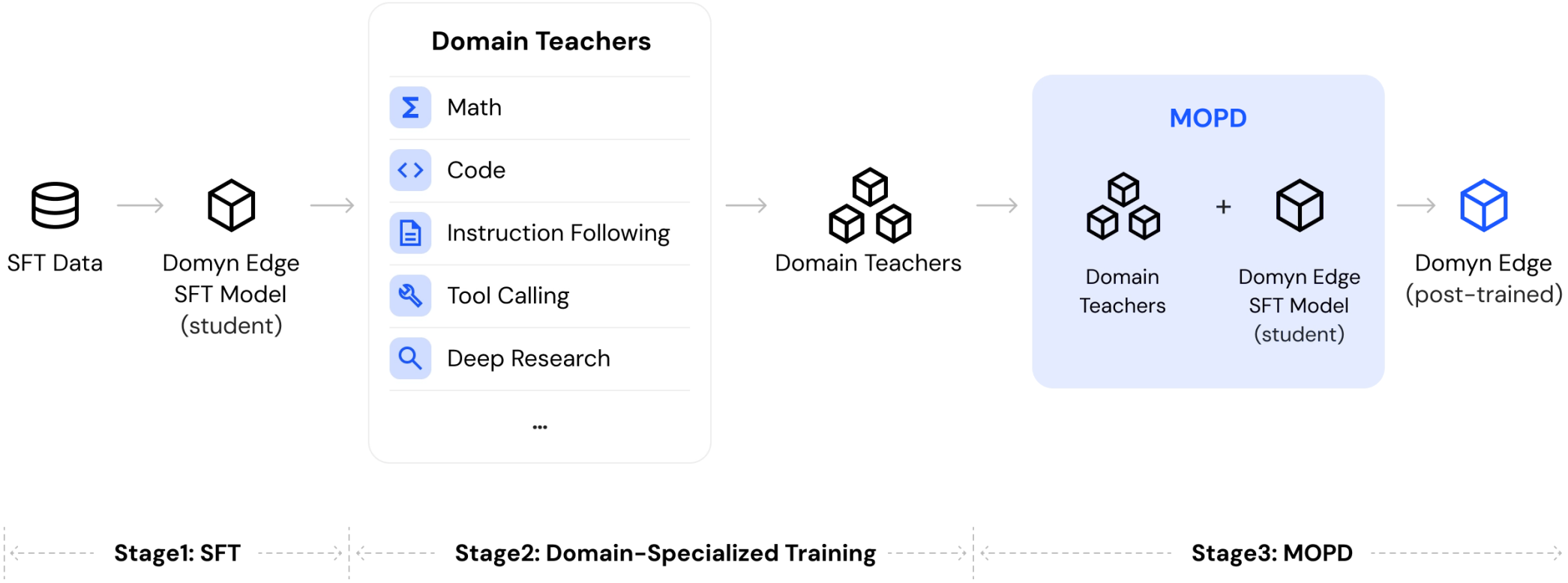


3198 11 734 11 1115 11

Prediction

#	probs	next token ID	predicted next token
0	46.44%	1440	four
1	7.48%	290	and
2	7.31%	1936	five
3	2.66%	393	or
4	2.54%	2237	six
5	2.09%	1115	three
6	1.86%	3863	maybe
7	1.62%	345	you
8	1.23%	257	a
9	0.92%	530	one

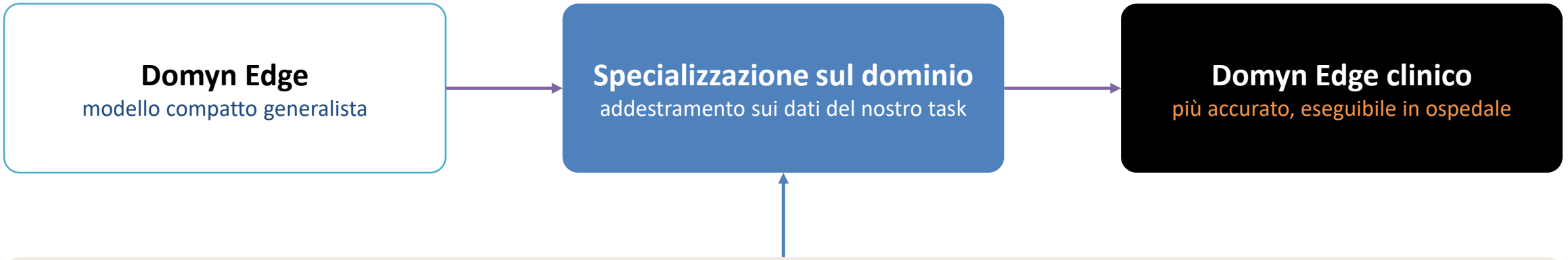
Domyn Edge


SFT Data
Domyn Edge SFT Model (student)
Domain Teachers
Math
Code
Instruction Following
Tool Calling
Deep Research
...
Domain Teachers
MOPD
Domain Teachers
+
Domyn Edge SFT Model (student)
Domyn Edge (post-trained)
Stage1: SFT
Stage2: Domain-Specialized Training
Stage3: MOPD

PROSPETTIVE · OPPORTUNITÀ

Specializzare Domyn Edge sul dominio clinico

Un modello compatto addestrato sul nostro task può guadagnare accuratezza restando eseguibile in ospedale



INGREDIENTI GIÀ DISPONIBILI DAL PROGETTO

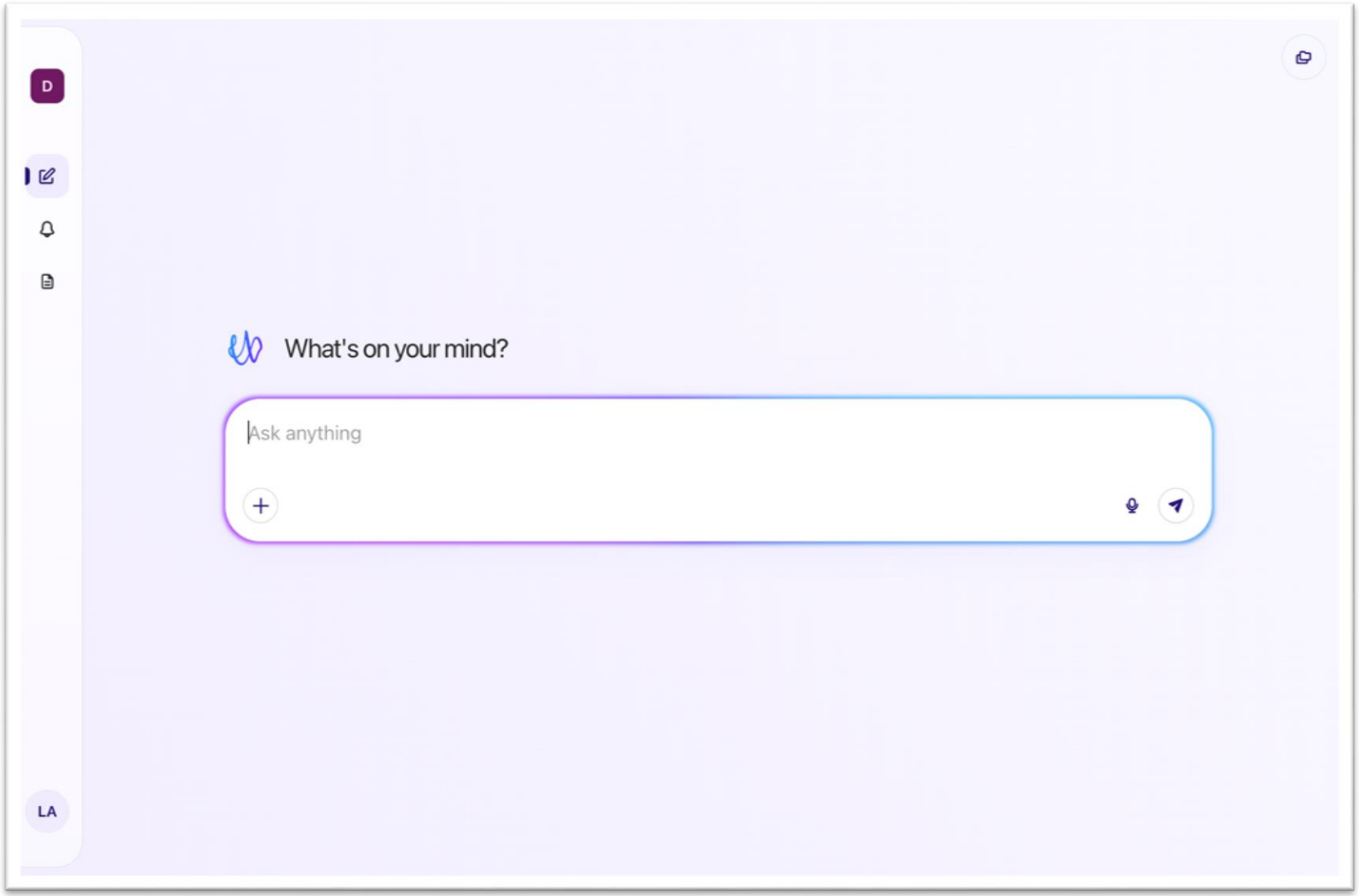
Prompt degli esperti
definizioni e criteri per ogni variabile

Note cliniche annotate
ground truth per addestrare e valutare

Output del modello grande
etichette motivate, utili come dati di distillazione

Obiettivo: avvicinare l'accuratezza del modello grande, mantenendo il deployment locale in ospedale

Demo



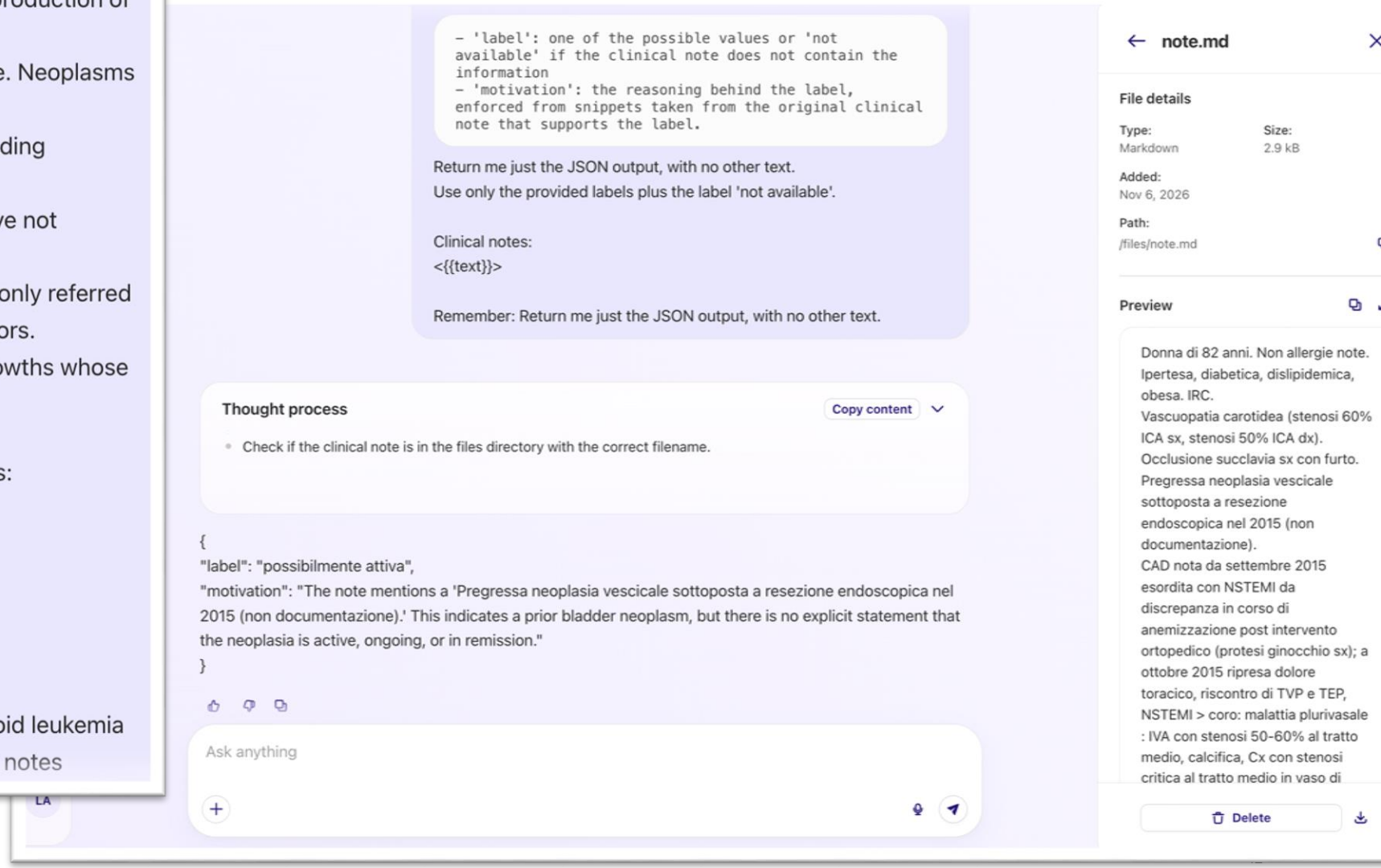
Your task is to determine if the patient has an active neoplasia.

Definitions:

- Neoplasia: The process that leads to the formation or production of a neoplasm.
- Neoplasm: An abnormal and excessive growth of tissue. Neoplasms are classified into four main groups:
  - Benign neoplasms: Non-cancerous growths, including adenoma-like tumors.
  - In situ neoplasms: Early-stage neoplasms that have not invaded surrounding tissues.
  - Malignant neoplasms: Cancerous growths (commonly referred to as cancer), including adenocarcinoma-like tumors.
  - Neoplasms of uncertain or unknown behavior: Growths whose behavior is unclear.

Common abbreviations and acronyms for neoplasms/tumors:

- CA: Cancer
- CA in situ: Carcinoma in situ
- BMT: Bone marrow tumor
- HCC: Hepatocellular carcinoma
- NSCLC: Non-small cell lung cancer
- ALL / AML: Acute lymphoblastic leukemia / Acute myeloid leukemia
- K: Common shorthand used for cancer in some clinical notes



You are provided with a clinical note. Your task is to determine the patient's level of consciousness at the time of assessment using the AVPU scale.

Classify the patient as one of the following:

- 'A' (Alert): Fully awake and aware of surroundings; typical behaviours of this state are any of: opens eyes spontaneously, follows commands, and tracks objects.
- 'V' (Verbal): Does not open eyes spontaneously but responds meaningfully to verbal stimuli.
- 'P' (Pain): Does not respond to verbal stimuli but reacts to painful stimuli (e.g., pinch or sternum rub) with moaning, movement, or crying.
- 'U' (Unresponsive): Does not respond to any verbal or painful stimuli.

Important:

- Only consider the patient's condition at the time of assessment. If more than one condition is present, consider the last one in chronological order.
- The note may not provide explicit information, or it may describe past events rather than assessment status.
- If the information is missing or only refers to the past and cannot reliably indicate the patient's condition at assessment, indicate this clearly.

EVALUATE DIRECT AND INDIRECT EVIDENCE FOR YOUR CONCLUSIONS THOROUGHLY and report it in the motivation

Return the output in a JSON format with the following structure:

```
- 'label': one of the possible values or 'not available' if the clinical note does not contain the information
- 'motivation': the reasoning behind the label, enforced from snippets taken from the original clinical note that supports the label
```

Return me just the JSON output, with no other text.
Use only the provided labels plus the label 'not available'.

Clinical note:
<{{text}}>

Remember: Return me just the JSON output, with no other text.

{
"label": "A",
"motivation": "The note states the patient was 'allertata per dispnea e astenia dal risveglio' (alerted for dyspnea and asthenia from waking up), indicating she was awake and responsive, and later reports resolution of symptoms ('Al mio arrivo la pz riferisce risoluzione della sintomatologia'), showing she remained conscious and able to communicate."
}

Ask anything

note.md

File details

Type: Markdown
Size: 2.9 kB
Added: Nov 6, 2026
Path: /files/note.md

Preview

Donna di 82 anni. Non allergie note.
Ipertesa, diabetica, dislipidemica, obesa. IRC.
Vascuopatia carotidea (stenosi 60% ICA sx, stenosi 50% ICA dx).
Occlusione succlavia sx con furto.
Pregressa neoplasia vescicale sottoposta a resezione endoscopica nel 2015 (non documentazione).
CAD nota da settembre 2015 esordita con NSTEMI da discrepanza in corso di anemizzazione post intervento ortopedico (protesi ginocchio sx); a ottobre 2015 ripresa dolore toracico, riscontro di TVP e TEP, NSTEMI > coro: malattia plurivasale : IVA con stenosi 50-60% al tratto medio, calcifica, Cx con stenosi critica al tratto medio in vaso di

Delete

Grazie per l’attenzione!

