HACK APERTUS 2026 · TRACK 2A · CHALLENGE BY OST

GeaNaVid

Every claim checked against the official voting booklet. Each Yes or No comes with the booklet line behind it.

Gea · Na · Vid — Romansh for Yes · No · Blank



▶ Demo video

Built on Apertus v1.5

Team Apertool · solo builder

THE PROBLEM

Voters hear the claim. The answer is on page 4 of 40.

THE CLAIM

„… die Bevölkerung seit Einführung der Personenfreizügigkeit im Jahr 2002 um rund 500'000 Personen gewachsen ist.“

“… the population has grown by about 500,000 since free movement began in 2002.”

OFFICIAL BOOKLET · PAGE 4

„Seit Einführung der Personenfreizügigkeit im Jahr 2002 ist die Bevölkerung um rund 1,7 Millionen Personen gewachsen.“



OST demo set, label 2; Federal Council explanations for 14 June 2026, German edition, PDF page 4.

THE TASK

Booklet + claim in, a ballot mark out.

Voting booklet
PDF · German, French or Italian

Claim
one sentence · DE, FR or IT

→



GeaNaVid
CLI · Apertus v1.5

→

Gea	entailment	0
Na	contradiction	2
Vid	neutral	1

+ the passages it rests on · tokens used · seconds taken

JUDGED ON

Macro-F1 on a hidden benchmark · **evidence quality** · **input tokens** · **inference time**

All 9 claim/booklet language combinations. Labels 0/1/2 and judging criteria as set by OST in the Hack Apertus challenge brief.

THE NAME

Three answers, in Romansh.



Gea

Yes

entailment



Na

No

contradiction



Vid

Blank

neutral

Every Gea and every Na comes with the booklet line that decides it. A Vid quotes nothing.

Gea, Na and vid(as) as printed in Romansh voting booklets of 1994 and 1995 (VotingBooklets-v1, Stüssi & Vamvas, SwissText 2026, CC BY 4.0).

THE CONTEXT

Narrow first, then let Apertus read.

1 **Top-5 passages of the right proposal, with page neighbours**

the vote title narrows the booklet, BM25 picks the passages, the section the claim names first

≈ 1,100 tokens of passages

CLAIM IN ANOTHER LANGUAGE? Apertus translates it into the booklet language first

2 **Translation**

one short call, then the same search

+ ≈ 300 tokens

full **The whole booklet**

the baseline OST recommends

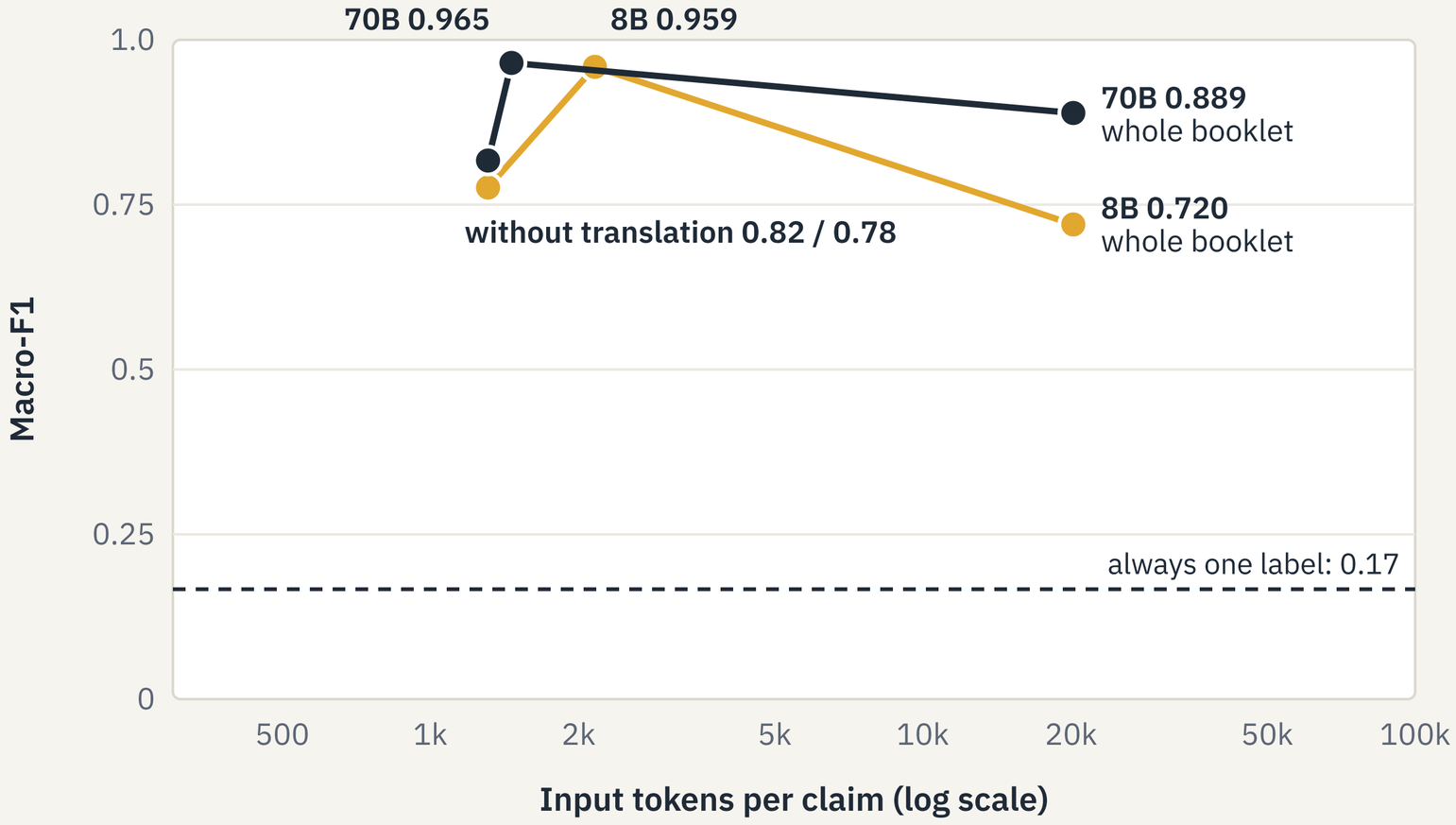
20,242 tokens

Per claim: 1,461 input tokens with 70B, 2,159 with 8B, instead of 20,242.

Mean input tokens per claim as reported by the API, OST v1.1 test split (493 pairs), Apertus v1.5, 14 October 2026.

THE TRADE-OFF

Under a ninth of the tokens. Better verdicts.



Apertus 70B SUBMITTED

0.965 · 1,461 tokens

1.3 s mean · 2.2 s p95

Apertus 8B

0.959 · 2,159 tokens

0.37 s mean · 0.55 s p95

Beyond the test split

Held-out booklets 0.932 / 0.874. Older booklets 1998–2018: 0.890 / 0.836 (70B / 8B).

OST v1.1 test, 493 pairs; 95 % CI 0.944–0.983 / 0.933–0.980; time: cold, 50 cases; caveats: report §4.

ACROSS LANGUAGES

Claim in Italian, booklet in German. The evidence stays German.

	DE	FR	IT
DE	0.98 0.98	0.94 0.98	0.91 0.91
FR	0.96 0.98	1.00 0.98	0.86 0.91
IT	1.00 0.94	0.96 0.97	1.00 0.98

Macro-F1 per pair: top Apertus 8B, below 70B. Rows: claim language · columns: booklet language. Framed: same language.

0.70 → 0.94 8B

0.74 → 0.94 70B

macro-F1 on the 330 cross-lingual test claims, before and after Apertus translates the claim (ablation step; full defaults 0.943 / 0.958).

> The search runs on the translation, the evidence is quoted from the booklet as it is. No language field is needed: the claim language is detected.

OST v1.1 test, 38–74 claims per pair (cells ± 0.05); same / cross language 0.993 / 0.943 (8B), 0.980 / 0.958 (70B).

EVIDENCE

Every Yes or No points to a line in the booklet.

Na `contradiction · label 2`

Claim (Italian): the reduced VAT rate falls from 2.5 to 2.4 per cent

> „Der dafür geltende reduzierte Mehrwertsteuersatz steigt von 2,5 auf 2,6 Prozent; …“

`25.09.2022 · DE · page 28 · AHV 21: additional financing by VAT`

`input tokens 2,794  output tokens 52  Apertus 8B`

Quotes, not paraphrase

The evidence is booklet sentences, verbatim, each with its PDF page, from the section the claim names, chosen after the verdict without another model call.

Evidence first, label second

Apertus names the passages before it gives the verdict, in strict JSON.

Not voting advice

GeaNaVid says what the booklet says. It never says how to vote.

OST case v1-1484 (dev split). By OST's scoring rule a quote lies in the gold passage in 96.3 % (70B) / 96.9 % (8B) of 324 cases.

OUR ASK

Score it on the hidden booklets as submitted, and read the lines it quotes.

Niklas Wichter

Team Apertool · solo builder · one of four tools at apertool.ch

VIDEO	**geanavid.apertool.ch/#video · 2 min**
RUN	make test · make run · make ost (Docker, 70B)
REPO	github.com/NWichter/Hack_Apertus_2A_GeaNaVid
REPORT	geanavid.apertool.ch/report.pdf

Scan: watch the 2-minute demo
`geanavid.apertool.ch`

Hack Apertus 2026, Track 2A, challenge by OST. GeaNaVid is an independent entry, not a product of OST or the Federal Chancellery.