AI tools for research

Dr. Thé Van Luong – Research service at UNIL
Head of AI and IT systems for research
the-van.luong@unil.ch

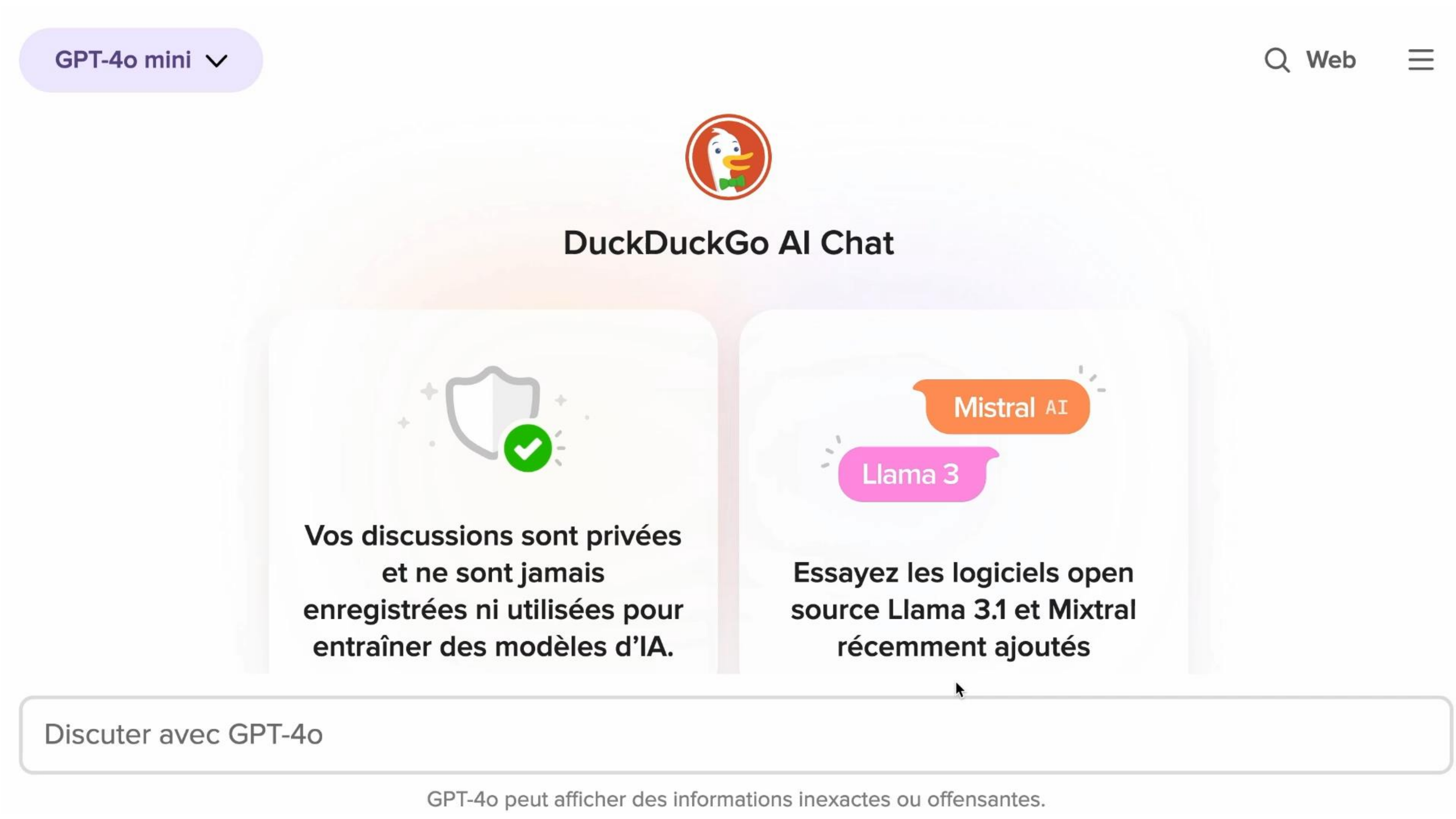
UNIL | Université de Lausanne
Service de la recherche

Overview: Usage of AI in research and publications

- Opportunities and challenges
- Institutional AI projects at UNIL

AI in everyday life

- **Virtual assistants and chatbots**: ChatGPT, Siri and Alexa
- **Content creation and editing**: images, videos and augmented reality
- **Processing and conversion**: transcription, image recognition and automatic translation
- **Recommendations**: analysis, behavior tracking, ads and entertainment
- **Automation and smart devices**: emails, calendars and smart homes
- **Social, wellness and health**: dating apps and medical assistants
- **Navigation and mobility**: GPS and autonomous vehicles
- **Security and finance**: fraud detection and budget management

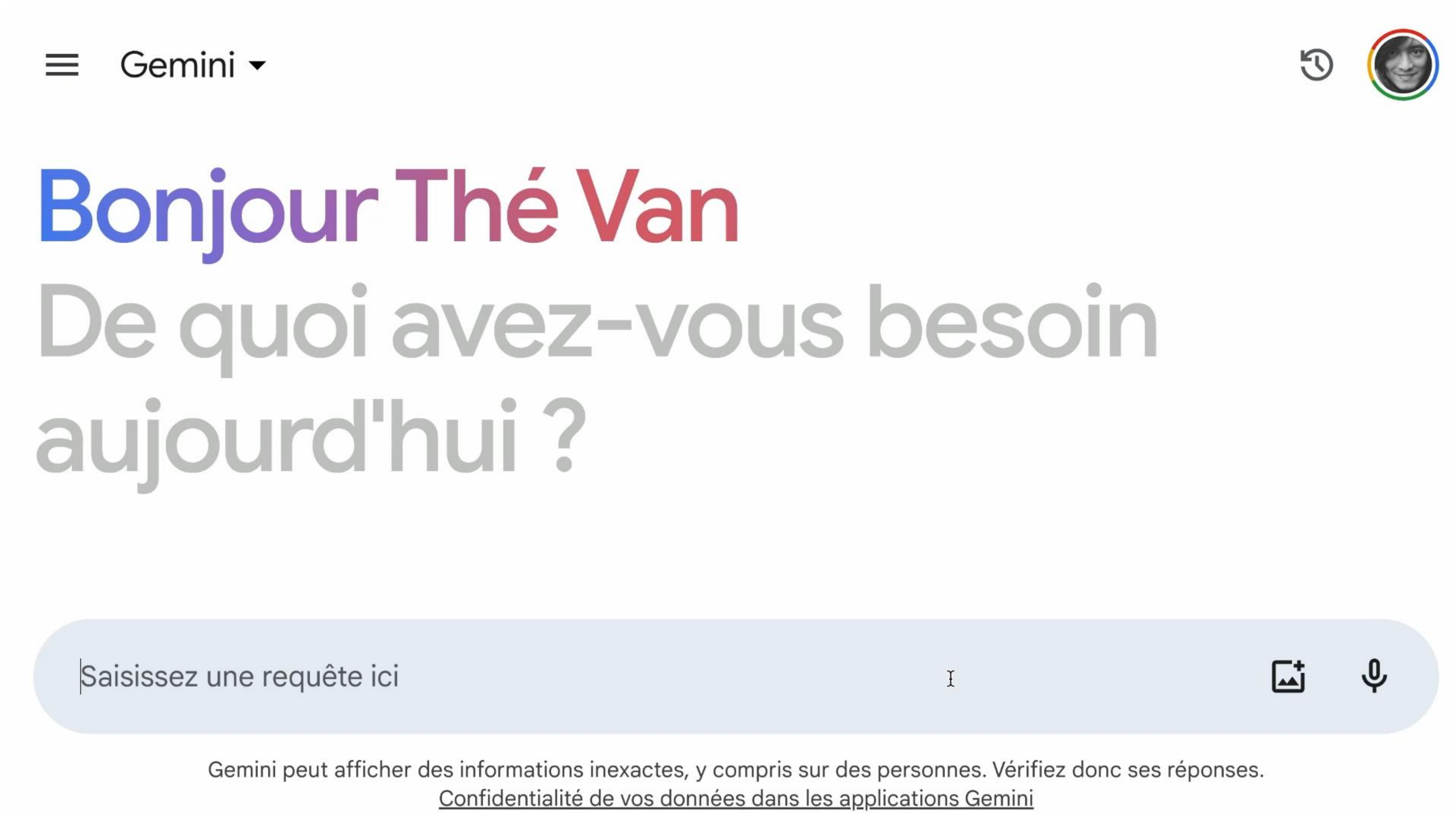


DuckDuckGo AI Chat

Vos discussions sont privées et ne sont jamais enregistrées ni utilisées pour entraîner des modèles d'IA.

Essayez les logiciels open source Llama 3.1 et Mixtral récemment ajoutés

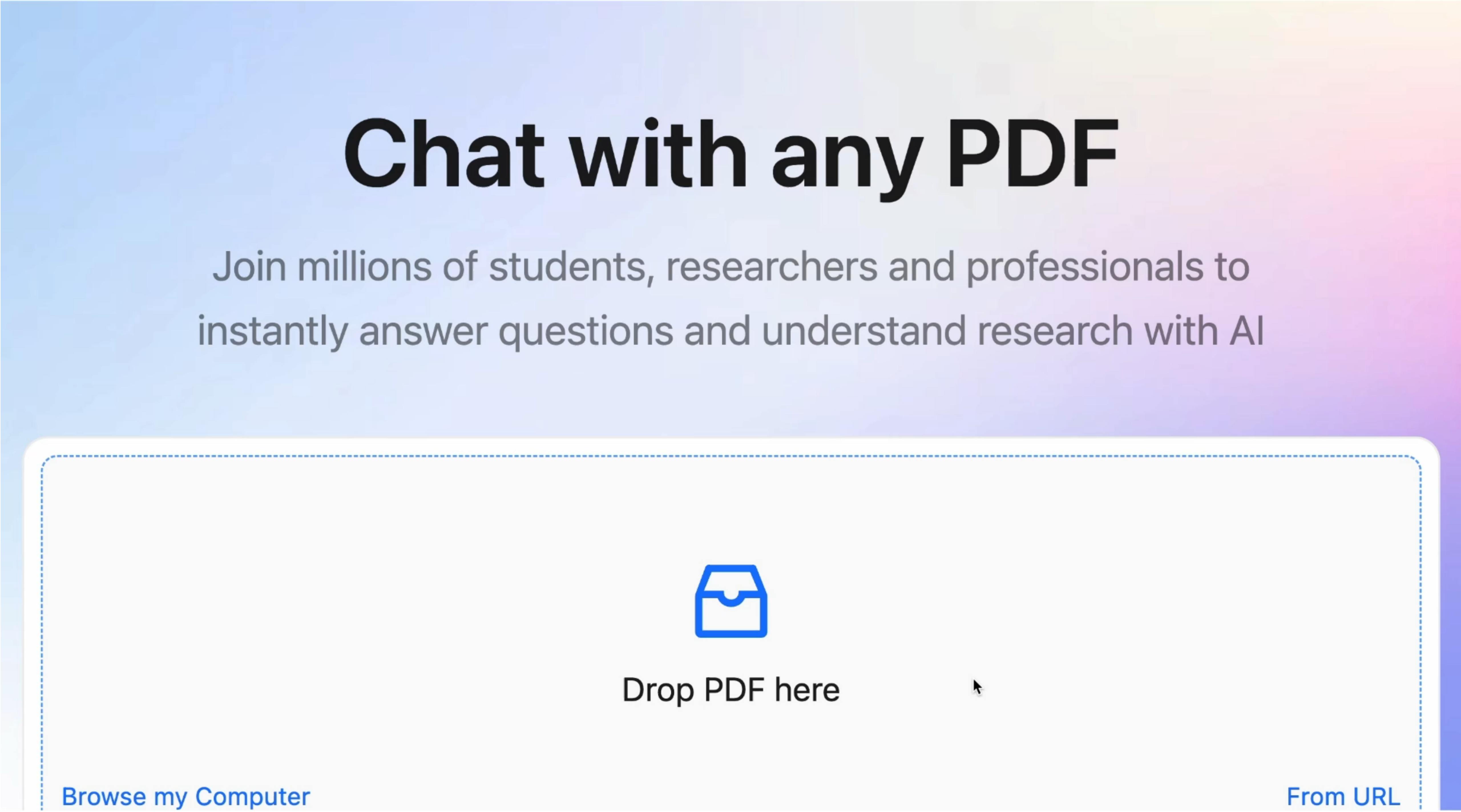
GPT-4o peut afficher des informations inexactes ou offensantes.



Bonjour Thé Van

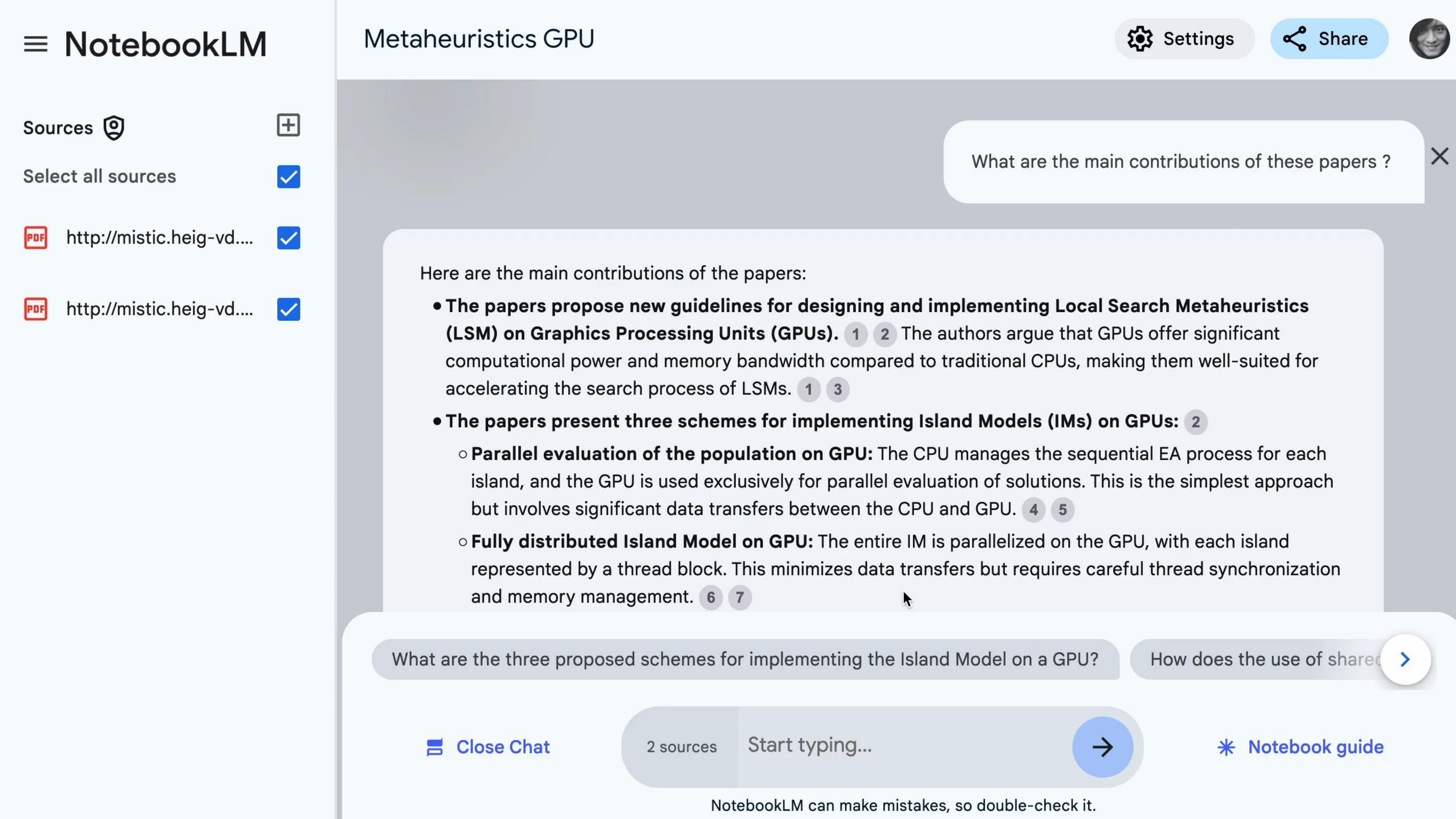
De quoi avez-vous besoin aujourd'hui ?

Gemini peut afficher des informations inexactes, y compris sur des personnes. Vérifiez donc ses réponses.
Confidentialité de vos données dans les applications Gemini



Chat with any PDF

Join millions of students, researchers and professionals to instantly answer questions and understand research with AI



What are the main contributions of these papers ?

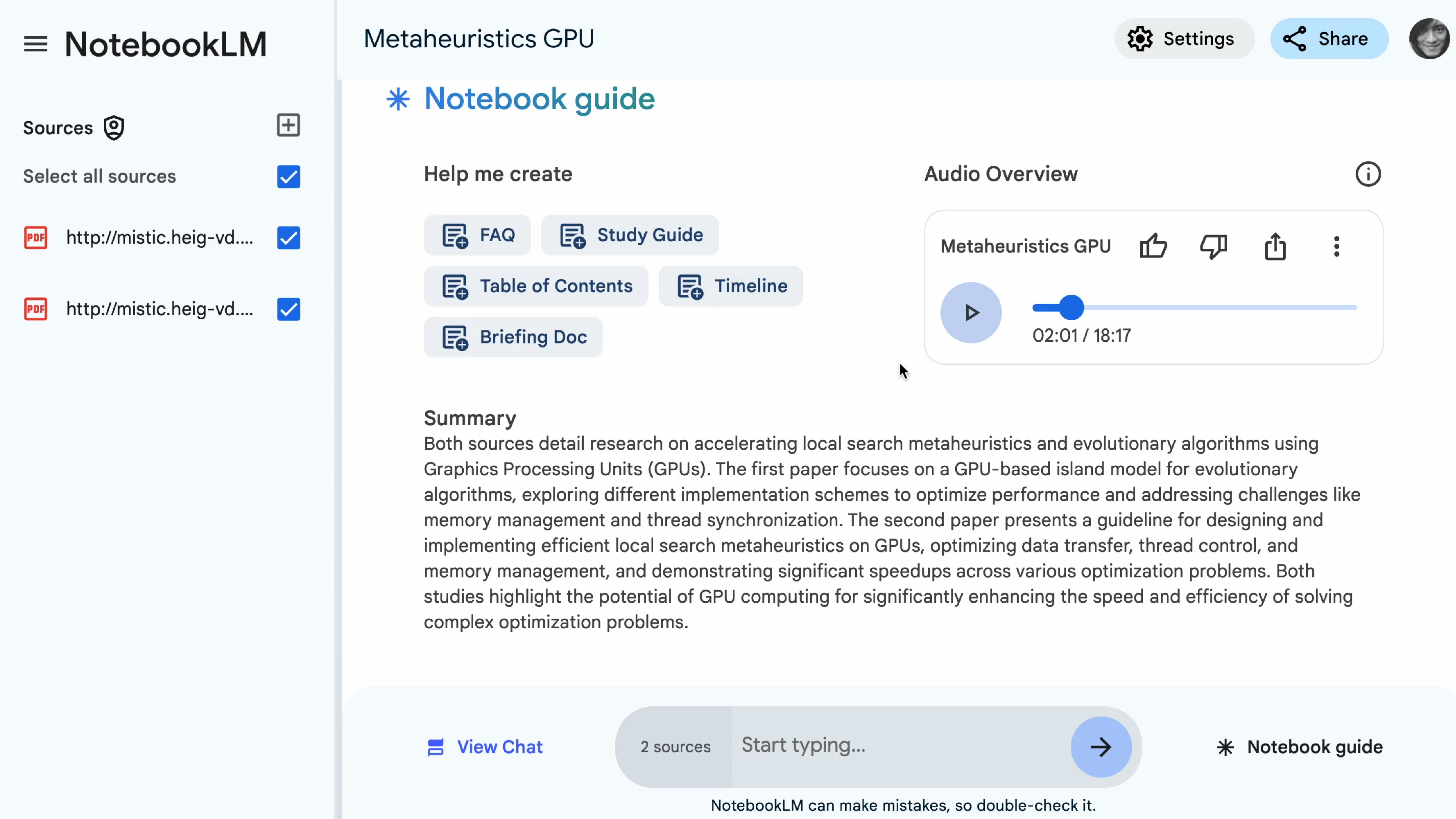
Here are the main contributions of the papers:

- **The papers propose new guidelines for designing and implementing Local Search Metaheuristics (LSM) on Graphics Processing Units (GPUs).** [1] [2] The authors argue that GPUs offer significant computational power and memory bandwidth compared to traditional CPUs, making them well-suited for accelerating the search process of LSMs. [1] [3]
- **The papers present three schemes for implementing Island Models (IMs) on GPUs:** [2]
  - **Parallel evaluation of the population on GPU:** The CPU manages the sequential EA process for each island, and the GPU is used exclusively for parallel evaluation of solutions. This is the simplest approach but involves significant data transfers between the CPU and GPU. [4] [5]
  - **Fully distributed Island Model on GPU:** The entire IM is parallelized on the GPU, with each island represented by a thread block. This minimizes data transfers but requires careful thread synchronization and memory management. [6] [7]

What are the three proposed schemes for implementing the Island Model on a GPU?

How does the use of share

NotebookLM can make mistakes, so double-check it.

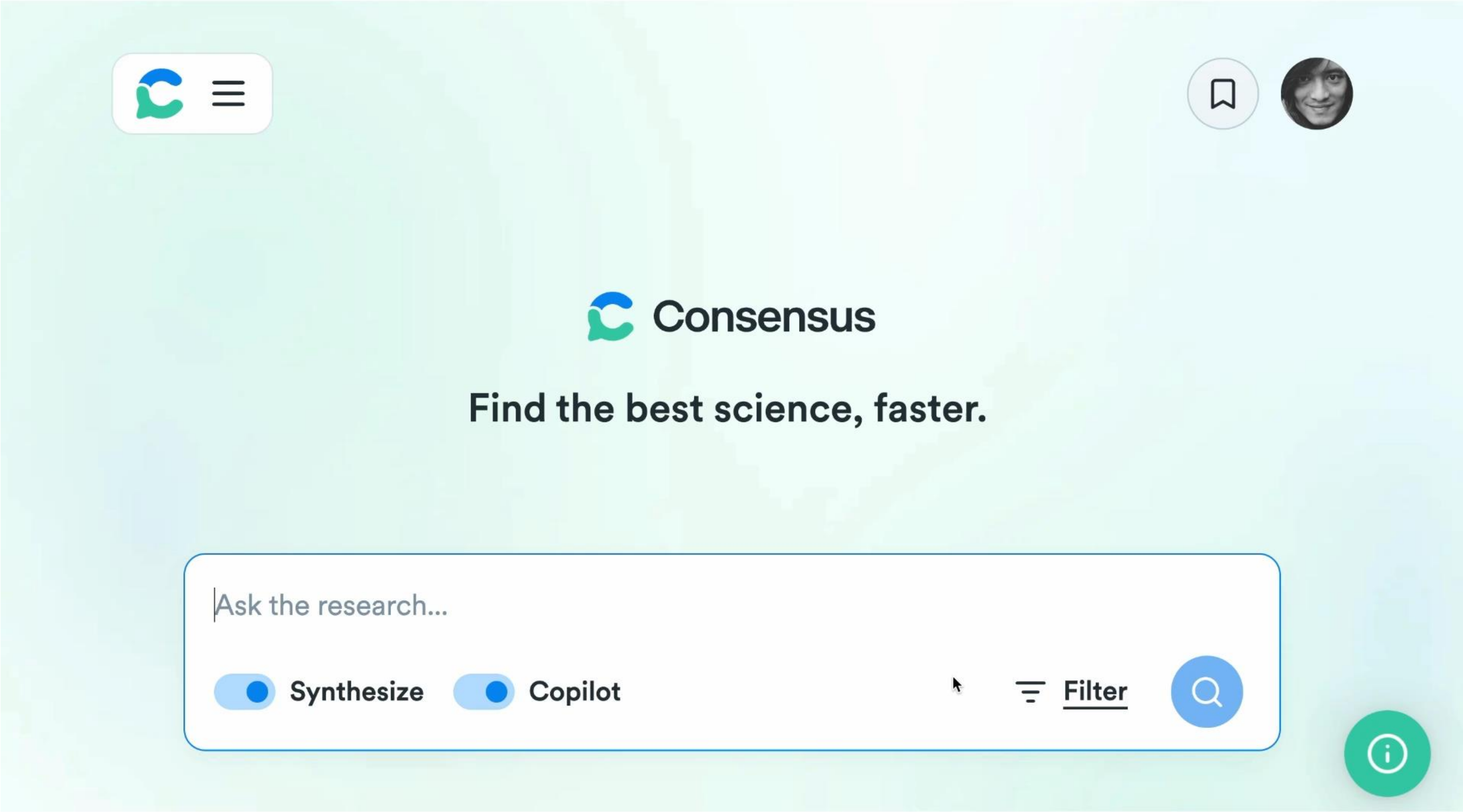


Notebook guide

Summary

Both sources detail research on accelerating local search metaheuristics and evolutionary algorithms using Graphics Processing Units (GPUs). The first paper focuses on a GPU-based island model for evolutionary algorithms, exploring different implementation schemes to optimize performance and addressing challenges like memory management and thread synchronization. The second paper presents a guideline for designing and implementing efficient local search metaheuristics on GPUs, optimizing data transfer, thread control, and memory management, and demonstrating significant speedups across various optimization problems. Both studies highlight the potential of GPU computing for significantly enhancing the speed and efficiency of solving complex optimization problems.

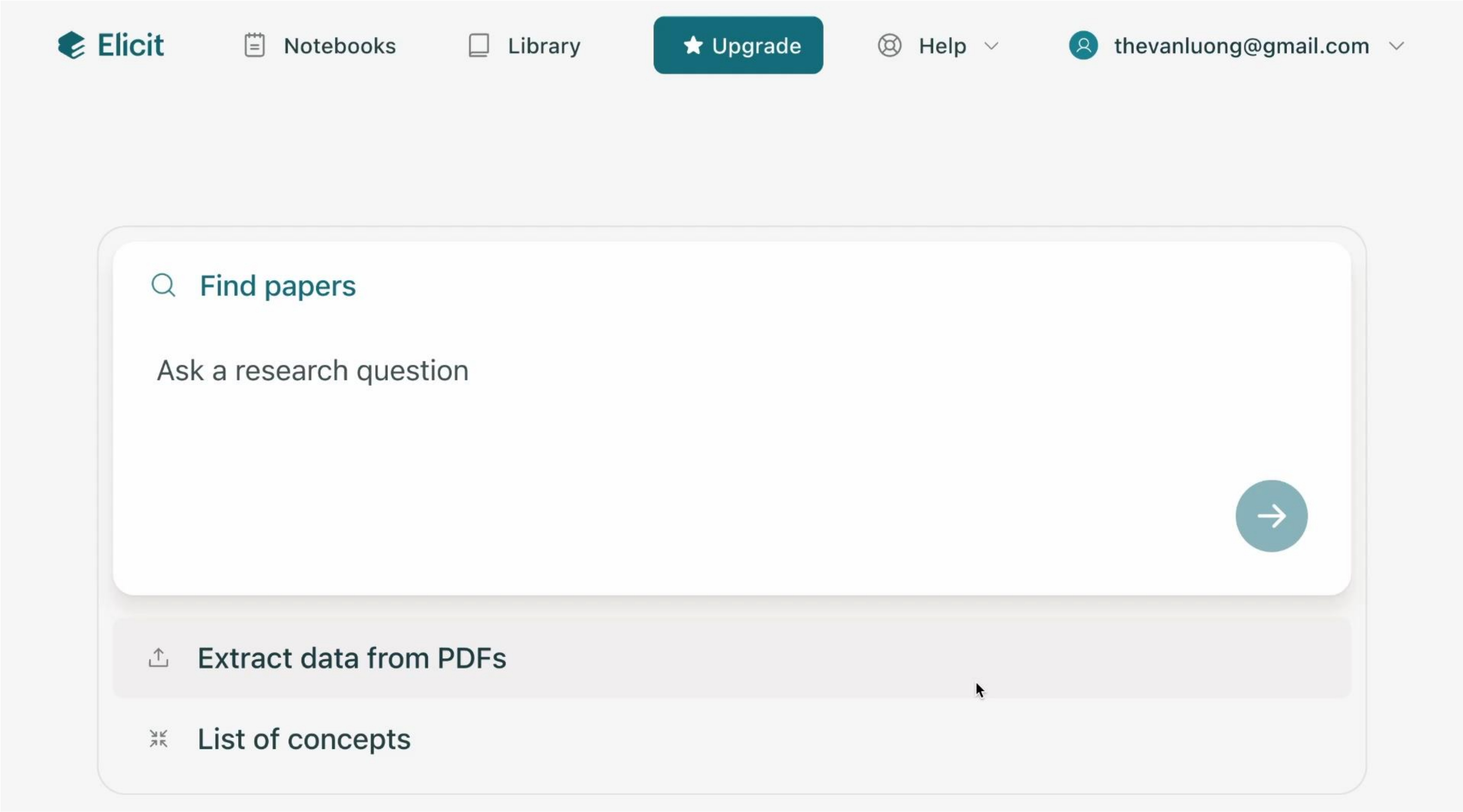
NotebookLM can make mistakes, so double-check it.

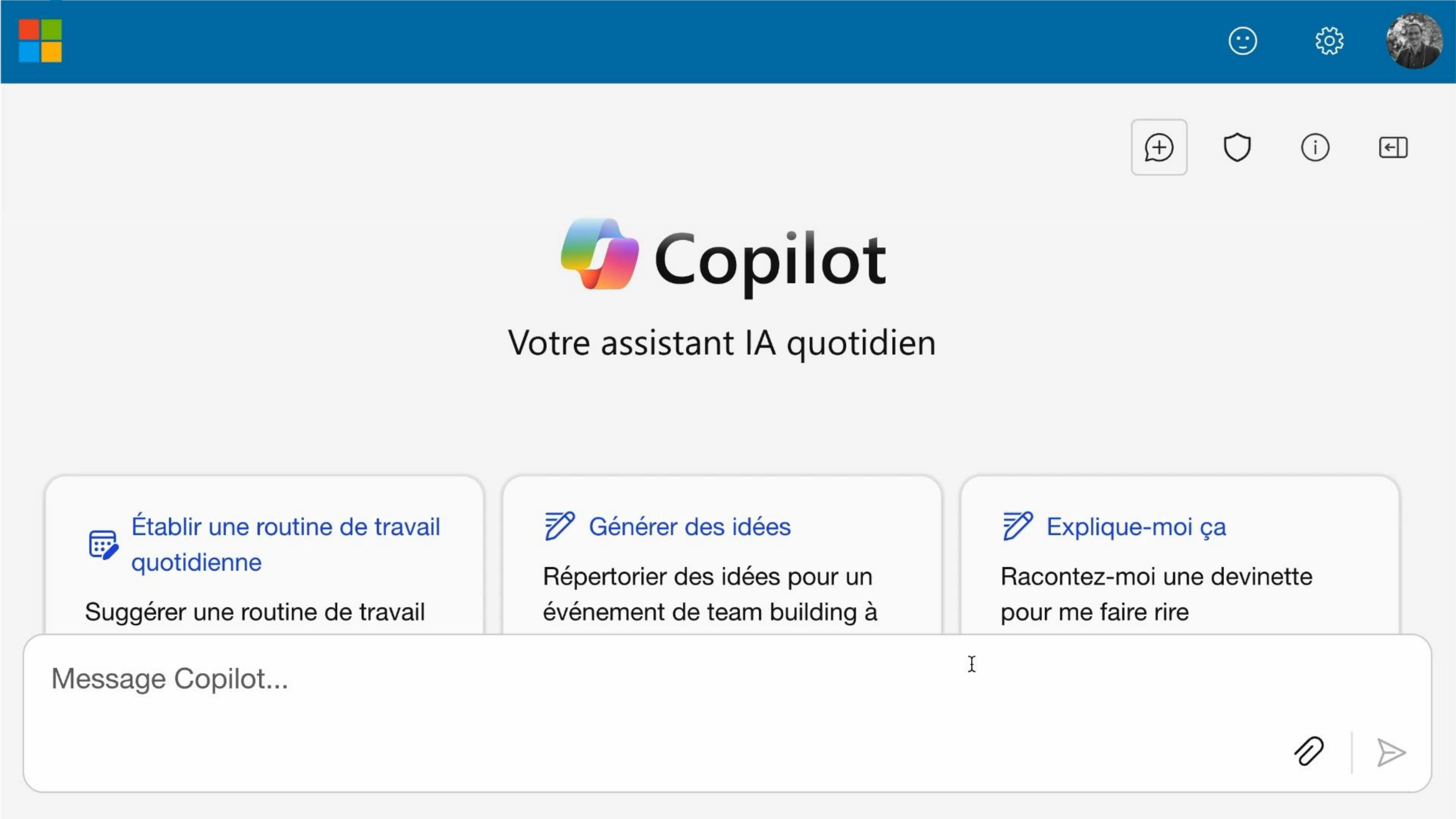


Find the best science, faster.

● Synthesize

● Copilot

Elicit
Notebooks
Library
Upgrade
Help
thevanluong@gmail.com
Find papers
Ask a research question
Extract data from PDFs
List of concepts



Copilot

Votre assistant IA quotidien

Établir une routine de travail quotidienne

Suggérer une routine de travail

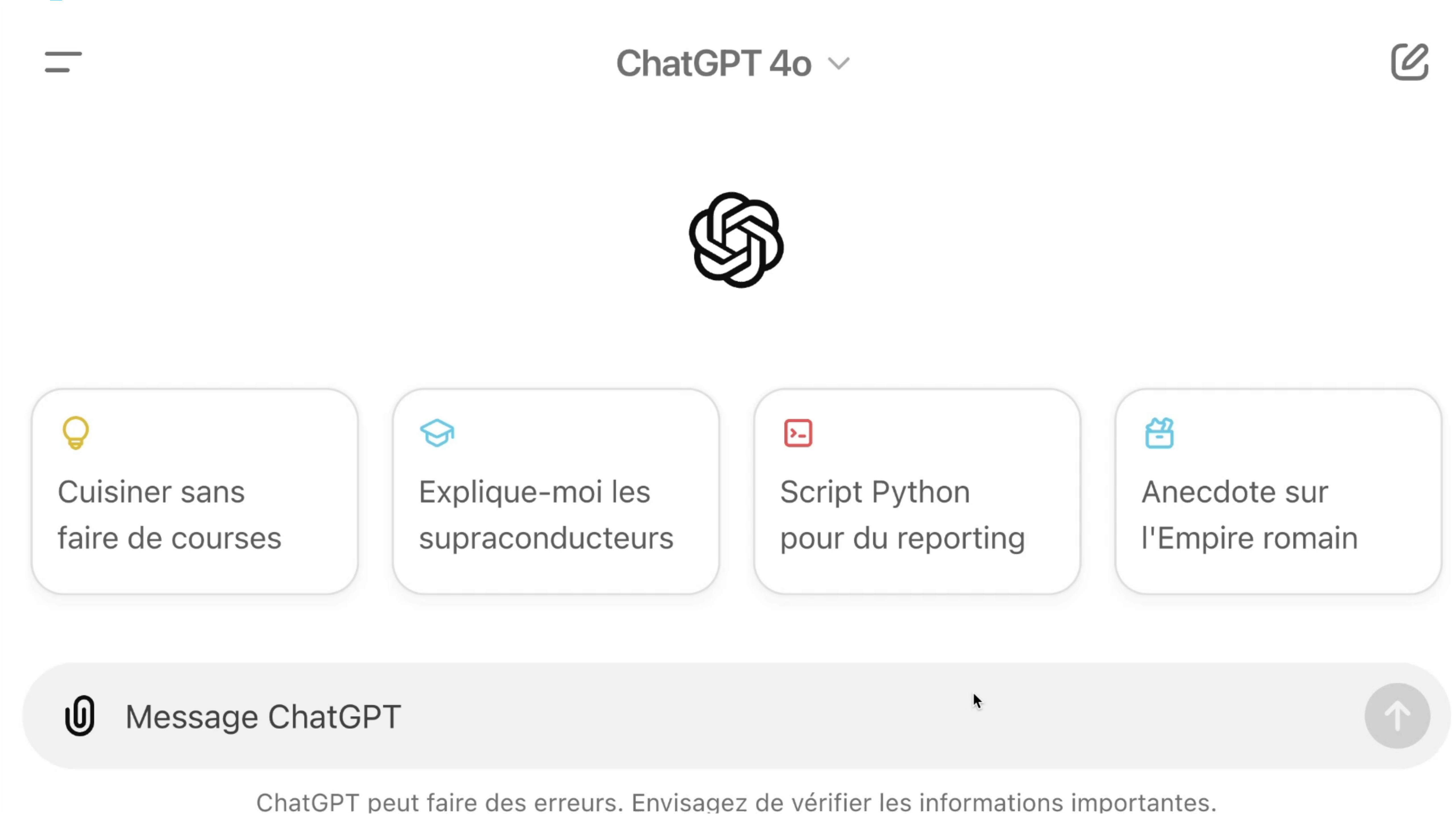
Générer des idées

Répertorier des idées pour un événement de team building à

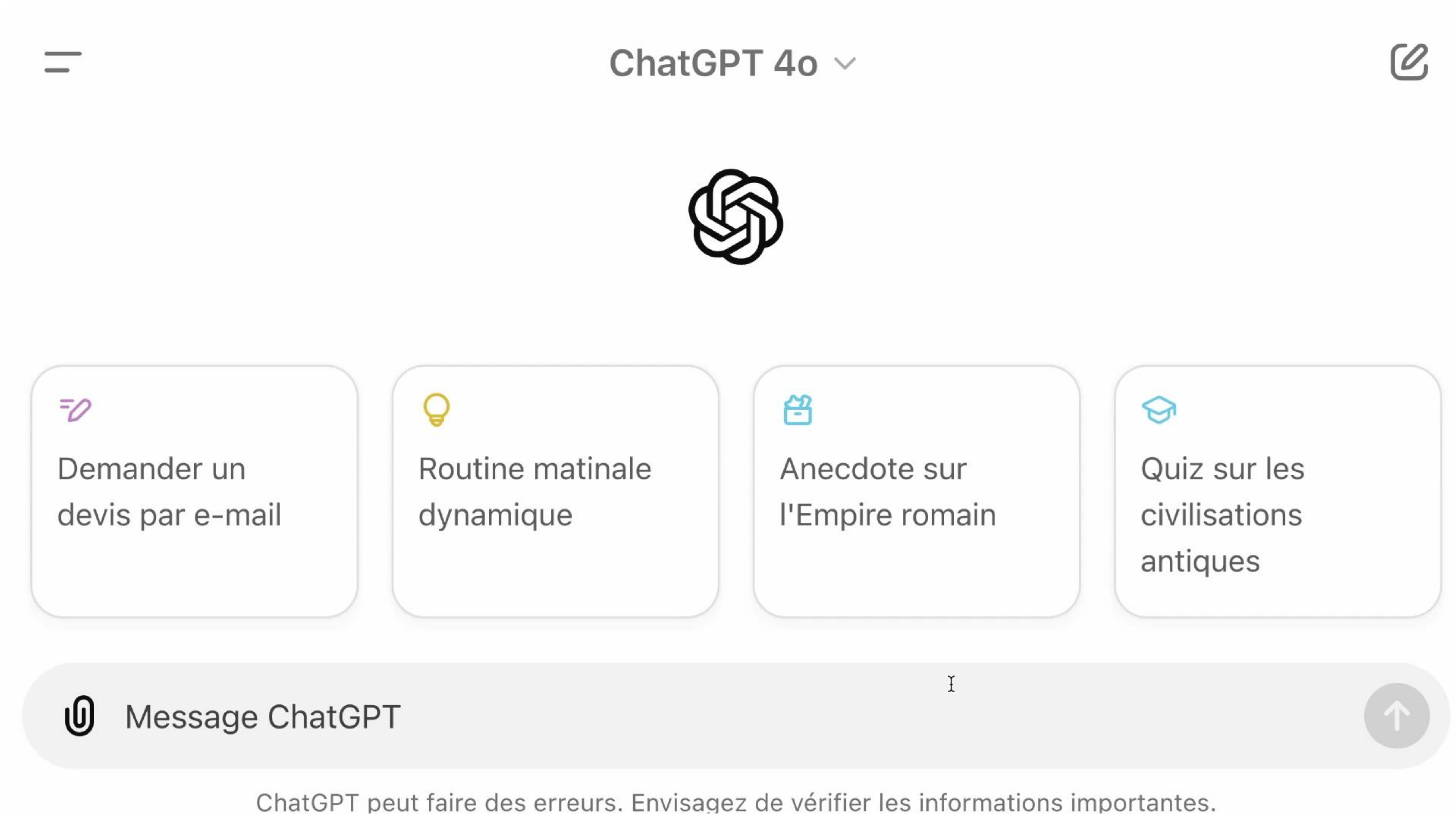
Explique-moi ça

Racontez-moi une devinette pour me faire rire

Message Copilot...


ChatGPT 4o
Cuisiner sans faire de courses
Explique-moi les supraconducteurs
Script Python pour du reporting
Anecdote sur l'Empire romain
Message ChatGPT

ChatGPT peut faire des erreurs. Envisagez de vérifier les informations importantes.



Demander un devis par e-mail

Routine matinale dynamique

Anecdote sur l'Empire romain

Quiz sur les civilisations antiques

ChatGPT peut faire des erreurs. Envisagez de vérifier les informations importantes.

AI FAQ **guidelines** (1/2): academic **usages**

- Research and publications
  Information retrieval, (re)writing, summarizing documents, literature review, generating images, programming, data generation and analysis
- Writing and editing administrative documents
  Reports, emails, memos or funding request
- Educational material creation and course preparation
  Lecture notes, presentations, learning modules, quizzes (Wooclap) or case studies
- Completing assignments and academic tasks
  Essay writing, preparing presentations or creating course summaries
- Organization, planning and project management
  Task organization, timeline planning or resource management

AI FAQ guidelines (2/2): issues and risks

Ethical issues and pitfalls

- Plagiarism, intellectual property, data confidentiality, dependency risks, skill deterioration and energy consumption
- **Statistical generation text**: inaccuracies, context understanding, training bias, content integrity and source referencing
- Proliferation of low-quality articles, fake data, misinformation, and the risk of losing of over personal identity and content

Best practices

- **Data protection**: avoid confidential data, disable model training features in AI tools and ensure minimal anonymization of data
- **Academic integrity**: mention in documents when AI tools have been used
- **Responsability:** you have an expertise and a control on the output

AI initiatives (research + education)

I. High-level, strategic and Direction support

- AI institutional website: UNIL positioning + AI FAQ – unil.ch/numerique
- AI taskforce (advisory commission) + AI strategic unit
- Survey campaign about AI usage from the UNIL community
- Join the Swiss AI initiative (ETHZ and EPFL) and the Alps cluster (~10’000 GPUs)
- AI blog to showcase the AI expertise at UNIL – unil.ch/iaunil
- Update of directives 0-3 Code of Ethics (and student form) and 6-9 Data Protection

II. Support for researchers: resources, awareness-raising and training ressources

- Awareness-raising and trainings (events and faculty seminars)
- Tutorials and resources (padlets and guides)
- Network animation: Group Teams Intelligence artificielle UNIL
- Index card: 10 commandments of AI in research
- Biannual forums on AI and research

III. Development of AI pilot projects (IRIS, CompaSciences and UNIL model on documents)

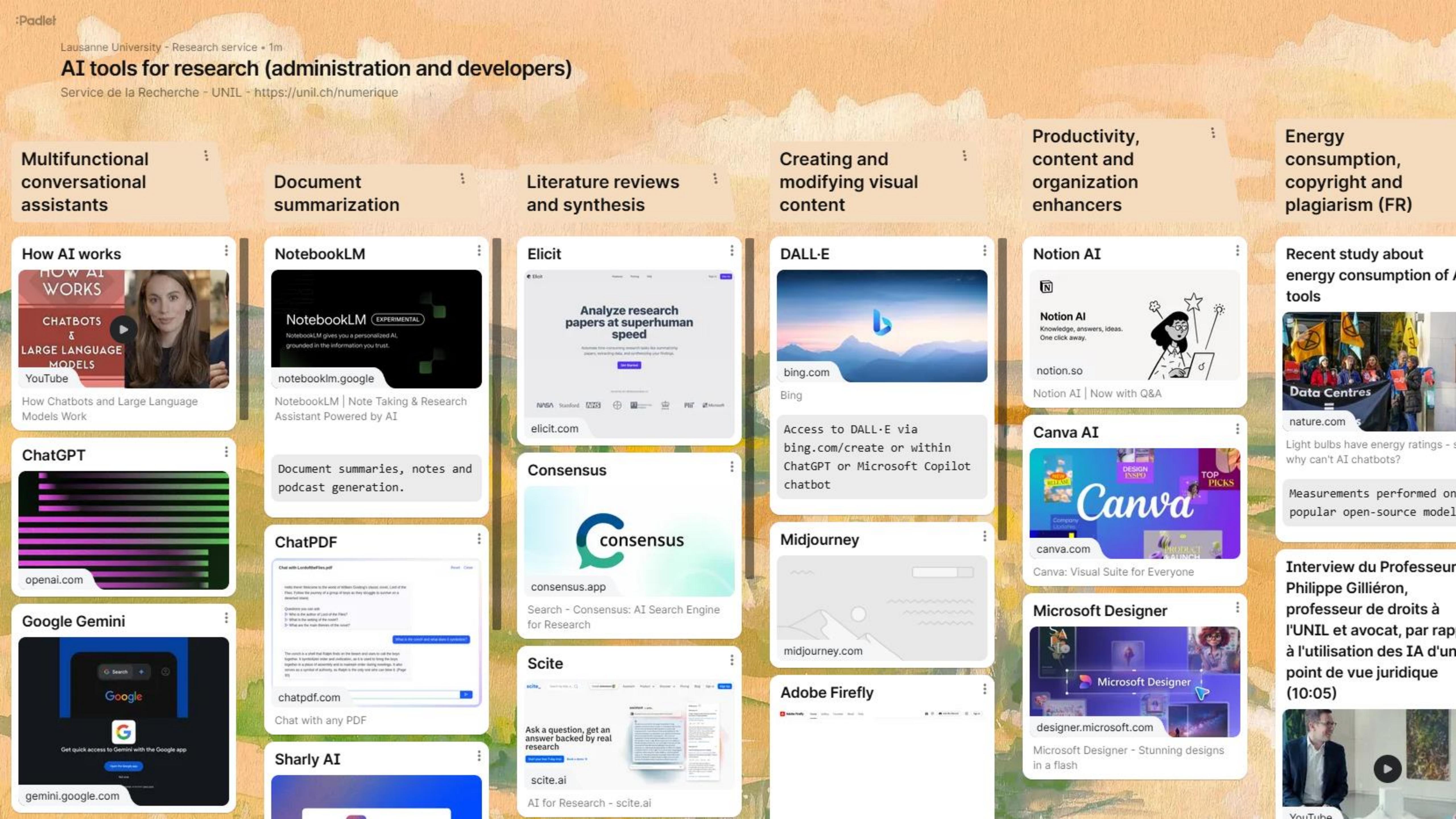
Available AI tools at UNIL

https://copilot.microsoft.com

- You can log in with your UNIL email address!
- OpenAI models (paid version) with different training data and features
- Microsoft Copilot for universities offers additional data protection (not used for training) but data handling practices are unclear

https://corv.unil.ch

- Transcription tool based on local Whisper models and PyAnnote
- Executed locally on Ci DCSR infrastructures. Audio files are not saved
- Best-effort queue system where jobs are submitted and processed as resources allow
- UNIL VPN is required



Padlet

Lausanne University - Research service • 1m

AI tools for research (administration and developers)

Service de la Recherche - UNIL - https://unil.ch/numerique

Multifunctional conversational assistants

How AI works

HOW AI WORKS

CHATBOTS & LARGE LANGUAGE MODELS

YouTube

How Chatbots and Large Language Models Work

ChatGPT

openai.com

Google Gemini

Google

Get quick access to Gemini with the Google app

gemini.google.com

Document summarization

NotebookLM

NotebookLM EXPERIMENTAL

NotebookLM gives you a personalized AI, grounded in the information you trust.

notebooklm.google

NotebookLM | Note Taking & Research Assistant Powered by AI

```
Document summaries, notes and
podcast generation.
```

ChatPDF

chatpdf.com

Chat with any PDF

Sharly AI

Literature reviews and synthesis

Elicit

Analyze research papers at superhuman speed

elicit.com

Consensus

consensus

consensus.app

Search - Consensus: AI Search Engine for Research

Scite

Ask a question, get an answer backed by real research

scite.ai

AI for Research - scite.ai

Creating and modifying visual content

DALL·E

bing.com

Bing

```
Access to DALL·E via
bing.com/create or within
ChatGPT or Microsoft Copilot
chatbot
```

Midjourney

midjourney.com

Adobe Firefly

Productivity, content and organization enhancers

Notion AI

Notion AI

Knowledge, answers, ideas. One click away.

notion.so

Notion AI | Now with Q&A

Canva AI

Canva

canva.com

Canva: Visual Suite for Everyone

Microsoft Designer

Microsoft Designer

designer.microsoft.com

Microsoft Designer - Stunning designs in a flash

Energy consumption, copyright and plagiarism (FR)

Recent study about energy consumption of tools

Data Centres

nature.com

Light bulbs have energy ratings - why can't AI chatbots?

```
Measurements performed on
popular open-source model
```

Interview du Professeur Philippe Gilliéron, professeur de droits à l'UNIL et avocat, par rap à l'utilisation des IA d'un point de vue juridique (10:05)

YouTube

Resources and perspectives

Dr. Thé Van Luong – Research service at UNIL

the-van.luong@unil.ch

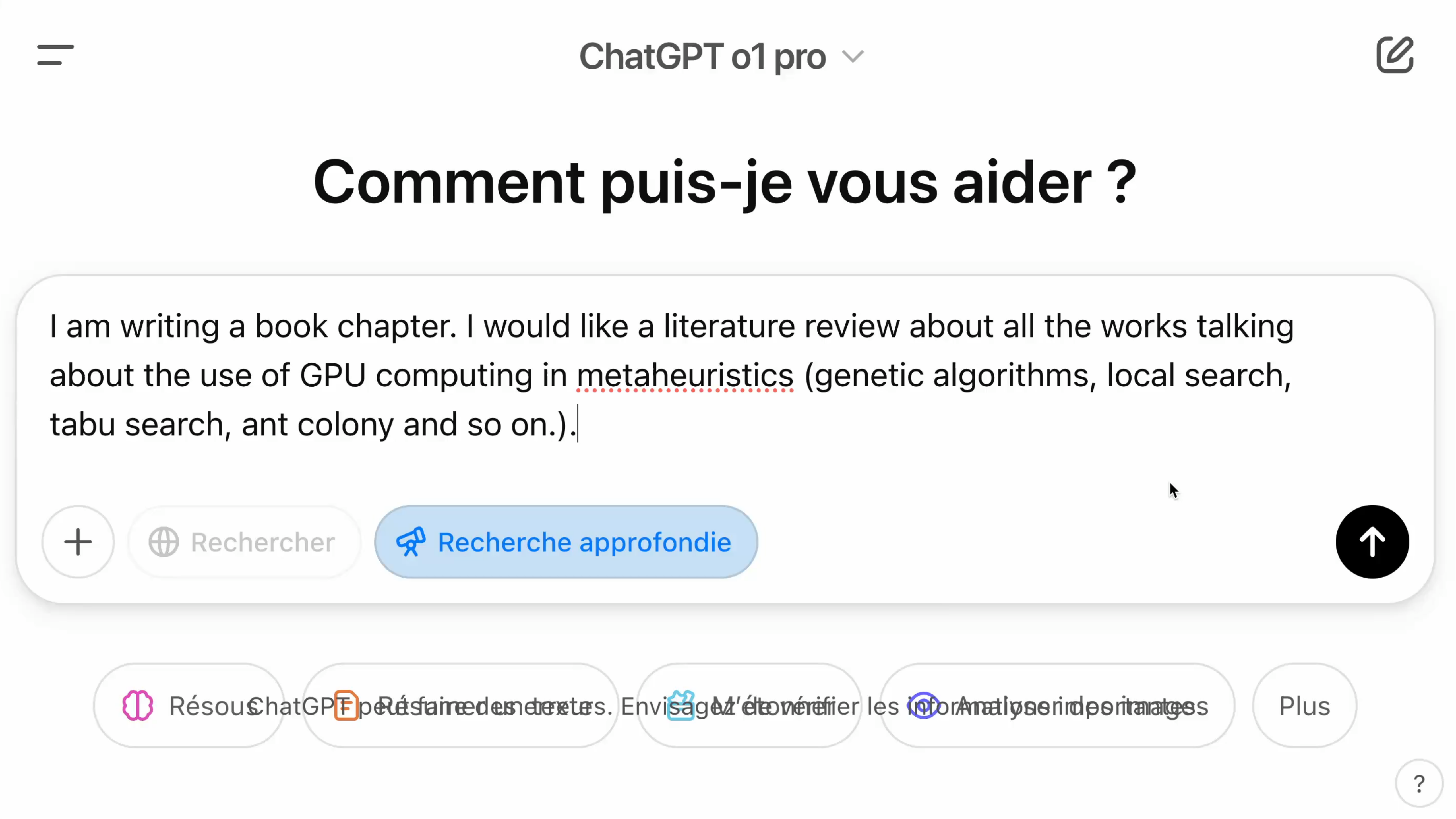
UNIL | Université de Lausanne

Service de la recherche

AI tools: padlet.com/AI_research and **copilot.microsoft.com** (UNIL login)

AI on the horizon:

- Multimodal AI: speech, photos and video
- Reasoning models and agents (e.g. ChatGPT-o3, DeepSeek-R1, DeepResearch, Operator)



Comment puis-je vous aider ?

I am writing a book chapter. I would like a literature review about all the works talking about the use of GPU computing in metaheuristics (genetic algorithms, local search, tabu search, ant colony and so on.).

ChatGPT peut faire des erreurs. Envisagez de vérifier les informations importantes.

Resources and perspectives

Dr. Thé Van Luong – Research service at UNIL
the-van.luong@unil.ch

UNIL | Université de Lausanne
Service de la recherche

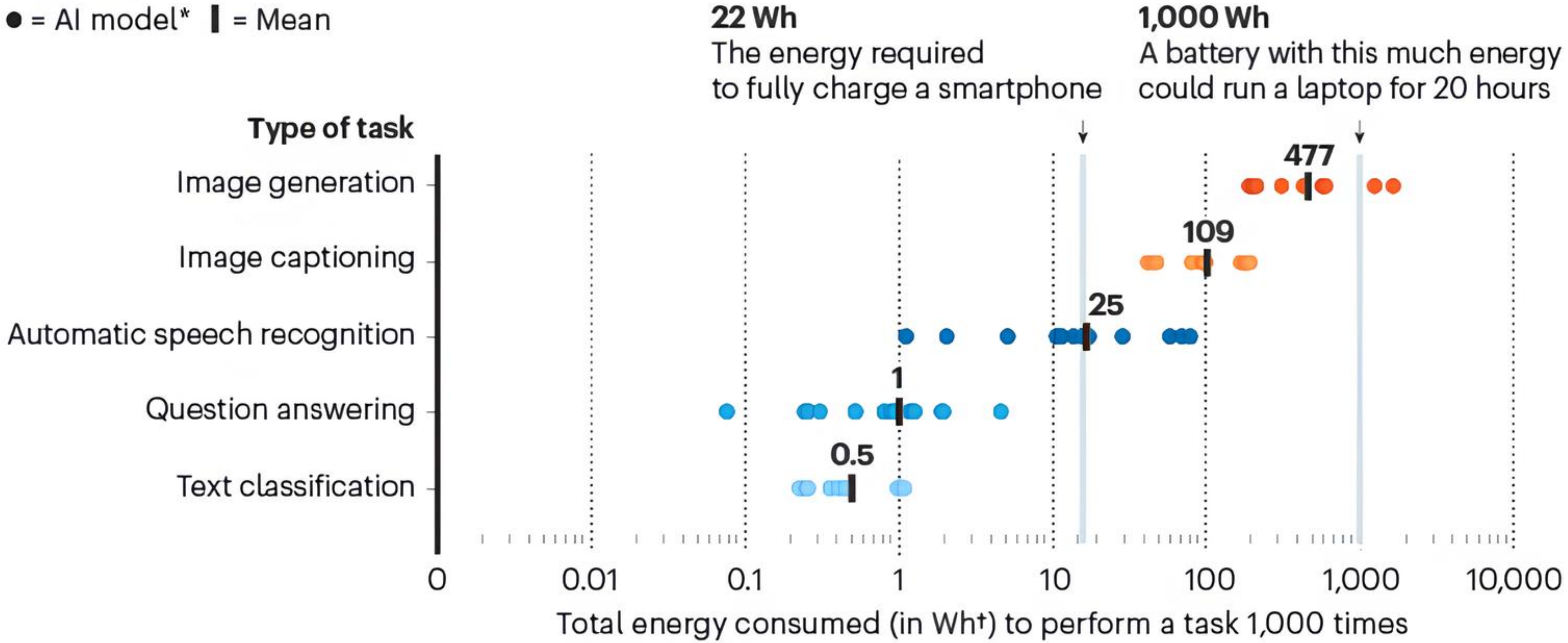
AI tools: padlet.com/AI_research

AI on the horizon:
- Multimodal AI: speech, photos and video
- Reasoning models and agents (e.g. ChatGPT-o3, DeepSeek-R1, DeepResearch, Operator)
- AI skills increasingly required in recruitment
- AI tools integrated into everyday social platforms
- On-device models (e.g. smartphones) for enhanced privacy and energy efficiency

Appendix

AI'S ENERGY FOOTPRINT

The power consumed by artificial intelligence (AI) tools varies greatly depending on the task. An AI model that provides answers to queries is much less energy-intensive than one that generates images from text prompts, for example. And the data show that even AI models of the same type can vary widely in energy consumption.



22 Wh
The energy required to fully charge a smartphone

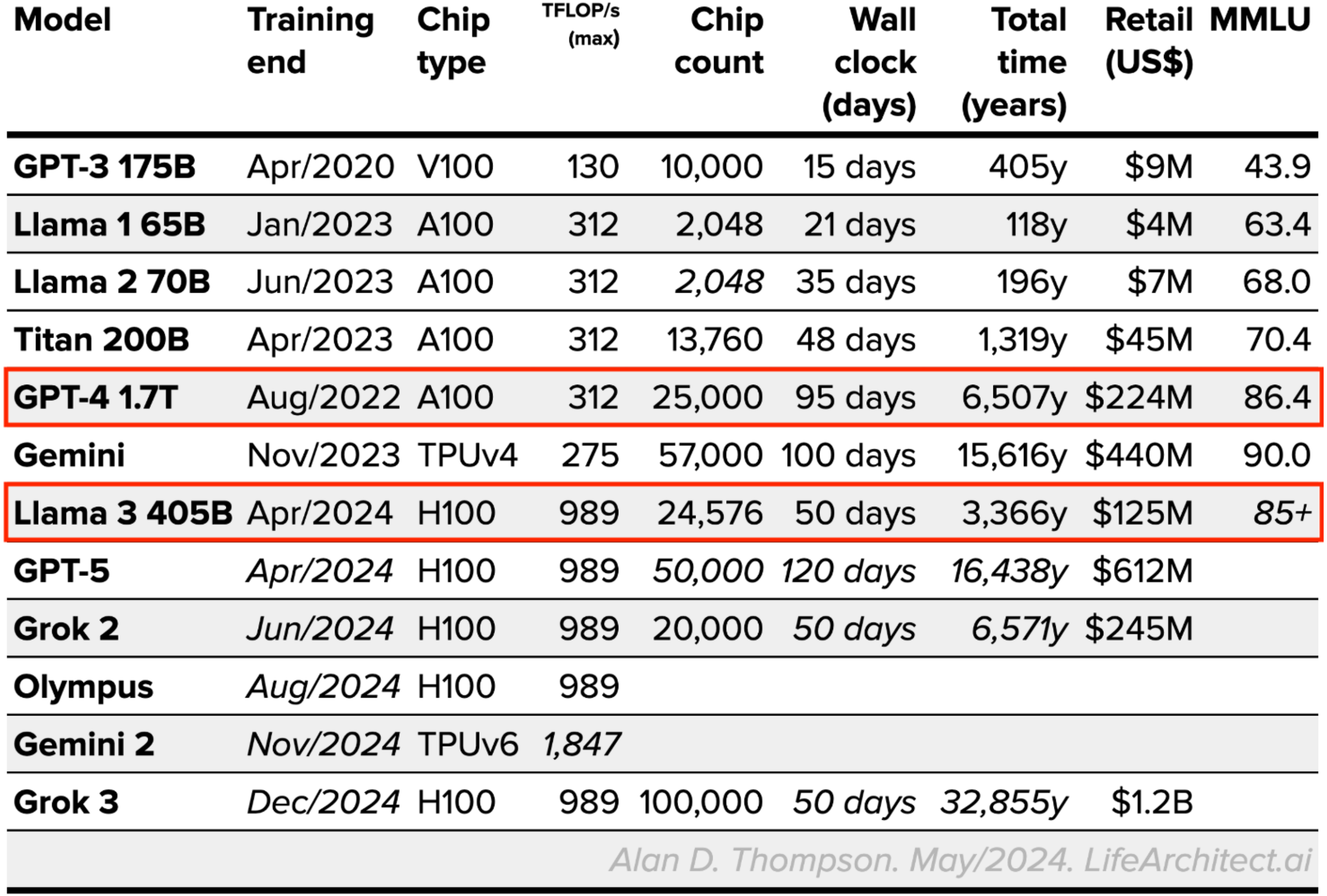
1,000 Wh
A battery with this much energy could run a laptop for 20 hours

*Tests conducted on 20 popular open-source models. Each dot represents one model.
†1 Watt-hour represents power consumption of 1 W extended over 1 hour.

©nature

S. Luccioni et al. Light bulbs have energy ratings—so why can't AI chatbots? Aug/2024

- Question answering: 10x more energy required than a google search.

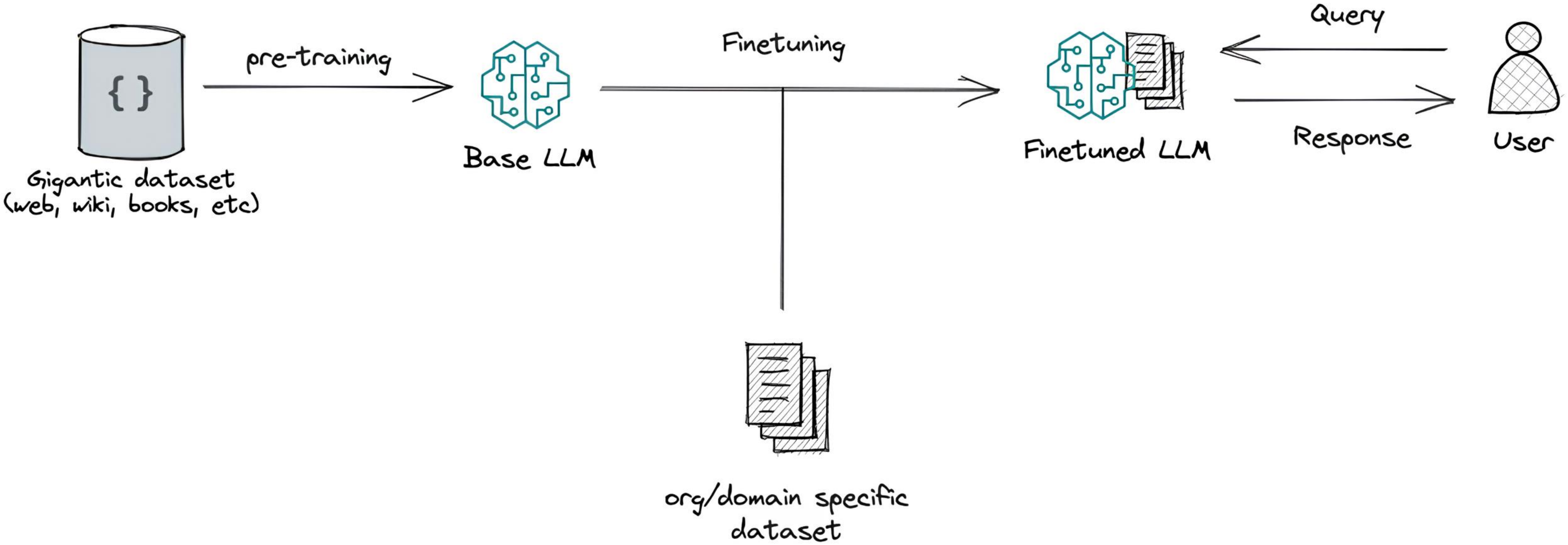


Model	Training end	Chip type	TFLOP/s (max)	Chip count	Wall clock (days)	Total time (years)	Retail (US$)	MMLU
GPT-3 175B	Apr/2020	V100	130	10,000	15 days	405y	$9M	43.9
Llama 1 65B	Jan/2023	A100	312	2,048	21 days	118y	$4M	63.4
Llama 2 70B	Jun/2023	A100	312	*2,048*	35 days	196y	$7M	68.0
Titan 200B	Apr/2023	A100	312	13,760	48 days	1,319y	$45M	70.4
GPT-4 1.7T	Aug/2022	A100	312	25,000	95 days	6,507y	$224M	86.4
Gemini	Nov/2023	TPUv4	275	57,000	100 days	15,616y	$440M	90.0
Llama 3 405B	Apr/2024	H100	989	24,576	50 days	3,366y	$125M	*85+*
GPT-5	*Apr/2024*	H100	989	*50,000*	*120 days*	*16,438y*	$612M	
Grok 2	*Jun/2024*	H100	989	20,000	*50 days*	*6,571y*	$245M	
Olympus	*Aug/2024*	H100	989					
Gemini 2	*Nov/2024*	TPUv6	*1,847*					
Grok 3	*Dec/2024*	H100	989	100,000	*50 days*	*32,855y*	$1.2B	

Alan D. Thompson. May/2024. LifeArchitect.ai

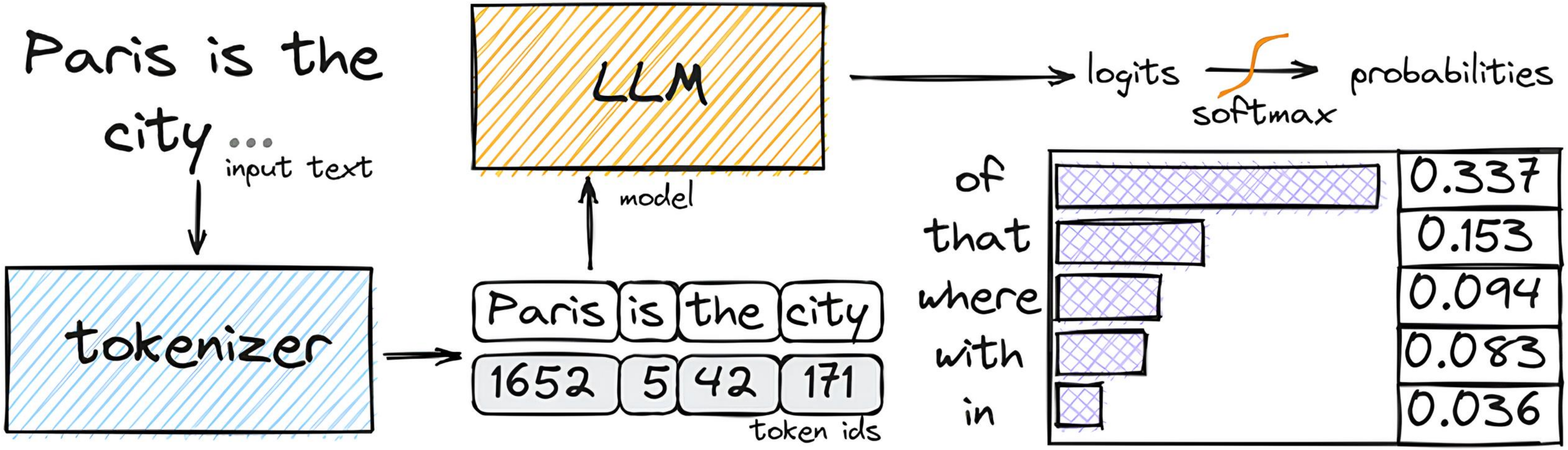
Table. Model training compute (see working, with sources[8]).

How **large langage models** are used


{ }
Gigantic dataset (web, wiki, books, etc)
pre-training
Base LLM
Finetuning
org/domain specific dataset
Finetuned LLM
Query
Response
User

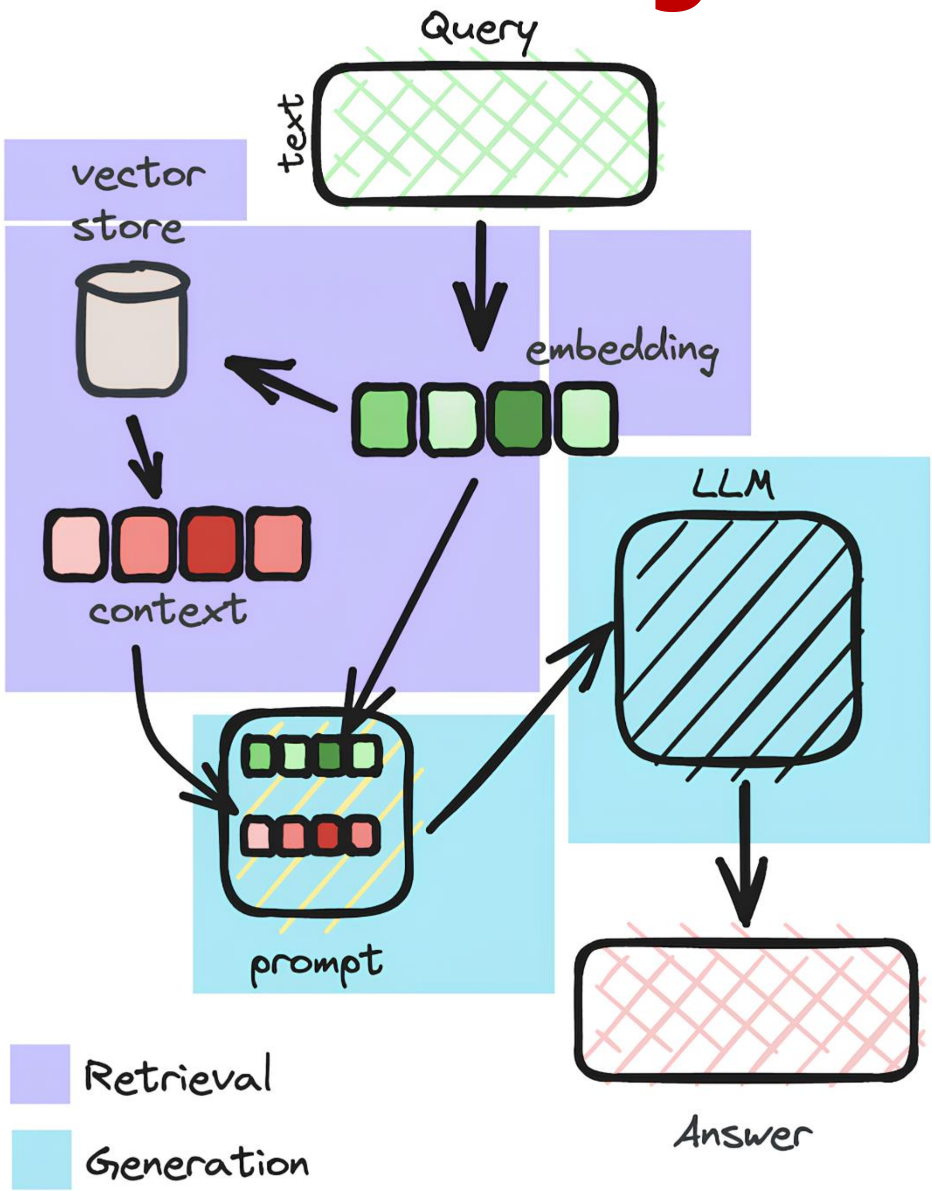
- **Training**: the LLM is pre-trained on a massive dataset to create a versatile base LLM
- **Fine-tuning**: the base LLM is specialized with domain-specific data
- **Inference**: process of generating results from new input

Decoding and generating text



- Tokenization: converts input text into tokens
- Model processing: tokens become embeddings (representation) and the model predicts the next tokens by generating scores
- Softmax function: converts scores into probabilities to determine the most likely next word

Retrieved-Augmented Generation (global)



- RAG enhances LLM responses by retrieving additional context from databases or the web using a vector store
- The supplementary context is combined with a query to form a prompt that the LLM uses to generate a more accurate answer.