



Hugging Face

Hugging Face 勉強会

この資料は下記GitHubの一部を再利用して作っています。

<https://github.com/nlp-with-transformers/notebooks>

是非ともこの本を手にとって勉強してください

目次

- ・ 自己紹介
- ・ 勉強会の目的
- ・ Hugging Faceを使おうと思ったわけ
- ・ Hugging Faceとは
- ・ Hugging Faceサイトの構成と見方
- ・ デモンストレーション
 - ・ デモンストレーションの前に
 - ・ Google Colabで使ってみる
- ・ UseCaseを考えてみよう
- ・ まとめ

本日の想定レベル：☆☆○○○

Python : ちょっとわかる

LLM : ちょっとわかる

英語 : ほとんどわかんない

自己紹介

- ・ザ・JTC企業で製造現場の効率化の支援
(システム導入/開発/AI活用)
- ・1974年生まれの50歳
- ・愛媛県今治市出身の名古屋市在住
- ・元自転車選手のフィドル (🎻) 弾き



Shinji Ikeda



Community of
Deep Learning Evangelists
CDLE

勉強会の私の目的

Hugging Faceを使う仲間作り

- なかなか日本語の資料がない
- 日本語の書籍が少ない
- Udemyにも講座がない
- あわよくばGENIACみたいなのに参加したい

[GENIAC 採択プロジェクト] 多様な日本語能力の向上を目指した公開の基盤モデル開発 | 東京大学松尾研究室 - Matsuo Lab (u-tokyo.ac.jp)



'23企業協働メンバーが二人参加！！(カワイイ)

Hugging Faceを使おうと思ったわけ

- Transformerの理解をしたくて買った本
→ 大半がTransformersの本だった。
- いろんなモデルがここ一箇所で見られる/試せるというより、
メモリーに乗り切らない学習データを扱えるところ

メモリーマッピング

- メモリーに学習データを読み込むのではなく、ディスク上のファイルを直接読み込むことでメモリーの節約を行う

ストリーミング

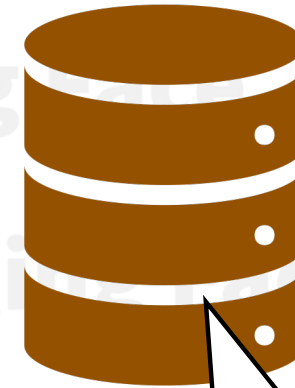
- 大量の学習データをローカルPCに置くのではなく、その時に必要な分だけをダウンロードしながら学習を進めていく



Hugging Faceとは 1



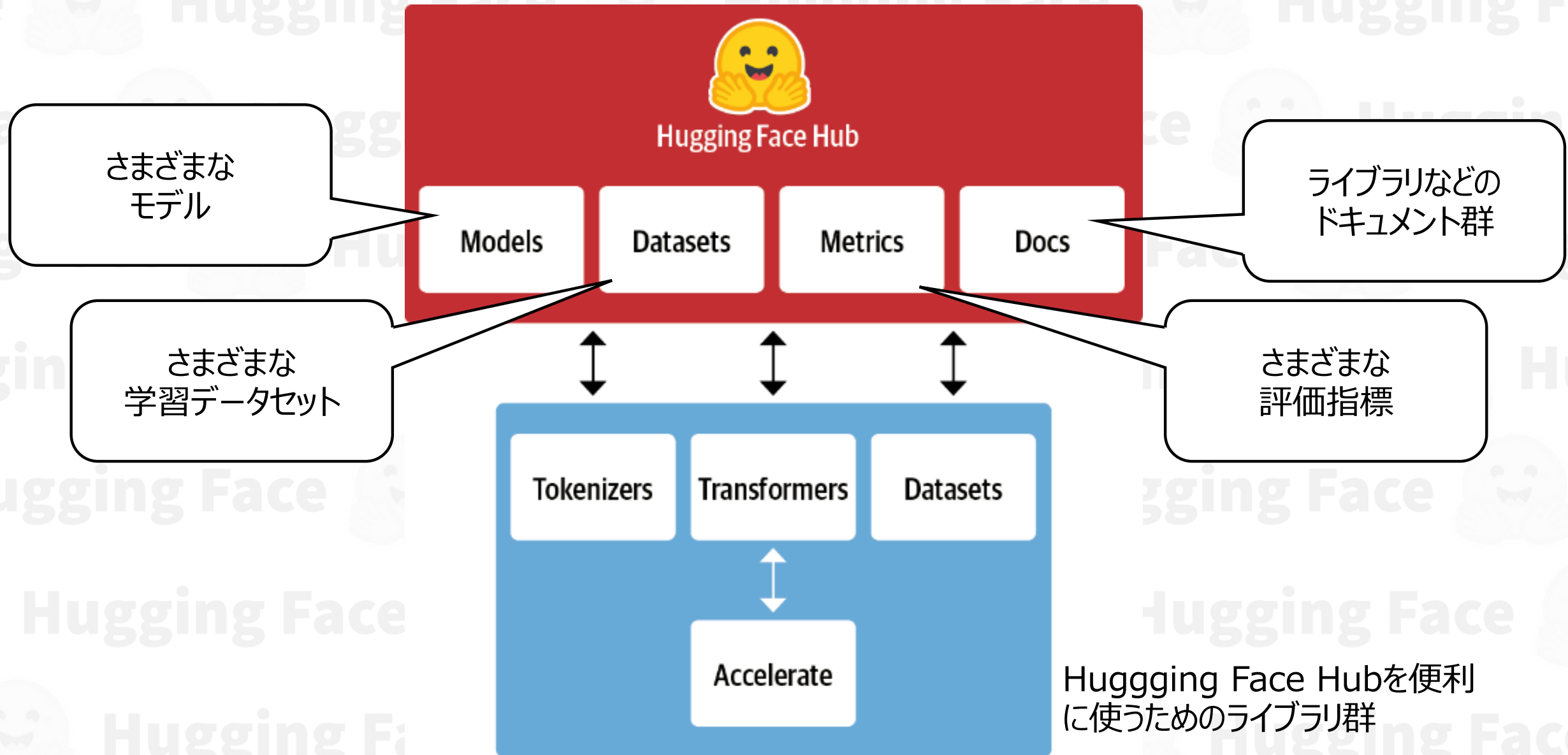
さまざまな
モデル



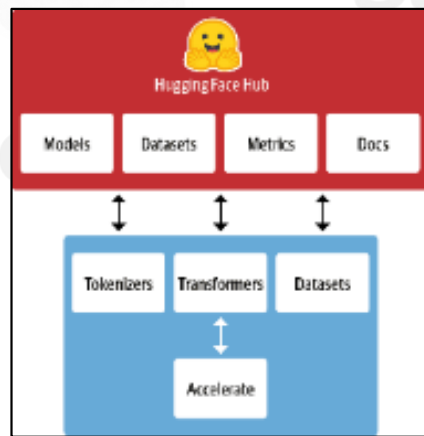
さまざまな
データセット

モデルやデータセット、使い方などを共有し、容易に活用
(AIモデルのGitHub、辞書みたいなもの)

Hugging Faceとは 2



Hugging Faceサイトの構成と見方



Organizations

- Create New

Resources

- Hub guide
- Transformers doc
- Forum
- Tasks
- Learn

Dark theme

Hugging Face Search models, datasets, users...

Models Datasets Spaces Posts Docs Solutions Pricing

Hugging Face is way more fun with friends and colleagues! [Join an organization](#)

Following

- ikedachin
- Profile
- Inbox (0)
- Settings
- Get Pro

Organizations

- Create New

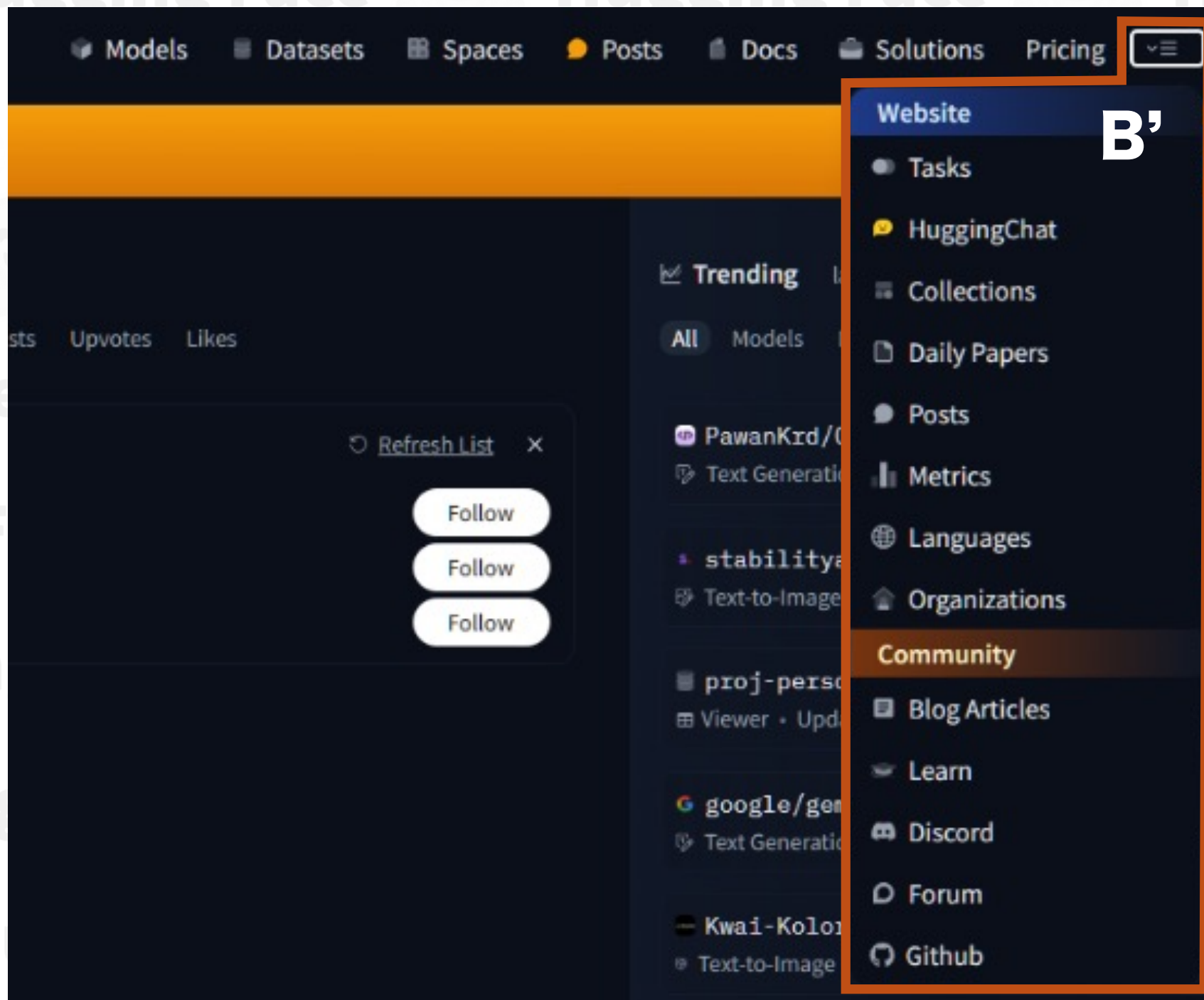
Resources

- Hub guide
- Transformers doc
- Forum
- Tasks
- Learn

Dark theme

Trending last 7 days

- PawanKrd/CosmosRP-8k
- stabilityai/stable-diffusion-3-medium
- proj-persona/PersonaHub
- google/gemma-2-9b
- meta-llama/Meta-Llama-3-8B
- BAAI/Infinity-Instruct
- Open LLM Leaderboard 2
- Salesforce/xlam-function-calling-60k



デモンストレーション

- 結局、使ってみた方が理解が早い！ w
- 「Textbooks Are All You Need」の論文で有名になったPhi-1の派生
“**Phi-1.5** (Microsoft)” を使ってみよう
- [論文はこちら](#)
- [HuggingfaceのPhi-1.5のページ](#)
- [Google Colab](#)

デモンストレーションの前に 1

Phi-1.5のページを見てみる

The image shows a screenshot of the Hugging Face website for the model `microsoft/phi-1_5`. The interface is dark-themed. At the top, there's a navigation bar with 'Model card', 'Files and versions', and 'Community 87'. Below this, a banner message says 'Hugging Face is way more fun with friends and colleagues!'. The main header shows the model name 'microsoft/phi-1_5' with a 'like' button and '1.3k' likes. Below the header, there's a row of tags: 'Text Generation', 'Transformers', 'Safetensors', 'English', 'phi', 'nlp', 'code', 'Inference Endpoints', 'text-generation-inference', 'arxiv:2309.05463', and 'License: mit'. The main content area is divided into two columns. The left column has a 'Model card' tab selected, showing a 'Model Summary' section. The right column has tabs for 'Train', 'Deploy', and 'Use this model'. The 'Model Summary' section contains text about the model's training and performance. The 'Train', 'Deploy', and 'Use this model' tabs are highlighted with a red box and a red circle with the number '2'. The 'Model Summary' section is highlighted with a red box and a red circle with the number '3'. The 'Model card' tab is highlighted with a red box and a red circle with the number '1'. The 'Train', 'Deploy', and 'Use this model' tabs are also highlighted with a red box and a red circle with the number '2'.

Model card Files and versions Community 87

Hugging Face Search models, datasets, users... Models Datasets Spaces Posts Docs Solutions Pricing

Hugging Face is way more fun with friends and colleagues! Join an organization

microsoft/phi-1_5 like 1.3k

Text Generation Transformers Safetensors English phi nlp code Inference Endpoints text-generation-inference arxiv:2309.05463 License: mit

Model card Files and versions Community 87

Model Summary

The language model Phi-1.5 is a Transformer with 1.3 billion parameters. It was trained using the same data sources as phi-1, augmented with a new data source that consists of various NLP synthetic texts. When assessed against benchmarks testing common sense, language understanding, and logical reasoning, Phi-1.5 demonstrates a nearly state-of-the-art performance among models with less than 10 billion parameters.

We did not fine-tune Phi-1.5 either for instruction following or through reinforcement learning

Downloads last month 85,902

Safetensors Model size 1.42B params Tensor type FP16

Inference API

Text Generation Examples

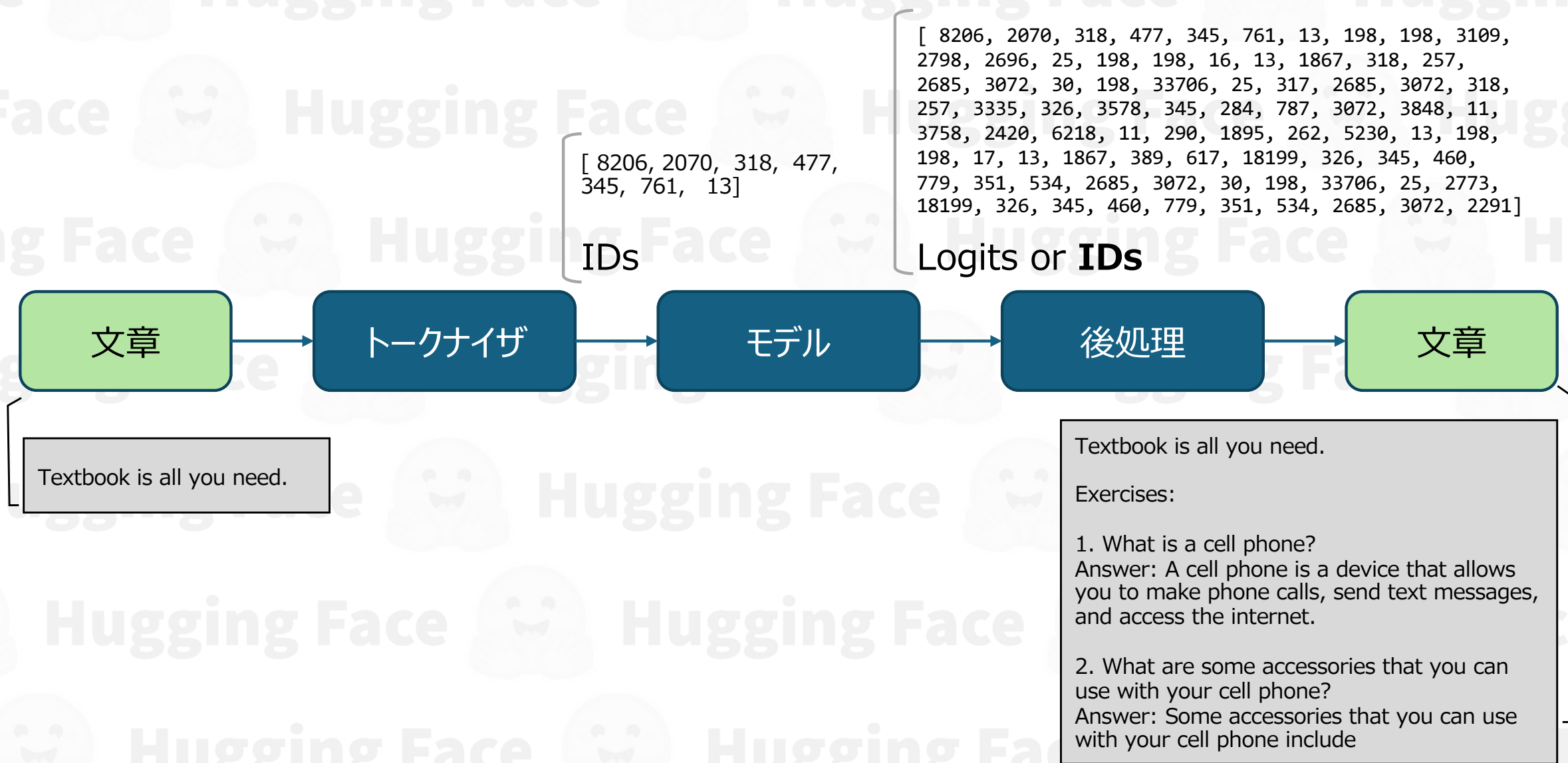
My name is Lewis and I like to

Train Deploy Use this model

https://huggingface.co/microsoft/phi-1_5

デモンストレーションの前に 2

LLMの処理をイメージする



GoogleColabで使ってみよう！

[Google Colab](#) ← click

[HuggingfaceのPhi-1.5のページ](#) ← click

UseCaseを考えてみよう

- ローカルLLMに活路？
 - 社内特化？特定の目的特化？

- 一般的でない専門用語

視差呼称、はつり、デレッキ、うま

逃げ、刺す、芋掘り、三味線...



まとめ

- ちょっと難しそうだけど、なんやら便利そう！
- 日本語の資料が少ないから、みんなで頑張ろうぜ！

みんなでHuggingFaceを使ってみよう！

