



Frontier AI. In your hands.

Configurable AI for all builders

Sopra Steria - Enablement day

June 2025

Agenda of the day

Time	Session
09h30-10h15	Update on Mistral and Latest Product : Kick off the session with an overview of the latest updates and new features in Mistral products.
10h15-11h15	Fine Tuning : Dive into the fine-tuning capabilities of Mistral tools
11h15-12h15	Document AI/OCR : Explore the Document AI and OCR functionalities within Mistral
12h15-14h00	Lunch Time
14h00-15h00	Introduction of Agentic API SDK : Get introduced to the new Agentic API SDK
15h00-17h00	Build the Best Agentic Workflow - Hackathon : Participate in a hackathon to build the best Agentic workflow
17h00-17h30	Hackathon projects presentations



Update on Mistral AI and Latest Product



Mistral today

260 Mistralers, and growing

4 Offices: Paris - London - Palo Alto - Singapore

1.1 B€ funded

15 State of the Art Models

1 Platform

Worldwide market coverage and customers

Partners:



Customers: BNP, Axa, CMA-CGM, EDF, Veolia, Cisco, Belfius, Dassault, Manufacturing, Governments, Defense, digital natives,

...

Mistral AI

Key metrics

270 Mistralers

4 Offices

15 State of the Art Models

Worldwide market

Key industries

Financial institutions

Public Sector

Defense

Advanced Industries

Use cases

Transformative use cases

Advanced customization

Agentic workflow

Coding assistant

Constrained deployment

Configurable AI for all builders and businesses

Customizable

25+ world class and benchmark-setting models, a comprehensive GenAI platform and 2 cutting-edge interfaces to customize, fine-tune and build on

Private and portable

Deploy the platform anywhere—on-premises, cloud, edge, devices, data centers, and more

Deeply engaged

Hands-on assistance from the world's best AI engineers and scientists across deployment, solutioning, safety, and beyond.

Science, Product and Solution

Close loop between science, product and solution to support transformative use cases and at scale deployment of GenAI use cases

Applied AI Services

Model Customization | Custom solutions | Production support



Le Chat

Enterprise AI assistant



Mistral Code

Coding copilot plug-in

La Plateforme

Custom frontier models | AI infrastructure | Administration APIs | Devtools

SaaS

Hybrid

Self-deployed

Edge

The Mistral AI model landscape

Premier models



Our **frontier-class reasoning model**



Mistral Medium 3

Our flagship model, for your **most sophisticated needs**



Ministrail 3B/8B

Our most **efficient / powerful edge models**



Mistral Moderation

Our **moderation service** to detect **harmful text content**



Our **low-latency code model** for code generation



Mistral Saba

Our first model for **Arabic & South Asian languages**



Mistral Embed

Our **embedding model** to make your data accessible



Mistral OCR

The world's **best OCR model** to-date

Open-source models

Apache 2.0



Our **enterprise-grade small model**, cost efficient and fast



Magistral Small

Our **small reasoning model**



Devstral Small

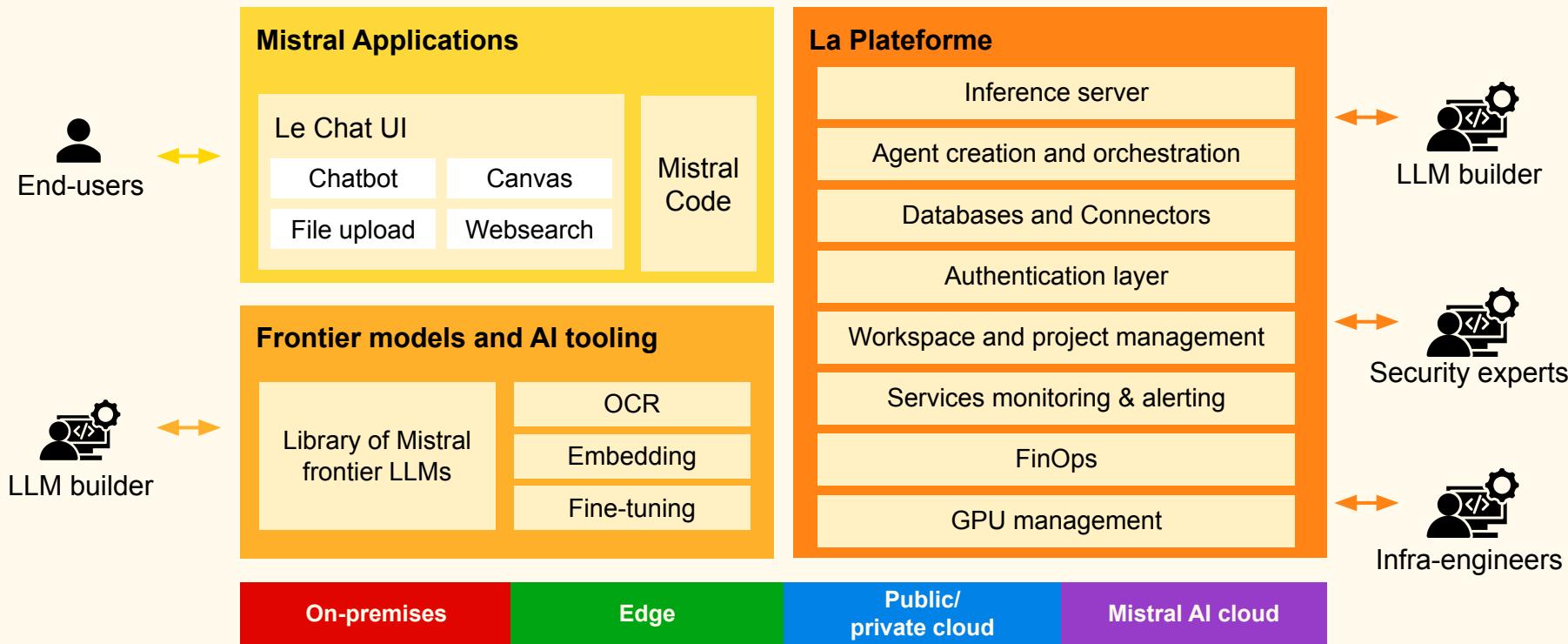
Our **agentic LLM** for software engineering tasks



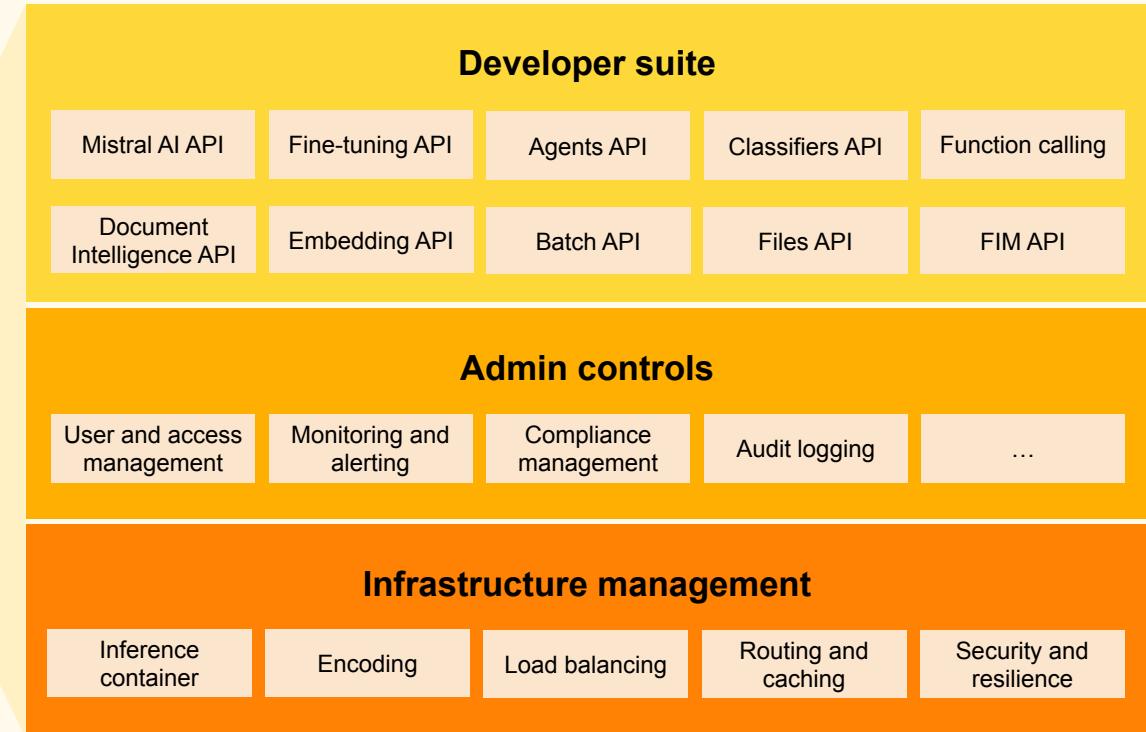
Codestral Mamba

Our coding model dedicated to **new research**

The Mistral AI Platform ecosystem and users

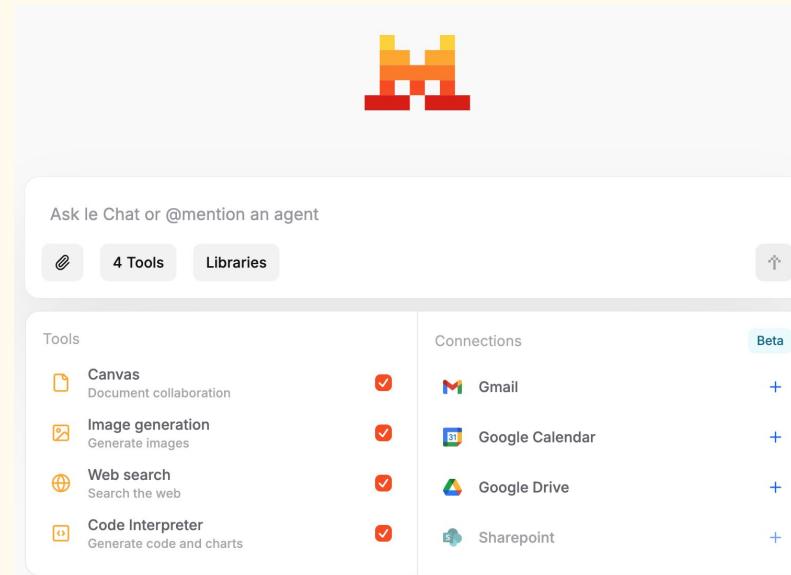


La Plateforme



Le Chat enterprise for individual productivity

1. **General AI assistance**
2. **Knowledge** integration via connectors
3. **Document library**
4. **Agent creation** in no code mode
5. **File upload** and analysis
6. **Code execution** sandbox
7. **Canvas** for creative ideation
8. **Personalization**, image generation, and much more

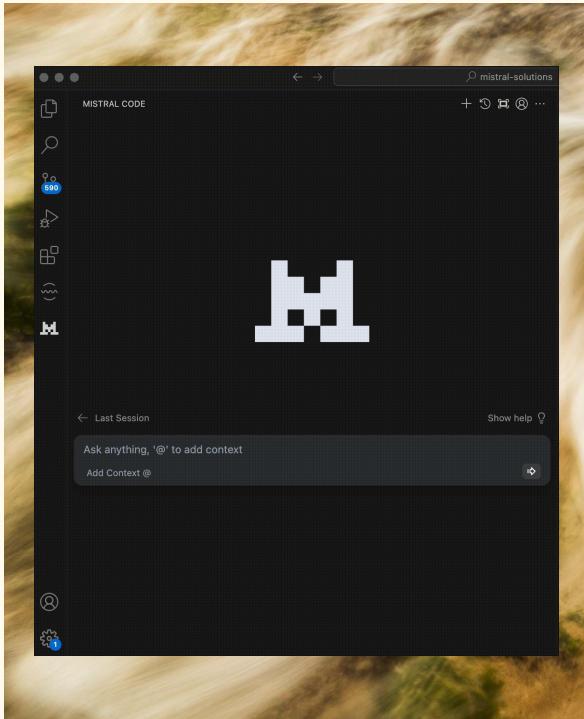


The screenshot shows the Mistral AI interface with a yellow header bar. The main area features a large orange 'M' logo at the top right. Below it is a search bar with the placeholder 'Ask le Chat or @mention an agent'. Underneath the search bar are three buttons: a pencil icon, '4 Tools', and 'Libraries'. To the right of these buttons is an upward arrow icon. The main content area is divided into two sections: 'Tools' on the left and 'Connections' on the right. The 'Tools' section contains four items: 'Canvas' (Document collaboration), 'Image generation' (Generate images), 'Web search' (Search the web), and 'Code Interpreter' (Generate code and charts). Each tool has a checked checkbox to its right. The 'Connections' section contains five items: 'Gmail', 'Google Calendar', 'Google Drive', and 'Sharepoint', each with a plus sign to its right. A 'Beta' label is located in the top right corner of the 'Connections' section.



Deep dive next

Four ways to use Mistral Code



 **Chat**

Ask in natural language coding related questions

 **Context integrations**

Add files, terminal commands, git commit to your prompt

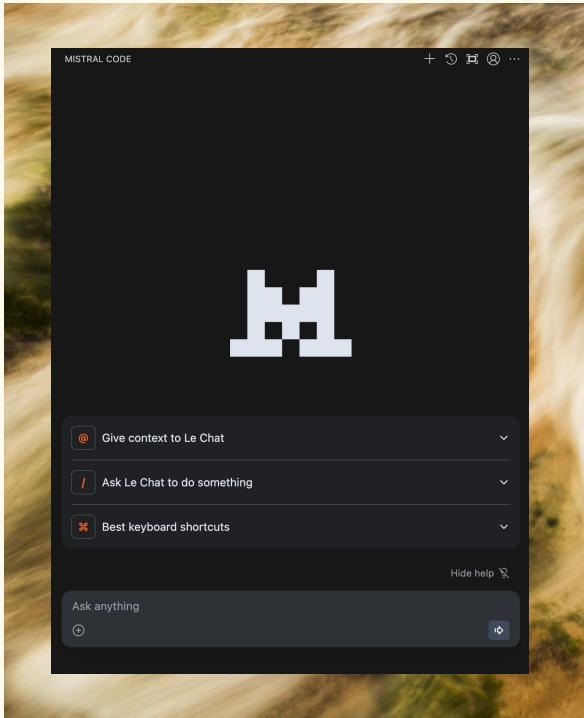
 **Auto-Completion**

Augment developer with live auto-completion (FIM)

 **Action automations**

Automate repetitive tasks in your codebase

Four ways to use Mistral Code



Chat

Ask in natural language coding related questions



Context integrations

Add files, terminal commands, git commit and more to your prompt



Auto-Completion

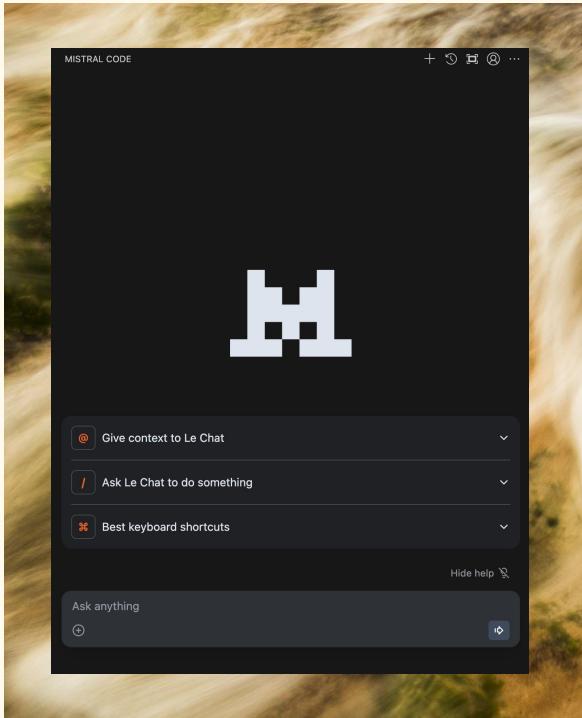
Augment developer with live auto-completion (FIM)



Action automations

Automate repetitive tasks in your codebase

Mistral Code is a unique and reliable code assistant



State of the Art model

- **Performant & efficient:** Codestral is the fastest and best-in-class Fill in the middle in the market

Enhanced features

- **Fine-tuning** codestral on your repositories
- **Optimized auto-complete** with smart indexation of your codebase
- **Context enhanced** by other files, git commits, latest messages from terminal, detected problems
- **Speculative decoding** for faster experience in auto-complete

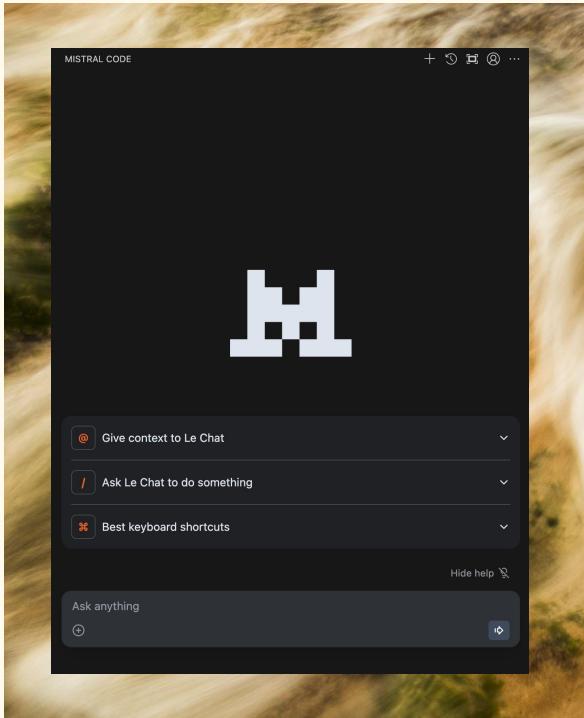
Deployment & integration

- **Hosted where you want:** on prem, VPC on the cloud
- **La Plateforme:** user management, access control and usage monitoring
- **Available** for VSCode and JetBrains

Support

- Our team **maintains the plugin**
- Support for **customization** (tailored context providers or automations)

Mistral Code available features



Available features

- **Optimized auto-complete** with smart indexation of your codebase
- **Speculative decoding** for faster experience in auto-complete
- **Context Providers** : Terminal, Git Diff, Files, Problems (ex: SonarQube reports, etc.)
- **Off-the-shelf Automations** : Commit message, Shell command, auto apply code from chats, Optimize code, Fix code, Write docstrings, Write comments
- **Observability metrics** in Mistral Console (% of auto-generated code, usage reporting, etc.)
- **Keyboard shortcuts** for seamless experience: debug, insert suggestions, accept, reject diff, etc.

Coming soon

- **RAG** on your external documentation & codebase with native code embedding
- **Custom agents** for action automation and code tasks
- **Fine-tuning** codestral directly on your repositories

We've delivered for our customers

Find more stories on:
<https://mistral.ai/customers>



BNP PARIBAS

Use case: Comprehensive internal generative AI platform

Adoption: 30k+ users, 100+ use cases across banking, markets, retail

Metrics: ~50% cheaper than Llama and other alternatives

Business value: Est. >\$30M / yr productivity unlocked

Helsing

Use case: Vision-Language-Action models, enhancing defence platforms' ability to understand their environment, communicate with operators, and make rapid, reliable decisions in complex scenarios. This initiative fosters human-AI collaboration on the battlefield and revolutionizing defence operations.

Qonto

Use case: Banking Advisor Agent for Customer Support

Adoption: >5000 tickets automated per month

Metrics: 88% accuracy, 80% decrease in cost per ticket

Business value: Est. ~\$500k / yr. saved



Use case: Enterprise AI agents for customer service and internal RAG applications

Adoption: 30k+ active users

Metrics: 20%+ uplift in daily productivity with task automation

Business value: Est. >\$50M / yr. saved

Carrefour The Carrefour logo, which is a red stylized arrow pointing to the left.

Use case: Automatic customer review response

Adoption: 3k+ stores, 1.3M monthly reviews

Metrics: 80%+ reduction in review response time, 80%+ satisfaction

Business value: Est. >\$10M / yr. saved



Use case: SQL assistant Cortex Analyst

Adoption: Est. 100+ Snowflake customers in production

Metrics: ~90%+ accuracy, ~2X single-shot SQL generation

Business value: Est. >\$3M productivity unlocked

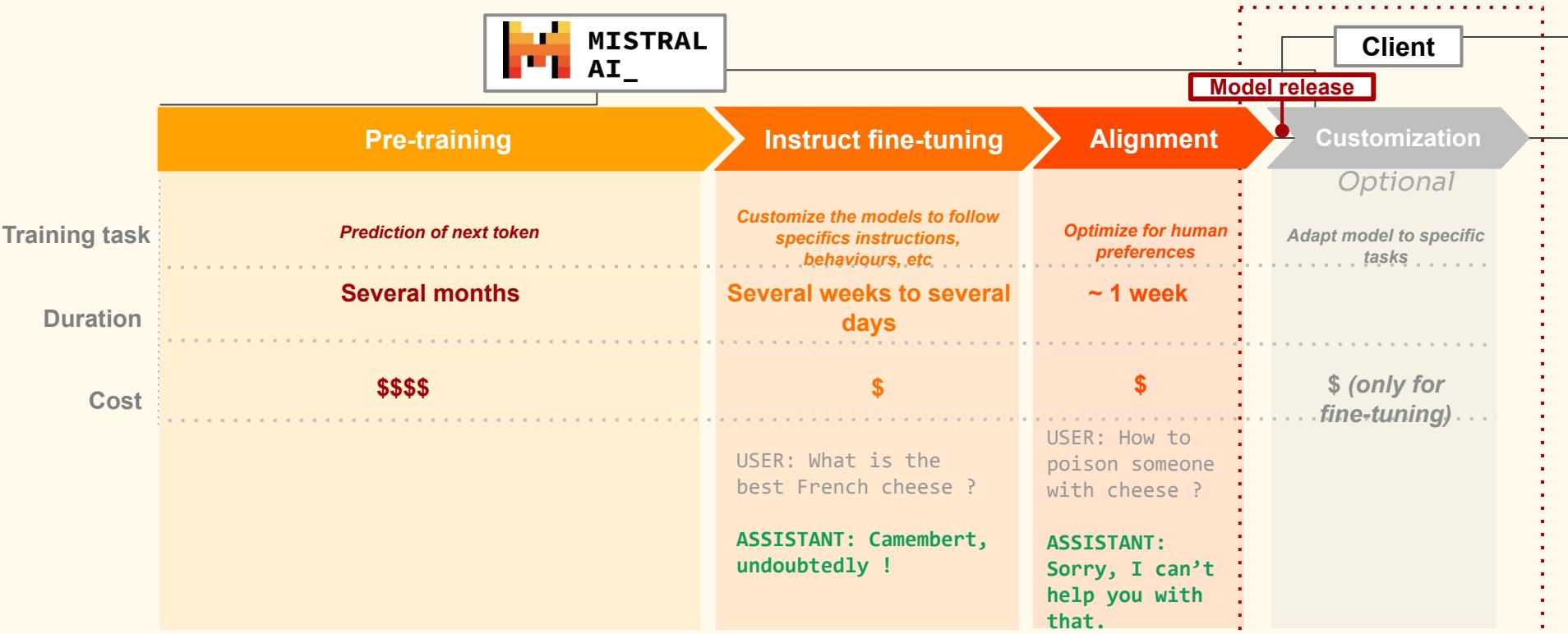


Mistral AI

Fine-Tuning

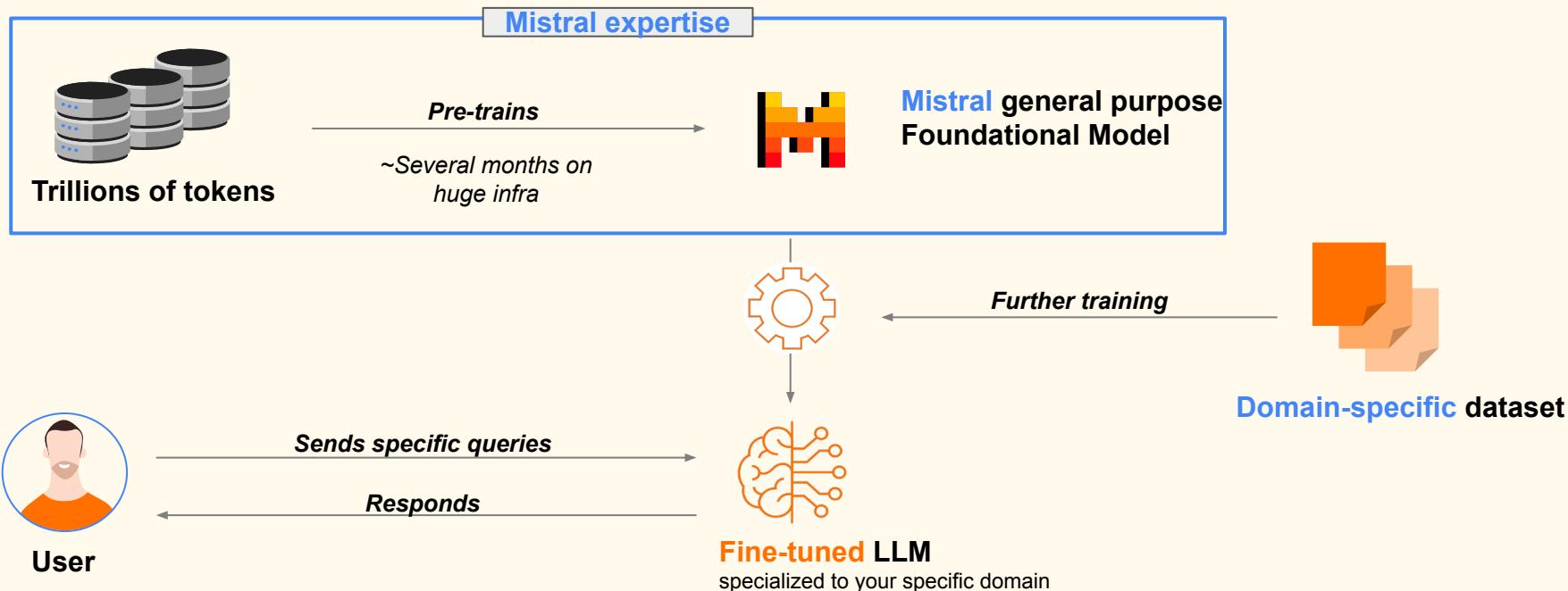


It takes several months to build a large language model

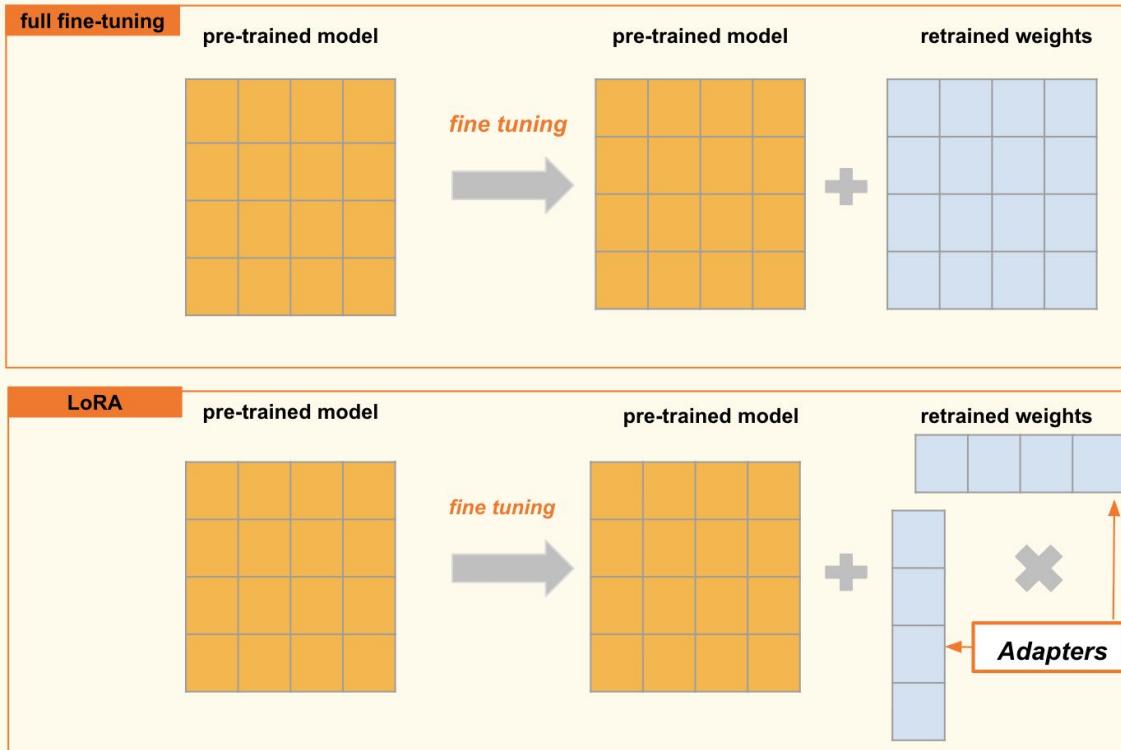


Fine-tuning is a process of specialization that involves adapting a pre-trained LLM with a domain-specific dataset

How it works



LoRA is a cost-effective method to fine-tune LLMs



LoRA retrains about 1% of the total model weights.

No additional infrastructure cost for fine-tuning.

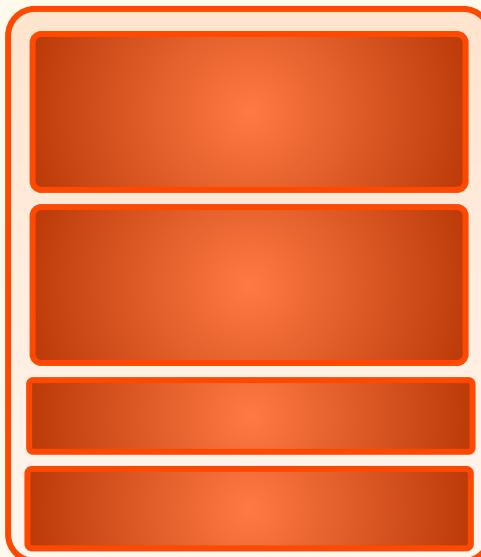
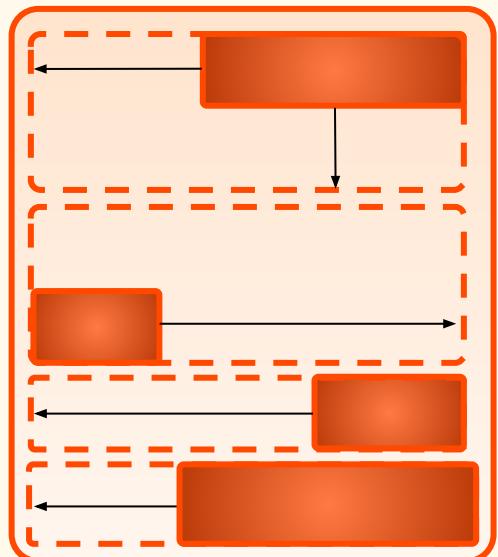
Fast and cost-effective

When fine-tune a generalist LLM ?

1. Overperform on a specific task

Mistral AI helps you customize an LLM on the domains that matters ...

... And reach frontier performance on your use case



Famous example : Codestral

We customized Mistral Small (22B parameters) on coding tasks with our **proprietary dataset**, resulting in turbocharged performance on coding

- **+23 pts** on HumanEval benchmark
- **+19 pts** on MBPP

HumanEval : solve coding problems

MBPP : Mostly Basic Python Programming

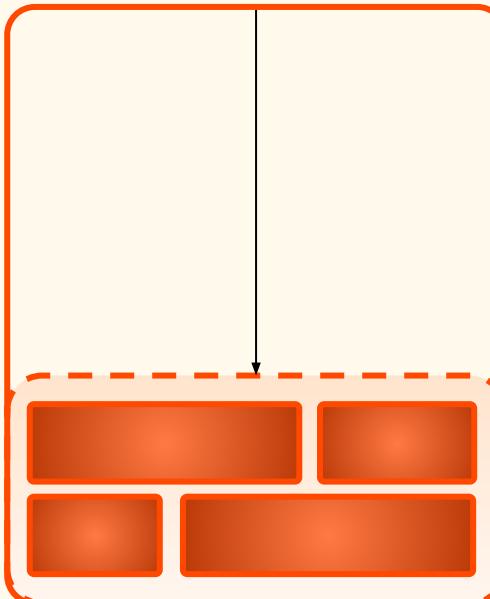
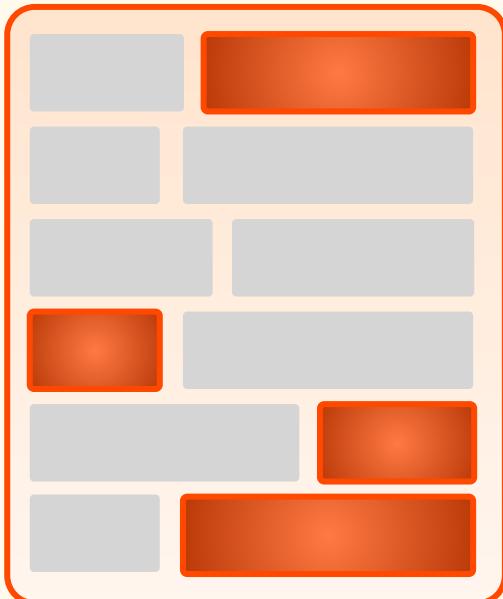
Domain specialization

When fine-tune a generalist LLM ?

2. *Reach the same result with a smaller model*

Mistral AI helps you keep the domains that matter ...

... and build a smaller LLM
turbo-charged for your use case



Leverage your **proprietary data** to achieve the same accuracy with a **model 2-3x times smaller**

- **Lower GPU footprint**
- **Better latency**
- **Lower cost per token**

Overcoming potential fine tuning limits

Edge cases

The model has not been sufficiently trained on **rare scenarios** that were not observed during training, which can lead to poor performance or errors in these specific situations.

Solution approaches:

- Increase the **diversity** of training data by including **representative examples** of all possible cases
- Use **synthetic data generation techniques** to create examples of edge cases if real data is insufficient

Overfitting

The model **fits too precisely** to the **training data**, leading to a decrease in performance on other tasks that may occasionally be required of it.

Solution approaches:

- Increase the **size** of the **training data**
- Use **cross-validation techniques** to evaluate the model's performance on unseen data.

Data drift

The distribution of the data used for **training** differs significantly from the **distribution of the data** the model encounters in **production**, which can degrade the model performance.

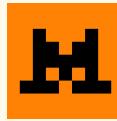
Solution approaches:

- Implement **continuous monitoring systems** (LLMops¹) for data to detect changes in data distribution in real-time.
- Periodically retrain the model** with new data to adapt to changes in data distribution.

1. e.g. : Braintrust, Langfuse, Langsmith

Two ways to fine tune **Mistral Models**

Via La Plateforme



Head to our Fine Tuning API or directly within la Plateforme, upload and validate your training data. Run the job, and when completed, you can access your custom model via your own model API endpoint.



Via the Codebase

Mistral-finetune is a light-weight codebase that enables memory-efficient and performant finetuning of Mistral's open-source models.

Dataset format (Text Fine Tuning)

Data must be stored in JSON Lines (.jsonl) files, which allow storing multiple JSON objects, each on a new line. Multiple structures are allowed :

Default Instruct :

Conversational data between user and assistant, which can be one-turn or multi-turn.

```
{"messages": [{"role": "user", "content": "..."}, {"role": "assistant", "content": "..."}, ...]}  
{"messages": [{"role": "user", "content": "..."}, {"role": "assistant", "content": "..."}, ...]}  
{"messages": [{"role": "user", "content": "..."}, {"role": "assistant", "content": "..."}, ...]}  
{"messages": [{"role": "user", "content": "..."}, {"role": "assistant", "content": "..."}, ...]}  
...  
...
```

Dataset format (Text Fine Tuning)

Data must be stored in JSON Lines (.jsonl) files, which allow storing multiple JSON objects, each on a new line. Multiple structures are allowed :

Function-calling Instruct :

Conversational data with tool usage

```
{
  "messages": [
    {
      "role": "system",
      "content": "You are a helpful assistant with access to the following functions to help the user. You can use the functions if needed."
    },
    {
      "role": "user",
      "content": "Can you help me generate an anagram of the word 'listen'?"
    },
    {
      "role": "assistant",
      "tool_calls": [
        {
          "id": "TX92Jm8Zi",
          "type": "function",
          "function": {
            "name": "generate_anagram",
            "arguments": "{\"word\": \"listen\"}"
          }
        }
      ]
    },
    {
      "role": "tool",
      "content": "{\"anagram\": \"silent\"}",
      "tool_call_id": "TX92Jm8Zi"
    },
    {
      "role": "assistant",
      "content": "The anagram of the word 'listen' is 'silent'."
    }
  ]
}
```

Dataset format (Classifier Fine Tuning)

Data must be stored in JSON Lines (.jsonl) files, which allow storing multiple JSON objects, each on a new line. Multiple structures are allowed :

Single Target : single label classification (eg. 'sentiment')

```
{  
    "text": "I love this product!",  
    "labels": {  
        "sentiment": ["positive", "neutral"]  
    }  
}
```

Multi-Target : if you want to classify different aspects of the same content independently (eg. 'sentiment' + 'is_english')

```
{  
    "text": "I love this product!",  
    "labels": {  
        "sentiment": "positive", // positive/neutral/negative  
        "is-english": "yes" // yes/no, boolean  
    }  
}
```

Fine tune Mistral Models from La Plateforme UI

1. Choose model type (completion or classification)

Aperçu

Jérémie Demos
Default Workspace
Jérémie Adjadj

Plateforme

- Agents
- Modèles fine-tunés
- Jeux de données
- Batches
- Playground

API

- Évaluations

Espace de travail

- Général
- Cles API
- Utilisation
- Membres

Codestral

Codestral

Affiner le modèle

Choisissez votre jeu de données, vos paramètres, votre validation et lancez facilement un entraînement sur vos données.

Type de modèle

Completion

Jeux de données d'entraînement

+ Ajouter un fichier

Lignes invalides autorisées (pourcentage)

0

Modèle de base

Mistral 3B

- ✓ Mistral 3B
- Mistral 7B
- Mistral Nemo
- Mistral Small
- Mistral Medium
- Mistral Large 2
- Pixtral
- Codestral

Mistral 3B

Mistral 3B est un modèle 3B à la pointe de la technologie, qui peut être utilisé ou ajusté pour diverses applications, de l'orchestration de workflows d'agents à la création de travailleurs spécialisés

Entraînement : 0,50 € / 1M tokens

Taux d'apprentissage

0,00001

Epochs

3

2. Choose the base model to finetune

3. Upload your training data (must be a .jsonl file and respect a specific format)

You can also tune the learning rate, the number of epochs, and add a maximum invalid rows threshold

Fine tune Mistral Models from La Plateforme UI

Once the model is fine-tuned, different pieces of information are available:

- a model card,
- a summary of the training job,
- as well as the names of the training and validation files.

The screenshot shows a detailed view of a fine-tuned model card for 'ft:mistral-small-latest'. The card includes sections for the model card, deployment details, training job summary, hyperparameters, and training/validation files.

Modèle fine-tuné
ft:mistral-small-latest:d3417809:20250613:dd0a987c
Pas de description.

Carte modèle

- Modèle de base
- Appel de fonction
- Remplissage au milieu
- Date de création

Déploiement

- API
- ft:m... (highlighted)
- Créer un agent
- Archiver

mistral-small-latest

No
No

13/06/2025 18:02:21

Job d'entraînement

	\$ Coût	13/06/2025 14:10:23
Tokens d'entraînement	60 945 580 tokens	
Date de création		

Voir plus de détails sur le job d'entraînement.

Paramètres d'entraînement

	Epochs	10
Taux d'apprentissage		0.0001

Fichiers d'entraînement

cba5f374-b59b-40b4-acea-65cdc183a6ea rater_finetuning_dataset.jsonl	21.2 MB(1540 lignes)
--	----------------------

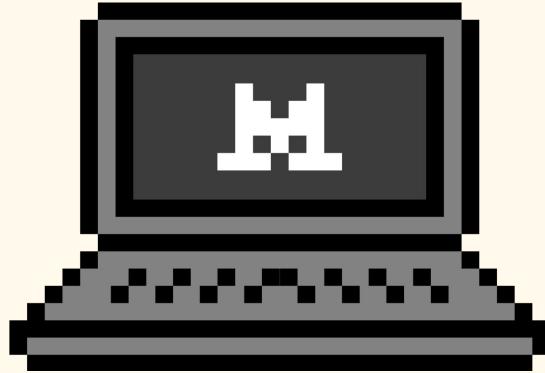
Fichiers de validation

fedbc025-a779-461d-b9c5-413a30bc9b35 rater_finetuning_dataset_valid.jsonl	1.5 MB(96 lignes)
--	-------------------

You can either deploy your fine tuned model as a La Plateforme Agent

Or directly copy its id and use it via API calls

Fine tune **Mistral Models** from the Fine-Tuning API



See the `text_fine_tuning.ipynb` and
`product_classification.ipynb` notebooks



No matter which fine-tuning job creation method you choose, the fine-tuned model will be accessible via API requests or when creating agents in La Plateforme !

Document AI / OCR



Extracting data from documents is hard

Multiple formats of documents

Complex layouts

Multiple complex domains



PDF document



Office document



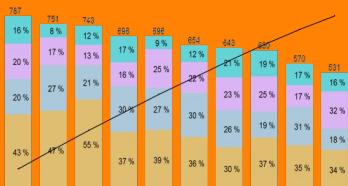
Scanned document

Extracting data from documents is hard

Complex layouts

(In millions, except percentages)	2022	2021
Revenue		
Productivity and Business Processes	\$ 63,364	\$ 53,915
Intelligent Cloud	75,251	60,080
More Personal Computing	59,655	54,093
Total	\$ 198,270	\$ 168,088
Operating Income		
Productivity and Business Processes	\$ 29,687	\$ 24,351
Intelligent Cloud	32,721	26,126
More Personal Computing	20,975	19,439
Total	\$ 83,383	\$ 69,916

Tables



Charts



Columns, Images, ...

Signatures

$$\begin{aligned} \sin^2 B + \cos^2 B &= 1 \\ a^2 + b^2 + c^2 &= 0 \\ a\left(a^2 + \frac{b^2}{a^2} + \frac{c^2}{a^2}\right) &< 0 \\ a^2 + 2 \cdot \frac{b^2}{2a} + \left(\frac{c^2}{a^2}\right)^2 - \left(\frac{b^2}{2a}\right)^2 + \frac{c^2}{a^2} &= 0 \\ \left(1 + \frac{b^2}{2a^2}\right)^2 - \frac{b^4 - 4a^2c^2}{4a^4} &= 0 \\ \left(1 + \frac{b}{2a}\right)^2 - \frac{(b^2 - 4a^2c^2)}{4a^2} &= 0 \end{aligned}$$

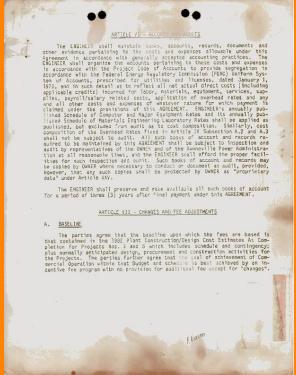
Handwriting



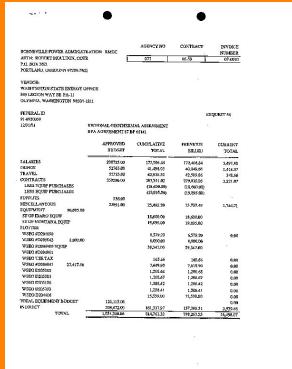
Poor scan quality

Extracting data from documents is hard

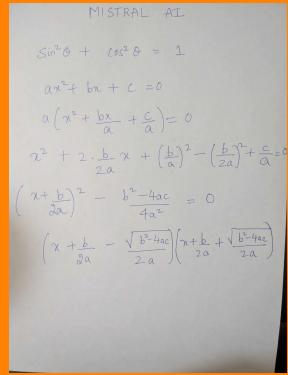
Multiple complex domains



Contracts



Invoices



Notes

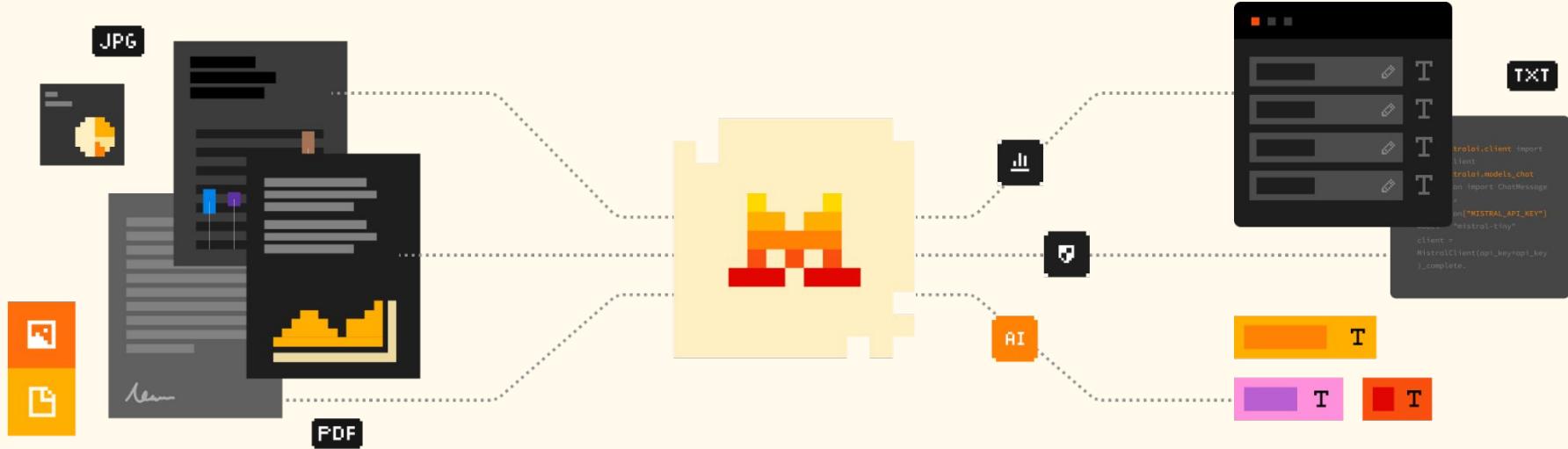
Huge volumes to process

Real-world examples from our customers

- Banking: 1.2B documents/year
 - Government: 200M documents/year

Real-time processing

Introducing Document AI



World's **most accurate OCR**

Extract text, images, ... from documents

Versatile annotations

Understand your OCR outputs

Introducing Document AI

Mistral OCR will extract all the information in the document with the highest level of accuracy

The raw output data is often messy

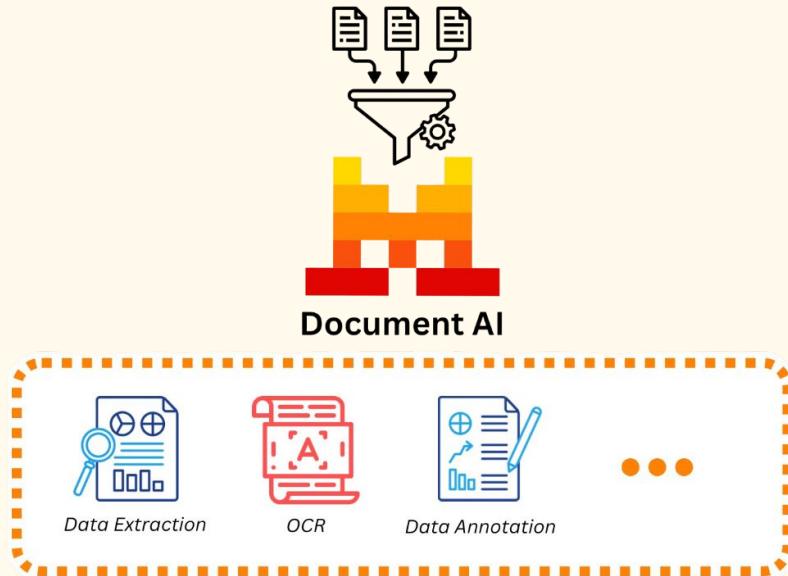


Often, only 1% of the information on an OCR is actually useful. Our **Annotations** help you capture what really matters

Introducing Document AI

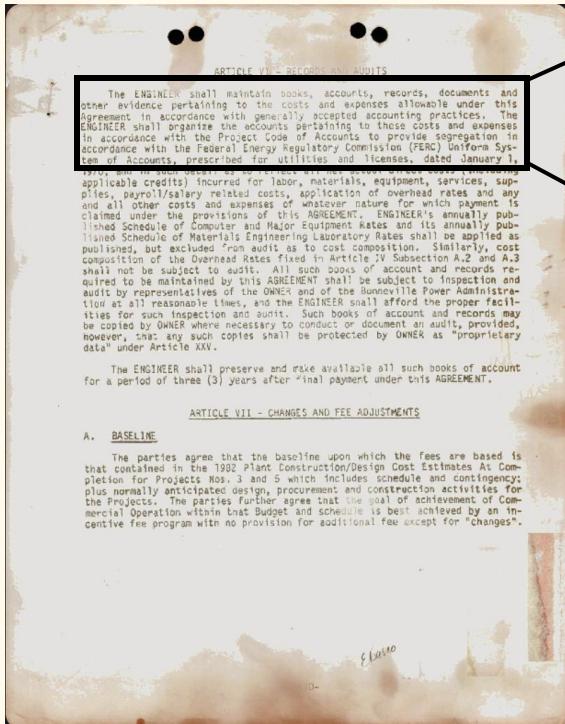
Mistral OCR will extract all the information in the document with the highest level of accuracy

The raw output data is often messy



Often, only 1% of the information on an OCR is actually useful. Our **Annotations** help you capture what really matters

What is Optical Character Recognition (OCR) ?



ARTICLE VI - RECORDS AND AUDITS

The ENGINEER shall maintain books, accounts, records, documents and other evidence pertaining to the costs and expenses allowable under this Agreement in accordance with generally accepted accounting practices. The ENGINEER shall organize the accounts pertaining to these costs and expenses in accordance with the Project Code of Accounts to provide segregation in accordance with the Federal Energy Regulatory Commission (FERC) Uniform System of Accounts, prescribed for utilities and licenses, dated January 1, 1970.

The ENGINEER shall maintain books, accounts, records, documents and other evidence pertaining to the costs and expenses allowable under this Agreement in accordance with generally accepted accounting practices. The ENGINEER shall organize the accounts pertaining to these costs and expenses in accordance with the Project Code of Accounts to provide segregation in accordance with the Federal Energy Regulatory Commission (FERC) Uniform System of Accounts, prescribed for utilities and licenses, dated January 1, 1970.

The ENGINEER shall maintain books, accounts, records, documents and other evidence pertaining to the costs and expenses allowable under this Agreement in accordance with generally accepted accounting practices. The ENGINEER shall organize the accounts pertaining to these costs and expenses in accordance with the Project Code of Accounts to provide segregation in accordance with the Federal Energy Regulatory Commission (FERC) Uniform System of Accounts, prescribed for utilities and licenses, dated January 1, 1970.

The ENGINEER shall maintain books, accounts, records, documents and other evidence pertaining to the costs and expenses allowable under this Agreement in accordance with generally accepted accounting practices. The ENGINEER shall organize the accounts pertaining to these costs and expenses in accordance with the Project Code of Accounts to provide segregation in accordance with the Federal Energy Regulatory Commission (FERC) Uniform System of Accounts, prescribed for utilities and licenses, dated January 1, 1970.

The ENGINEER shall preserve and make available all such books of account for a period of three (3) years after final payment under this AGREEMENT.

ARTICLE VII - CHANGES AND FEE ADJUSTMENTS

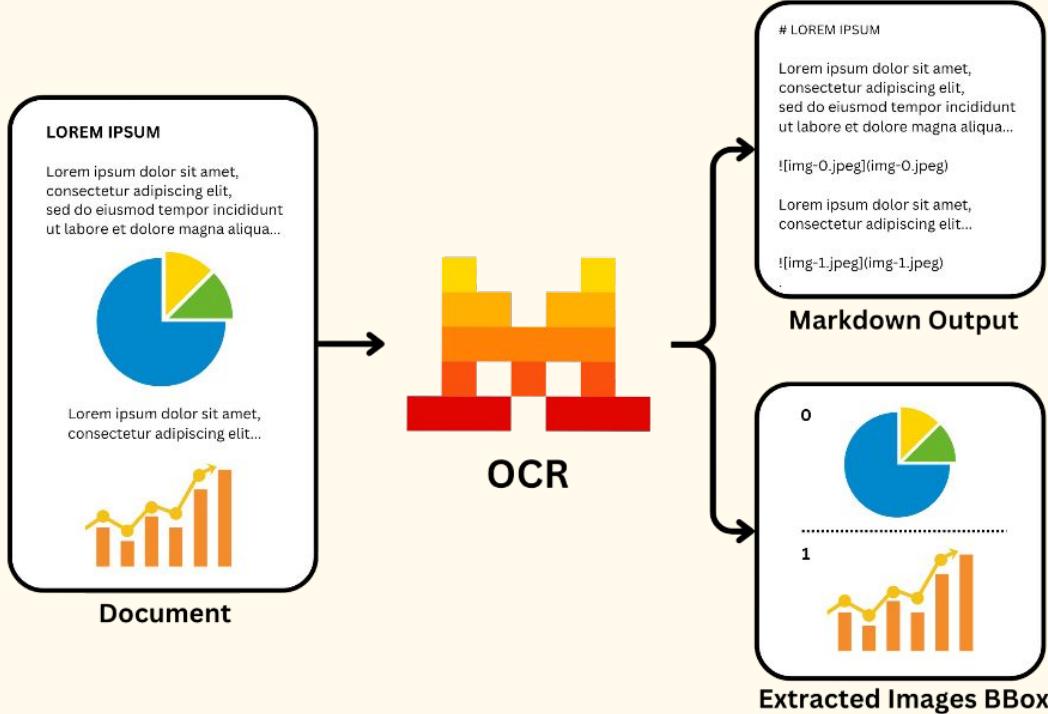
A. BASELINE

The parties agree that the baseline upon which the fees are based is that contained in the 1989 Plant Construction/Design Cost Estimates At Completed Projects No. 3 and which includes schedule and contingencies plus normally anticipated design, procurement and construction activities for the Projects. The parties further agree that the goal of achievement of Commercial Operation within that Budget and schedule is best achieved by an incentive fee program with no provision for additional fee except for "changes".

ARTICLE VI - RECORDS AND AUDITS

The ENGINEER shall maintain books, accounts, records, documents and other evidence pertaining to the costs and expenses allowable under this Agreement in accordance with generally accepted accounting practices. The ENGINEER shall organize the accounts pertaining to these costs and expenses in accordance with the Project Code of Accounts to provide segregation in accordance with the Federal Energy Regulatory Commission (FERC) Uniform System of Accounts, prescribed for utilities and licenses, dated January 1, 1970.

What is Optical Character Recognition (OCR) ?



Key features:

- Extracts text content while maintaining document structure and hierarchy
- Preserves formatting like headers, paragraphs, lists and tables
- Returns results in markdown format for easy parsing and rendering
- Handles complex layouts including multi-column text and mixed content
- Supports multiple document formats including:
 - `image_url`: png, jpeg/jpg, avif and more...
 - `document_url`: pdf, pptx, docx and more...

You need intelligence to make sense of raw OCR output

Raw OCR output

BONNEVILLE POWER ADMINISTRATION RMGC
 ATTN: ROBERT MOULTON, COTR
 P.O. BOX 3621
 PORTLAND, OREGON 97204-3621

VENDOR:
 WASHINGTON STATE ENERGY OFFICE
 809 LEGION WAY SE FA-11
 OLYMPIA, WASHINGTON 98504-1211

FEDERAL ID:
 91-0950059
 12/31/91

AGENCY NO	CONTRACT	INVOICE NUMBER
077	K6-59	07-0010

REGIONAL GEOTHERMAL ASSESSMENT
 BPA AGREEMENT 87 BP 66141

REQUEST #4

	APPROVED BUDGET	CUMULATIVE TOTAL	PREVIOUS BILLING	CURRENT TOTAL
SALARIES	230725.00	177,994.64	172,406.84	5,497.80
FRINGE	52763.00	41,496.03	40,049.66	1,446.37
TRAVEL	5323.00	42,938.52	40,599.65	2,438.66
CONTRACTS	359286.00	283,351.02	279,939.95	3,521.07
LESS EQUIP PURCHASES		(18,600.00)	(18,600.00)	
LESS EQUIP PURCHASES		(13,895.00)	(13,895.00)	
SUPPLIES	536.00			
MISC/ANEOUS	23891.00	23,482.20	23,737.49	1,744.71
EQUIPMENT	90,695.00			
ST OF IDAHO EQUIP	15,600.00	18,600.00		
ST OF MONTANA EQUIP	19,895.00	19,895.00		
MOULTON				
WSEO #D2041890		6,579.20	6,579.20	0.00
WSEO #D2057452	8,000.00		8,000.00	0.00
WSEO #D2054949 EQUIP		39,247.00	39,347.00	
WSEO #D2049090				
WSEO USE TAX	165.66	165.66	0.00	
WSEO D2054949	27,417.06		7,619.90	0.00
WSEO D205382				
WSEO D205383				
WSEO D205384				
WSEO D205385				
WSEO D205386				
WSEO D205387				
WSEO D205388				
WSEO D205389				
WSEO D205390				
TOTAL EQUIPMENT BUDGET	126,112.06		0.00	
INDIRECT	208,672.00	161,337.97	157,398.51	3,939.46
TOTAL	1,051,308.06	844,761.32	798,263.25	16,498.07

BONNEVILLE POWER ADMINISTRATION RMGC | AGENCY NO | CONTRACT | INVOICE NUMBER |
 --- | --- | --- | --- | --- |
 ATTN: ROBERT MOULTON, COTR | 077 | 86-59 | 07-0010 |
 P.O. BOX 3621 | | | |
 PORTLAND, OREGON 97208-3621 | | | |

VENDOR: WASHINGTON STATE ENERGY OFFICE 809 LEGION WAY SE FA-11 OLYMPIA, WASHINGTON 98504-1211

FEDERAL ID 91-0950059 12/31/91 REGIONAL GEOTHERMAL ASSESSMENT BPA AGREEMENT 87 BP 66141

	APPROVED BUDGET	CUMULATIVE TOTAL	PREVIOUS BILLED	CURRENT TOTAL
SALARIES	230725.00	177,994.64	172,406.84	5,497.80
FRINGE	52763.00	41,496.03	40,049.66	1,446.37
TRAVEL	5323.00	42,938.52	40,599.65	2,438.66
CONTRACTS	359286.00	283,351.02	279,939.95	3,521.07
LESS EQUIP PURCHASES		(18,600.00)	(18,600.00)	
LESS EQUIP PURCHASES		(13,895.00)	(13,895.00)	
SUPPLIES	536.00			
MISC/ANEOUS	23891.00	23,482.20	23,737.49	1,744.71
EQUIPMENT	90,695.00			
ST OF IDAHO EQUIP	15,600.00	18,600.00	18,600.00	
ST OF MONTANA EQUIP	19,895.00	19,895.00	19,895.00	
MOULTON				
WSEO #D2041890		6,579.20	6,579.20	0.00
WSEO #D2057452	8,000.00		8,000.00	0.00
WSEO #D2054949 EQUIP		39,247.00	39,347.00	
WSEO #D2049090				
WSEO USE TAX	165.66	165.66	0.00	
WSEO #D204890				
WSEO #D2054942	8,000.00		8,000.00	8,000.00
WSEO #D204960 EQUIP			39,347.00	39,347.00
WSEO #D204901				
WSEO USE TAX	165.66	165.66	0.00	
WSEO #D2048941	27,417.06		7,619.90	7,619.90
WSEO D205382				
WSEO D205383				
WSEO D205384				
WSEO D205385				
WSEO D205386				
WSEO D205387				
WSEO D205388				
WSEO D205389				
WSEO D205390				
TOTAL EQUIPMENT BUDGET	126,112.06		0.00	
INDIRECT	208,672.00	161,337.97	157,398.51	3,939.46
TOTAL	1,051,308.06	844,761.32	798,263.25	16,498.07

You often need 1% of the information on a page

AGENCY NO		CONTRACT	INVOICE NUMBER
BONNEVILLE POWER ADMINISTRATION RMGC ATTN: ROBERT MOULTON, COTR P.O. BOX 3621 PORTLAND, OREGON 97204-3621	077	K6-59	07-0010
Contract number			
VENDOR: WASHINGTON STATE ENERGY OFFICE 301 LEGION WAY SE EA-11 OLYMPIA, WASHINGTON 98501-1211			
Vendor's name			
FEDERAL ID 91495009 120101		REQUEST #4 REGIONAL GEOTHERMAL ASSESSMENT BPA AGREEMENT #7 BP 6341	
What you buy			
	APPROVED BUDGET	CUMULATIVE TOTAL	PREVIOUS BILLED
			CURRENT TOTAL
SALARIES	230721.00	177,994.64	172,498.64
RUNGE	5262.00	41,494.03	40,049.65
TRAVEL	5323.00	42,538.52	40,559.65
CONTRACTS	359286.00	281,351.02	279,939.93
LESS EQUIP PURCHASES		(18,600.00)	3,521.07
LESS EQUIP PURCHASES		(13,695.00)	(13,895.00)
SUPPLIES	536.00		
MECHANICALS	23891.00	23,482.20	23,737.49
EQUIPMENT	90,695.00		1,744.71
ST OF IDAHO EQUIP		15,630.00	18,000.00
ST OF MONTANA EQUIP		15,630.00	19,000.00
PILOT FUEL			
WSIB #1201890		6,579.20	6,579.20
WSIB #1201642	8,000.00	8,000.00	8,000.00
WSIB #1201449 EQUIP		39,247.00	39,347.00
WSIB #1201490			
WSIB SHT TAX		165.66	165.66
WSIB #1201440	27,417.00	7,649.90	7,619.90
WSIB #1201392		1,208.68	1,208.68
WSIB #1201383		1,208.68	1,208.68
WSIB #1201166		1,286.41	1,286.41
WSIB #1201393		1,286.41	1,286.41
WSIB #1201405		15,539.00	15,539.00
TOTAL EQUIPMENT BUDGET	116,112.06		0.00
INURST	208,672.00	161,337.97	157,598.51
TOTAL	1,051,308.06	844,761.32	795,263.25
			16,498.07

Total amount

DocumentAI Annotation extracts the information you need

AGENCY NO		CONTRACT	INVOICE NUMBER
077		86-59	07-00010
BONNEVILLE POWER ADMINISTRATION RMGC ATTN: ROBERT MOULTON, COTR P.O. BOX 3621 PORTLAND, OREGON 97204-3621			
VENOR: WASHINGTON STATE ENERGY OFFICE 301 LEGION WAY SE EA-11 OLYMPIA, WASHINGTON 98501-1211			
FEDERAL ID: 91495009 120101			
REQUEST #4			
REGIONAL GEOTHERMAL ASSESSMENT BPA AGREEMENT #7 BP 6341			
APPROVED BUDGET	CUMULATIVE TOTAL	PREVIOUS BILLED	CURRENT TOTAL
SALARIES	230725.00	177,994.64	172,495.64
FRINGE	5262.00	41,494.03	40,049.65
TRAVEL	5323.00	42,538.52	40,559.65
CONTRACTS	359286.00	28,351.02	27,939.93
LESS EQUIP PURCHASES		(18,600.00)	3,521.07
LESS EQUIP PURCHASES		(13,695.00)	(13,895.00)
SUPPLIES	536.00		
MECHANICALS	23891.00	23,482.20	23,737.49
EQUIPMENT	90,695.00		1,744.71
ST OF IDAHO EQUIP	15,630.00	18,000.00	
ST OF MONTANA EQUIP	15,630.00	19,000.00	
PLANTER			
WSB# 412041890	6,579.20	6,579.20	0.00
WSB# 412057452	8,000.00		
WSB# 412044907 RQD/P	39,247.00	39,347.00	
WSB# 412049051			
WSB# 412041840	165.66	165.66	0.00
WSB# 412057450	7,649.90	7,619.90	0.00
WSB# 412053932	1,208.68	1,208.68	0.00
WSB# 412053833	1,208.68	1,208.68	0.00
WSB# 412051166	1,286.42	1,286.42	0.00
WSB# 412053933	1,286.41	1,286.41	0.00
WSB# 412054056	15,539.00	15,539.00	0.00
TOTAL EQUIPMENT BUDGET	116,112.06		0.00
INURCT	208,672.00	161,337.97	157,598.51
TOTAL	1,051,308.06	844,761.32	795,263.25
			16,498.07

Contract number

Invoice number

Vendor's name

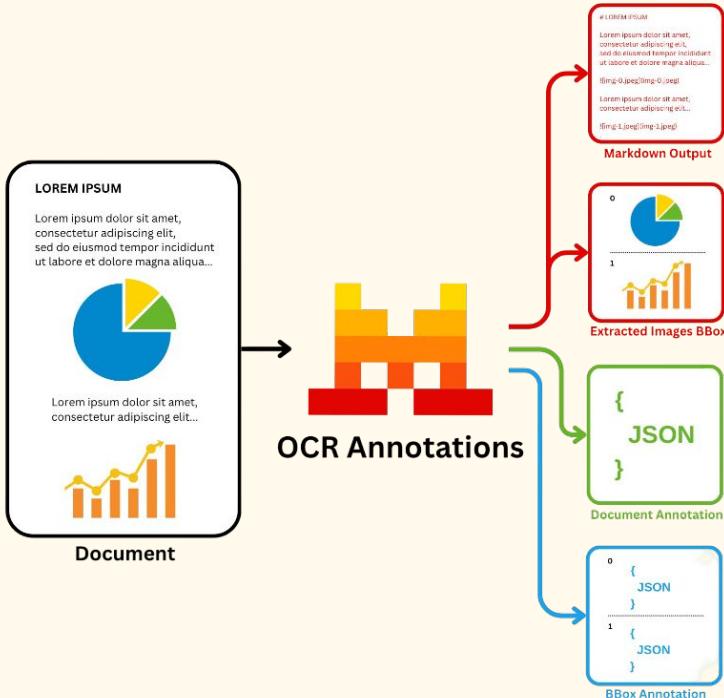
What you buy

Total amount

Desired output

```
{  
  "contract_number": "86-59",  
  "invoice_number": "07-00010",  
  "vendor": "Washington State Energy  
Office",  
  "description": "Regional Geothermal  
Assessment",  
  "agency_no": "077"  
}
```

DocumentAI Annotation extracts the information you need

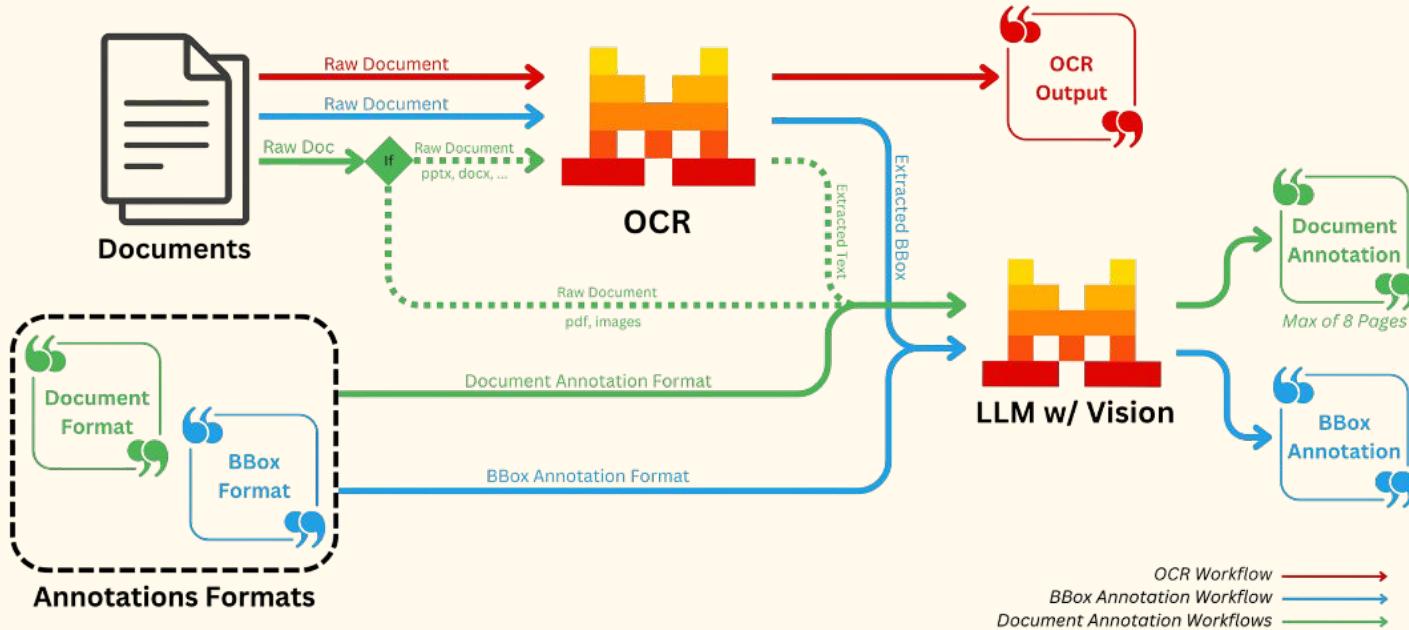


Document AI API offers two types of annotations:

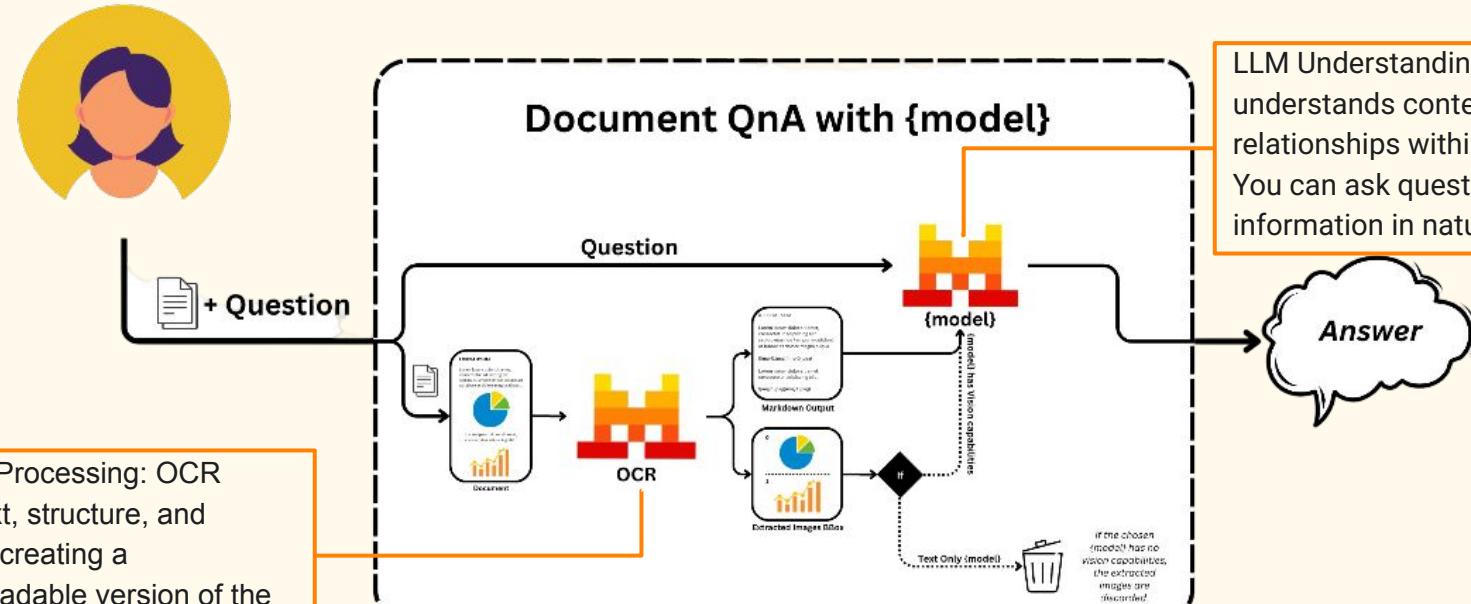
- **bbox_annotation**: gives you the annotation of the bboxes extracted by the OCR model (charts/ figures etc) based on user requirement and provided bbox/image annotation format. The user may ask to describe/caption the figure for instance.
- **document_annotation**: returns the annotation of the entire document based on the provided document annotation format.

DocumentAI Annotation : How it works

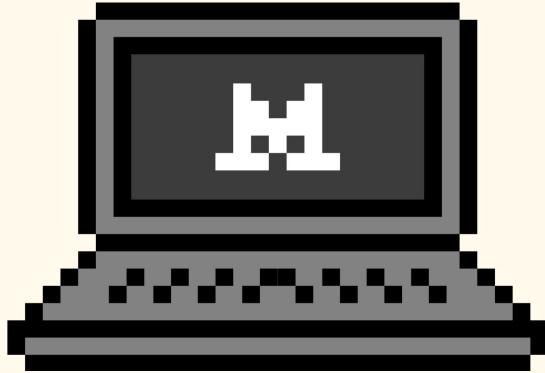
Annotations Workflow



DocumentAI QnA : Combine OCR with LLM capabilities to enable natural language interaction with document



Let's see **DocumentAI** in practice !



**See the Document_QnA.ipynb and
Document_Annotation.ipynb notebooks !**

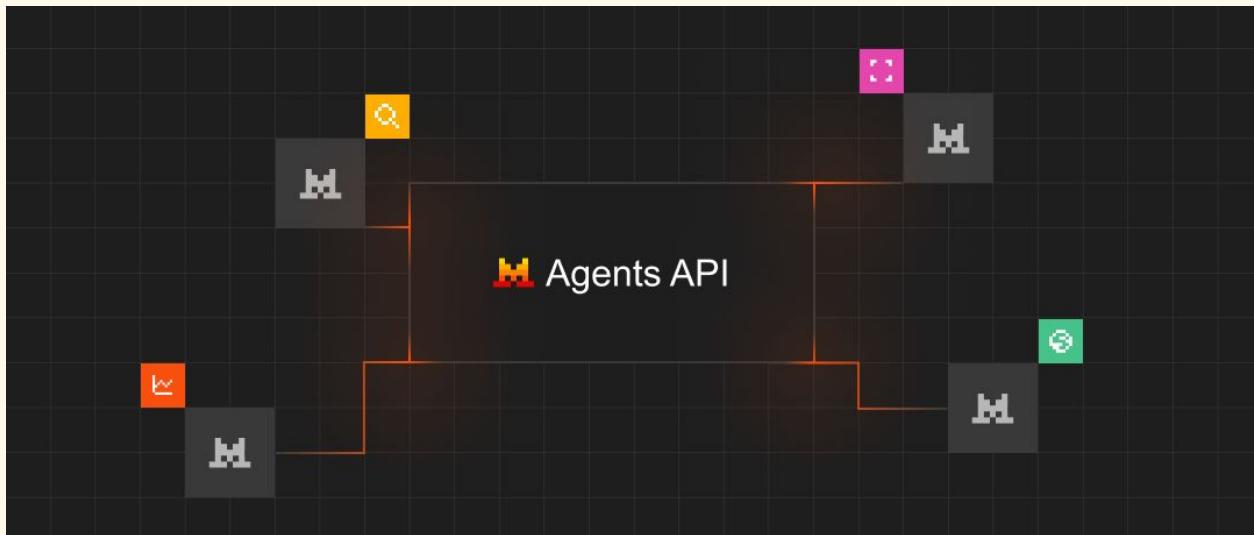


Agent API



Introducing Agents API : making AI more capable, useful, and an active problem-solver

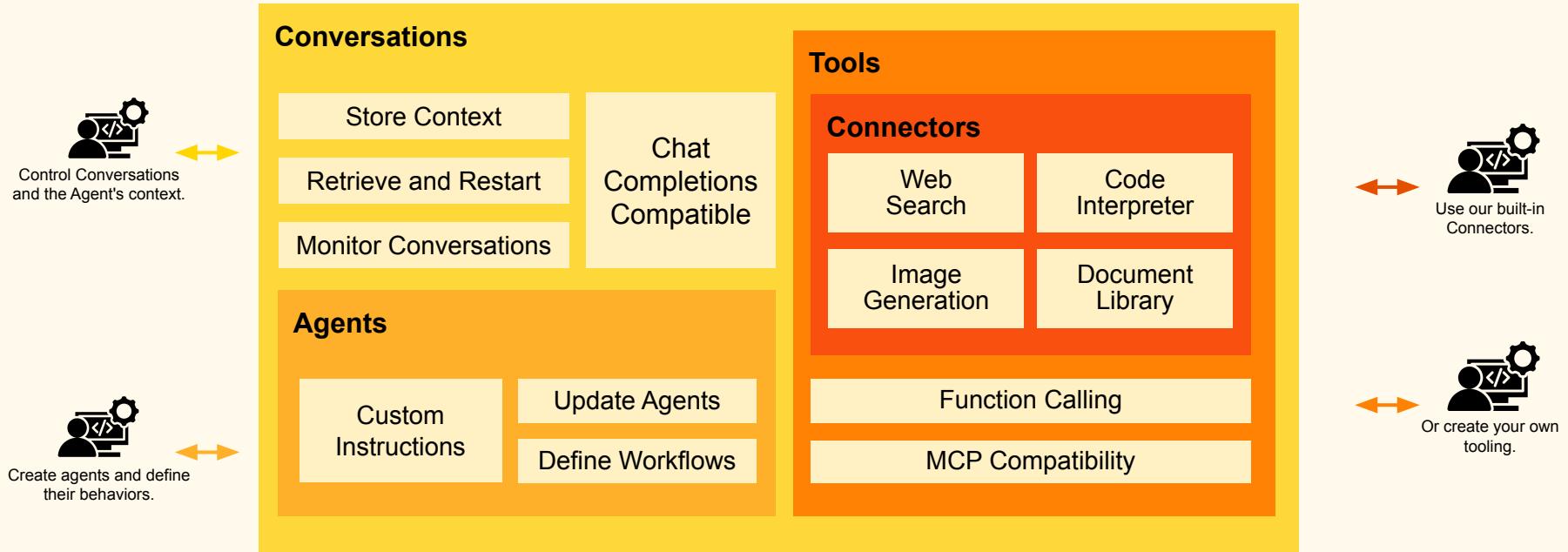
Traditional language models excel at generating text but are limited in their ability to perform actions or maintain context...



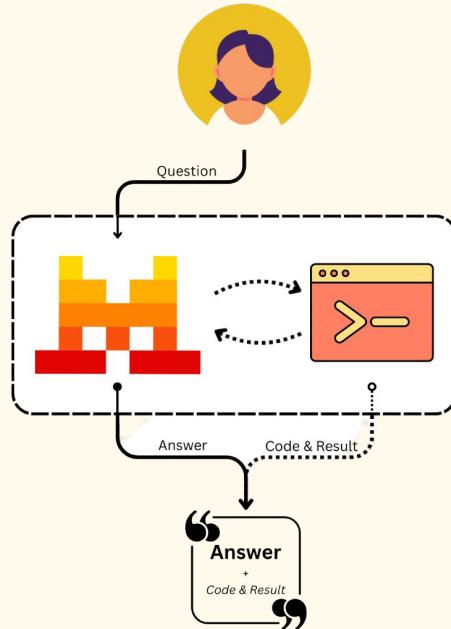
Our new Agents API addresses these limitations by combining Mistral's powerful language models with:

- Built-in connectors for code execution, web search, image generation, and MCP tools
- Persistent memory across conversations
- Agentic orchestration capabilities

The Mistral AI Agents ecosystem



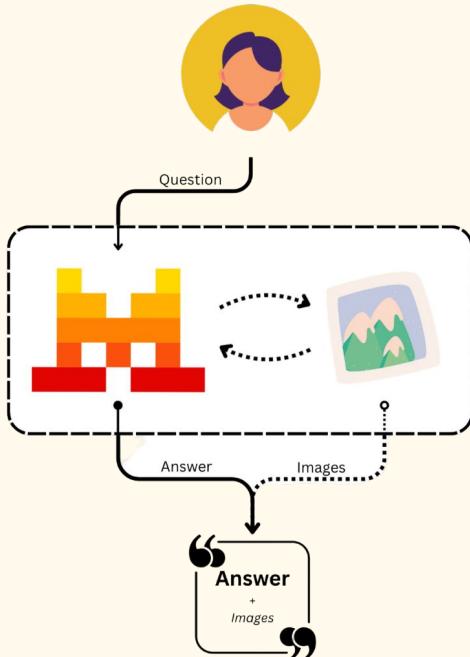
Create an agent with built-in connectors and MCP tools - Code Interpreter



The Agents API can use the code execution connector, empowering developers to create agents that execute Python code in a secure sandboxed environment.

This enables agents to tackle a wide range of tasks, including mathematical calculations and analysis, data visualization and plotting, and scientific computing.

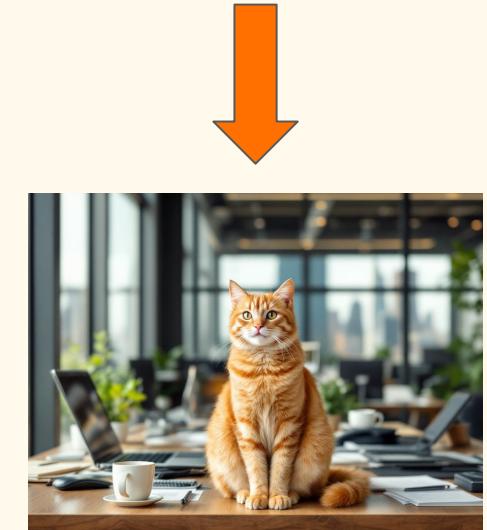
Create an agent with built-in connectors and MCP tools - Image Generation



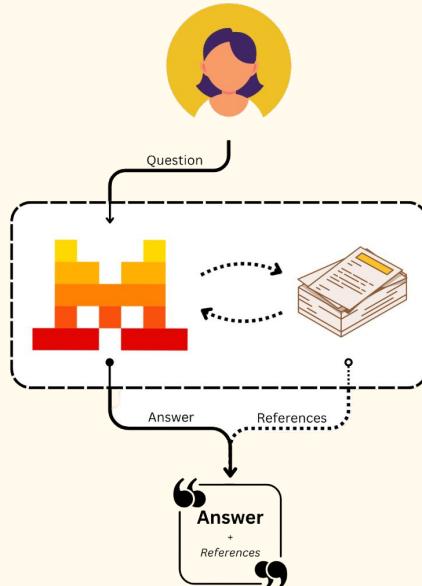
The image generation connector tool, powered by Black Forest Lab FLUX1.1 [pro] Ultra, enables agents to create images for diverse applications.

This feature can be leveraged for various use cases such as generating visual aids for educational content, creating custom graphics for marketing materials, or even producing artistic images.

"Generate an orange cat in an office."



Create an agent with built-in connectors and MCP tools - Document Library (as of June 25)

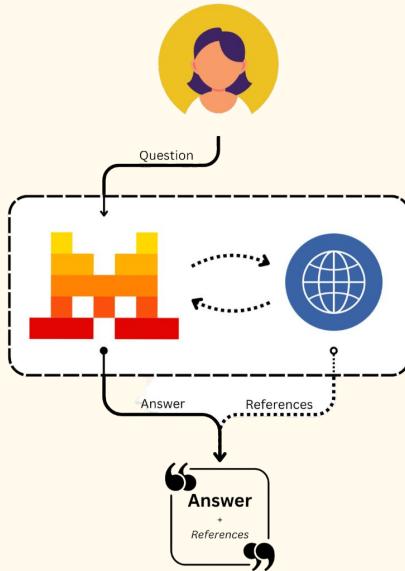


Document Library is a built-in connector tool that enables agents to access documents from Mistral Cloud.

It powers the integrated RAG-in-a-box functionality, strengthening agents' knowledge by leveraging the content of user-uploaded documents.

Currently, you can only create libraries via Le Chat. To enable the Agent to access the library, you have to be an Org admin and share it with the Organization.

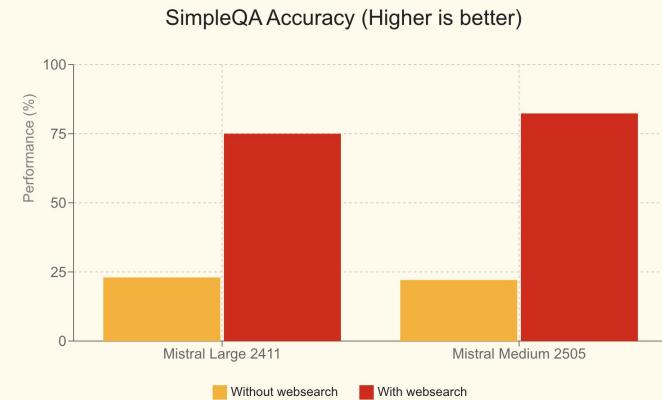
Create an agent with built-in connectors and MCP tools - Web Search



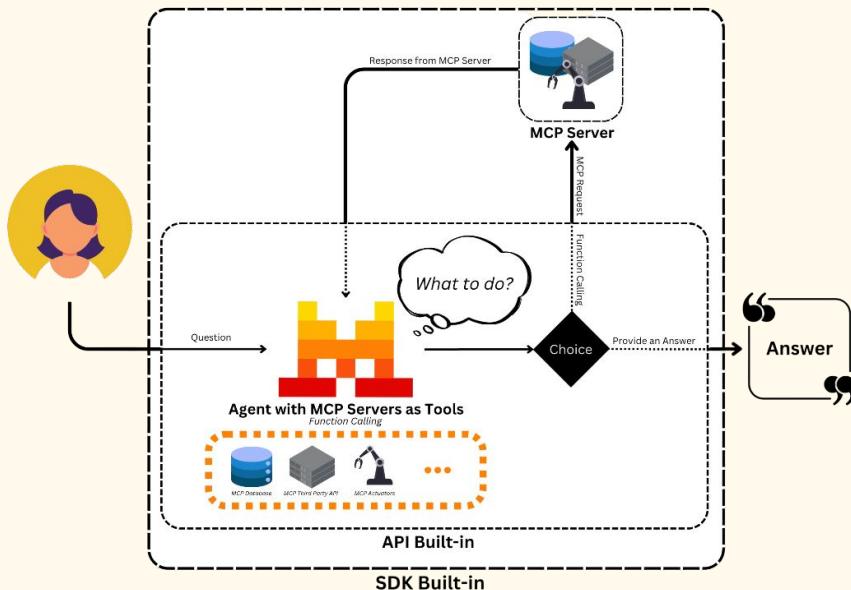
The Agents API offers web search as a connector, enabling developers to combine Mistral models with diverse, up-to-date information from web search, reputable news, and other sources.

There are two versions:

- `web_search`: A simple web search tool that enables access to a search engine.
- `web_search_premium`: A more complex web search tool that enables access to both a search engine and to news articles via integrated news provider verification.



Create an agent with built-in connectors and MCP tools - MCP Tools



The Agents API SDK can also leverage tools built on the Model Context Protocol (MCP)—an open, standardized protocol that enables seamless integration between agents and external systems.

MCP tools provide a flexible and extensible interface for agents to access real-world context, including APIs, databases, user data, documents, and other dynamic resources.

You can either use a **local MCP Server**, or a **Remote MCP Server with or without Authentication**.

For more information on configuring and deploying your own MCP Server, refer to the [Model Context Protocol documentation](#).

Agents API : Persistent memory across conversations

Once your agent is created, you can **start** conversations at any point while keeping the same conversation persistent.

- There are two ways to start a conversation:
 1. **With an Agent:** Create a conversation with a specific agent_id to leverage its specialized capabilities.

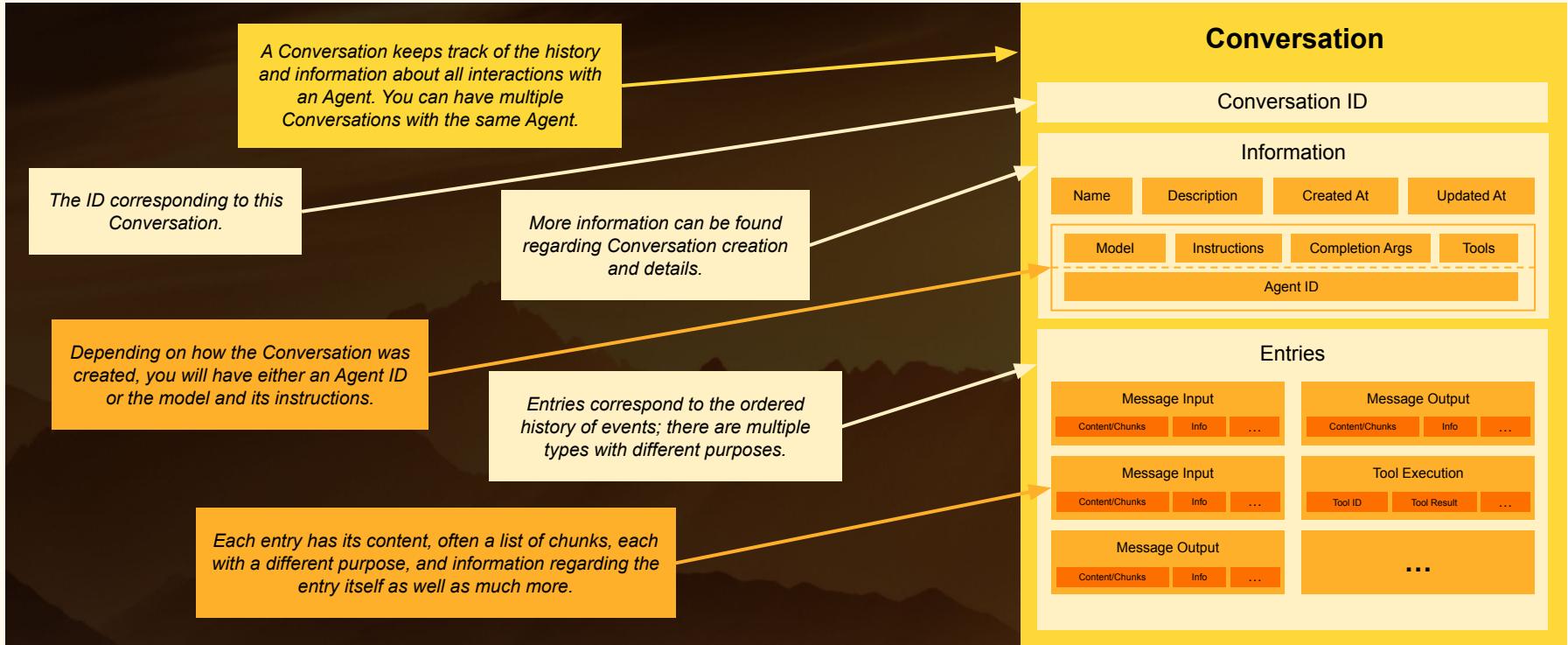
```
response = client.beta.conversations.start(  
    agent_id=simple_agent.id,  
    inputs="Who is Albert Einstein?",  
    #store=False  
)
```

2. **Direct Access:** Start a conversation by directly specifying the model and completion parameters, providing quick access to built-in connectors.

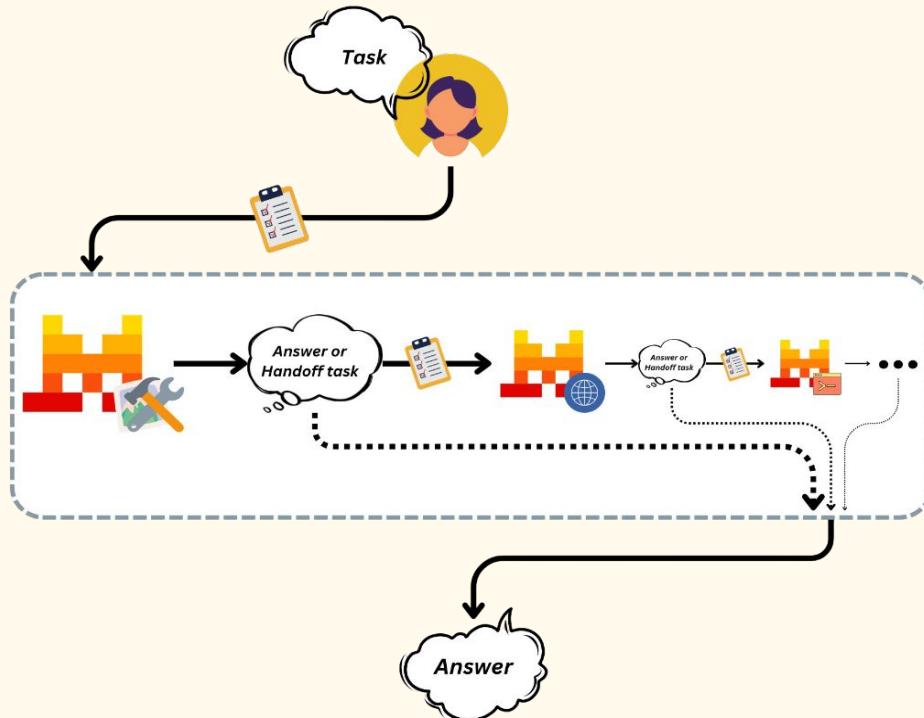
```
response = client.beta.conversations.start(  
    model="mistral-medium-latest",  
    inputs=[{"role": "user", "content": "Who is Albert Einstein?"}],  
    #store=False  
)
```

Each conversation maintains a structured history through conversation entries, ensuring that the context is preserved across interactions. You can continue, retrieve or restart conversations.

Conversations, manage and monitor context



Agents API : Agent orchestration with Handoffs



When creating and using Agents, often with access to specific tools, there are moments where it is desired to call other Agents mid-action.

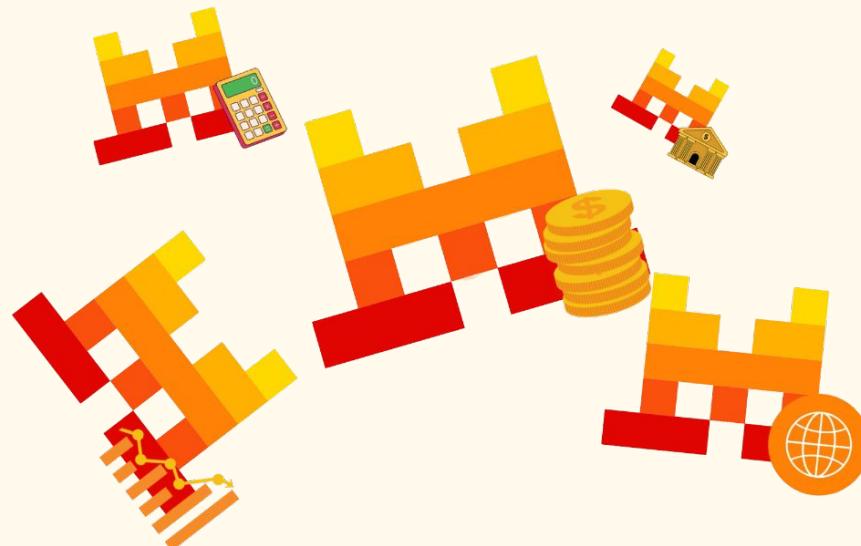
To elaborate and engineer workflows for diverse tasks that you may want automated, this ability to give Agents tasks or hand over a conversation to other agents is called **Handoffs**.

Agents API : Agent orchestration with Handoffs

1st step : Create multiple Agents

When creating a workflow powered by Handoffs, we first need to create all the Agents that our workflow will use. There is no limit to how many chained Handoffs a workflow can have.

You are free to create multiple Agents using diverse tools, models and handoffs, and orchestrate your own workflow using these Agents.



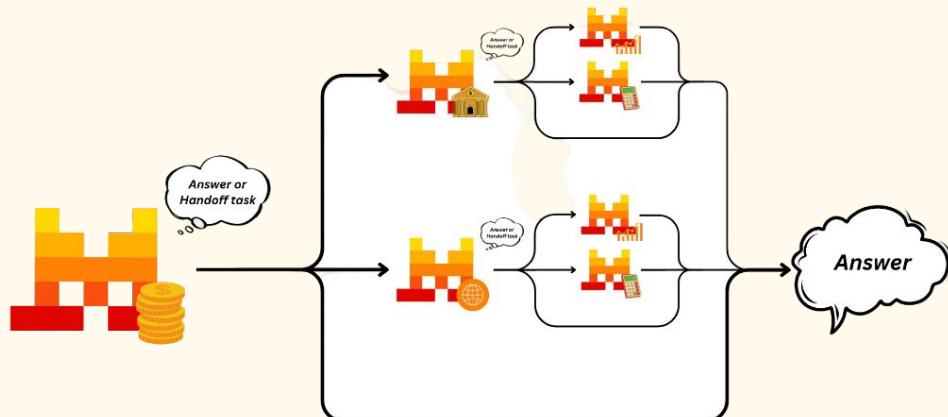
Agents API : Agent orchestration with Handoffs - Example

Agent name	Model	Description	Tools
finance_agent	Mistral Large	Agent used to answer financial related requests	/
web_search_agent	Mistral Large	Agent that can search online for any information if needed	Web_search (built-in connector)
ecb_interest_rate_agent	Mistral Large	Can find the current interest rate of the European central bank	Get_european_central_bank_interest_rate (function)
graph_agent	Mistral Large	Agent used to create graphs using the code interpreter tool	Code_interpreter (built-in connector)
calculator_agent	Mistral Large	Agent used to make detailed calculations	/

Agents API : Agent orchestration with Handoffs

2nd step : Define Handoffs Responsibilities

Example with **5 agents** (`finance_agent`, `ecb_interest_rate_agent`, `web_search_agent`, `graph_agent`, `calculator_agent`)



```
# Allow the finance_agent to handoff the conversation to the ecb_interest_rate_agent or
# web_search_agent
finance_agent = client.beta.agents.update(
    agent_id=finance_agent.id,
    handoffs=[ecb_interest_rate_agent.id, web_search_agent.id]
)

# Allow the ecb_interest_rate_agent to handoff the conversation to the graph_agent or
# calculator_agent
ecb_interest_rate_agent = client.beta.agents.update(
    agent_id=ecb_interest_rate_agent.id,
    handoffs=[graph_agent.id, calculator_agent.id]
)

# Allow the web_search_agent to handoff the conversation to the graph_agent or calculator_agent
web_search_agent = client.beta.agents.update(
    agent_id=web_search_agent.id,
    handoffs=[graph_agent.id, calculator_agent.id]
)
```

Agents API : Agent orchestration with Handoffs

3rd step : Let's run our Agentic Workflow

"Fetch the current US bank interest rate and calculate the compounded effect if investing for the next 10y"



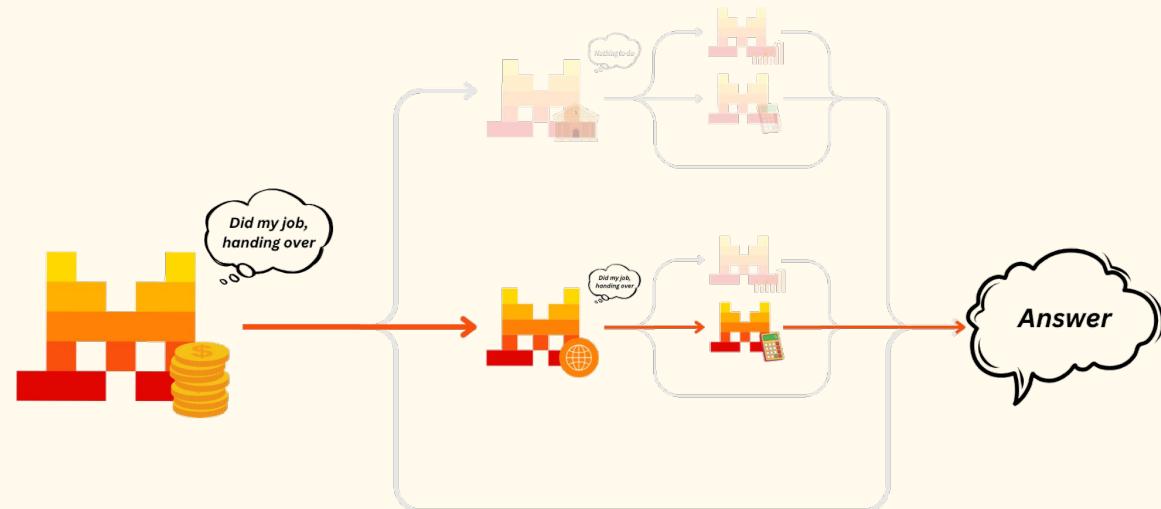
tool.execution :
Web Search

"The current US bank interest rate is 4.50 percent"



agent.handoff :
calculator-agent

"The future value of the investment after 10 years is \$1,540.10."



Agents API : Your turn ! 3pm-5pm

It's your turn to build agentic workflows!

In **teams of 6**, you'll have **2 hours** to work on a topic of your choice and build a use case that showcases the **capabilities of the Agent SDK**.

Feel free to use **built-in connectors**, create **new tools**, and more!

At 5:00 PM, each team will give a short **5-minute presentation** (with or without slides) to share what they built with the rest of the group!

You can create a dedicated directory in the Git repo and get started 

Useful Links

- [Mistral AI documentation](#)
- [Help Center](#)
- [Mistral AI Github / Agents API Cookbooks](#)
- [API Reference](#)
- [Discord Community](#)

