



LLM-powered chatbots for document question answering

AI Workshop - ODATIS
Jan 2026, Marseille

Redouane Lguensat ^{1,2}

¹ Institut Pierre Simon-Laplace (IPSL)

² IRD (French National Research Institute for Sustainable Development)



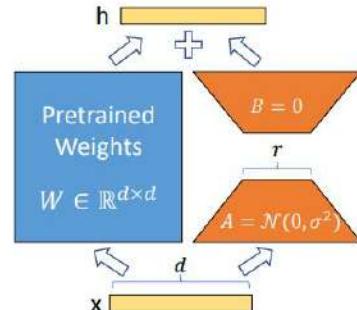
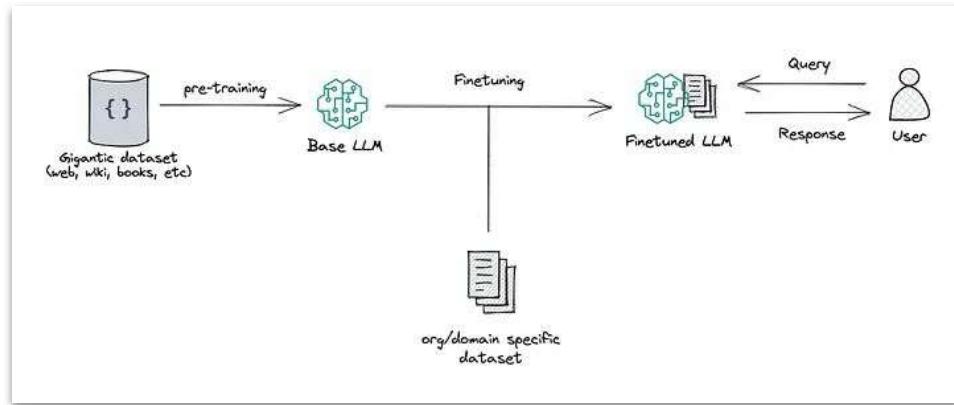
Augmenting LLMs with your own data

Fine tuning LLMs

First option: Fine tuning ! but

- Fine tuning involves training the LLM
- Expensive in terms of compute/financial cost
- You stick to the particular LLM you finetuned
- You need to refinetune if you would like to add more data

it is a good option if you have high confidence on your LLM and if you have the technical skills and knowledge needed



💡 PEFT: State-of-the-art Parameter-Efficient Fine-Tuning.
huggingface.co/docs/peft

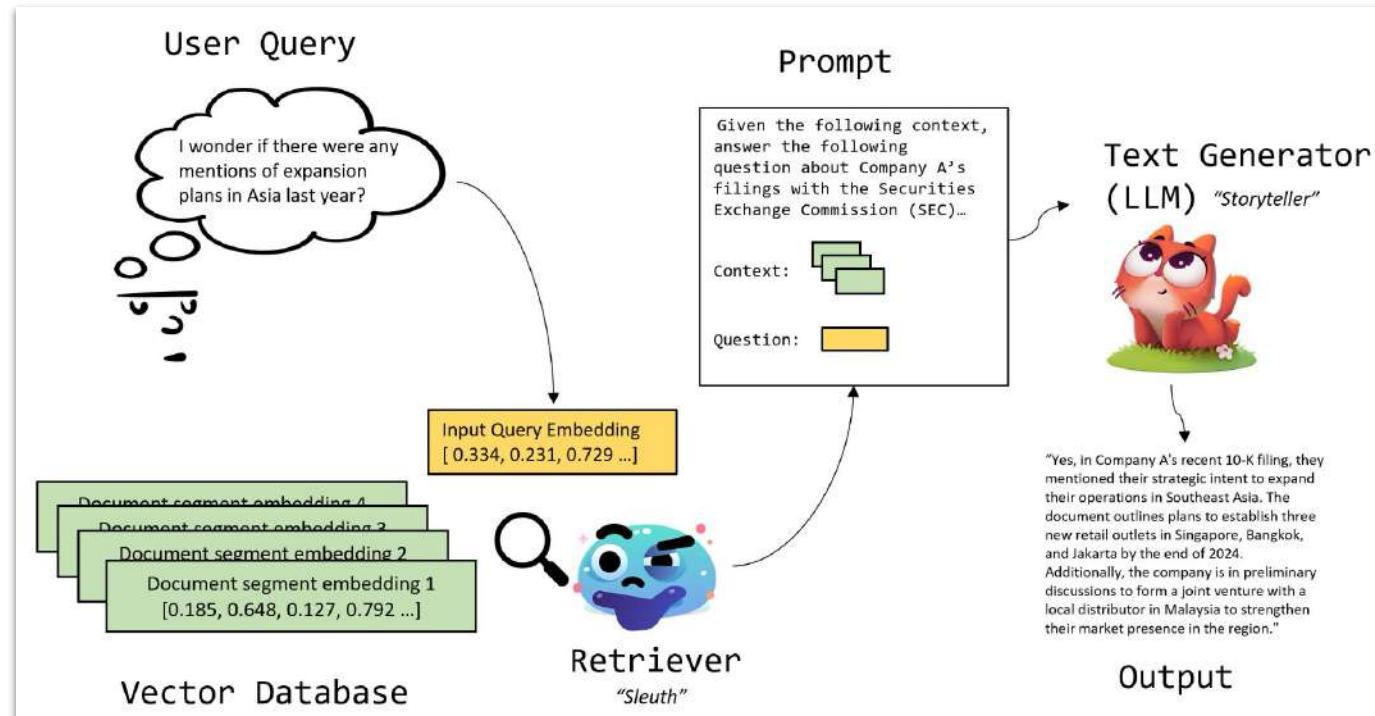


Augmenting LLMs with private data

RAG: Retrieval Augmented Generation (Lewis et al. 2020)

Second option: RAG

- Simple, intuitive, not very costly
- Flexible, you can plug and play any LLM
- Needs some Prompt engineering



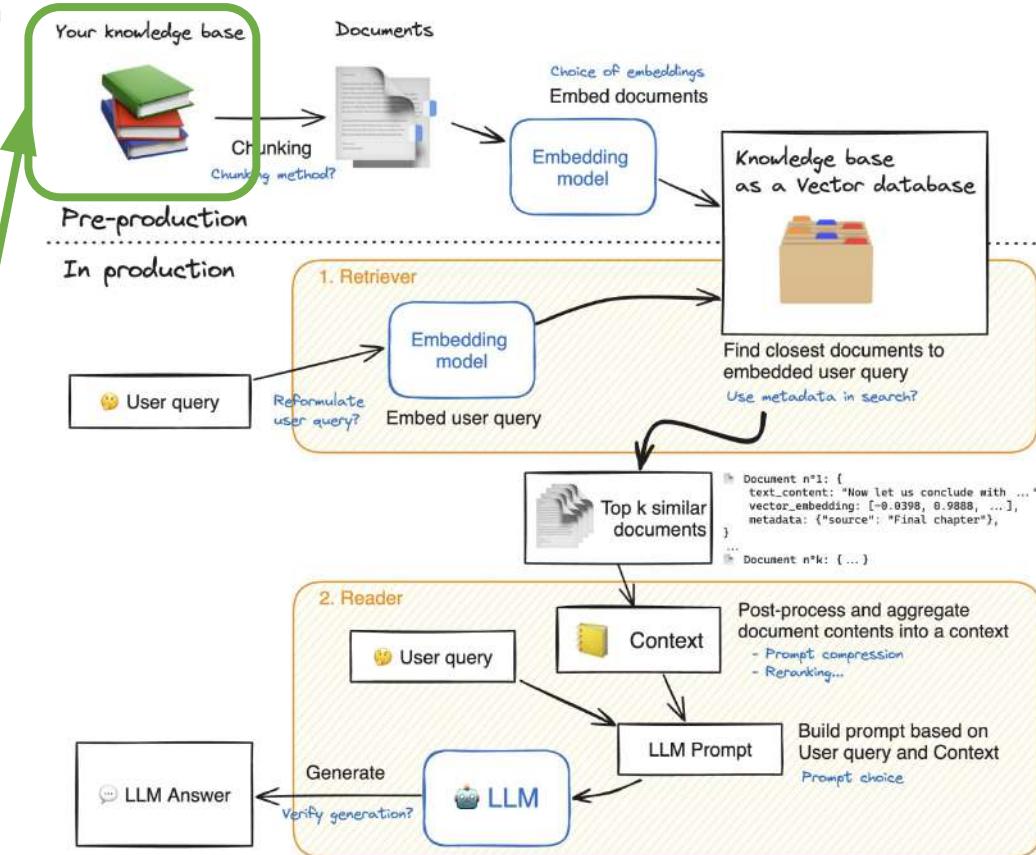
<https://www.determined.ai/blog/rag>



Demo with RAG on the NEMO documentation

General Idea

Latex files from the NEMO forge
<https://forge.nemo-ocean.eu/nemo/doc/manuals>
+
Rst files from the user guide
<https://forge.nemo-ocean.eu/nemo/doc/guide>





RAG: Retrieval Augmented Generation

Option 1: create your GPT with OpenAI

The screenshot shows the OpenAI GPTs interface. At the top right, there are buttons for "My GPTs" and "+ Create". The main heading is "GPTs" with the subtext: "Discover and create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills." A search bar shows the query "Lguensat". Below the search bar, a card for "NEMOfinder" is displayed, created by Redouane Lguensat. The card includes a small image of a globe, the name "NEMOfinder", a description "Answers questions using the NEMO v4.2 model", and the creator's name "By Redouane Lguensat".

Featured
Curated top picks from this week

 Adobe Express Stand out with Adobe Express. Quickly and easily make impactful social posts, images, videos, flyers... By adobe.com	 Code Copilot Code Smarter, Build Faster—With the Expertise of a 10x Programmer by Your Side. By promptscraft.com
 Social Butterfly-AI 2.0 AI Assistant for content development and social media strategy. Nuanced, guided... By digitalmogul.co	 Software Architect GPT Builds new software architecture documents by understanding user requirements and design... By V B Wickramasinghe



RAG: Retrieval Augmented Generation

Option 1: create your GPT with OpenAI

Advantages / Drawbacks

- VERY simple and easy to create, you only need to upload your document
- Shared link (you can access mine here <https://chat.openai.com/g/g-PFUCFXueM-nemofinder>) but only if you and your colleagues have a ChatGPT Plus subscription (20\$ per month)
- Based on one of the best Embedding + LLM at the moment but closed source



You
What are the key components of NEMO?

NEMOfinder
The NEMO (Nucleus for European Modelling of the Ocean) system is a comprehensive ocean modeling framework designed for research and forecasting in oceanography and climatology. Its key components encompass a range of ocean-related engines that collectively address physical, biogeochemical, and sea-ice processes. The primary components of NEMO include:

1. **Ocean Dynamics (DYN):** This component handles the three-dimensional velocity field, sea surface height, and the thermodynamics of the ocean, solving the primitive equations using various numerical schemes. It includes the treatment of Coriolis force, advection, hydrostatic, and non-hydrostatic pressure gradients, and surface pressure gradient schemes.
2. **Ocean Tracers (TRA):** This module manages the advection and diffusion of passive tracers like temperature and salinity, which are essential for understanding the ocean's thermodynamics and biogeochemical properties. It encompasses tracer advection schemes, lateral and vertical diffusion, external forcing, bottom boundary layer processes, and the equation of state for seawater.
3. **Sea Ice Modeling (SI3 or LIM):** NEMO includes a sea ice component, which can be either the Sea Ice Integrated System (SI3) or the Louvain-la-Neuve Sea Ice Model (LIM). These components simulate sea-ice dynamics, thermodynamics, including brine inclusions and subgrid-scale thickness variations, essential for polar and global ocean studies.
4. **Biogeochemistry (TOP-PISCES or BFM):** This module focuses on the ocean's biogeochemical processes, including nutrient cycling, carbon fluxes, and ecosystem dynamics. It can be run offline or coupled with the physical modules, using either the Pelagic Interaction Scheme for Carbon and Ecosystem Studies (PISCES) or the Biogeochemical Flux Model (BFM).
5. **Surface Boundary Condition (SBC):** This component handles the interaction between the ocean surface and the atmosphere, including the fluxes of momentum, heat, and freshwater. It includes

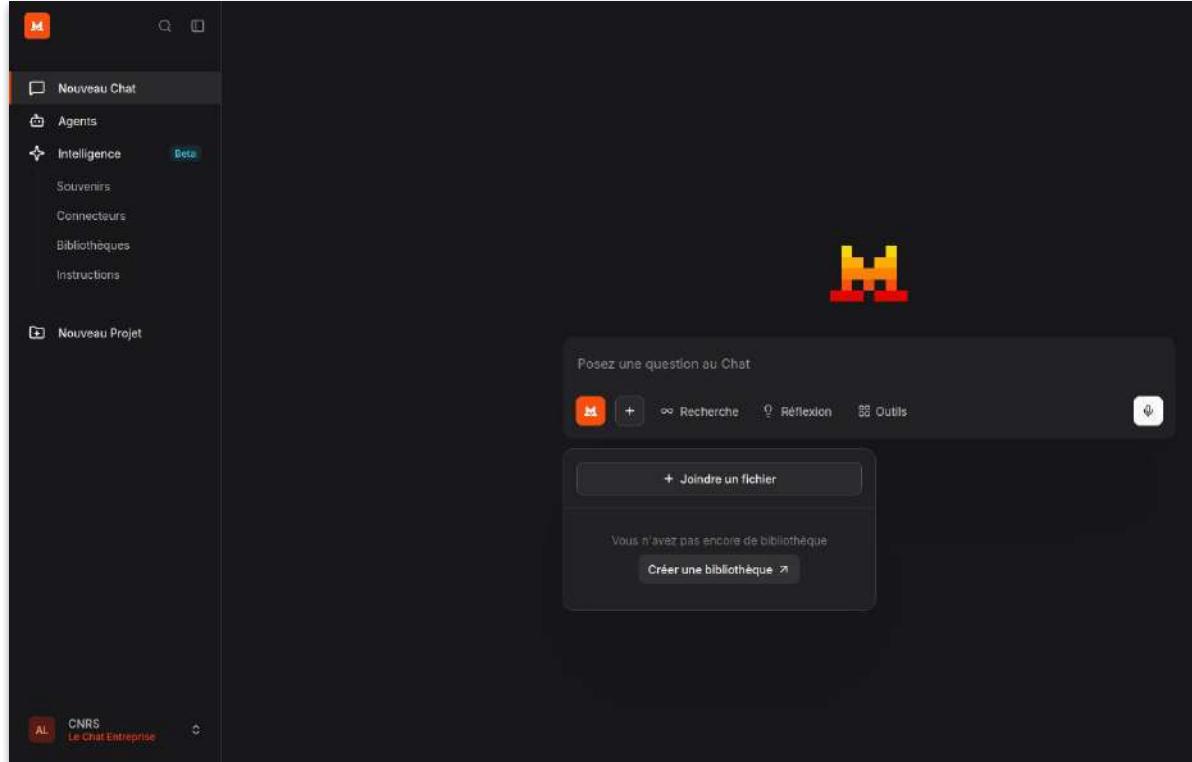
Message NEMOfinder...

ChatGPT can make mistakes. Consider checking important information.



RAG: Retrieval Augmented Generation

Option 2: use Emmy the CNRS assistant (based on MistralAI)





RAG: Retrieval Augmented Generation

Option 2: use Emmy the CNRS assistant (based on MistralAI)



Emmy* est un agent conversationnel en intelligence artificielle générative,
respectueux de vos données personnelles et professionnelles.

Bonjour Redouane LGUENSAT

Vous n'avez pas accès à Emmy

L'accès à Emmy est réservé aux agents CNRS
et aux directeurs d'unités.



RAG: Retrieval Augmented Generation

Option 3: rely on Open Source solutions and models



You can do from scratch (like me two years ago...)

- Needs a lot more work
- This field changes **TOO much** and **TOO quickly**, each month a new model !
- **Free !** (if you decide to not use APIs)
- Still not reaching the performance of closed source models but the gap is closing...

ESPRI-chat v0.0.3

What are the different options of bathymetry in NEMO?

Based on the provided passages, there are three options for bathymetry in NEMO:

1. No bathymetry: This option is specified by setting `nn_bathy` to 0. In this case, the model will not read any bathymetry data and will assume a flat ocean floor.
2. Predefined bathymetry: This option is specified by setting `nn_bathy` to 1. In this case, the model will read a bathymetry file called "bathy_meter.nc" in NetCDF format. This file provides the ocean depth at each grid point of the model grid, which is typically obtained by interpolating a standard bathymetry product (such as ETOP02) onto the horizontal ocean mesh.
3. Domain with a bump of topography: This option is specified by setting `nn_bathy` to 1 and using the "EEL-R5" configuration. In this case, the model will create a domain with a seamount (a bump of topography) located at the center of the domain. This option is intended for use in periodic or open boundary channels.

Note that defining the bathymetry also defines the coastline: where the bathymetry is zero, no wet levels are defined (all levels are masked).

Type a message... Submit

Clear

Retry Undo

Examples

What are the different options of bathymetry in NEMO? | What is AGRIF? | How can I add a new diagnostic to NEMO?

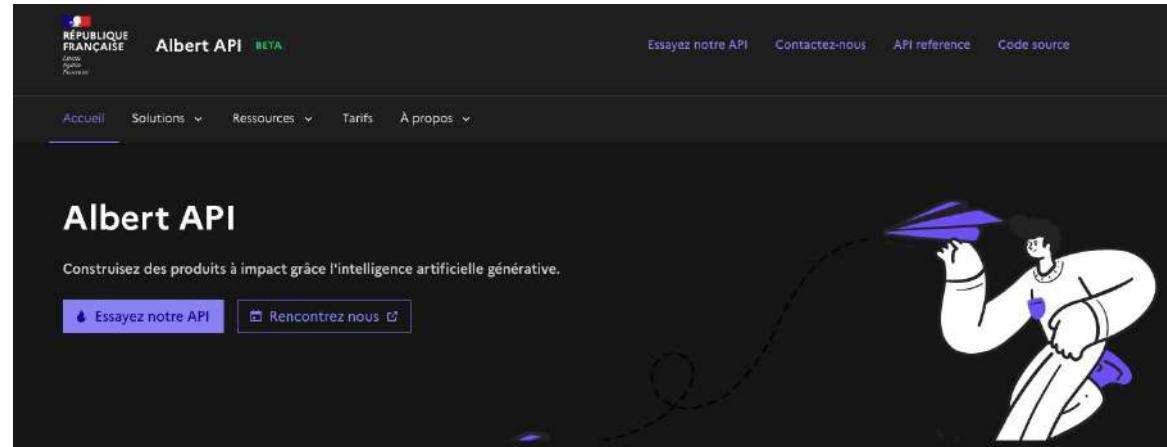


RAG: Retrieval Augmented Generation

Option 3: rely on Open Source solutions and models

Free API access to the best Open Source LLMs !

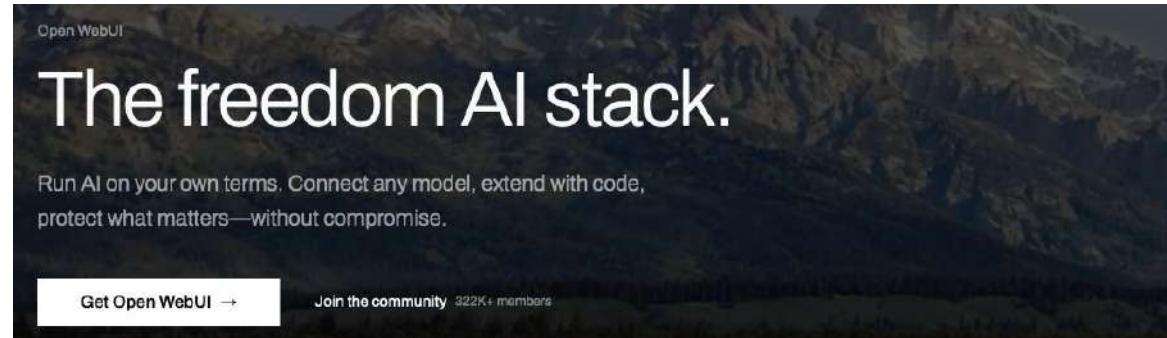
- **Albert API** by DINUM  (big thanks to them for the quality work)
- Free for all the French public servants
- Privacy preserving (SecNumCloud certified)



The screenshot shows the Albert API homepage. At the top, there is a French flag and the text "RÉPUBLIQUE FRANÇAISE". To the right, it says "Albert API" and "BETA". On the far right, there are links for "Essayez notre API", "Contactez-nous", "API reference", and "Code source". Below this is a navigation bar with "Accueil", "Solutions", "Ressources", "Tarifs", and "À propos". The main content area has a dark background with the text "Albert API" and "Construisez des produits à impact grâce l'intelligence artificielle générative.". Below this are two buttons: "Essayez notre API" and "Rencontrez nous". To the right of the text, there is a white cartoon illustration of a person in a white lab coat flying a purple paper airplane.

Open Source platform with built-in RAG !

- **Open WebUI** is a user-friendly self-hosted AI platform designed to operate entirely offline.



The screenshot shows the Open WebUI homepage. The background is a dark image of a forest. At the top, it says "Open WebUI". The main title is "The freedom AI stack.". Below the title, the text reads "Run AI on your own terms. Connect any model, extend with code, protect what matters—without compromise.". At the bottom, there are two buttons: "Get Open WebUI" and "Join the community 322K+ members".



RAG: Retrieval Augmented Generation

Option 3: rely on Open Source solutions and models 🏆

The screenshot shows a dark-themed interface for 'Assistant NEMO'. On the left is a sidebar with various links: 'Open WebUI', 'New Chat', 'Search', 'Notes', 'Workspace', 'Folders', 'Chats', 'Today', 'NEMO Core Components' (with 'Bathymetry Types in NEMO' expanded), 'Previous 30 days', 'ESPRI Research Overview', '2025', 'Accès resto CROUS', 'Contacter les RH IPSL', 'Directeur de l'IPSL', and 'Conversation Assistance'. The main area has a header 'what are the different types of bathymetry I can use?'. Below it, the 'Assistant NEMO' section says 'Thought for 1 seconds' and provides the following text:

You can choose between two broad ways of defining the bathymetry in NEMO:

- Analytical (hand-defined) bathymetry** – set directly in the model configuration (`DOMAINcfg` → `zgr_bathy`).
 - If `nn_bathy ≤ 0` you are using an analytic definition.
 - `nn_bathy = 0` gives a flat-bottom domain (the total depth is obtained from the vertical coordinate transformation).Other non-positive values can be used for predefined analytic bathymetries.
- File-based bathymetry from a NetCDF file** – the model reads depth fields supplied by the user.
 - If `nn_bathy > 0` the bathymetry must be provided in NetCDF files.
 - `nn_bathy = 1` (and similarly `nn_bathy = 2`) requires a bathymetry variable (`cn_bath` in `cn_topo`) and, when ice-shelf cavities are activated (`ln_isfcav = .true.`), an optional ice-shelf draft variable (`cn_isfd` in `cn_fisfd`). These files must contain data on every point of the 2-D model grid, with positive depth values in meters; land is represented by zero or negative values.

So the options are:

- Flat-bottom analytic bathymetry (`nn_bathy = 0`).
- Other analytic bathymetries defined by hand (`nn_bathy ≤ 0` but $\neq 0$).
- NetCDF-based bathymetry (`nn_bathy > 0`), with the necessary variables (`cn_topo`, `cn_fisfd`, etc.) and optional ice-shelf draft data.

These possibilities are described in the NEMO documentation [apdx_DOMAINcfg.tex](#).

At the bottom, there is a 'Send a Message' input field, a '+' button, a 'Capture d'écran' button, a download icon, a microphone icon, and a speaker icon.



THANK YOU

Redouane Lguensat
rlguensat@ipsl.fr

Website: redouanelg.github.io

X: @redouanelg