

# **The Invalsi Benchmark**

**12/04/2024**

**Giovanni Puccetti**

# Invalsi Benchmark: Why?

## Problems:

1. Lack of Large Language Models benchmarks in Italian, even more so for Math and Language understanding;
2. Current benchmarks are translations of English benchmarks in Italian: Translating benchmarks from English is particularly hard for these two tasks;

## Solution:

Build natively Italian benchmarks for Language and Mathematical Understanding

# Invalsi Benchmark: How?

The Invalsi are a series of questionnaires that are periodically undertaken by students of age between 6 and 18 (1st grade to 13th) in all Italian schools. There are 3 different tests, Italian Language, Mathematics and English. We focus on the first 2.

We collect questions from these tests to create two hand-curated datasets:

- Invalsi MATE: Math and Logic;
- Invalsi ITA: Language Understanding and Comprehension in Italian.

**Invalsi MATE**

# Invalsi MATE: Question Types

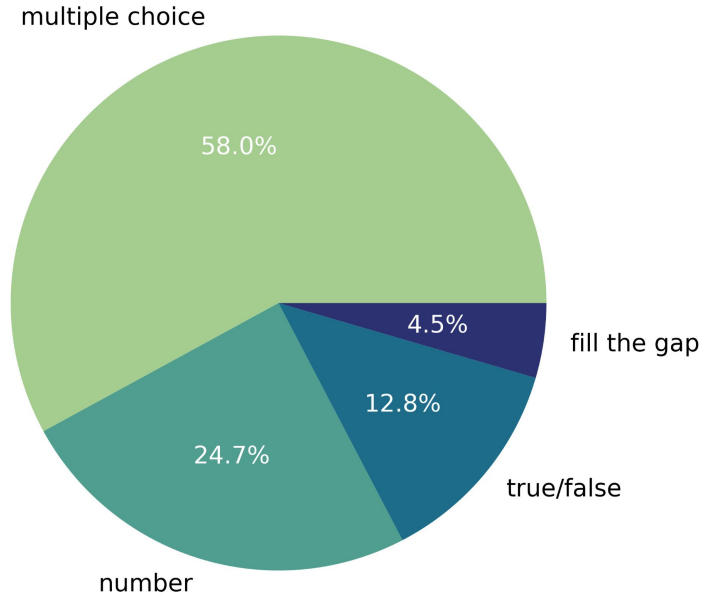
Type	Typical Answer	Quantity
Multiple Choice	A, B, C, D	243
Number	1234, 0.2, 10%	104
True/False	V, F	54
Fill the gap	how can you . . . . the gap	31
Total	-	<b>422</b>

# Invalsi MATE: Examples

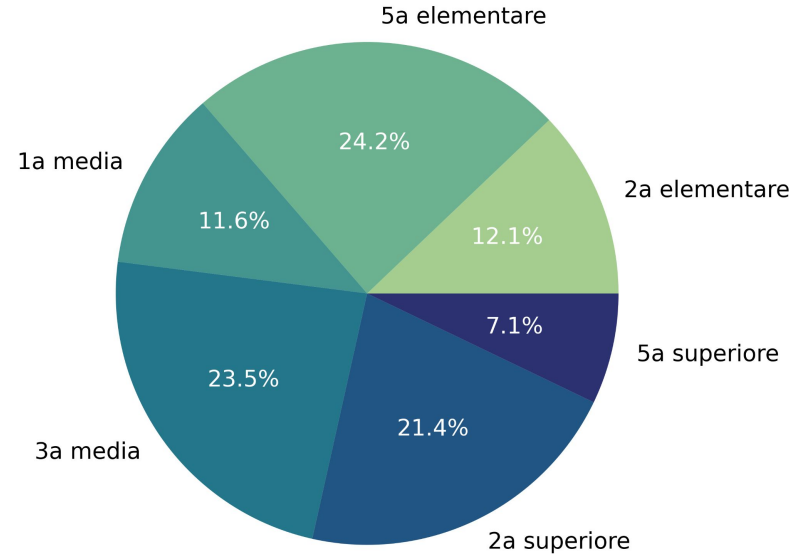
Multiple Choices	True - False
<p><b>Testo</b> Elisa è uscita da casa questa mattina alle ore 8:15. Elisa è rientrata nel pomeriggio alle ore 1:15</p> <p><b>Domanda</b> Quanto tempo è stata fuori casa Elisa? A. 5 ore B. 7 ore C. 9 ore</p>	<p><b>Testo</b> Se moltiplichi per 2 un numero naturale e dal risultato sottrai 1, ottieni sempre un numero pari.</p> <p><b>Domanda</b> Vero o Falso?</p>
Number	Fill the gap
<p><b>Testo</b> Filippo dice: per trovare il numero della mia maglietta aggiungi una decina e sei unità al numero 4.</p> <p><b>Domanda</b> Qual è il numero della maglietta di Filippo?</p>	<p><b>Testo</b> Luca lancia due dadi a sei facce non truccati.</p> <p><b>Domanda</b> Completa la frase inserendo una delle seguenti espressioni.</p> <ul style="list-style-type: none"><li>- maggiore della</li><li>- minore della</li><li>- uguale alla</li></ul> <p>La probabilità che la somma dei punti sia 12 è .... probabilità che la somma sia 2.</p>

# Invalsi MATE: Composition

Question Percentage by Type



Question Percentage by Grade



# Invalsi MATE: Evaluation Strategy 1

TESTO:

A ogni giro di trenino salgono 10 bambini.

Il trenino parte sempre al completo.

Sara aspetta in fila il suo turno e prima di lei ci sono 26 bambini.

DOMANDA:

Quando salirà Sara sul trenino?

A. Al secondo giro, B. Al terzo giro, C. Al decimo giro

### Risposta:

Ragioniamo passo passo:

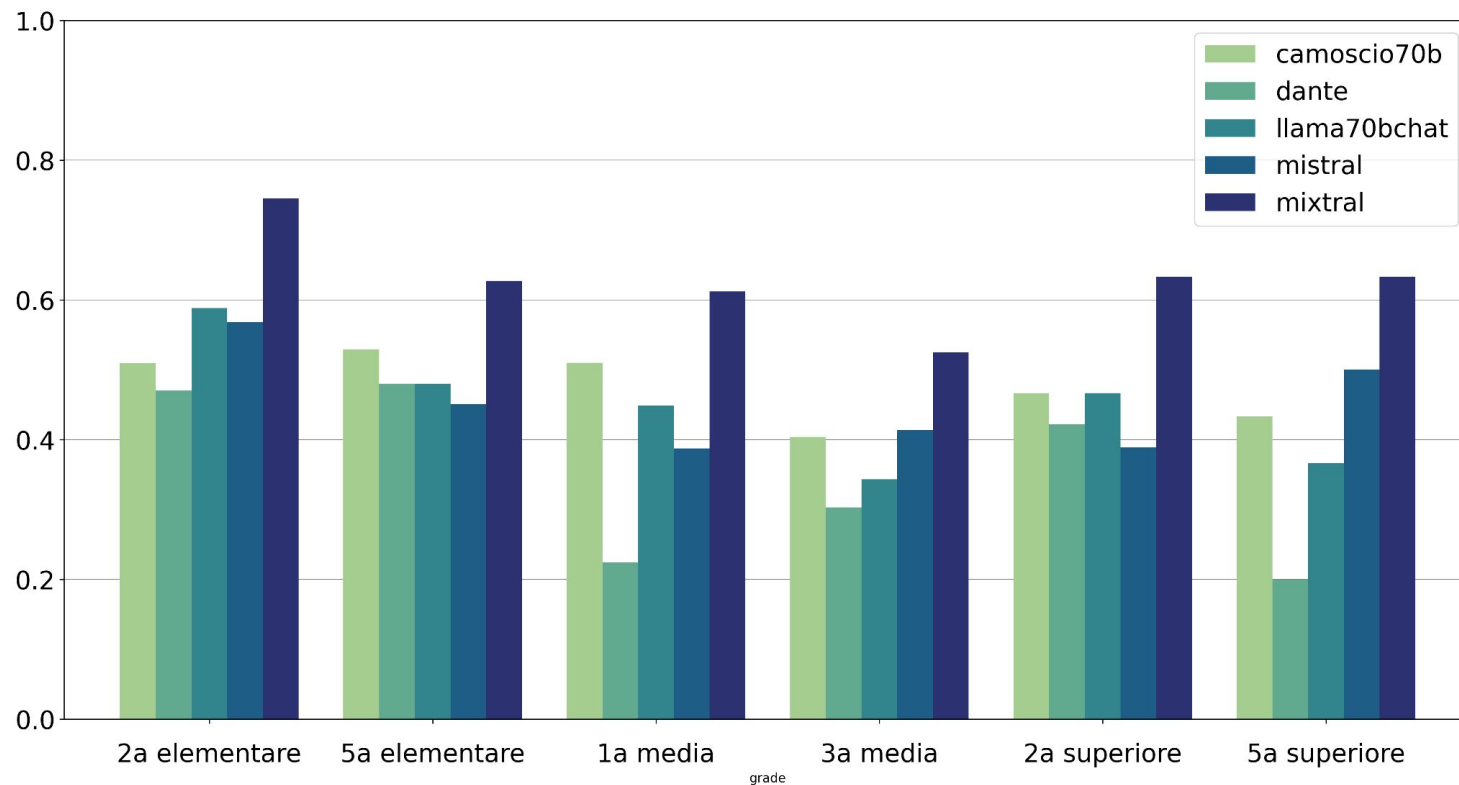
1. A ogni giro di trenino salgono 10 bambini.
2. Il trenino parte sempre al completo.
3. Sara aspetta in fila il suo turno e prima di lei ci sono 26 bambini.
4. Quindi, Sara salirà sul trenino al terzo giro.



# Invalsi MATE: Language Models Accuracy

Model	Accuracy	Completa Frase Accuracy	Multipla Accuracy	Numero Accuracy	Vero/Falso Accuracy
Mixtral	61.76	36.84	62.7	62.5	64.81
Camoscio 2 70b (ours)	47.51	31.58	47.13	49.04	51.85
LLaMA 70b chat	44.66	36.84	43.03	44.23	55.56
Mistral	43.94	21.05	43.85	42.31	55.56
DanteLLM	37.53	5.26	39.75	33.65	46.3

# Invalsi MATE: Accuracy by Grade



# Invalsi MATE: Evaluation Strategy 2

Di seguito è riportata una istruzione che descrive un task. Scrivete una risposta che completi in modo appropriato la richiesta.

### Istruzione:

Dato un testo (TESTO) scegli la risposta alla domanda (DOMANDA) seguente tra quelle disponibili e dai una risposta finale.

### Input:

TESTO:

A ogni giro di trenino salgono 10 bambini. \nIl trenino parte sempre al completo.\nSara aspetta in fila il suo turno e prima di lei ci sono 26 bambini.

DOMANDA:

Quando salirà Sara sul trenino?

A. Al secondo giro

B. Al terzo giro

C. Al decimo giro

RISPOSTA:

Label	Likelihood	Picked
"A"	-2.62	false
"B"	-1.25	true
"C"	-1.91	false
"D"	-5.30	false

# Invalsi MATE: Language Models Accuracy

Model	Multipia Accuracy	Vero/Falso Accuracy
<i>Command R +</i>	51.03	59.26
Mixtral	49.38	66.67
Camoscio 2 70b (ours)	42.8	66.67
LLaMantino 70b chat	40.74	53.70
LLaMa 70b chat	39.51	48.15
DanteLLM	32.28	63.01
Mistral	30	57.41

**Invalsi ITA**

# Invalsi ITA: Question Types

Type	Typical Answer	Quantity
Multiple Choice	A, B, C, D	978
Binary	True/False, Si/No	139
Open Question	This is what happened	111
Other	N/A	53
Total	-	1281

# Invalsi ITA: Example

## POLIPETTO HA UN PROBLEMA!

Polipetto viveva in una graziosa casetta nel mare blu. Dalla finestra poteva ammirare un panorama meraviglioso e pescare tutto il cibo che voleva. Una mattina però, di ritorno dalla sua nuotata quotidiana, Polipetto trovò una grossa coda che bloccava l'ingresso della sua casetta. "Oh no!" esclamò. "Che coda enorme!" E, impaurito, nuotò via. Doveva assolutamente parlarne con un amico. Paguro lo ascoltò e sospirò: "Bene, bene, un intruso. Che t'importa: trasferisciti da un'altra parte. Il mare è pieno di opportunità." Polipetto scosse la testa: non gli sembrava un buon consiglio... Le Meduse – che avevano sentito tutto – gli suggerirono: "C'è solo una cosa da fare: sbarazzarsi dell'intruso. Caccialo fuori!" Polipetto pensò all'enorme coda: no, non era affatto una buona idea. Quindi, nuotò via per chiedere consiglio a Balena. Enorme com'era avrebbe avuto certamente una grande idea. Ma Balena non fu per nulla d'aiuto. Non aveva una casa e, per di più, non ne aveva alcun bisogno. La sua casa era il mare e di queste faccende non si preoccupava affatto. Polipetto scappò via. Proprio allora s'imbatté in Pesce Scorpione. "Che cosa vuoi?" ringhiò Pesce Scorpione, mostrando i suoi aculei per impressionare Polipetto. "Niente, signore," bisbigliò Polipetto. "Ma c'è un intruso, uno sconosciuto che ha occupato la mia casetta senza chiedere niente, e non so proprio..." "Liberatene..." disse Pesce Scorpione fissando Polipetto "...mangiandotelo!" Polipetto deglutì forte, annuì e nuotò via più veloce che poté. Ormai, tutti gli animali marini sapevano del suo problema e tutti gli consigliavano una soluzione diversa. "Pungilo," disse Pesce Ago. "Sono d'accordo," disse Pesce Riccio. "Oooh ..." gemette Polipetto. Cominciò ad avere la nausea: così tanti pesci e tante soluzioni gli avevano fatto venire il mal di mare. I pesci parlavano tutti insieme. "E tu, come pensi di risolvere il problema?" sussurrò, infine, il grande Oceano a Polipetto. "Cosa farai?" "Io?" "Sì, cosa farai?" "Potrei andare dalla coda... e chiederle gentilmente, ma molto gentilmente, di andarsene..." E così Polipetto tornò pian piano alla sua casetta. Là, picchietto con prudenza sulla coda. "Mio caro signore," disse gentilmente, "vorreste essere così cortese da..." "Aiuto!" singhiozzò una vocina. "Per favore, tiratemi fuori di qui! Sono incastrata da così tanto tempo!" Una coda così grande e una vocina così sottile! Con tutti i suoi otto tentacoli, Polipetto afferrò la coda e tirò più forte che poté. "AIUTO!" gridò. "Qualcuno ci aiuti!" E tutti gli altri pesci corsero ad aiutarlo. "Oh," arrossì Polipetto. "Se solo avessi saputo che eri una sirena... Questo cambia tutto!"

*(Testo tratto e adattato da: E. van Lieshout, E. van Os, illustrazioni di M. van Hout, Polipetto ha un problema, Il Castello, Milano, 2010)*

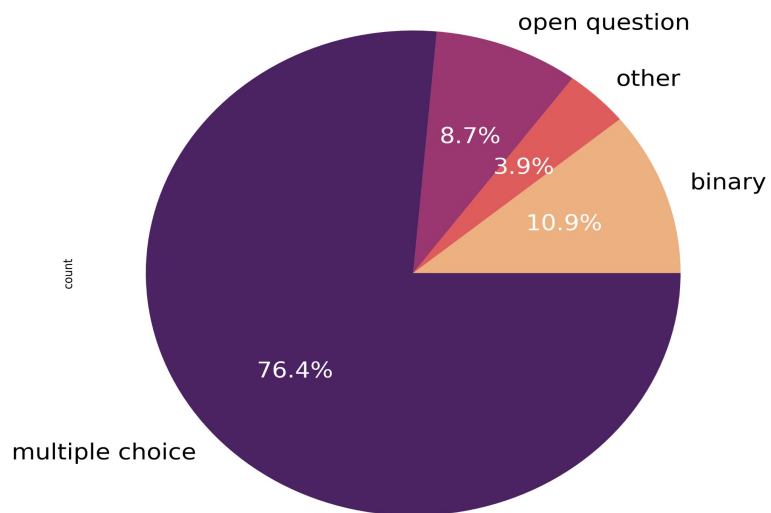
# Invalsi ITA: Examples

Multiple Choices	Binary
<p><b>Testo</b> Il titolo dice che Polipetto ha un problema e l'inizio del racconto spiega di che cosa si tratta.</p> <p><b>Domanda</b> Qual è il problema di Polipetto? A. Non gli piace più la sua casa B. Non può più entrare in casa sua C. La sua casa si è riempita di animali D. La sua casa non è più ordinata come prima</p>	<p><b>Testo</b> Grazie all'incontro con il grande Oceano qualcosa cambia in Polipetto.</p> <p><b>Domanda</b> Com'è Polipetto PRIMA di parlare con Oceano e DOPO avere parlato con lui? Metti una crocetta per ogni riga. A. Polipetto si fa coraggio: Prima/Dopo B. Polipetto è confuso: Prima/Dopo C. Polipetto pensa che gli altri siano più bravi di lui: Prima/Dopo D. Polipetto si fida della sua idea</p>
Open Question	Other
<p><b>Testo</b> Che cosa ci dice il racconto a proposito dell'escursione? Completa la sintesi che segue, inserendo la parola appropriata in ogni spazio.</p> <p><b>Domanda</b> "Durante l'escursione alcuni ragazzi non si accorgono di quanto camminano perché cantano e scherzano. All'arrivo, dopo esattamente cinque ore di strada. ... (1) si siede perché è affaticato e i ... (2) gli fanno male; posa vicino a sé lo zaino con la ... (3) dentro. La signora Salici trova straordinario il lago, mentre ai ... (4) non sembra niente di speciale."</p>	<p><b>Testo</b> "Doveva assolutamente parlarne con un amico". Nel testo non c'è scritto che cosa ha detto Polipetto a questo amico, ma dal racconto si può capire.</p> <p><b>Domanda</b> Che cosa può avergli detto Polipetto?</p>

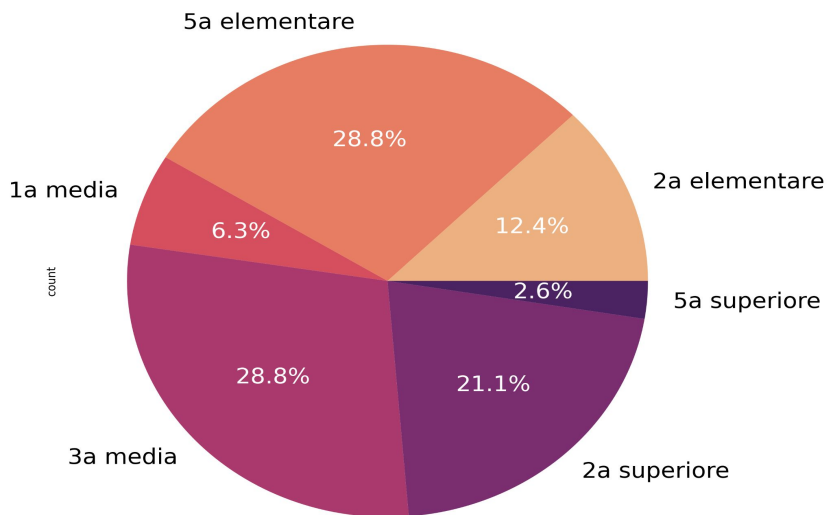


# Invalsi ITA: Examples

## Question Percentage by Type



## Question Percentage by Grade



# Invalsi ITA: Accuracy

Model	Multipla Acc.
Mixtral	81.78
Camoscio 2 70b (ours)	78.30
LLaMa 70b chat	73.69
Mistral	60.18

**The End :)**