CNEXT AI · TEST REPORT · ONE-OFF

Model comparison on three tasks

CUSTOMER
Muster AG, Bern (fictional)

TESTED
14 to 20 October 2026

TASKS
Summarising contracts, questions on internal rules, code review

TEST DATA
Anonymised documents and a test repository

4	3	120	2
models compared	tasks	test cases	models recommended

When you receive this document: Before the choice of model, for your approval.

Summary

We tested two cloud models and two open models on three Muster AG tasks, with 120 test cases from anonymised documents. For contracts and code we recommend a cloud model with governed data processing. For questions on internal rules, an open model on your own hardware answered just as accurately and stays inside your network.

Results

CASES	TASK AND RESULT	RECOMMENDATION
40	Summarising contracts: cloud model A found every notice period, open model C missed two in appendices.	Cloud model A
50	Questions on internal rules: all four answered with a source, open model C as accurately as the cloud models.	Open model C, in your network
30	Code review: cloud models A and B found the planted bugs, the open models about half of them.	Cloud model A or B

To settle before use

- Legal reviews the data processing agreement for the cloud model.
- The open model needs a server with a GPU; IT has the quote.

Next steps

With your decision we set up the chosen models and rerun the test cases after every model change. The real report names each model and version.

Every name, system and number in this document is invented. Your report contains your environment on the day of the check.