day5

January 6, 2025

1 A full business solution

1.1 Now we will take our project from Day 1 to the next level

1.1.1 BUSINESS CHALLENGE:

Create a product that builds a Brochure for a company to be used for prospective clients, investors and potential recruits.

We will be provided a company name and their primary website.

See the end of this notebook for examples of real-world business applications.

And remember: I'm always available if you have problems or ideas! Please do reach out.

[1]:
```
# imports
# If these fail, please check you're running from an 'activated' environment␣
 ↪with (llms) in the command prompt

import os
import requests
import json
from typing import List
from dotenv import load_dotenv
from bs4 import BeautifulSoup
from IPython.display import Markdown, display, update_display
from openai import OpenAI
```

[2]:
```
# Initialize and constants

load_dotenv(override=True)
api_key = os.getenv('OPENAI_API_KEY')

if api_key and api_key.startswith('sk-proj-') and len(api_key)>10:
    print("API key looks good so far")
else:
    print("There might be a problem with your API key? Please visit the␣
 ↪troubleshooting notebook!")

MODEL = 'gpt-4o-mini'
```

```
openai = OpenAI()
```

API key looks good so far

[3]:
```
# A class to represent a Webpage

# Some websites need you to use proper headers when fetching them:
headers = {
 "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36␣
 ↪(KHTML, like Gecko) Chrome/117.0.0.0 Safari/537.36"
}

class Website:
    """
    A utility class to represent a Website that we have scraped, now with links
    """

    def __init__(self, url):
        self.url = url
        response = requests.get(url, headers=headers)
        self.body = response.content
        soup = BeautifulSoup(self.body, 'html.parser')
        self.title = soup.title.string if soup.title else "No title found"
        if soup.body:
            for irrelevant in soup.body(["script", "style", "img", "input"]):
                irrelevant.decompose()
            self.text = soup.body.get_text(separator="\n", strip=True)
        else:
            self.text = ""
        links = [link.get('href') for link in soup.find_all('a')]
        self.links = [link for link in links if link]

    def get_contents(self):
        return f"Webpage Title:\n{self.title}\nWebpage Contents:\n{self.
 ↪text}\n\n"
```

[4]:
```
ed = Website("https://edwarddonner.com")
ed.links
```

[4]:
```
['https://edwarddonner.com/',
 'https://edwarddonner.com/outsmart/',
 'https://edwarddonner.com/about-me-and-about-nebula/',
 'https://edwarddonner.com/posts/',
 'https://edwarddonner.com/',
 'https://news.ycombinator.com',
 'https://nebula.io/?utm_source=ed&utm_medium=referral',
 'https://www.prnewswire.com/news-releases/wynden-stark-group-acquires-nyc-
venture-backed-tech-startup-untapt-301269512.html',
```

```
 'https://patents.google.com/patent/US20210049536A1/',
 'https://www.linkedin.com/in/eddonner/',
 'https://edwarddonner.com/2024/12/21/llm-resources-superdatascience/',
 'https://edwarddonner.com/2024/12/21/llm-resources-superdatascience/',
 'https://edwarddonner.com/2024/11/13/llm-engineering-resources/',
 'https://edwarddonner.com/2024/11/13/llm-engineering-resources/',
 'https://edwarddonner.com/2024/10/16/from-software-engineer-to-ai-data-
scientist-resources/',
 'https://edwarddonner.com/2024/10/16/from-software-engineer-to-ai-data-
scientist-resources/',
 'https://edwarddonner.com/2024/08/06/outsmart/',
 'https://edwarddonner.com/2024/08/06/outsmart/',
 'https://edwarddonner.com/',
 'https://edwarddonner.com/outsmart/',
 'https://edwarddonner.com/about-me-and-about-nebula/',
 'https://edwarddonner.com/posts/',
 'mailto:hello@mygroovydomain.com',
 'https://www.linkedin.com/in/eddonner/',
 'https://twitter.com/edwarddonner',
 'https://www.facebook.com/edward.donner.52']
```

1.2 First step: Have GPT-4o-mini figure out which links are relevant

1.2.1 Use a call to gpt-4o-mini to read the links on a webpage, and respond in structured JSON.

It should decide which links are relevant, and replace relative links such as "/about" with "https://company.com/about".
We will use "one shot prompting" in which we provide an example of how it should respond in the prompt.

This is an excellent use case for an LLM, because it requires nuanced understanding. Imagine trying to code this without LLMs by parsing and analyzing the webpage - it would be very hard!

Sidenote: there is a more advanced technique called "Structured Outputs" in which we require the model to respond according to a spec. We cover this technique in Week 8 during our autonomous Agentic AI project.

[5]:
```
link_system_prompt = "You are provided with a list of links found on a webpage.␣
 ↪\
You are able to decide which of the links would be most relevant to include in␣
 ↪a brochure about the company, \
such as links to an About page, or a Company page, or Careers/Jobs pages.\n"
link_system_prompt += "You should respond in JSON as in this example:"
link_system_prompt += """
{
    "links": [
        {"type": "about page", "url": "https://full.url/goes/here/about"},
        {"type": "careers page": "url": "https://another.full.url/careers"}
```

```
    ]
}
"""
```

[6]:
```
print(link_system_prompt)
```

```
You are provided with a list of links found on a webpage. You are able to decide
which of the links would be most relevant to include in a brochure about the
company, such as links to an About page, or a Company page, or Careers/Jobs
pages.
You should respond in JSON as in this example:
{
    "links": [
        {"type": "about page", "url": "https://full.url/goes/here/about"},
        {"type": "careers page": "url": "https://another.full.url/careers"}
    ]
}
```

[7]:
```
def get_links_user_prompt(website):
    user_prompt = f"Here is the list of links on the website of {website.url} -␣
 ↪"
    user_prompt += "please decide which of these are relevant web links for a␣
 ↪brochure about the company, respond with the full https URL in JSON format. \
Do not include Terms of Service, Privacy, email links.\n"
    user_prompt += "Links (some might be relative links):\n"
    user_prompt += "\n".join(website.links)
    return user_prompt
```

[8]:
```
print(get_links_user_prompt(ed))
```

```
Here is the list of links on the website of https://edwarddonner.com - please
decide which of these are relevant web links for a brochure about the company,
respond with the full https URL in JSON format. Do not include Terms of Service,
Privacy, email links.
Links (some might be relative links):
https://edwarddonner.com/
https://edwarddonner.com/outsmart/
https://edwarddonner.com/about-me-and-about-nebula/
https://edwarddonner.com/posts/
https://edwarddonner.com/
https://news.ycombinator.com
https://nebula.io/?utm_source=ed&utm_medium=referral
https://www.prnewswire.com/news-releases/wynden-stark-group-acquires-nyc-
venture-backed-tech-startup-untapt-301269512.html
https://patents.google.com/patent/US20210049536A1/
https://www.linkedin.com/in/eddonner/
https://edwarddonner.com/2024/12/21/llm-resources-superdatascience/
```

```
https://edwarddonner.com/2024/12/21/llm-resources-superdatascience/
https://edwarddonner.com/2024/11/13/llm-engineering-resources/
https://edwarddonner.com/2024/11/13/llm-engineering-resources/
https://edwarddonner.com/2024/10/16/from-software-engineer-to-ai-data-scientist-
resources/
https://edwarddonner.com/2024/10/16/from-software-engineer-to-ai-data-scientist-
resources/
https://edwarddonner.com/2024/08/06/outsmart/
https://edwarddonner.com/2024/08/06/outsmart/
https://edwarddonner.com/
https://edwarddonner.com/outsmart/
https://edwarddonner.com/about-me-and-about-nebula/
https://edwarddonner.com/posts/
mailto:hello@mygroovydomain.com
https://www.linkedin.com/in/eddonner/
https://twitter.com/edwarddonner
https://www.facebook.com/edward.donner.52
```

[9]:
```python
def get_links(url):
    website = Website(url)
    response = openai.chat.completions.create(
        model=MODEL,
        messages=[
            {"role": "system", "content": link_system_prompt},
            {"role": "user", "content": get_links_user_prompt(website)}
      ],
        response_format={"type": "json_object"}
    )
    result = response.choices[0].message.content
    return json.loads(result)
```

[10]:
```python
# Anthropic has made their site harder to scrape, so I'm using HuggingFace..

huggingface = Website("https://huggingface.co")
huggingface.links
```

[10]:
```
['/',
 '/models',
 '/datasets',
 '/spaces',
 '/posts',
 '/docs',
 '/enterprise',
 '/pricing',
 '/login',
 '/join',
 '/blog/smolagents',
```

```
'/deepseek-ai/DeepSeek-V3',
'/PowerInfer/SmallThinker-3B-Preview',
'/deepseek-ai/DeepSeek-V3-Base',
'/black-forest-labs/FLUX.1-dev',
'/hexgrad/Kokoro-82M',
'/models',
'/spaces/osanseviero/gemini-coder',
'/spaces/JeffreyXiang/TRELLIS',
'/spaces/lllyasviel/iclight-v2',
'/spaces/Kwai-Kolors/Kolors-Virtual-Try-On',
'/spaces/akhaliq/anychat',
'/spaces',
'/datasets/agibot-world/AgiBotWorld-Alpha',
'/datasets/fka/awesome-chatgpt-prompts',
'/datasets/PowerInfer/QWQ-LONGCOT-500K',
'/datasets/cfahlgren1/react-code-instructions',
'/datasets/OpenLeecher/lmsys_chat_1m_clean',
'/datasets',
'/join',
'/pricing#endpoints',
'/pricing#spaces',
'/pricing',
'/enterprise',
'/enterprise',
'/enterprise',
'/enterprise',
'/enterprise',
'/enterprise',
'/enterprise',
'/allenai',
'/facebook',
'/amazon',
'/google',
'/Intel',
'/microsoft',
'/grammarly',
'/Writer',
'/docs/transformers',
'/docs/diffusers',
'/docs/safetensors',
'/docs/huggingface_hub',
'/docs/tokenizers',
'/docs/peft',
'/docs/transformers.js',
'/docs/timm',
'/docs/trl',
'/docs/datasets',
```

```
 '/docs/text-generation-inference',
 '/docs/accelerate',
 '/models',
 '/datasets',
 '/spaces',
 '/tasks',
 'https://ui.endpoints.huggingface.co',
 '/chat',
 '/huggingface',
 '/brand',
 '/terms-of-service',
 '/privacy',
 'https://apply.workable.com/huggingface/',
 'mailto:press@huggingface.co',
 '/learn',
 '/docs',
 '/blog',
 'https://discuss.huggingface.co',
 'https://status.huggingface.co/',
 'https://github.com/huggingface',
 'https://twitter.com/huggingface',
 'https://www.linkedin.com/company/huggingface/',
 '/join/discord']
```

[11]:
```
get_links("https://huggingface.co")
```

[11]:
```
{'links': [{'type': 'about page', 'url': 'https://huggingface.co/huggingface'},
  {'type': 'careers page', 'url': 'https://apply.workable.com/huggingface/'},
  {'type': 'enterprise page', 'url': 'https://huggingface.co/enterprise'},
  {'type': 'blog page', 'url': 'https://huggingface.co/blog'},
  {'type': 'models page', 'url': 'https://huggingface.co/models'},
  {'type': 'datasets page', 'url': 'https://huggingface.co/datasets'},
  {'type': 'spaces page', 'url': 'https://huggingface.co/spaces'},
  {'type': 'contact page', 'url': 'https://discuss.huggingface.co'},
  {'type': 'GitHub page', 'url': 'https://github.com/huggingface'},
  {'type': 'LinkedIn page',
   'url': 'https://www.linkedin.com/company/huggingface/'}]}
```

1.3 Second step: make the brochure!

Assemble all the details into another prompt to GPT4-o

[12]:
```
def get_all_details(url):
    result = "Landing page:\n"
    result += Website(url).get_contents()
    links = get_links(url)
    print("Found links:", links)
    for link in links["links"]:
```

```
        result += f"\n\n{link['type']}\n"
        result += Website(link["url"]).get_contents()
    return result
```

[13]:
```
print(get_all_details("https://huggingface.co"))
```

```
Found links: {'links': [{'type': 'about page', 'url':
'https://huggingface.co/huggingface'}, {'type': 'careers page', 'url':
'https://apply.workable.com/huggingface/'}, {'type': 'blog', 'url':
'https://huggingface.co/blog'}, {'type': 'enterprise page', 'url':
'https://huggingface.co/enterprise'}, {'type': 'pricing page', 'url':
'https://huggingface.co/pricing'}]}
Landing page:
Webpage Title:
Hugging Face - The AI community building the future.
Webpage Contents:
Hugging Face
Models
Datasets
Spaces
Posts
Docs
Enterprise
Pricing
Log In
Sign Up
NEW
smolagents - a smol library to build great agents
Use models from the HF Hub in LM Studio
Use Ollama with GGUF Models from the HF Hub
The AI community building the future.
The platform where the machine learning community collaborates on models,
datasets, and applications.
Trending on
this week
Models
deepseek-ai/DeepSeek-V3
Updated
8 days ago
•
71.7k
•
1.32k
PowerInfer/SmallThinker-3B-Preview
Updated
about 11 hours ago
•
6.37k
```

•
262
deepseek-ai/DeepSeek-V3-Base
Updated
8 days ago
•
8.36k
•
1.16k
black-forest-labs/FLUX.1-dev
Updated
Aug 16, 2024
•
1.17M
•
7.77k
hexgrad/Kokoro-82M
Updated
12 minutes ago
•
986
•
241
Browse 400k+ models
Spaces
Running
534

Gemini Coder
Running
on
Zero
2.59k

TRELLIS
Scalable and Versatile 3D Generation from images
Running
on
Zero
1.33k

IC Light V2
Running
on
CPU Upgrade
6.57k

Kolors Virtual Try-On

Running
on
CPU Upgrade
1.2k

Anychat
Browse 150k+ applications
Datasets
agibot-world/AgiBotWorld-Alpha
Updated
4 days ago
•
6.75k
•
144
fka/awesome-chatgpt-prompts
Updated
about 21 hours ago
•
5.63k
•
6.73k
PowerInfer/QWQ-LONGCOT-500K
Updated
11 days ago
•
437
•
68
cfahlgren1/react-code-instructions
Updated
8 minutes ago
•
301
•
62
OpenLeecher/lmsys_chat_1m_clean
Updated
6 days ago
•
381
•
39
Browse 100k+ datasets
The Home of Machine Learning
Create, discover and collaborate on ML better.
The collaboration platform
Host and collaborate on unlimited public models, datasets and applications.

Move faster
With the HF Open source stack.
Explore all modalities
Text, image, video, audio or even 3D.
Build your portfolio
Share your work with the world and build your ML profile.
Sign Up
Accelerate your ML
We provide paid Compute and Enterprise solutions.
Compute
Deploy on optimized
Inference Endpoints
or update your
Spaces applications
to a GPU in a few clicks.
View pricing
Starting at $0.60/hour for GPU
Enterprise
Give your team the most advanced platform to build AI with enterprise-grade
security, access controls and
dedicated support.
Getting started
Starting at $20/user/month
Single Sign-On
Regions
Priority Support
Audit Logs
Resource Groups
Private Datasets Viewer
More than 50,000 organizations are using Hugging Face
Ai2
Enterprise
non-profit
•
378 models
•
1.84k followers
AI at Meta
Enterprise
company
•
2.06k models
•
4.03k followers
Amazon Web Services
company
•
21 models

•
2.5k followers
Google
company
•
913 models
•
6.25k followers
Intel
company
•
217 models
•
2.1k followers
Microsoft
company
•
351 models
•
6.56k followers
Grammarly
company
•
10 models
•
106 followers
Writer
Enterprise
company
•
17 models
•
191 followers
Our Open Source
We are building the foundation of ML tooling with the community.
Transformers
137,230
State-of-the-art ML for Pytorch, TensorFlow, and JAX.
Diffusers
26,987
State-of-the-art diffusion models for image and audio generation in PyTorch.
Safetensors
2,983
Simple, safe way to store and distribute neural networks weights safely and quickly.
Hub Python Library
2,205
Client library for the HF Hub: manage repositories from your Python runtime.

Tokenizers
9,227
Fast tokenizers, optimized for both research and production.
PEFT
16,887
Parameter efficient finetuning methods for large models.
Transformers.js
12,520
State-of-the-art Machine Learning for the web. Run Transformers directly in your browser, with no need for a server.
timm
32,811
State-of-the-art computer vision models, layers, optimizers, training/evaluation, and utilities.
TRL
10,514
Train transformer language models with reinforcement learning.
Datasets
19,436
Access and share datasets for computer vision, audio, and NLP tasks.
Text Generation Inference
9,560
Toolkit to serve Large Language Models.
Accelerate
8,134
Easily train and use PyTorch models with multi-GPU, TPU, mixed-precision.
System theme
Website
Models
Datasets
Spaces
Tasks
Inference Endpoints
HuggingChat
Company
About
Brand assets
Terms of service
Privacy
Jobs
Press
Resources
Learn
Documentation
Blog
Forum
Service Status
Social

GitHub
Twitter
LinkedIn
Discord

about page
Webpage Title:
huggingface (Hugging Face)
Webpage Contents:
Hugging Face
Models
Datasets
Spaces
Posts
Docs
Enterprise
Pricing
Log In
Sign Up
Hugging Face
Enterprise
company
Verified
https://huggingface.co
huggingface
huggingface
Activity Feed
Follow
10,700
AI & ML interests
The AI community building the future.
Recent Activity
nielsr
updated
a dataset
about 10 hours ago
huggingface/community-science-merged
FL33TW00D-HF
new
activity
about 11 hours ago
huggingface/documentation-images:
pos-enc-update-v2
FL33TW00D-HF
updated
a dataset

about 11 hours ago
huggingface/documentation-images
View all activity
Team members
223
+189
+176
+155
+145
+125
Organization Card
Community
About org cards
Hi!
We are on a mission to democratize
good
machine learning, one commit at a time.
If that sounds like something you should be doing, why don't you
join us
!
For press enquiries, you can
contact our team here
.
Collections
1
DistilBERT release
Original DistilBERT model, checkpoints obtained from using teacher-student
learning from the original BERT checkpoints.
distilbert/distilbert-base-cased
Fill-Mask
•
Updated
May 6, 2024
•
277k
•
36
distilbert/distilbert-base-uncased
Fill-Mask
•
Updated
May 6, 2024
•
10.8M
•
•
593
distilbert/distilbert-base-multilingual-cased

Fill-Mask
•
Updated
May 6, 2024
•
374k
•
156
distilbert/distilbert-base-uncased-finetuned-sst-2-english
Text Classification
•
Updated
Dec 19, 2023
•
6.52M
•
•
660
spaces
23
Sort:
Recently updated
pinned
Running
55

Number Tokenization Blog
Running
497

Open Source Ai Year In Review 2024
What happened in open-source AI this year, and what’s next?
Running
44

Inference Playground
Running
203

paper-central
Running
on
TPU v5e
7

Keras Chatbot Battle
Running
101

Modelcard Creator
Expand 23
spaces
models
18
Sort:
Recently updated
huggingface/test-gating-group-2
Updated
20 days ago
huggingface/test-gating-group-1
Updated
20 days ago
huggingface/timesfm-tourism-monthly
Updated
28 days ago
•
34
•
1
huggingface/CodeBERTa-language-id
Text Classification
•
Updated
Mar 29, 2024
•
1.89k
•
54
huggingface/falcon-40b-gptq
Text Generation
•
Updated
Jun 14, 2023
•
19
•
12
huggingface/autoformer-tourism-monthly
Updated
May 24, 2023
•
1.19k
•
9
huggingface/distilbert-base-uncased-finetuned-mnli
Text Classification

•
Updated
Mar 22, 2023
•
679
•
2
huggingface/informer-tourism-monthly
Updated
Feb 24, 2023
•
1.05k
•
5
huggingface/time-series-transformer-tourism-monthly
Updated
Feb 23, 2023
•
1.52k
•
18
huggingface/the-no-branch-repo
Text-to-Image
•
Updated
Feb 10, 2023
•
70
•
3
Expand 18
models
datasets
32
Sort:
Recently updated
huggingface/community-science-merged
Viewer
•
Updated
about 10 hours ago
•
5.14k
•
171
•
1
huggingface/documentation-images

Viewer
•
Updated
about 11 hours ago
•
50
•
2.15M
•
45
huggingface/paper-central-data
Viewer
•
Updated
3 days ago
•
124k
•
664
•
8
huggingface/transformers-metadata
Viewer
•
Updated
13 days ago
•
1.52k
•
304
•
15
huggingface/community-science-paper-v2
Viewer
•
Updated
15 days ago
•
5.03k
•
282
•
7
huggingface/diffusers-metadata
Viewer
•
Updated
19 days ago

•
62
•
384
•
4
huggingface/policy-docs
Updated
19 days ago
•
891
•
6
huggingface/my-distiset-3f5a230e
Updated
Nov 21, 2024
•
9
huggingface/cookbook-images
Viewer
•
Updated
Nov 14, 2024
•
1
•
31.4k
•
6
huggingface/vllm-metadata
Updated
Oct 8, 2024
•
10
Expand 32
datasets
System theme
Company
TOS
Privacy
About
Jobs
Website
Models
Datasets
Spaces
Pricing
Docs

careers page
Webpage Title:
Hugging Face - Current Openings
Webpage Contents:

blog
Webpage Title:
Hugging Face - Blog
Webpage Contents:
Hugging Face
Models
Datasets
Spaces
Posts
Docs
Enterprise
Pricing
Log In
Sign Up
Blog, Articles, and discussions
New Article
Everything
community
guide
open source collab
partnerships
research
NLP
Audio
CV
RL
ethics
Diffusion
Game Development
RLHF
Leaderboard
Case Studies
Introducing smolagents: simple agents that write actions in code.
By
m-ric
December 31, 2024
•

231
Community Articles
view all
Fine-tune SmolLM's on custom synthetic data
By
prithivMLmods
•
1 day ago
•
10
Building Effective Agents with Anthropic’s Best Practices and smolagents
By
Sri-Vigneshwar-DJ
•
2 days ago
•
2
AI in 2025: A Combinatorial Explosion of Possibilities, but NOT AGI
By
Kseniase
•
2 days ago
•
1
Superposition in Transformers: A Novel Way of Building Mixture of Experts
By
BenChaliah
•
2 days ago
•
12
**Building a System That Can Build Systems: Toward a Self-Replicating Ecosystem
Framework**
By
adityagaharawar
•
3 days ago
•
1
Fine-tune a SmolLM on domain-specific synthetic data from a LLM
By
davidberenstein1957
•
3 days ago
•
10
ScreenSpot-Pro: GUI Grounding for Professional High-Resolution Computer Use
By

Ziyang
•
3 days ago
•
9
Process Reinforcement through Implicit Rewards
By
ganqu
•
4 days ago
•
9
LLM Comparison/Test: DeepSeek-V3, QVQ-72B-Preview, Falcon3 10B, Llama 3.3 70B, Nemotron 70B in my updated MMLU-Pro CS benchmark
By
wolfram
•
4 days ago
•
30
Finetuning Falcon 7b in a hybrid distributed fashion
By
Neo111x
•
7 days ago
•
4
Debate Championship for LLMs
By
as-cle-bert
•
7 days ago
•
4
Fine-tune ModernBERT for text classification using synthetic data
By
davidberenstein1957
•
7 days ago
•
19
#2: Your Go-To Vocabulary to Navigate the World of AI Agents and Agentic Workflows
By
Kseniase
•
10 days ago
•

8
Unlocking the Power of Reasoning: Introducing CriticalThinker-LLaMA-3.1-8B-GGUF and Its Groundbreaking Dataset
By
theeseus-ai
•
10 days ago
#1: Open-endedness and AI Agents - A Path from Generative to Creative AI?
By
Kseniase
•
12 days ago
•
5
Deriving DPO's Loss
By
hba123
•
13 days ago
•
24
#81: Key AI Concepts to Follow in 2025
By
Kseniase
•
14 days ago
•
19
Introducing KaibanJS v0.13.0: Structured Output for Smarter Workflows
By
darielnoel
•
14 days ago
•
1
FineWeb2-C: Help Build Better Language Models in Your Language
By
davanstrien
•
14 days ago
•
11
Tags generation dataset
By
zino36
•
17 days ago
•

3
Visualize and understand GPU memory in PyTorch
By
qgallouedec
December 24, 2024
•
129
Controlling Language Model Generation with NVIDIA's LogitsProcessorZoo
By
ariG23498
December 23, 2024
•
30
Evaluating Audio Reasoning with Big Bench Audio
By
mhillsmith
December 20, 2024
guest
•
15
Finally, a Replacement for BERT: Introducing ModernBERT
By
bwarner
December 19, 2024
guest
•
441
Bamba: Inference-Efficient Hybrid Mamba2 Model
By
Linsong-C
December 18, 2024
guest
•
37
Welcome the Falcon 3 Family of Open Models!
By
FalconLLM
December 17, 2024
•
109
Benchmarking Language Model Performance on 5th Gen Xeon at GCP
By
MatrixYao
December 17, 2024
•
3
Introducing the Synthetic Data Generator - Build Datasets with Natural Language
By

davidberenstein1957
December 16, 2024
•
72
LeMaterial: an open source initiative to accelerate materials discovery and research
By
AlexDuvalinho
December 10, 2024
guest
•
32
Hugging Face models in Amazon Bedrock
By
pagezyhf
December 9, 2024
•
10
Open Preference Dataset for Text-to-Image Generation by the Community
By
davidberenstein1957
December 9, 2024
•
49
Welcome PaliGemma 2 - New vision language models by Google
By
merve
December 5, 2024
•
123
"How good are LLMs at fixing their mistakes? A chatbot arena experiment with Keras and TPUs
By
martin-gorner
December 5, 2024
•
12
Rethinking LLM Evaluation with 3C3H: AraGen Benchmark and Leaderboard
By
alielfilali01
December 4, 2024
guest
•
29
Previous
1
2
3

…
37
Next
Community Articles
view all
Fine-tune SmolLM's on custom synthetic data
By
prithivMLmods
•
1 day ago
•
10
Building Effective Agents with Anthropic’s Best Practices and smolagents
By
Sri-Vigneshwar-DJ
•
2 days ago
•
2
AI in 2025: A Combinatorial Explosion of Possibilities, but NOT AGI
By
Kseniase
•
2 days ago
•
1
Superposition in Transformers: A Novel Way of Building Mixture of Experts
By
BenChaliah
•
2 days ago
•
12
**Building a System That Can Build Systems: Toward a Self-Replicating Ecosystem
Framework**
By
adityagaharawar
•
3 days ago
•
1
Fine-tune a SmolLM on domain-specific synthetic data from a LLM
By
davidberenstein1957
•
3 days ago
•
10

ScreenSpot-Pro: GUI Grounding for Professional High-Resolution Computer Use
By
Ziyang
•
3 days ago
•
9
Process Reinforcement through Implicit Rewards
By
ganqu
•
4 days ago
•
9
LLM Comparison/Test: DeepSeek-V3, QVQ-72B-Preview, Falcon3 10B, Llama 3.3 70B, Nemotron 70B in my updated MMLU-Pro CS benchmark
By
wolfram
•
4 days ago
•
30
Finetuning Falcon 7b in a hybrid distributed fashion
By
Neo111x
•
7 days ago
•
4
Debate Championship for LLMs
By
as-cle-bert
•
7 days ago
•
4
Fine-tune ModernBERT for text classification using synthetic data
By
davidberenstein1957
•
7 days ago
•
19
#2: Your Go-To Vocabulary to Navigate the World of AI Agents and Agentic Workflows
By
Kseniase
•

10 days ago
•
8
Unlocking the Power of Reasoning: Introducing CriticalThinker-LLaMA-3.1-8B-GGUF and Its Groundbreaking Dataset
By
theeseus-ai
•
10 days ago
#1: Open-endedness and AI Agents - A Path from Generative to Creative AI?
By
Kseniase
•
12 days ago
•
5
Deriving DPO's Loss
By
hba123
•
13 days ago
•
24
#81: Key AI Concepts to Follow in 2025
By
Kseniase
•
14 days ago
•
19
Introducing KaibanJS v0.13.0: Structured Output for Smarter Workflows
By
darielnoel
•
14 days ago
•
1
FineWeb2-C: Help Build Better Language Models in Your Language
By
davanstrien
•
14 days ago
•
11
Tags generation dataset
By
zino36
•

17 days ago
•
3
System theme
Company
TOS
Privacy
About
Jobs
Website
Models
Datasets
Spaces
Pricing
Docs

enterprise page
Webpage Title:
Enterprise Hub - Hugging Face
Webpage Contents:
Hugging Face
Models
Datasets
Spaces
Posts
Docs
Enterprise
Pricing
Log In
Sign Up
Enterprise Hub
Enterprise-ready version of the world's leading AI platform
Subscribe to
Enterprise Hub
for $20/user/month with your Hub organization
Give your organization the most advanced platform to build AI with enterprise-grade security, access controls,
dedicated support and more.
Single Sign-On
Connect securely to your identity provider with SSO integration.
Regions
Select, manage, and audit the location of your repository data.
Audit Logs
Stay in control with comprehensive logs that report on actions taken.
Resource Groups
Accurately manage access to repositories with granular access control.

Token Management
Centralized token control and custom approval policies for organization access.
Analytics
Track and analyze repository usage data in a single dashboard.
Advanced Compute Options
Increase scalability and performance with more compute options like ZeroGPU for Spaces.
Private Datasets Viewer
Enable the Dataset Viewer on your private datasets for easier collaboration.
Advanced security
Configure organization-wide security policies and default repository visibility.
Billing
Control your budget effectively with managed billing and yearly commit options.
Priority Support
Maximize your platform usage with priority support from the Hugging Face team.
Join the most forward-thinking AI organizations
Everything you already know and love about Hugging Face in Enterprise mode.
Subscribe to
Enterprise Hub
or
Talk to sales
NVIDIA
Enterprise
company
•
244 models
•
8.77k followers
Nerdy Face
Enterprise
company
•
1 model
•
269 followers
Cerebras
Enterprise
company
•
17 models
•
427 followers
Writer
Enterprise
company
•
17 models
•

191 followers
Fidelity Investments
Enterprise
company
•
118 followers
Jusbrasil
Enterprise
company
•
78 followers
Mistral AI_
Enterprise
company
•
21 models
•
3.73k followers
IBM Granite
Enterprise
company
•
70 models
•
777 followers
Virtusa Corporation
Enterprise
company
•
52 followers
creditkarma
Enterprise
company
•
36 followers
Ekimetrics
Enterprise
company
•
48 followers
Johnson & Johnson
Enterprise
company
•
37 followers
Widn AI
Enterprise
company

•
28 followers
Adyen
Enterprise
company
•
36 followers
Meta Llama
Enterprise
company
•
57 models
•
17.7k followers
Orange
Enterprise
company
•
4 models
•
152 followers
Arm
Enterprise
company
•
120 followers
Bloomberg
Enterprise
company
•
2 models
•
137 followers
Deutsche Telekom AG
Enterprise
company
•
7 models
•
114 followers
Lightricks
Enterprise
company
•
3 models
•
448 followers
Technology Innovation Institute

Enterprise
company
•
64 models
•
1.1k followers
Chegg Inc
Enterprise
company
•
78 followers
Liberty Mutual
Enterprise
company
•
42 followers
HiddenLayer
Enterprise
company
•
49 followers
Kakao Corp.
Enterprise
company
•
42 followers
Gsk-tech
Enterprise
company
•
33 followers
BCG X
Enterprise
company
•
29 followers
Qualcomm
Enterprise
company
•
153 models
•
378 followers
Shopify
Enterprise
company
•
377 followers

Together
Enterprise
company
•
28 models
•
474 followers
ServiceNow-AI
Enterprise
company
•
111 followers
Toyota Research Institute
Enterprise
company
•
8 models
•
93 followers
H2O.ai
Enterprise
company
•
71 models
•
365 followers
AI at Meta
Enterprise
company
•
2.06k models
•
4.03k followers
Aledade Inc
Enterprise
company
•
54 followers
Stability AI
Enterprise
company
•
96 models
•
10.8k followers
Nutanix
Enterprise
company

•
246 models
•
40 followers
Arcee AI
Enterprise
company
•
130 models
•
296 followers
Gretel.ai
Enterprise
company
•
8 models
•
76 followers
Mercedes-Benz AG
Enterprise
company
•
84 followers
HOVER External
Enterprise
company
•
26 followers
BRIA AI
Enterprise
company
•
28 models
•
1.51k followers
Compliance & Certifications
GDPR Compliant
SOC 2 Type 2
System theme
Website
Models
Datasets
Spaces
Tasks
Inference Endpoints
HuggingChat
Company
About

Brand assets
Terms of service
Privacy
Jobs
Press
Resources
Learn
Documentation
Blog
Forum
Service Status
Social
GitHub
Twitter
LinkedIn
Discord

pricing page
Webpage Title:
Hugging Face - Pricing
Webpage Contents:
Hugging Face
Models
Datasets
Spaces
Posts
Docs
Enterprise
Pricing
Log In
Sign Up
Pricing
Leveling up AI collaboration and compute.
Users and organizations already use the Hub as a collaboration platform,
we're making it easy to seamlessly and scalably launch ML compute directly from
the Hub.
HF Hub
Collaborate on Machine Learning
Host unlimited public models, datasets
Create unlimited orgs with no member limits
Access the latest ML tools and open source
Community support
Forever
Free
PRO
Pro Account

Unlock advanced HF features
ZeroGPU and Dev Mode for Spaces
Higher rate limits for serverless inference
Get early access to upcoming features
Show your support with a Pro badge
Subscribe for
$9
/month
Enterprise Hub
Accelerate your AI roadmap
SSO and SAML support
Select data location with Storage Regions
Precise actions reviews with Audit logs
Granular access control with Resource groups
Centralized token control and approval
Dataset Viewer for private datasets
Advanced compute options for Spaces
Deploy Inference on your own Infra
Managed billing with yearly commits
Priority support
Starting at
$20
per user per month
Spaces Hardware
Upgrade your Space compute
Free CPUs
Build more advanced Spaces
7 optimized hardware available
From CPU to GPU to Accelerators
Starting at
$0
/hour
Inference Endpoints
Deploy models on fully managed infrastructure
Deploy dedicated Endpoints in seconds
Keep your costs low
Fully-managed autoscaling
Enterprise security
Starting at
$0.032
/hour
Need support to accelerate AI in your organization? View our
Expert Support
.
Hugging Face Hub
free
The HF Hub is the central place to explore, experiment, collaborate and build
technology with Machine

Learning.
Join the open source Machine Learning movement!
→
Sign Up
Create with ML
Packed with ML features, like model eval, dataset viewer and much more.
Collaborate
Git based and designed for collaboration at its core.
Play and learn
Learn by experimenting and sharing with our awesome community.
Build your ML portfolio
Share your work with the world and build your own ML profile.
Spaces Hardware
Starting at $0
Spaces are one of the most popular ways to share ML applications and demos with the world.
Upgrade your Spaces with our selection of custom on-demand hardware:
→
Get started with Spaces
Name
CPU
Memory
Accelerator
VRAM
Hourly price
CPU Basic
2 vCPU
16 GB
-
-
FREE
CPU Upgrade
8 vCPU
32 GB
-
-
$0.03
Nvidia T4 - small
4 vCPU
15 GB
Nvidia T4
16 GB
$0.40
Nvidia T4 - medium
8 vCPU
30 GB
Nvidia T4
16 GB

$0.60
1x Nvidia L4
8 vCPU
30 GB
Nvidia L4
24 GB
$0.80
4x Nvidia L4
48 vCPU
186 GB
Nvidia L4
96 GB
$3.80
1x Nvidia L40S
8 vCPU
62 GB
Nvidia L4
48 GB
$1.80
4x Nvidia L40S
48 vCPU
382 GB
Nvidia L4
192 GB
$8.30
8x Nvidia L40S
192 vCPU
1534 GB
Nvidia L4
384 GB
$23.50
Nvidia A10G - small
4 vCPU
15 GB
Nvidia A10G
24 GB
$1.00
Nvidia A10G - large
12 vCPU
46 GB
Nvidia A10G
24 GB
$1.50
2x Nvidia A10G - large
24 vCPU
92 GB
Nvidia A10G
48 GB

$3.00
4x Nvidia A10G - large
48 vCPU
184 GB
Nvidia A10G
96 GB
$5.00
Nvidia A100 - large
12 vCPU
142 GB
Nvidia A100
80 GB
$4.00
TPU v5e 1x1
22 vCPU
44 GB
Google TPU v5e
16 GB
$1.20
TPU v5e 2x2
110 vCPU
186 GB
Google TPU v5e
64 GB
$4.75
TPU v5e 2x4
220 vCPU
380 GB
Google TPU v5e
128 GB
$9.50
Custom
on demand
on demand
on demand
on demand
on demand
Spaces Persistent Storage
All Spaces get ephemeral storage for free but you can upgrade and add persistent storage at any time.
Name
Storage
Monthly price
Small
20 GB
$5
Medium
150 GB

$25
Large
1 TB
$100
Building something cool as a side project? We also offer community GPU grants.
Inference Endpoints
Starting at $0.033/hour
Inference Endpoints (dedicated) offers a secure production solution to easily deploy any ML model on dedicated
and autoscaling infrastructure, right from the HF Hub.
→
Learn more
CPU
instances
Provider
Architecture
vCPUs
Memory
Hourly rate
aws
Intel Sapphire Rapids
1
2GB
$0.03
2
4GB
$0.07
4
8GB
$0.13
8
16GB
$0.27
azure
Intel Xeon
1
2GB
$0.06
2
4GB
$0.12
4
8GB
$0.24
8
16GB
$0.48

gcp
Intel Sapphire Rapids
1
2GB
$0.05
2
4GB
$0.10
4
8GB
$0.20
8
16GB
$0.40
Accelerator
instances
Provider
Architecture
Topology
Accelerator Memory
Hourly rate
aws
Inf2
Neuron
x1
14.5GB
$0.75
x12
760GB
$12.00
gcp
TPU
v5e
1x1
16GB
$1.20
2x2
64GB
$4.75
2x4
128GB
$9.50
GPU
instances
Provider
Architecture
GPUs
GPU Memory

Hourly rate
aws
NVIDIA T4
1
14GB
$0.50
4
56GB
$3.00
aws
NVIDIA L4
1
24GB
$0.80
4
96GB
$3.80
aws
NVIDIA L40S
1
48GB
$1.80
4
192GB
$8.30
8
384GB
$23.50
aws
NVIDIA A10G
1
24GB
$1.00
4
96GB
$5.00
aws
NVIDIA A100
1
80GB
$4.00
2
160GB
$8.00
4
320GB
$16.00
8

640GB
$32.00
gcp
NVIDIA T4
1
16GB
$0.50
gcp
NVIDIA L4
1
24GB
$0.70
4
96GB
$3.80
gcp
NVIDIA A100
1
80GB
$3.60
2
160GB
$7.20
4
320GB
$14.40
8
640GB
$28.80
gcp
NVIDIA H100
1
80GB
$10.00
2
160GB
$20.00
4
320GB
$40.00
8
640GB
$80.00
Pro Account
PRO
A monthly subscription to access powerful features.
→
Get Pro

($9/month)
ZeroGPU
: Get 5x usage quota and highest GPU queue priority
Spaces Hosting
: Create ZeroGPU Spaces with A100 hardware
Spaces Dev Mode
: Fast iterations via SSH/VS Code for Spaces
Dataset Viewer
: Activate it on private datasets
Inference API
: Get x20 higher rate limits on Serverless API
Blog Articles
: Publish articles to the Hugging Face blog
Social Posts
: Share short updates with the community
Features Preview
: Get early access to upcoming
features
PRO
Badge
:
Show your support on your profile
System theme
Website
Models
Datasets
Spaces
Tasks
Inference Endpoints
HuggingChat
Company
About
Brand assets
Terms of service
Privacy
Jobs
Press
Resources
Learn
Documentation
Blog
Forum
Service Status
Social
GitHub
Twitter
LinkedIn
Discord

[25]:
```
system_prompt = "You are an assistant that analyzes the contents of several␣
 ↪relevant pages from a company website \
and creates a short brochure about the company for prospective customers,␣
 ↪investors and recruits. Respond in markdown.\
Include details of company culture, customers and careers/jobs if you have the␣
 ↪information."

# Or uncomment the lines below for a more humorous brochure - this demonstrates␣
 ↪how easy it is to incorporate 'tone':

# system_prompt = "You are an assistant that analyzes the contents of several␣
 ↪relevant pages from a company website \
# and creates a short humorous, entertaining, jokey brochure about the company␣
 ↪for prospective customers, investors and recruits. Respond in markdown.\
# Include details of company culture, customers and careers/jobs if you have␣
 ↪the information."
```

[15]:
```
def get_brochure_user_prompt(company_name, url):
    user_prompt = f"You are looking at a company called: {company_name}\n"
    user_prompt += f"Here are the contents of its landing page and other␣
 ↪relevant pages; use this information to build a short brochure of the␣
 ↪company in markdown.\n"
    user_prompt += get_all_details(url)
    user_prompt = user_prompt[:5_000] # Truncate if more than 5,000 characters
    return user_prompt
```

[16]:
```
get_brochure_user_prompt("HuggingFace", "https://huggingface.co")
```

```
Found links: {'links': [{'type': 'about page', 'url':
'https://huggingface.co/about'}, {'type': 'careers page', 'url':
'https://apply.workable.com/huggingface/'}, {'type': 'enterprise page', 'url':
'https://huggingface.co/enterprise'}, {'type': 'pricing page', 'url':
'https://huggingface.co/pricing'}, {'type': 'blog page', 'url':
'https://huggingface.co/blog'}, {'type': 'contact page', 'url':
'https://huggingface.co/join'}]}
```

[16]:
```
"You are looking at a company called: HuggingFace\nHere are the contents of its
landing page and other relevant pages; use this information to build a short
brochure of the company in markdown.\nLanding page:\nWebpage Title:\nHugging
Face - The AI community building the future.\nWebpage Contents:\nHugging
Face\nModels\nDatasets\nSpaces\nPosts\nDocs\nEnterprise\nPricing\nLog In\nSign
Up\nNEW\nsmolagents - a smol library to build great agents\nUse models from the
HF Hub in LM Studio\nUse Ollama with GGUF Models from the HF Hub\nThe AI
community building the future.\nThe platform where the machine learning
community collaborates on models, datasets, and applications.\nTrending on\nthis
```

week\nModels\ndeepseek-ai/DeepSeek-V3\nUpdated\n8 days ago\n•\n71.7k\n•\n1.32k\nPowerInfer/SmallThinker-3B-Preview\nUpdated\nabout 11 hours ago\n•\n6.37k\n•\n262\ndeepseek-ai/DeepSeek-V3-Base\nUpdated\n8 days ago\n•\n8.36k\n•\n1.16k\nblack-forest-labs/FLUX.1-dev\nUpdated\nAug 16, 2024\n•\n1.17M\n•\n7.77k\nhexgrad/Kokoro-82M\nUpdated\n12 minutes ago\n•\n986\n•\n241\nBrowse 400k+ models\nSpaces\nRunning\n534\n \nGemini Coder\nRunning\non\nZero\n2.59k\n \nTRELLIS\nScalable and Versatile 3D Generation from images\nRunning\non\nZero\n1.33k\n \nIC Light V2\nRunning\non\nCPU Upgrade\n6.57k\n \nKolors Virtual Try-On\nRunning\non\nCPU Upgrade\n1.2k\n \nAnychat\nBrowse 150k+ applications\nDatasets\nagibot-world/AgiBotWorld-Alpha\nUpdated\n4 days ago\n•\n6.75k\n•\n144\nfka/awesome-chatgpt-prompts\nUpdated\nabout 21 hours ago\n•\n5.63k\n•\n6.73k\nPowerInfer/QWQ-LONGCOT-500K\nUpdated\n11 days ago\n•\n437\n•\n68\ncfahlgren1/react-code-instructions\nUpdated\n8 minutes ago\n•\n301\n•\n62\nOpenLeecher/lmsys_chat_1m_clean\nUpdated\n6 days ago\n•\n381\n•\n39\nBrowse 100k+ datasets\nThe Home of Machine Learning\nCreate, discover and collaborate on ML better.\nThe collaboration platform\nHost and collaborate on unlimited public models, datasets and applications.\nMove faster\nWith the HF Open source stack.\nExplore all modalities\nText, image, video, audio or even 3D.\nBuild your portfolio\nShare your work with the world and build your ML profile.\nSign Up\nAccelerate your ML\nWe provide paid Compute and Enterprise solutions.\nCompute\nDeploy on optimized\nInference Endpoints\nor update your\nSpaces applications\nto a GPU in a few clicks.\nView pricing\nStarting at $0.60/hour for GPU\nEnterprise\nGive your team the most advanced platform to build AI with enterprise-grade security, access controls and\n\t\t\tdedicated support.\nGetting started\nStarting at $20/user/month\nSingle Sign-On\nRegions\nPriority Support\nAudit Logs\nResource Groups\nPrivate Datasets Viewer\nMore than 50,000 organizations are using Hugging Face\nAi2\nEnterprise\nnon-profit\n•\n378 models\n•\n1.84k followers\nAI at Meta\nEnterprise\ncompany\n•\n2.06k models\n•\n4.03k followers\nAmazon Web Services\ncompany\n•\n21 models\n•\n2.5k followers\nGoogle\ncompany\n•\n913 models\n•\n6.25k followers\nIntel\ncompany\n•\n217 models\n•\n2.1k followers\nMicrosoft\ncompany\n•\n351 models\n•\n6.56k followers\nGrammarly\ncompany\n•\n10 models\n•\n106 followers\nWriter\nEnterprise\ncompany\n•\n17 models\n•\n191 followers\nOur Open Source\nWe are building the foundation of ML tooling with the community.\nTransformers\n137,230\nState-of-the-art ML for Pytorch, TensorFlow, and JAX.\nDiffusers\n26,987\nState-of-the-art diffusion models for image and audio generation in PyTorch.\nSafetensors\n2,983\nSimple, safe way to store and distribute neural networks weights safely and quickly.\nHub Python Library\n2,205\nClient library for the HF Hub: manage repositories from your Python runtime.\nTokenizers\n9,227\nFast tokenizers, optimized for both research and production.\nPEFT\n16,887\nParameter efficient finetuning methods for large models.\nTransformers.js\n12,520\nState-of-the-art Machine Learning for the web. Run Transformers directly in your browser, with no need for a server.\ntimm\n32,811\nState-of-the-art computer vision models, layers, optimizers, training/evaluation, and utilities.\nTRL\n10,514\nTrain transformer

```
language models with reinforcement learning.\nDatasets\n19,436\nAccess and share
datasets for computer vision, audio, and NLP tasks.\nText Generation
Inference\n9,560\nToolkit to serve Large Language
Models.\nAccelerate\n8,134\nEasily train and use PyTorch models with multi-GPU,
TPU, mixed-precision.\nSystem
theme\nWebsite\nModels\nDatasets\nSpaces\nTasks\nInference
Endpoints\nHuggingChat\nCompany\nAbout\nBrand assets\nTerms of service\nPrivacy\
nJobs\nPress\nResources\nLearn\nDocumentation\nBlog\nForum\nService
Status\nSocial\nGitHub\nTwitter\nLinkedIn\nDiscord\n\n\n\nabout page\nWebpage
Title:\nabout (Sergei)\nWebpage Contents:\nHugging
Face\nModels\nDatasets\nSpaces\nPosts\nDocs\nEnterprise\nPricing\nLog In\nSign
Up\nSergei\nabout\nFollow\nRenumathi's profile picture\nKalaipriya's profile
picture\nselvivincent's profile picture\n3\n\t\t\t\t\tfollowers\n·\n0
following\nAI & ML interests\nNone yet\nOrganizations\nNone yet\nmodels\nNone
public yet\ndatasets\nNone public yet\nSystem theme\nCompany\nTOS\nPrivacy\nAbou
t\nJobs\nWebsite\nModels\nDatasets\nSpaces\nPricing\nDocs\n\n\n\ncareers
page\nWebpage Title:\nHugging Face - Current Openings\nWebpage
Contents:\n\n\n\n\nenterprise page\nWebpage Title:\nEnterprise Hub - Hugging F"
```

[17]:
```python
def create_brochure(company_name, url):
    response = openai.chat.completions.create(
        model=MODEL,
        messages=[
            {"role": "system", "content": system_prompt},
            {"role": "user", "content": get_brochure_user_prompt(company_name,␣
 ↪url)}
        ],
    )
    result = response.choices[0].message.content
    display(Markdown(result))
```

[18]:
```python
create_brochure("HuggingFace", "https://huggingface.com")
```

```
Found links: {'links': [{'type': 'about page', 'url':
'https://huggingface.com/#about'}, {'type': 'careers page', 'url':
'https://apply.workable.com/huggingface/'}, {'type': 'company page', 'url':
'https://huggingface.com/enterprise'}, {'type': 'pricing page', 'url':
'https://huggingface.com/pricing'}, {'type': 'blog page', 'url':
'https://huggingface.com/blog'}]}
```

Hugging Face: The AI Community Building the Future

Who We Are

Hugging Face is a collaborative platform designed for the machine learning community. Our missi

Our Offerings

- **Models**: Explore a vast repository of over **400,000 models** including state-of-the-art t
- **Datasets**: Access more than **100,000 datasets** tailored for tasks in natural language pr

```
- **Spaces**: Create and run applications effortlessly within our platform. Power your innovati
- **Enterprise Solutions**: We cater to over **50,000 organizations** globally, including tech

## Company Culture
At Hugging Face, we foster a culture of collaboration and innovation. Our team is passionate ab

## Our Customers
We proudly support a diverse array of clients, from non-profits to top-tier tech companies such
- **Google**
- **Microsoft**
- **Amazon Web Services**
- **Intel**
- **Grammarly**

No matter the sector, our solutions are tailored to empower users across the globe.

## Careers at Hugging Face
Join us in our journey to create groundbreaking AI technology! We are constantly on the lookout

### Current Openings
Visit our [Jobs Page](https://huggingface.co/jobs) for a list of available positions and oppor

## Join Our Community
- **Get Started**: Sign up at [Hugging Face](https://huggingface.co) to begin your AI journey.
- **Connect with Us**: Follow us on [Twitter](https://twitter.com/huggingface), [LinkedIn](htt

Together, let's build the future of AI!
```

This brochure presents an overview of Hugging Face, detailing who they are, their offerings, company culture, customers, and career opportunities. It aims to attract potential customers, investors, and recruits by emphasizing the collaborative nature and innovative spirit of the company.

1.4 Finally - a minor improvement

With a small adjustment, we can change this so that the results stream back from OpenAI, with the familiar typewriter animation

[19]:
```
def stream_brochure(company_name, url):
    stream = openai.chat.completions.create(
        model=MODEL,
        messages=[
            {"role": "system", "content": system_prompt},
            {"role": "user", "content": get_brochure_user_prompt(company_name,␣
 ↪url)}
          ],
        stream=True
    )
```

```
    response = ""
    display_handle = display(Markdown(""), display_id=True)
    for chunk in stream:
        response += chunk.choices[0].delta.content or ''
        response = response.replace("```","").replace("markdown", "")
        update_display(Markdown(response), display_id=display_handle.display_id)
```

[20]:
```
stream_brochure("HuggingFace", "https://huggingface.co")
```

```
Found links: {'links': [{'type': 'about page', 'url':
'https://huggingface.co/huggingface'}, {'type': 'careers page', 'url':
'https://apply.workable.com/huggingface/'}, {'type': 'pricing page', 'url':
'https://huggingface.co/pricing'}, {'type': 'enterprise page', 'url':
'https://huggingface.co/enterprise'}, {'type': 'docs page', 'url':
'https://huggingface.co/docs'}, {'type': 'blog page', 'url':
'https://huggingface.co/blog'}, {'type': 'community discussion', 'url':
'https://discuss.huggingface.co'}, {'type': 'GitHub page', 'url':
'https://github.com/huggingface'}, {'type': 'LinkedIn page', 'url':
'https://www.linkedin.com/company/huggingface/'}, {'type': 'Twitter page',
'url': 'https://twitter.com/huggingface'}]}
```

2 Hugging Face: The AI Community Building The Future

Welcome to **Hugging Face**, where we are dedicated to creating a collaborative platform for the machine learning community. Our mission is to foster innovation and accessibility in AI technologies, empowering individuals and organizations to harness the power of machine learning.

2.1 About Us

Hugging Face serves as a home for everything Machine Learning. With over **400,000 models** and **100,000 datasets**, our platform enables users to create, discover, and collaborate seamlessly on machine learning projects. We offer tools that facilitate the handling of various modalities including text, image, video, audio, and 3D.

2.1.1 Our Community

More than **50,000 organizations** currently rely on Hugging Face, including industry giants like:

- **Amazon Web Services**
- **Microsoft**
- **Google**
- **Intel**

Our users collaborate and exchange knowledge, fostering a vibrant community that thrives on shared learning and breakthrough ideas.

2.2 Company Culture

At Hugging Face, we believe in the power of community. We encourage our team and network of users to share insights, tools, and applications. Our culture revolves around inclusivity, support

for open-source initiatives, and continuous learning. We are dedicated to building an environment where innovation is nurtured and every voice is heard.

2.3 Our Products

- **Models:** Explore and utilize a vast library of machine learning models tailored for various tasks.
- **Datasets:** Access and share datasets for numerous applications and research needs.
- **Spaces:** Deploy and run applications easily in a collaborative environment.
- **Enterprise Solutions:** Tailored services for companies needing advanced security and dedicated support.

2.3.1 Pricing

We provide flexible pricing options starting from **$0.60/hour for GPU compute** services and **$20/user/month for our Enterprise solutions**.

2.4 Careers at Hugging Face

Join a team of passionate individuals moving the AI field forward! We are constantly looking for innovative thinkers, problem solvers, and team players who are ready to embark on a journey to shape the future of AI. Explore our job openings and discover the possibilities of a career at Hugging Face.

2.5 Connect with Us

Join our community on various platforms: - GitHub - Twitter - LinkedIn - Discord

We're excited to have you become a part of our journey in shaping the future of AI.

[23]:
```
# Try changing the system prompt to the humorous version when you make the␣
↪Brochure for Hugging Face:

stream_brochure("HuggingFace", "https://huggingface.co")
```

```
Found links: {'links': [{'type': 'about page', 'url':
'https://huggingface.co/huggingface'}, {'type': 'careers page', 'url':
'https://apply.workable.com/huggingface/'}, {'type': 'enterprise page', 'url':
'https://huggingface.co/enterprise'}, {'type': 'pricing page', 'url':
'https://huggingface.co/pricing'}, {'type': 'blog page', 'url':
'https://huggingface.co/blog'}, {'type': 'community page', 'url':
'https://discuss.huggingface.co'}, {'type': 'GitHub page', 'url':
'https://github.com/huggingface'}, {'type': 'Twitter page', 'url':
'https://twitter.com/huggingface'}, {'type': 'LinkedIn page', 'url':
'https://www.linkedin.com/company/huggingface/'}]}
```

3 Welcome to Hugging Face: The AI Playground!

3.1 Who Are We?

At Hugging Face, we're not just an AI company; we're the AI community building the future (and having a ton of fun while doing it)! Whether you're a budding machine learner, a dataset wizard, or a debugging ninja, we've got a place for you!

3.2 What Do We Do?

Forget about cars and rocket ships; we're all about **Models, Datasets, and Spaces** - sounds like a sci-fi movie, right? Here's what's trending this week:

- **DeepSeek-V3:** The model that seeks deep — like a philosopher in a coffee shop!
- **PowerInfer/SmallThinker-3B-Preview:** Small in name but mighty in intelligence!

With over **400k+ models** and **150k+ applications**, you're sure to find a gem!

3.3 Why Join Us?

- **Collaboration for the win!** Get your ML team together and host unlimited public models, datasets, and applications. Who needs a Classroom when you can create the coolest AI projects from anywhere?
- **Accelerate your career.** With our HF Open Source stack, you'll grow faster than a computer can run a 3D model!
- **Join 50,000+ organizations** that trust us, including AI at Meta, Google, and…hey, even Grammarly! Can't forget about those grammar heroes, right?

3.4 Company Culture

- **Work hard, code harder:** At Hugging Face, our idea of fun includes brainstorming AI solutions while sipping coffee or decoding complicated algorithms (or both!).
- **Diversity is our strength:** We welcome everyone to our quirky family—from PhD geniuses to AI newcomers. Who says you can't teach an old bot new tricks?

3.5 Careers: Be Part of Something Great!

- **We're hiring!** Want to help us build the future? Check out our jobs page! Whether you're into engineering or marketing, we're on the lookout for talent. Just be prepared for spontaneous discussions about neural networks and the latest memes from the AI community!

3.6 Join the Hugging Face Community!

Ready to embrace the future of AI? Sign up today and become part of the family that recognizes the power of hugging…well, virtually!

So don your best coding socks, grab your favorite snack, and let's build some AI magic together!

Get in Touch! - Website - GitHub - Twitter - Discord

Remember, at Hugging Face, no one gets left behind—except maybe for that old model you never used.

Business applications

In this exercise we extended the Day 1 code to make multiple LLM calls, and generate a document.

This is perhaps the first example of Agentic AI design patterns, as we combined multiple calls to LLMs. This will feature more in Week 2, and then we will return to Agentic AI in a big way in Week 8 when we build a fully autonomous Agent solution.

Generating content in this way is one of the very most common Use Cases. As with summarization, this can be applied to any business vertical. Write marketing content, generate a product tutorial from a spec, create personalized email content, and so much more. Explore how you can apply content generation to your business, and try making yourself a proof-of-concept prototype.

Before you move to Week 2 (which is tons of fun)

Please see the week1 EXERCISE notebook for your challenge for the end of week 1. This will give you some essential practice working with Frontier APIs, and prepare you well for Week 2.

A reminder on 2 useful resources

1. The resources for the course are available here. 2. I'm on LinkedIn here and I love connecting with people taking the course!

Finally! I have a special request for you

My editor tells me that it makes a MASSIVE difference when students rate this course on Udemy - it's one of the main ways that Udemy decides whether to show it to others. If you're able to take a minute to rate this, I'd be so very grateful! And regardless - always please reach out to me at ed@edwarddonner.com if I can help at any point.

[ ]: