

AI-based Front-End Generation

Tools & Services.

Internship at
iO Digital

Internship Assignment Document

Luuk Briels
467020
0.0.1

Table of Contents.

1.	Context	4
1.1	Introduction	4
1.2	Objective	4
1.3	Models and services to compare	4
2.	Functionalities and Capabilities	5
2.1	Theme understanding	5
2.2	HTML/CSS understanding	11
3.	Technical Considerations	15
3.1	API integration	15
3.2	Limitations	18
3.3	Processing speed	21
4.	Cost and Support	22
4.1	Pricing models	22
4.2	Support	24
5.	Additional Considerations	30
5.1	Data privacy	30
5.2	Security	33
5.3	Licensing	36
6.	Overview	37
7.	Sources	40

1.1 Introduction

This research focuses on evaluating the capabilities, technical aspects, development complexities, and costs associated with various AI solutions. By analyzing the different AI models and services, an overview has been made which gives insights into the positives and negatives of each of the different solutions for this project.

1.2 Objective

The goal of this research is to help identify the most effective and suitable AI models and services to use in the development of a tool that is capable of creating temporary theme pages, such as those used during Christmas or Black Friday.

1.3 Models and services to compare

To achieve the objective of identifying the most suitable AI solution, a selection of models and services will be investigated and compared. This section will show the key contenders that will be compared.

Large language models

- GPT-4
- LLaMA 2 70B
- Code LLaMA
- Mistral Large
- Gemini
- Command
- Perplexity 70B
- Claude-3 (Not available in the Netherlands as of writing)

Image generation models

- DALL-E 3
- Stable Diffusion XL
- OpenJourney
- Playground V2.5



2.1 Theme understanding

An important factor in selecting the optimal AI solution for this project is its ability to understand and interpret various promotional themes. This section looks into the theme understanding capabilities of the models under consideration. Does it know specific imagery, color schemes, and language patterns associated with certain promotional events?

To test this understanding, each of the models will be given the following prompts (Both in English and in Dutch):

Language models

1. "Write a very short description of a Christmas promotion on a website selling winter clothing."
2. "Rewrite the following sentence to fit the Back-to-School theme, include items/objects associated with the theme: 'Browse our latest arrivals!'"
3. "List five keywords related to a Summer Sale theme."

Image models

1. "Create an image that is Christmas related."
2. "Create an image that is Easter related."
3. "Create an image that is Summer related."

Based on the results, each of the models will be determined on if it is capable of theme understanding. Next are two different tables (one for language models, and one for image models) which show the results of the tests.

Language models



Model	Prompt Nr.	Output	Passed
GPT-4	1	"Embrace the spirit of the season with our Christmas promotion! Stay warm and stylish with our cozy winter clothing collection. From snug sweaters to fashionable jackets, find the perfect ensemble to celebrate in style. Enjoy special discounts and festive offers throughout the holiday season. 'Tis the season to stay fashionable and festive!"	YES
	2	"Explore our fresh stock of backpacks, textbooks, and stationery supplies!"	YES
	3	"Sunshine", "Beach", "Heatwave", "Savings", "Vacation"	YES




LLaMA 2 70B	1	"Get ready for the cold weather with our winter clothing sale! Use code SLEIGHTHESEASON at checkout to receive 20% off all coats, gloves, hats, and scarves. Plus, spend over \$50 and get a free pair of warm winter socks. Don't miss out - shop now and stay warm and stylish all winter long!"	YES
	2	"Get ready to head back to school in style! Check out our latest arrivals, including stylish backpacks, sleek laptops, and all the school supplies you'll need to succeed. Shop now and start the school year off right!"	YES
	3	"Summer Sale", "Discounts", "Deals", "Savings", "Clearance"	YES
Mistral Large	1	"Embrace the Holiday Chill! Enjoy 30% off on our cozy winter clothing collection. Stay warm and stylish this Christmas. Limited time offer. Shop now!"	YES
	2	"Explore our fresh selection of back-to-school essentials, including stylish backpacks, vibrant notebooks, and innovative tech gadgets! Dive in and gear up for a successful academic year!"	YES
	3	"Discounts", "Beachwear", "Barbecue", "Sun Protection", "Summer Clearance"	YES
Gemini	1	"Wrap up warmth this Christmas! Get 25% off all winter wear at [Your Website Name]. Cozy up with new jackets, scarves, and more. Sale ends December 24th!"	YES
	2	"Unpack your potential! Explore our newest backpacks, notebooks, and pencils."	YES
Command	1	"Jingle all the way through our Winter Sale! Warm up to 20% off select jackets and apparel throughout the store. Pour yourself some eggnog, put on a festive record, and get in the holiday spirit with us! Offers end December 26th."	YES
	2	"We've got loads of new supplies and accessories in our Back-to-School shop, perfect for kicking off the new semester in style! Check out our fresh backpacks, colorful pencils, inventive gadgets, and updated textbooks for all your classroom needs."	YES
	3	"Markdowns", "Discounts", "Deals", "Offers", "Savings"	YES
Perplexity 70B	1	"The website is offering a Christmas promotion on winter clothing, featuring discounts on a wide range of cozy and stylish apparel perfect for the holiday season. Get ready to stay warm and fashionable this winter with great deals on jackets, sweaters, scarves, and more!"	YES
	2	"Discover our new Back-to-School essentials, including backpacks, notebooks, and pencils."	YES

Capabilities.

L.




Image models




Model	Prompt Nr.	Output	Passed
DALL-E 3	1		YES
	2		YES
	3		YES

Stable Diffusion XL	1		YES
	2		YES
	3		YES

Capabilities.

L.

Open-Journey	1		YES
	2		YES
	3		YES

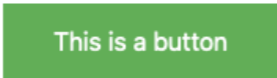


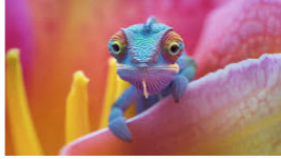
Play-ground V2.5	1		YES
	2		YES
	3		YES

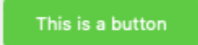


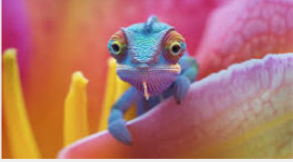


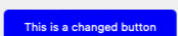
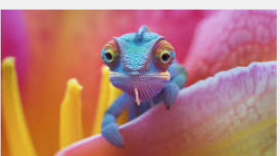
2.2 HTML/CSS understanding

To test the HTML and CSS capabilities, a variety of different HTML and CSS related prompts will be used for evaluation.

The following tasks will be given to perform:



1. "Make a simple HTML & CSS template with the following elements: heading: "This is a heading", sub text: "This is the text below the heading", green button: "This is a button"
2. "Change the template so that the first heading says: "This is a changed heading" and the buttons says: "This is a changed button" and make the button blue"
3. "Add an image to template that is aligned next to the text on the right side"



Model	Prompt Nr.	Output	Passed
GPT-4	1	<p>This is a heading</p> <p>This is the text below the heading</p> 	YES
	2	<p>This is a changed heading</p> <p>This is the text below the heading</p> 	YES
	3	<p>This is a changed heading</p> <p>This is the text below the heading</p>  	YES

LLaMA 2 70B	1	<p>This is a heading</p> <p>This is the text below the heading</p> 	YES
	2	<p>This is a changed heading</p> <p>This is the text below the heading</p> 	YES
	3	<p>This is a changed heading</p> <p>This is the text below the heading</p>  	NO
Code LLa-MA 70B	1	<p>This is a heading</p> <p>This is the text below the heading</p> 	YES
	2	<p>This is a changed heading</p> <p>This is the text below the heading</p> 	YES
	3	<p>This is a changed heading</p> <p>This is the text below the heading</p>  	NO

Capabilities.

L.

Model	Prompt Nr.	Output	Passed
Mistral Large	1	<p>This is a heading</p> <p>This is the text below the heading</p> <p>This is a button</p>	YES
	2	<p>This is a changed heading</p> <p>This is the text below the heading</p> <p>This is a changed button</p>	YES
	3	<p>This is a changed heading</p> <p>This is the text below the heading </p> <p>This is a changed button</p>	YES
Gemini	1	<p>This is a heading</p> <p>This is the text below the heading.</p> <p>This is a button</p>	YES
	2	<p>This is a changed heading</p> <p>This is the text below the heading.</p> <p>This is a changed button</p>	YES
	3	<p>This is a changed heading This is the text below the heading. </p>	NO

Command	1	<p>This is a heading</p> <p>This is the text below the heading</p> <p>This is a button</p>	YES
	2	<p>This is a changed heading</p> <p>This is the text below the heading</p> <p>This is a changed button</p>	YES
	3	<p>This is a changed heading</p> <p>This is the text below the heading </p> <p>This is a changed button</p>	YES
Perplexity 70B	1	<p>This is a heading</p> <p>This is the text below the heading</p> <p>This is a button</p>	YES
	2	<p>This is a changed heading</p> <p>This is the text below the heading</p> <p>This is a changed button</p>	YES
	3	<p>This is a changed heading</p> <p>This is the text below the heading </p> <p>This is a changed button</p>	YES

3.1 API Integration

It could be a great benefit if the model already has an API integration available to use. This removes the need to host and call the models on your own as that could require expensive hardware and electricity cost.

Language models

GPT-4

API Service	SDK	Languages
OpenAI	YES	Python, C#, C++, Go, Java, JavaScript, PHP, Ruby, Rust
Azure	YES	C#, Go, Java, JavaScript, Python

(OpenAI Platform, n.d. & Mrbullwinkle, 2023)

LLaMA 2 70B

Service	SDK	Languages
Lepton AI	YES	Python, C#, C++, Go, Java, JavaScript, PHP, Ruby, Rust
Groq	YES	Python, JavaScript
Fireworks	YES	Python
Deepinfra	YES	Python, C#, C++, Go, Java, JavaScript, PHP, Ruby, Rust
Anyscale	YES	Python
Together.ai	YES	Python, C#, C++, Go, Java, JavaScript, PHP, Ruby, Rust
AWS	YES	Python, C++, Go, Java, JavaScript, PHP, Ruby, Rust
Azure	NO	-

(Model APIs | Lepton AI, n.d. & GroqCloud, n.d. & Quickstart, n.d. & Use OpenAI API Clients With LLaMas | ML Models | Deep Infra, n.d. & QuickStart | Anyscale Docs, n.d. & Amazon Bedrock API Reference - Amazon Bedrock, n.d.)

Code LLaMA 70B

Service	SDK	Languages
Fireworks	YES	Python
Deepinfra	YES	Python, C#, C++, Go, Java, JavaScript, PHP, Ruby, Rust
Together.ai	YES	Python, C#, C++, Go, Java, JavaScript, PHP, Ruby, Rust

(Quickstart, n.d. & Use OpenAI API Clients With LLaMas | ML Models | Deep Infra, n.d. & Quickstart, n.d.)

Mistral Large

Service	SDK	Languages
Mistral	YES	Python, C#, C++, Go, Java, JavaScript, PHP, Ruby, Rust
Azure	NO	-

(Azure AI | Mistral AI Large Language Models, n.d. & Microsoft Azure Marketplace, n.d.)

Gemini

Service	SDK	Languages
Google Cloud	YES	Python, JavaScript, Java, Go

(Introduction to Language Model Classes in the Vertex AI SDK, n.d.)

Command

Service	SDK	Languages
Cohere	YES	Python
AWS	YES	Python

(About, n.d. & Cohere on AWS, n.d.)

Tech Considerations.

Perplexity 70B

Service	SDK	Languages
Perplexity	NO	-

(Getting Started With Pplx-api, n.d.)

Image models

DALL-E 3

Service	SDK	Languages
OpenAI	YES	Python, C#, C++, Go, Java, JavaScript, PHP, Ruby, Rust
Azure	YES	C#, Go, Java, JavaScript, Python

(Getting Started With Pplx-api, n.d.)

Stable Diffusion XL

Service	SDK	Languages
Stability AI	YES	Python, JavaScript
Novita AI	YES	Python

(Stability AI - Developer Platform, n.d. & novitalabs, n.d.)

OpenJourney

Service	SDK	Languages
Hugging Face	NO	-
Replicate	YES	JavaScript, Python, Go

(Client Libraries – Replicate, n.d.)

Playground V2.5

Service	SDK	Languages
Hugging Face	NO	-
Replicate	YES	JavaScript, Python, Go

(Client Libraries – Replicate, n.d.)

3.2 Limitations

GPT-4

API Service	Rate limits
OpenAI	500 requests per minute 10,000 requests per day 10,000 tokens per minute
Azure	10 requests per second

(OpenAI Platform, n.d. & PatrickFarley, 2024)

LLaMA 2 70B

Service	Rate limits
Lepton AI	10 per minute
Groq	20 requests per minute 25 requests per 10 minutes
Fireworks	600 requests per minute
Deepinfra	200 concurrent requests
Anyscale	30 Concurrent requests
Together.ai	100 requests per second
AWS	400 requests per second 300,000 tokens per minute
Azure	10 requests per second

(Model APIs | Lepton AI, n.d. & Fireworks.ai, 2016 & GroqCloud, n.d. & Rate Limits of the Deep Infra Endpoints | ML Models | Deep Infra, n.d. & Anyscale Endpoints, n.d. & Quotas for Amazon Bedrock - Amazon Bedrock, n.d. & PatrickFarley, 2024)

Mistral Large

Service	Rate limits
Mistral	5 requests per second 2 million tokens per minute 10,000 million tokens per month
Azure	10 requests per second

(Pricing and Rate Limits | Mistral AI Large Language Models, n.d. & PatrickFarley, 2024)

Tech Considerations.

Gemini

Service	Rate limits
Google Cloud	60 requests per minute

(Quickstart: Get Started With the Gemini API in Web Apps, n.d.)

Command

Service	Rate limits
Cohere	100 requests per minute
AWS	400 requests per second 300,000 tokens per minute

(Team, 2023 & PatrickFarley, 2024)

Perplexity 70B

Service	Rate limits
Perplexity	8 requests per 5 seconds 24 requests per minute 240 requests per hour

(Rate Limits, n.d.)

Image models

DALL-E 3

Service	Rate limits
OpenAI	5 img/min

(OpenAI Platform, n.d.)

Stable Diffusion XL

Service	Rate limits
Stability AI	150 requests per 10 seconds
Novita AI	-

(StabilityAI - Developer Platform, n.d. & Novita.ai API Documentation, n.d.)

OpenJourney

Service	Rate limits
Hugging Face	-
Replicate	10 requests per second

(How Does Replicate Work?, n.d.)

Playground V2.5

Service	Rate limits
Hugging Face	-
Replicate	10 requests per second

(How Does Replicate Work?, n.d.)

Tech Considerations.

3.3 Processing speed

Language models

Model	Avg. Throughput	Avg. Latency
GPT-4	29 tokens/second	0.6s
LLaMA 2 70B	70 tokens/second	0.3s
Code LLaMA 70B	-	-
Mistral Large	24 tokens/second	1.0s
Gemini	89 tokens/second	1.9s
Command	29 tokens/second	0.3s
Perplexity 70B	54 tokens/second	1.2s

(Comparison of AI Models Across Quality, Performance, Price | Artificial Analysis, n.d.)

Image models

Model	Avg. Throughput	Avg. Latency
DALL-E 3	21s/image	0.2s
Stable Diffusion XL	6s/image	0.4s
OpenJourney	3s/image	0.1s
Playground V2.5	5s/image	0.2s

(Based on my own tests, average time of 5 runs)



4.1 Pricing models

Another factor in choosing the appropriate model is cost. Therefore a pricing comparison between the models is made. For each model the pricing is assessed by the cost per 1M tokens, and cost per image.

Language models

GPT-4

Service	Input (Per 1M tokens)	Output (Per 1M tokens)
OpenAI	\$30	\$60
Azure	\$30	\$60

(Pricing, n.d. & Azure OpenAI Service - Prijzen | Microsoft Azure, n.d.)

LLaMA 2 70B

Service	Input (Per 1M tokens)	Output (Per 1M tokens)
Lepton AI	\$0.8	\$0.8
Groq	\$0.7	\$0.8
Fireworks	\$0.7	\$2.8
Deepinfra	\$0.7	\$0.9
Anyscale	\$1.0	\$1.0
Together.ai	\$0.9	\$0.9
AWS	\$1.95	\$2.56
Azure	\$1.54	\$1.77

(Pricing | Lepton AI, n.d. & Groq, Inc., 2024 & Fireworks.ai., 2016 & Simple Pricing | Machine Learning Infrastructure | Deep Infra, n.d. & Pricing | Anyscale Docs, n.d. & Together Pricing | the Most Powerful Tools at the Best Value, n.d. & AWS Marketplace: Meta Llama 2 Chat 70B (Amazon Bedrock Edition), n.d. & Microsoft Azure Marketplace, n.d.)

Code LLaMA 70B

Service	Input (Per 1M tokens)	Output (Per 1M tokens)
Fireworks	\$0.7	\$2.8
Deepinfra	\$0.7	\$0.9
Together.ai	\$0.9	\$0.9

(Fireworks.ai., 2016 & Simple Pricing | Machine Learning Infrastructure | Deep Infra, n.d. & Together Pricing | the Most Powerful Tools at the Best Value, n.d.)

Mistral Large

Service	Input (Per 1M tokens)	Output (Per 1M tokens)
Mistral	\$8	\$24
Azure	\$8	\$24

(Pricing and Rate Limits | Mistral AI Large Language Models, n.d. & Microsoft Azure Marketplace, n.d.)

Gemini

Service	Input (Per 1M tokens)	Output (Per 1M tokens)
Google Cloud	\$0.125	\$0.375

(Pricing for Generative AI on Vertex AI | Google Cloud, n.d.)

Command

Service	Input (Per 1M tokens)	Output (Per 1M tokens)
Cohere	\$1	\$2
AWS	\$1.5	\$2

(Pricing, n.d. & Build Generative AI Applications With Foundation Models - Amazon Bedrock Pricing - AWS, n.d.)

Perplexity 70B

Service	Input (Per 1M tokens)	Output (Per 1M tokens)
Perplexity	\$1	\$1

(Pricing, n.d.)

Image models

DALL-E 3

Service	Price per image
OpenAI	\$0.04 - \$0.12
Azure	\$0.04 - \$0.12

(Pricing, n.d. & Azure OpenAI Service - Pricing | Microsoft Azure, n.d.)



Stable Diffusion XL

Service	Price per image
Stability AI	\$0.002 - \$0.003
Novita AI	\$0.0015 - \$0.0126

(*Stability AI - Developer Platform, n.d. & Novita.ai Pricing, n.d.*)

OpenJourney

Service	Price per image
Hugging Face	Unlimited for \$9/month
Replicate	≈ \$0.0046

(*Prompthero/Openjourney · Hugging Face, n.d. & Prompthero/Openjourney – Replicate, n.d.*)

Playground V2.5

Service	Price per image
Hugging Face	Unlimited for \$9/month
Replicate	≈ \$0.0055

(*Playgroundai/playground-v2.5-1024px-aesthetic · Hugging Face, n.d. & Lucataco/playground-v2.5-1024px-aesthetic – Replicate, n.d.*)

4.2 Support

It is important that there are enough resources available that help to understand, develop, and resolve problems when using a certain model and service. Each model and service will be assessed by documentation and the community around it.

Language models

GPT-4

Community

Dedicated Forums: The primary forum for GPT-4 discussions exists within the OpenAI Developer Forum. Here, developers can find discussions on using the API, troubleshooting errors, and sharing creative use cases. However, the GPT-4 specific threads are still developing compared to established models like GPT-3. (*OpenAI Developer Forum, n.d.*)

External Communities: Discussions on other platforms are common like Reddit and Stackoverflow. However, independent communities or forums dedicated solely to GPT-4 haven't emerged yet.

Support

Official Documentation: OpenAI provides official documentation for the GPT-4 model. (*OpenAI Platform, n.d.*)

Limited Direct Support: OpenAI offers limited direct support for developers. For troubleshooting, developers primarily rely on the documentation and community forum.

LLaMA 2 70B

Community

No Dedicated Forums: There aren't currently any established developer communities or forums specifically focused on LLaMA 2 70B.

External Communities: Discussions about LLaMA 2 70B appear in broader AI development forums or communities that cover various large language models like Hugging Face and Reddit.

Support

No Direct Support: As of now, there are no direct support channels for LLaMA 2 70B developers.

Code LLaMA 70B

Community

No Dedicated Forums: There aren't established, independent developer communities solely for Code LLaMA 70B yet.

External Communities: Discussions are emerging on platforms like Reddit, where developers share experiences and explore the model's capabilities.

Support

No Direct Support: There are no official dedicated support channels from Meta for Code LLaMA 70B developers.

Cost and Support.



Mistral Large

Community

Limited Dedicated Forums: There are no dedicated forums, online communities, or developer groups specifically focused on Mistral Large. They do, however, have their own Discord server. *(Ai, n.d.)*

Limited External Communities: Mentions of Mistral Large in broader AI development circles are scarce.

Support

Official Documentation: Mistral AI provides official documentation for the Mistral Large API. However, this documentation doesn't go into much depth. *(Introduction | Mistral AI Large Language Models, n.d.)*

Limited Direct Support: Mistral doesn't offer any direct support, but they do however have their own Discord server and contact form where questions can be asked. *(Ai, n.d.)*

Gemini

Community

Potential Dedicated Forums: Google has a developer community forum that supports various AI tools and services. They also have a category "AI/ML", where Gemini is mainly talked about. *(Byakuren, n.d.)*

Support

Official Documentation: There is official documentation for the Gemini API, guiding developers on using the model's functionalities. *(Gemini Docs and API Reference, n.d.)*

Direct Support: Google Offers support for Google Cloud which is where Gemini also falls under. Support could also come from the broader Google AI developer community forum. *(Get Support, n.d.)*

Command

Community

Dedicated Forums: There are no dedicated forums, online communities, or developer groups specifically focused on Mistral Large. They do, however, have their own Discord server. *(Gomez, 2023)*

Support

Official Documentation: There is official documentation for the Cohere API, guiding developers on using the model's functionalities. *(Cohere AI, n.d.)*

Limited Direct Support: Command doesn't offer any direct support, but they can however be contacted via email, X (Twitter), or Discord. *(Home, n.d.)*

Perplexity 70B

Community

Dedicated Forums: Perplexity does have their own dedicated forum for Perplexity AI. However, activity in this forum is rather lacking. However, they do also have their own Discord server. *(Discussions, n.d.)*

Support

Official Documentation: There is official documentation for the Perplexity API, which guides developers on using the model's functionalities. *(Getting Started With Pplx-api, n.d.)*

Limited Direct Support: Perplexity doesn't offer any direct support, but they do however have their own Discord server and X (Twitter) where questions can be asked. *(Pplx-api, n.d.)*

Cost and Support.



Image models

DALL-E 3

Community

Dedicated Forums: The primary forum for DALL-E 3 discussions exists within the OpenAI Developer Forum. Here, developers can find discussions on using the API, troubleshooting errors, and sharing creative use cases. *(OpenAI Developer Forum, n.d.)*

External Communities: There are external communities that are aimed at DALL-E 3 like Reddit and Stackoverflow.

Support

Official Documentation: There is official documentation for the DALL-E 3 API, which guides developers on using the model's functionalities. *(OpenAI Platform, n.d.)*

Limited Direct Support: OpenAI offers limited direct support for developers. For troubleshooting, developers primarily rely on the documentation and community forum.

Stable Diffusion XL

Community

Limited Dedicated Forums: There are no dedicated forums, but there are dedicated discussions on the official Stable Diffusion profile on Hugging Face. There are also dedicated threads on Reddit. *(Stabilityai/Stable-diffusion · Discussions, n.d.)*

Support

Official Documentation: There is official documentation for the Stable Diffusion 3 API, which guides developers on using the model's functionalities. *(Stability AI - Developer Platform, n.d.)*

Direct Support: They do offer direct support via their form. They also have a Discord server. *(Stability AI - Developer Platform, n.d.)*

OpenJourney

Community

Dedicated Forums: There are no dedicated forums, but there are dedicated discussions on the official OpenJourney profile on Hugging Face. There are also dedicated threads on Reddit. *(Prompthero/Openjourney · Discussions, n.d.)*

External Communities: There are external communities that are aimed at OpenJourney like Reddit and Stackoverflow.

Support

No Official Documentation: There is no official documentation available on OpenJourney.

No Direct Support: There is no direct support for OpenJourney.

Playground V2.5

Community

Dedicated Forums: There are no dedicated forums, but there are dedicated discussions on the official OpenJourney profile on Hugging Face.

(Playgroundai/playground-v2.5-1024px-aesthetic · Discussions, n.d.)

Support

No Official Documentation: There is no official documentation available on Playground V2.5..

Direct Support: They don't offer direct support but they do have a Discord server with their own community.

(Free AI Image Generator: Art, Social Media, Marketing | Playground, n.d.)

5.1 Data privacy

How data is being handled by the different models is also something to take into account. If data is put in, what happens with this data?

Language models

GPT-4

Service	API data used for training	Data Compliances
OpenAI	NO	GDPR
Azure	NO	GDPR

(Security, n.d. & Jemash, n.d.)

LLaMA 2 70B

Service	API data used for training	Data Compliances
Lepton AI	Not mentioned	None mentioned
Groq	Not mentioned	None mentioned
Fireworks	Not mentioned	None mentioned
Deepinfra	Not mentioned	GDPR
Anyscale	Not mentioned	GDPR
Together.ai	NO	None mentioned
AWS	NO	GDPR
Azure	NO	GDPR

(Privacy Policy | Lepton AI, n.d. & Groq, Inc., 2024 & Privacy Notice, n.d. & DeepInfra Privacy Policy, n.d. & Privacy Notice | Anyscale, n.d. & Legal, n.d. & Build Generative AI Applications With Foundation Models - Amazon Bedrock FAQs - AWS, n.d. & Jemash, n.d.)

Code LLaMA 70B

Service	API data used for training	Data Compliances
Fireworks	Not mentioned	None mentioned
Deepinfra	Not mentioned	GDPR
Together.ai	NO	None mentioned

(Privacy Notice, n.d. & DeepInfra Privacy Policy, n.d. & Legal, n.d.)

Mistral Large

Service	API data used for training	Data Compliances
Mistral	YES, unless opted-out	GDPR
Azure	NO	GDPR

(Ai, n.d. & Jemash, n.d.)

Gemini

Service	API data used for training	Data Compliances
Google Cloud	YES, and may be manually reviewed	GDPR

(Gemini API Additional Terms of Service | Google AI for Developers, n.d.)

Command

Service	API data used for training	Data Compliances
Cohere	YES, unless opted-out	None mentioned
AWS	NO	GDPR

(Data Usage Policy, n.d. & Build Generative AI Applications With Foundation Models - Amazon Bedrock FAQs - AWS, n.d. & Jemash, n.d.)

Perplexity 70B

Service	API data used for training	Data Compliances
Perplexity	NO	None mentioned

(What Data Does Perplexity Collect About Me?, n.d.)

Image models

DALL-E 3

Service	API data used for training	Data Compliances
OpenAI	NO	GDPR

(Security, n.d.)

Stable Diffusion XL

Service	API data used for training	Data Compliances
Stability AI	Not mentioned	GDPR
Novita AI	Not mentioned	None mentioned

(Privacy Policy — Stability AI, n.d.)

OpenJourney

Service	API data used for training	Data Compliances
Hugging Face	NO	GDPR
Replicate	NO	GDPR

(Security, n.d. & Privacy Policy – Replicate, n.d.)

Playground V2.5

Service	API data used for training	Data Compliances
Hugging Face	NO	GDPR
Replicate	NO	GDPR

(Security, n.d. & Privacy Policy – Replicate, n.d.)

5.2 Security

How is are the services being protected and what compliances are in place that help to verify their integrity?

Language models

GPT-4

Service	Transport	Compliances
OpenAI	SSL	SOC 2, SOC 3
Azure	SSL	ISO 20000-1, ISO 22301, ISO 27001, ISO 27017, ISO 27018, ISO 27701, ISO 9001, SOC 1, SOC 2, SOC 3

(Security, n.d. & TerryLanfear, n.d.)

LLaMA 2 70B

Service	Transport	Compliances
Lepton AI	SSL	None mentioned
Groq	SSL	None mentioned
Fireworks	SSL	SOC 2, HIPAA
Deepinfra	SSL	None mentioned
Anyscale	SSL	None mentioned
Together.ai	SSL	None mentioned
AWS	SSL	CSA, CyberGRX, CyberVadis, ISO 14001, ISO 9001, ISO 20000, ISO 22301, ISO 27001, ISO 27017, ISO 27701, ISO 27018. SOC 1, SOC 2, SOC 3 HIPAA
Azure	SSL	ISO 20000-1, ISO 22301, ISO 27001, ISO 27017, ISO 27018, ISO 27701, ISO 9001, SOC 1, SOC 2, SOC 3

(Privacy Policy | Lepton AI, n.d.& Groq, Inc., 2024 & Privacy Notice, n.d. & Fireworks.Ai, 2023 & DeepInfra Privacy Policy, n.d. & Privacy Notice | Anyscale, n.d. & Legal, n.d. & Compliance Programs - Amazon Web Services (AWS), n.d. & TerryLanfear, n.d..)

Code LLaMA 70B

Service	Transport	Compliances
Fireworks	SSL	SOC 2, HIPAA
Deepinfra	SSL	None mentioned
Together.ai	SSL	None mentioned

(Privacy Notice, n.d. & DeepInfra Privacy Policy, n.d. & Legal, n.d.)

Mistral Large

Service	Transport	Compliances
Mistral	SSL	None mentioned
Azure	SSL	ISO 20000-1, ISO 22301, ISO 27001, ISO 27017, ISO 27018, ISO 27701, ISO 9001, SOC 1, SOC 2, SOC 3

(Ai, n.d. & TerryLanfear, n.d.)

Image models

DALL-E 3

Service	Transport	Compliances
OpenAI	SSL	SOC 2, SOC 3

(Security, n.d..)

Stable Diffusion XL

Service	Transport	Compliances
Stability AI	SSL	None mentioned
Novita AI	SSL	None mentioned

(Privacy Policy — Stability AI, n.d.)

OpenJourney

Service	Transport	Compliances
Hugging Face	SSL	SOC 2
Replicate	SSL	None mentioned

(Security, n.d.)

Playground V2.5

Service	Transport	Compliances
Hugging Face	SSL	SOC 2
Replicate	SSL	None mentioned

(Security, n.d.)

5.3 Licensing

If this project will be used commercially, it is important to know what rights you have using the different models.

Language models

Model	Commercial use
GPT-4	YES
LLaMA 2 70B	YES, on request
Code LLaMA 70B	YES, on request
Mistral Large	YES
Gemini	YES
Command	YES
Perplexity 70B	MAYBE, required to contact them first

(Terms of Use, n.d. & Download Llama, n.d. & Mistral, n.d. & Banks, 2023 & Commercial SAAS Agreement, n.d. & Frequently Asked Questions, n.d.)

Image models

Model	Commercial use
DALL-E 3	YES
Stable Diffusion XL	YES
OpenJourney	YES
Playground V2.5	YES

(Can I Sell Images I Create With DALL·E? | OpenAI Help Center, n.d. & Mostaque, 2023 & Prompthero/Openjourney · Hugging Face, n.d. & Pricing - Playground, n.d.)

Overview.

L.

6 Overview

■ No abnormalities
 ■ Notable abnormalities
 ■ Significant abnormalities

Language models

Model	GPT-4	LLaMA 2 70B	Code LLaMA 70B	Mistral Large	Gemini	Command	Perplexity 70B
Theme understanding	Green	Green	Green	Green	Green	Green	Green
HTML/CSS understanding	Green	Yellow	Yellow	Green	Yellow	Green	Green
API integration	Green	Green	Green	Green	Green	Yellow	Yellow
Processing speed	Green	Green	Green	Green	Green	Green	Green
Pricing	Yellow	Green	Green	Green	Green	Green	Green
Support	Green	Yellow	Yellow	Yellow	Green	Green	Green
Commercial use	Green	Yellow	Yellow	Green	Green	Green	Yellow

Services

Service	OpenAI	Azure	AWS	Google Cloud	Groq	Fireworks	Deepinfra	Anyscale	Together.ai	Mistral	Lepton AI	Cohere	Perplexity
Limitations	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Data Privacy	Green	Green	Green	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	Yellow	Red	Yellow
Security	Green	Green	Green	Green	Yellow	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Yellow

Overview.

L.

Image models

Model	DALL-E 3	Stable Diffusion XL	OpenJourney	Playground V2.5
Theme understanding				
API integration				
Processing speed				
Pricing				
Support				
Commercial use				

Services

Service	OpenAI	Stability AI	Novita AI	Hugging Face	Replicate
Limitations					
Data Privacy					
Security					

7 Sources

3.1 API Integration

OpenAI Platform. (n.d.). <https://platform.openai.com/docs/api-reference>

Mrbullwinkle. (2023, December 18). Azure OpenAI Service supported programming languages - Azure AI services. Microsoft Learn. <https://learn.microsoft.com/en-us/azure/ai-services/openai/supported-languages>

Model APIs | Lepton AI. (n.d.). https://www.lepton.ai/docs/public_models/model_apis

GroqCloud. (n.d.). <https://console.groq.com/docs/libraries>

Quickstart. (n.d.). Fireworks.ai. <https://readme.fireworks.ai/reference/installation>

Use OpenAI API clients with LLaMas | ML Models | Deep Infra. (n.d.). https://deepinfra.com/docs/advanced/openai_api

Quickstart. (n.d.). Together. <https://docs.together.ai/docs/inference-python>

QuickStart | Anyscale docs. (n.d.). <https://docs.anyscale.com/reference/python-sdk/quickstart>

Microsoft Azure Marketplace. (n.d.). <https://azuremarketplace.microsoft.com/en-us/marketplace/apps/000-000.mistral-ai-large-offer?tab=Overview>

Introduction to language model classes in the Vertex AI SDK. (n.d.). Generative AI on Vertex AI | Google Cloud. <https://cloud.google.com/vertex-ai/generative-ai/docs/sdk-for-llm/llm-sdk-overview>

About. (n.d.). Cohere AI. <https://docs.cohere.com/reference/about#sdks>

Cohere on AWS. (n.d.). Cohere AI. <https://docs.cohere.com/docs/cohere-on-aws>

Getting Started with pplx-api. (n.d.). Pplx-api. <https://docs.perplexity.ai/docs/getting-started>

Stability AI - Developer Platform. (n.d.). <https://platform.stability.ai/docs/getting-started/python-sdk>

novitalabs. (n.d.). GitHub - novitalabs/python-sdk: Python SDK for Stable Diffusion API (Doodle/Remove Background/Replace Object/Reimagine/Merge-Faces/Txt2Img/Img2Img/ControlNet/VAE). GitHub. <https://github.com/novitalabs/python-sdk>

Client libraries – Replicate. (n.d.). <https://replicate.com/docs/reference/client-libraries>

3.2 Limitations

OpenAI Platform. (n.d.). <https://platform.openai.com/docs/guides/rate-limits/usage-tiers?context=tier-one>

PatrickFarley. (2024, January 20). Autoscale AI services limits - Azure AI services. Microsoft Learn. <https://learn.microsoft.com/en-us/azure/ai-services/autoscale?tabs=portal>

Fireworks.ai. (2016). <https://fireworks.ai/pricing>

Model APIs | Lepton AI. (n.d.). https://www.lepton.ai/docs/overview/model_apis

GroqCloud. (n.d.). <https://console.groq.com/docs/rate-limits>

Anyscale Endpoints. (n.d.). Anyscale. <https://www.anyscale.com/endpoints>

Quotas for Amazon Bedrock - Amazon Bedrock. (n.d.). <https://docs.aws.amazon.com/bedrock/latest/userguide/quotas.html>

Pricing and rate limits | Mistral AI Large Language Models. (n.d.). <https://docs.mistral.ai/platform/pricing/>

Quickstart: Get started with the Gemini API in web apps. (n.d.). Google AI for Developers. https://ai.google.dev/tutorials/web_quickstart

Team, C. (2023, November 3). Introducing a free developer tier + simplified pricing. Context by Cohere. <https://txt.cohere.com/free-developer-tier-announcement/>

Rate limits. (n.d.). Pplx-api. <https://docs.perplexity.ai/docs/rate-limits>

Stability AI - Developer Platform. (n.d.). <https://platform.stability.ai/docs/api-reference>

Novita.ai API documentation. (n.d.). Theneo API Documentation. <https://docs.novita.ai/novita-ai>

How does Replicate work? (n.d.). Replicate. <https://replicate.com/docs/how-does-replicate-work>

3.3 Processing speed

Comparison of AI Models across Quality, Performance, Price | Artificial Analysis. (n.d.). <https://artificialanalysis.ai/models>

4.1 Pricing models

Pricing. (n.d.). <https://openai.com/pricing>

Azure OpenAI Service - Prijzen | Microsoft Azure. (n.d.). Microsoft Azure. <https://azure.microsoft.com/nl-nl/pricing/details/cognitive-services/openai-service/>

Pricing | Lepton AI. (n.d.). <https://www.lepton.ai/pricing>

Groq, Inc. (2024, February 28). Accelerating Systems with Real-time AI Solutions - Groq. Groq. <https://wow.groq.com/>

Fireworks.ai. (2016). <https://fireworks.ai/pricing>

Simple pricing | Machine Learning Infrastructure | Deep Infra. (n.d.). <https://deepinfra.com/pricing>

Pricing | Anyscale docs. (n.d.). <https://docs.endpoints.anyscale.com/pricing/>

Together Pricing | The most powerful tools at the best value. (n.d.). <https://www.together.ai/pricing>

AWS Marketplace: Meta Llama 2 Chat 70B (Amazon Bedrock Edition). (n.d.). <https://aws.amazon.com/marketplace/pp/prodview-7kcpngbt6eprg>

Microsoft Azure Marketplace. (n.d.). <https://azuremarketplace.microsoft.com/en/marketplace/apps/metagenai.meta-llama-2-70b-offer?tab=PlansAndPrice>

Pricing and rate limits | Mistral AI Large Language Models. (n.d.). <https://docs.mistral.ai/platform/pricing/>

Microsoft Azure Marketplace. (n.d.). <https://azuremarketplace.microsoft.com/en-us/marketplace/apps/000-000.mistral-ai-large-offer?tab=PlansAndPrice>

Pricing for Generative AI on Vertex AI | Google Cloud. (n.d.). Google Cloud. <https://cloud.google.com/vertex-ai/generative-ai/pricing>

Pricing. (n.d.). Cohere. <https://cohere.com/pricing>

Pricing. (n.d.). Pplx-api. <https://docs.perplexity.ai/docs/pricing>

Pricing. (n.d.). <https://openai.com/pricing#language-models>

Stability AI - Developer Platform. (n.d.). <https://platform.stability.ai/pricing>

Novita.ai pricing. (n.d.). <https://novita.ai/pricing>

Azure AI | Mistral AI Large Language Models. (n.d.). <https://docs.mistral.ai/cloud-deployment/azure/#going-further>

prompthero/openjourney · Hugging Face. (n.d.). <https://huggingface.co/prompthero/openjourney>

prompthero/openjourney – Replicate. (n.d.). <https://replicate.com/prompthero/openjourney>

playgroundai/playground-v2.5-1024px-aesthetic · Hugging Face. (n.d.). <https://huggingface.co/playgroundai/playground-v2.5-1024px-aesthetic>

lucataco/playground-v2.5-1024px-aesthetic – Replicate. (n.d.). <https://replicate.com/lucataco/playground-v2.5-1024px-aesthetic>

4.2 Support

OpenAI Developer Forum. (n.d.). OpenAI Developer Forum. <https://community.openai.com>

OpenAI Platform. (n.d.). <https://platform.openai.com/docs/guides/text-generation/>

Introduction | Mistral AI Large Language Models. (n.d.). <https://docs.mistral.ai/>

Byakuren. (n.d.). AI/ML. <https://www.googlecloudcommunity.com/gc/forums/filteredbylabelpage/board-id/cloud-ai-ml/label-name/Gemini>

Gemini Docs and API reference. (n.d.). Google AI for Developers. <https://ai.google.dev/docs>

Get support. (n.d.). Google Cloud. <https://cloud.google.com/vertex-ai/docs/support/getting-support>

Discussions. (n.d.). Pplx-api. <https://docs.perplexity.ai/discuss>

Getting Started with pplx-api. (n.d.). Pplx-api. <https://docs.perplexity.ai/docs/getting-started>

pplx-api. (n.d.). Pplx-api. <https://docs.perplexity.ai/>

Ai, M. (n.d.). Contact us. Mistral AI | Frontier AI in Your Hands. <https://mistral.ai/contact/>

Gomez, A. (2023, November 3). Building Co:mmunity . Context by Cohere. <https://txt.cohere.com/building-community/>

Cohere AI. (n.d.). <https://docs.cohere.com/>

Home. (n.d.). Cohere. <https://cohere.com/>

stabilityai/stable-diffusion · Discussions. (n.d.). <https://huggingface.co/spaces/stabilityai/stable-diffusion/discussions>

Stability AI - Developer Platform. (n.d.). <https://platform.stability.ai/docs/getting-started>

Stability AI - Developer Platform. (n.d.). <https://platform.stability.ai/support>

prompthero/openjourney · Discussions. (n.d.). <https://huggingface.co/prompthero/openjourney/discussions>

playgroundai/playground-v2.5-1024px-aesthetic · Discussions. (n.d.). <https://huggingface.co/playgroundai/playground-v2.5-1024px-aesthetic/discussions>

Free AI image generator: Art, Social Media, Marketing | Playground. (n.d.). Playground. <https://playground.com/>

5.1 Data Privacy

Security. (n.d.). <https://openai.com/security>

Jemash. (n.d.). *General Data Protection Regulation (GDPR)*. Microsoft Learn. <https://learn.microsoft.com/en-us/legal/gdpr>

Privacy Policy | Lepton AI. (n.d.). <https://www.lepton.ai/policies/privacy>

Groq, Inc. (2024, February 6). *Privacy Policy - Groq*. Groq. <https://wow.groq.com/privacy-policy/>

Privacy Notice. (n.d.). <https://fireworks.ai/privacy-policy>

DeepInfra Privacy Policy. (n.d.). <https://deepinfra.com/privacy>

Privacy Notice | Anyscale. (n.d.). Anyscale. <https://www.anyscale.com/privacy-policy>

Legal. (n.d.). <https://www.together.ai/privacy>

Build Generative AI Applications with Foundation Models - Amazon Bedrock FAQs - AWS. (n.d.). Amazon Web Services, Inc. <https://aws.amazon.com/bedrock/faqs/>

Ai, M. (n.d.). *Legal terms and conditions*. Mistral AI | Frontier AI in Your Hands. <https://mistral.ai/terms/>

Gemini API Additional Terms of Service | Google AI for Developers. (n.d.). Google AI for Developers. <https://ai.google.dev/terms#content-license>

Data Usage Policy. (n.d.). Cohere. <https://cohere.com/data-usage-policy>

What data does Perplexity collect about me? (n.d.). <https://www.perplexity.ai/hub/faq/what-data-does-perplexity-collect-about-me>

Privacy Policy — Stability AI. (n.d.). Stability AI. <https://stability.ai/privacy-policy>

Security. (n.d.). <https://huggingface.co/docs/hub/security>

Privacy policy – Replicate. (n.d.). <https://replicate.com/privacy>

5.2 Security

Security. (n.d.). <https://openai.com/security>

Privacy Policy | Lepton AI. (n.d.). <https://www.lepton.ai/policies/privacy>

Groq, Inc. (2024, February 6). *Privacy Policy - Groq*. Groq. <https://wow.groq.com/privacy-policy/>

Privacy Notice. (n.d.). <https://fireworks.ai/privacy-policy>

DeepInfra Privacy Policy. (n.d.). <https://deepinfra.com/privacy>

Privacy Notice | Anyscale. (n.d.). Anyscale. <https://www.anyscale.com/privacy-policy>

Legal. (n.d.). <https://www.together.ai/privacy>

Build Generative AI Applications with Foundation Models - Amazon Bedrock FAQs - AWS. (n.d.). Amazon Web Services, Inc. <https://aws.amazon.com/bedrock/faqs/>

Ai, M. (n.d.). *Legal terms and conditions*. Mistral AI | Frontier AI in Your Hands. <https://mistral.ai/terms/>

Cloud compliance & Regulations resources. (n.d.). Google Cloud. <https://cloud.google.com/compliance?hl=en>

Data Usage Policy. (n.d.). Cohere. <https://cohere.com/data-usage-policy>

What data does Perplexity collect about me? (n.d.). <https://www.perplexity.ai/hub/faq/what-data-does-perplexity-collect-about-me>

Privacy Policy — Stability AI. (n.d.). Stability AI. <https://stability.ai/privacy-policy>

Compliance Programs - Amazon Web Services (AWS). (n.d.). Amazon Web Services, Inc. <https://aws.amazon.com/compliance/programs/>

TerryLanfear. (n.d.). *Azure compliance documentation*. Microsoft Learn. <https://learn.microsoft.com/en-us/azure/compliance/>

Security. (n.d.). Cohere. <https://cohere.com/security>

Security. (n.d.). <https://huggingface.co/docs/hub/security>

5.3 Licensing

Terms of use. (n.d.). <https://openai.com/policies/terms-of-use>

Download Llama. (n.d.). *Llama.* <https://llama.meta.com/llama-downloads/>

Mistral. (n.d.). https://huggingface.co/docs/transformers/model_doc/mistral

Banks, J. (2023, December 18). It's time for developers and enterprises to build with Gemini Pro. Google. <https://blog.google/technology/ai/gemini-api-developers-cloud/>

Commercial SAAS Agreement. (n.d.). *Cohere.* <https://cohere.com/saas-agreement>

Frequently asked questions. (n.d.). *Pplx-api.* <https://docs.perplexity.ai/page/frequently-asked-questions>

Can I sell images I create with DALL·E? | OpenAI Help Center. (n.d.). <https://help.openai.com/en/articles/6425277-can-i-sell-images-i-create-with-dall-e>

Mostaque, E. (2023, November 8). Stable diffusion public release — stability AI. *Stability AI.* <https://stability.ai/news/stable-diffusion-public-release>

prompthero/openjourney · *Hugging Face.* (n.d.). <https://huggingface.co/prompthero/openjourney>

Pricing - playground. (n.d.). *Playground.* <https://playground.com/pricing>