MARLUC

CASE STUDY · MARLUC SYNOPSIS

National Library of Norway Mímir tests and language model licence

IMPLEMENTED 2024–2026 · GOVERNMENT AND PUBLIC SERVICES · NORWAY · BAND C · CLAIMED ONLY

VERSION 1.0 · PUBLISHED 25 NOVEMBER 2026 · UPDATED 25 NOVEMBER 2026 · FACTS CHECKED 25 NOVEMBER 2026

A national library tested whether copyright-protected books and newspapers improve Norwegian language models, then agreed a paid licence with rightsholders to train and share models on newspaper text.

ABOUT THIS SYNOPSIS

This is MarLuc's summary and grading of publicly available sources, for information only. It is not investment, legal, financial or professional advice, and should not be relied on for any decision. Readers must validate it against the original sources and, where appropriate, challenge its conclusions.

Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date. Check the original source before relying on any figure.

The impact band, proof marker and SEAM view are MarLuc's opinion. The organisation is not a MarLuc client and has not reviewed or endorsed this synopsis. Names and trademarks belong to their owners.

NOTICE WORDING LAST UPDATED 16 NOVEMBER 2026

AT A GLANCE

AI used: Norwegian generative language models trained on library collections under a collective licence.

Implemented: 2024–2026. Grade: band C, Process change, Claimed only.

Headline result, as reported: Library's own tests found newspapers and non-fiction lifted model scores; a NOK 45 million a year licence followed, with no results yet reported from models trained under it

1 · The case

The facts, as reported by the 3 sources listed at the end. Where a source is silent, the field says "Not reported".

Organisation

National Library of Norway (Nasjonalbiblioteket), with Kopinor. Norway's national library. Staff numbers: Not reported. Kopinor is the Norwegian copyright organisation that signed the agreement on behalf of rightsholder organisations. Canvas block: 8. Governance and risk. Also 4. Operations and process.

Business challenge

The challenge. In November 2023 the Norwegian rightsholders' organisation asked the government for compensation for the use of their material in generative AI training, by the library's report. The government asked the library for evidence to inform a compensation scheme. The agreement text describes a national aim of Norwegian language models built on Norwegian and Sami language, history and culture.

How it was addressed. The library led the Mímir consortium with the University of Oslo and NTNU, with computing from Sigma2. It trained 17 models on corpora with and without copyright-protected books and newspapers, and compared them on 28 tasks. Using those results, the library and Kopinor signed a licence on 19 December 2025 for the library to train, develop and share Norwegian language models on Norwegian press publications. The agreement's English text covers press publications "up to one year from the date of publication"; Kopinor's page describes the permitted content as newspaper text more than one year old. Per the academic commentary, the Norwegian state pays NOK 45 million a year under it. Rightsholders may opt out, and may list items for removal, with the list updated every three months. The library may share trained models royalty-free under a licence. [1] [2] [3]

Who implemented it

In-house (In-house): Led the Mímir tests and trains the models under the licence.
University of Oslo and NTNU (University or research lab): Consortium partners in the Mímir tests.
Mistral AI (Model or cloud provider): Source of the Mistral 7B architecture and weights used as a base.
Sigma2 (Model or cloud provider): Supplied computing resources for the project.
Kopinor signed the agreement for rightsholders; it is a counterparty, not an implementer.

Technology (SEAM pillar E)

SEAM types: T6 Open-source / locally hosted models. Deployment: Not reported. Vendors and products: Mistral AI, Mistral 7B architecture and weights (v0.1).

Where it was built in: Research models trained on the library's cleaned, deduplicated corpora of books, newspapers and web text. No live service is reported.

What changed

Before: Norwegian rightsholders had asked for compensation and the library had earlier agreements with some newspapers for language technology purposes. After: a collective licence covers Norwegian newspaper text for training and sharing language models, with remuneration, reservation rights and a monitoring duty on the library. Models trained under the licence and their use: Not reported. [1] [2] [3]

What it cost

Per the agreement, NOK 45 million a year for 2026, index-adjusted from 2027; per the academic commentary, the Norwegian state provides it. Mímir used 270,000 GPU-hours on the LUMI supercomputer, the Idunn cluster and Google TPUs; the cost of that computing is not reported.

What it saved

Not reported. The library's tests found that models trained with a mix of copyrighted and non-copyrighted text generally scored higher than models trained on non-copyrighted text alone. Average gains over the base model were 6.37% for newspapers and 6.73% for all in-copyright material. [1] [2] [3]

Governance and control

The agreement gives rightsholders an opt-out and a removal list every three months, requires the library to monitor breaches by licensees, sets contact meetings on how the agreement is working, and has Kopinor indemnify the library against external claims. These are contract terms by the parties; no outside body's review of the licence in use is reported. Staff oversight of the models themselves: Not reported.

S2 · Accountable owner · Claimed
The agreement names the National Library as responsible for monitoring compliance with the licensing terms for further use of the models. [2]

A1 · Safe use · Claimed
Rightsholders can opt out, and can list items for removal every three months; the library must review the list and remove the items. [2]

A2 · Human in the loop · Not reported
Not reported.

A3 · Monitoring and assurance · Claimed
The library must monitor breaches by licensees, and the parties hold regular contact meetings; the library's evaluation covered 28 tasks. [1] [2]

A4 · Vendor AI · Claimed
The report says warm-started models built on Mistral 7B are not fully auditable because their pre-training data is unknown; the agreement does not cover base-model pre-training. [1] [2]

A5 · Agent controls · Not reported
Not reported; the sources describe language models, not agents.

A6 · Standards and regulation · Claimed
The licence relies on extended collective licensing under the Norwegian Copyright Act, provided Kopinor is approved by the Norwegian Industrial Property Office as a representative organisation, and carries GDPR and data security clauses. [2]

What worked, what did not, what is unproven

WORKED

– The library's tests gave the parties a measured basis for pricing: models trained with a mix of copyrighted and non-copyrighted content generally scored higher.

– Newspapers alone gave an average gain of 6.37% over the base model; adding non-fiction books and newspapers gave 6.52%.

– The parties agreed a licence in about two years from the first request for compensation.

DID NOT

– Adding fiction books was the only configuration to reduce performance, by 1.40%.

– The report says the difference between the warm-started models is very small and further testing is needed to say whether it is statistically significant.

– The project did not aim to produce the best performing Norwegian model.

UNPROVEN

– Whether the NOK 45 million level and its distribution to rightsholders are fair.

– Whether enough rightsholders stay in the corpus: the library can ask to renegotiate if the remaining training set is no longer representative.

Timeline

Nov 2023 Norwegian rightsholders' organisation contacts the government to demand compensation for use of material in generative AI training [1]

Jan 2024 Library forms the Mímir consortium with the University of Oslo and NTNU [1]

Oct 2024 Mímir technical report, version 1.1, dated 24 October 2024 [1]

Dec 2025 Kopinor and the library sign an agreement allowing newspaper content to be used to train Norwegian and Sami language models [2]

2026 First year of annual remuneration of NOK 45,000,000 under the agreement [2]

2 · Grade and SEAM assessment

MarLuc's opinion, made on the evidence above. How cases are graded: marluc.co.uk/case-studies/how-we-rate

Level of change and impact

Claimed: Not applicable. Grade: C, Process change. Proof: Claimed only. Level of change: Process change.

Process change (Level 2, band C), as a judgement: a new, signed and priced process for sourcing training data is in place; no models trained under it are reported yet.

Graded C, no higher claim, as a judgement. The library built a new, signed process for sourcing training data: it measured the value of protected text, then agreed a priced licence with an opt-out and a removal list. The sources show the test results and the contract, not models trained and used under the licence, so benefits to users are unproven. A stricter reading that needs the process to be seen working would give band E, no results reported.

The SEAM view and lessons learnt

The result came from the library's own test and a negotiated contract. The AI's part is separable only for the test: the benchmark gains are the library's own and were not independently replicated. Before copying the approach, a leadership team should ask how it will measure the value of protected material it licenses, and what happens to the economics if many rightsholders opt out or use the removal list.

1. Measure the value of protected content with a test before pricing a licence.
2. Write opt-out, removal and renegotiation rights into the contract from the start.
3. Publish what the licensed models deliver so the price can be judged.

Evidence

Score 2 of 4. Self-reported results: the test results are the library's own; the agreement is a signed contract published by the counterparty.

Evaluation method. Self-reported library tests plus a signed agreement: the library's own benchmark comparison, not independently replicated; the licence terms come from the agreement text, with academic commentary.

Sources disagree. The agreement's English text covers press publications "up to one year from the date of publication"; Kopinor's page says the agreement allows newspaper content that is more than one year old. The agreement is dated Oslo, 19 December 2025.

Status. Ongoing: the agreement sets remuneration for 2026 and index-adjusts it from 2027. Models trained under it were not reported in the sources reviewed on 25 November 2026.

3 · Original sources

Read these in full. To challenge a fact or a grade, use the contact form at marluc.co.uk or the link on https://marluc.co.uk/case-studies/nb-mimir-norwegian-language-models

[1] National Library of Norway: The Mímir Project, technical report (short), version 1.1
https://www.nb.no/content/uploads/2025/12/The-Mimir-Project-Technical-Report-Short-eng-.pdf

[2] Kopinor and the National Library of Norway: Agreement on the use of copyright-protected material in the training of Norwegian language models
https://kopinor.no/wp-content/uploads/2026/02/Agreement-on-the-use-of-copyright-protected-material-in-the-training-of-Norwegian-language-models.pdf

[3] Kopinor: National Library agreements
https://kopinor.no/en/licences-for-the-use-of-works/national-library/

National Library of Norway (Nasjonalbiblioteket), with Kopinor website: https://www.nb.no

Version 1.0 · Published 25 November 2026 · Updated 25 November 2026 · Facts checked against sources 25 November 2026 · Downloaded 26 November 2026. This PDF may be out of date: the latest version, and a way to report an error or a better source, is at https://marluc.co.uk/case-studies/nb-mimir-norwegian-language-models

VERSION HISTORY

v1.0 25 November 2026 First published