



Hugging Face

Hugging Face勉強会

この資料は下記GitHubの一部を再利用、改変して作っています。

<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

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/ザ・JTC)
- ・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



Community of
Deep Learning Evangelists
CDLE

勉強会の私の目的

HuggingFaceを使う仲間作り

- なかなか日本語の資料がない
- 日本語の書籍があまりない
- 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に向けた公開の基盤モデル開発 | 東京大学松尾研究室 - [Matcha-Lab](https://matcha-lab.github.io/matcha-tokyo.ac.jp/) (tokyo.ac.jp)

'23企業協働メンバーが二人参加！！(カヨイ)





← 素人 (えっ

Hugging Faceを使おうと思ったわけ

- 興味をもったのは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

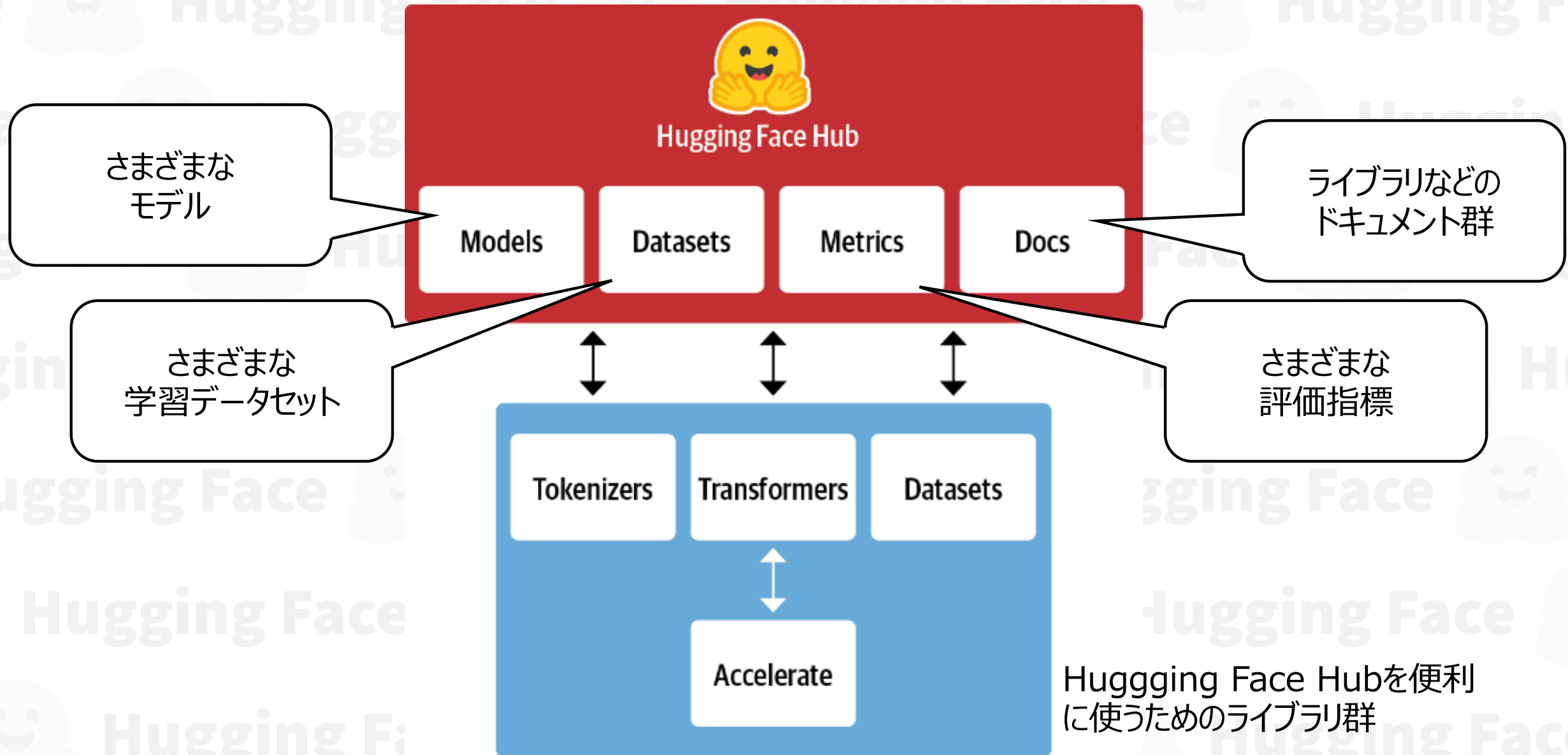
- メモリーに学習データを読み込むのではなく、ディスク上のファイルを直接読み込むことでメモリーの節約を行う

ストリーミング

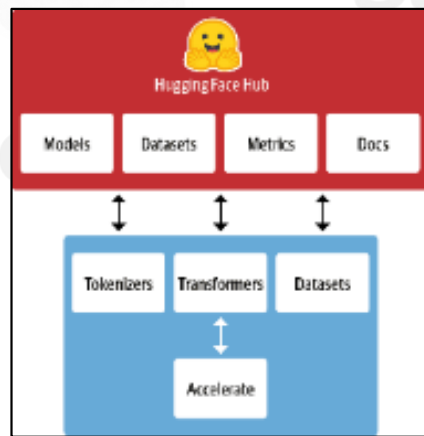
- 何テラもの学習データをローカルPCに置くのは現実的でない
- 必要な分だけをダウンロードしながら学習を進めていく



Hugging Faceとは



Hugging Faceサイトの構成と見方



Organizations

- Create New

Resources

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

Dark theme

Navigation Bar (B)

- Models
- Datasets
- Spaces
- Posts
- Docs
- Solutions
- Pricing

Left Sidebar (A)

- New
- ikedachin
 - Profile
 - Inbox (0)
 - Settings
 - Get Pro
- Organizations
 - Create New
- Resources
 - Hub guide
 - Transformers doc
 - Forum
 - Tasks
 - Learn
- Dark theme

Main Content Area

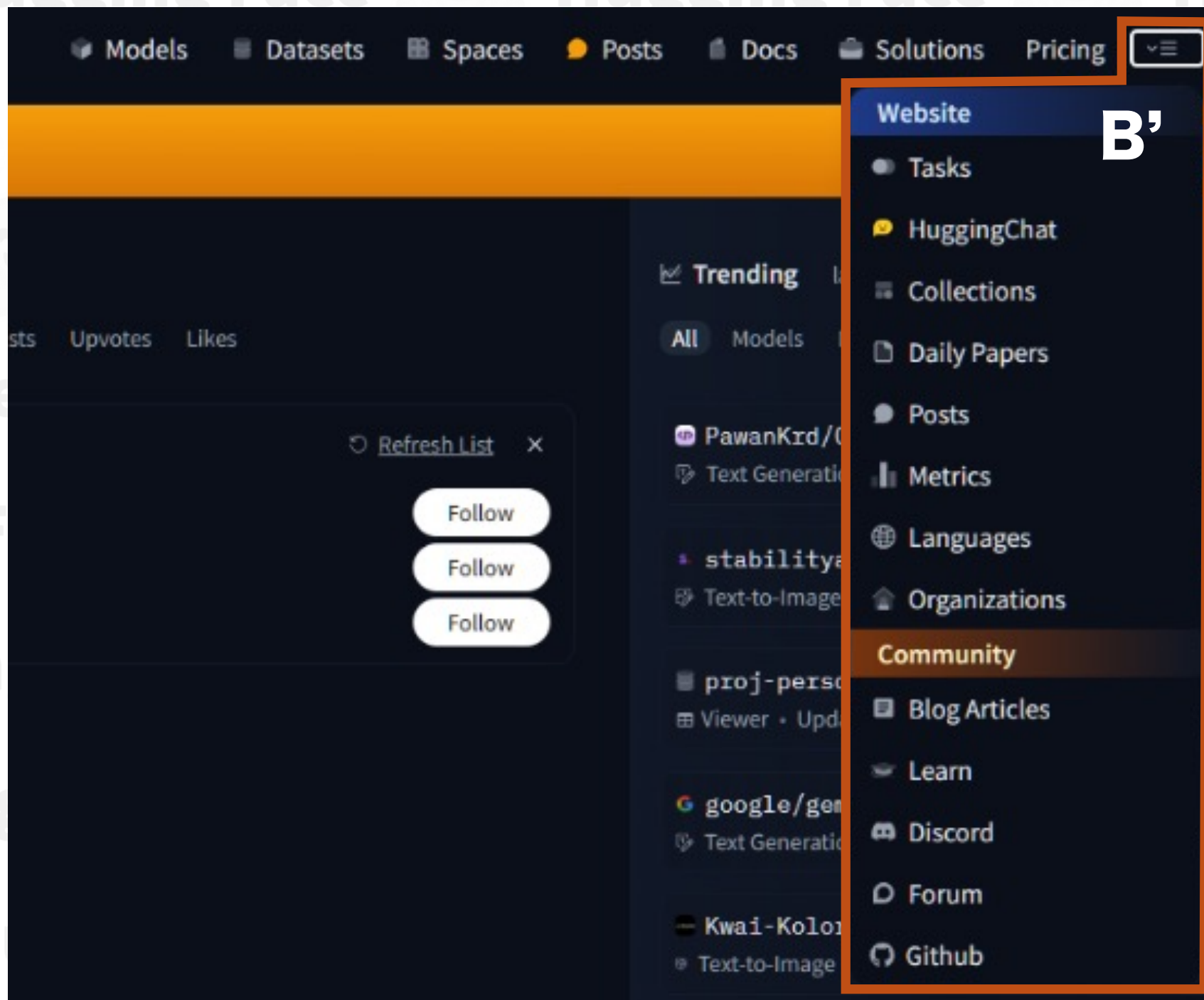
Following

Follow your favorite AI creators

- ilyasvick - Creator of ControlNet and other image gen gems
- ylecun - The godfather of deep learning
- bwang0911 - Research on sentence embedding models

Trending

- PawanKrd/CosmosRP-8k
- stabilityai/stable-diffusion-3-medium
- proj-persona/PersonaHub
- google/gemma-2-9b
- meta-llama/Meta-Llama-3-8B
- BAAI/Infinity-Instruct
- 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)

The image shows a screenshot of the Hugging Face website for the model `microsoft/phi-1_5`. Three specific areas are highlighted with orange boxes and numbered circles:

- Annotation 1:** Points to the top navigation bar of the model card, which includes links for "Model card", "Files and versions", and "Community 87".
- Annotation 2:** Points to the bottom right of the model card, highlighting the "Train", "Deploy", and "Use this model" buttons.
- Annotation 3:** Points to the "Model Summary" section, which contains the following text:

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.

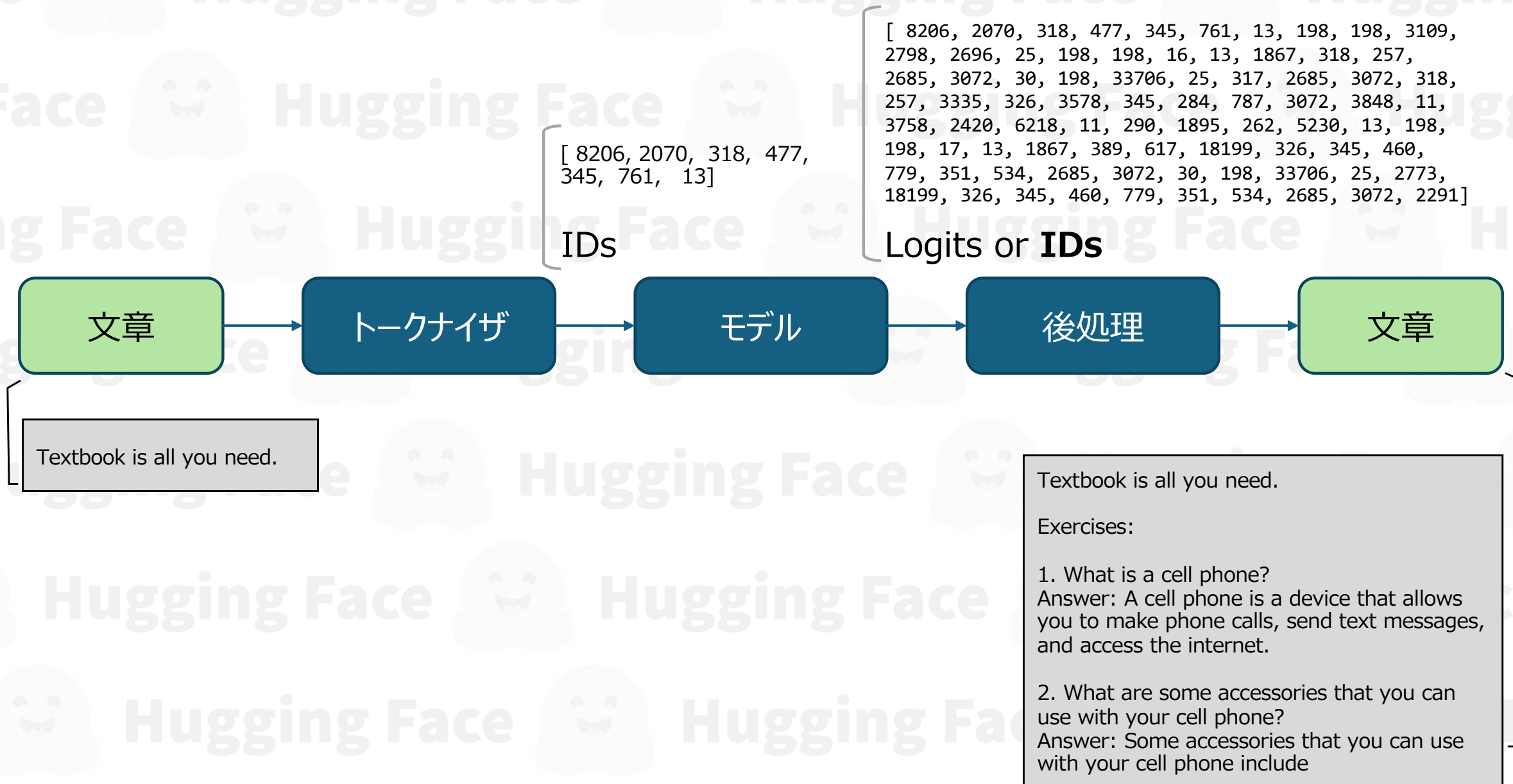
We did not fine-tune Phi-1.5 either for instruction following or through reinforcement learning

The main content area of the model card also displays various details: the model name `microsoft/phi-1_5`, a "like" count of 1.3k, tags for "Text Generation", "Transformers", "Safetensors", "English", "phi", "nlp", "code", "Inference Endpoints", "text-generation-inference", "arxiv:2309.05463", and "License: mit". It also shows a "Downloads last month" of 85,902, a line graph, and a section for "Inference API" with a text input field containing "My name is Lewis and I like to".

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
使ってみよか！**

[Google Colab](#) ← click

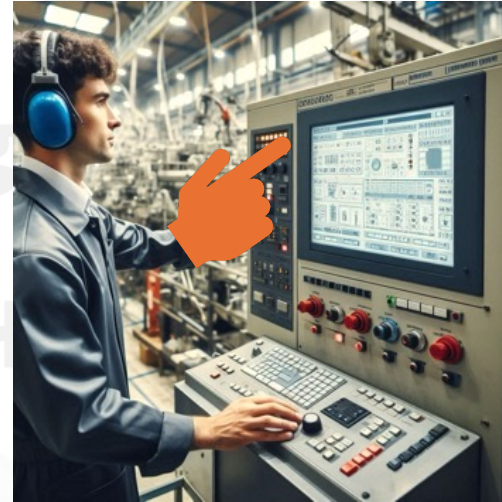
UseCaseを考えてみよう

- ローカルLLMに活路？
 - 社内特化？特定の目的特化？

- 業界の専門用語

視差呼称、はつり、デレッキ、うま

逃げ、刺す、芋掘り、三味線・・・



©KINAN Racing Team / Syunsuke FUKUMITSU

まとめ

- なんやら便利そう
- 日本語テキストが少ない
- ハードルが高く見えるけど、案外そうでもなさそう？
- **みんなで使ってみようぜ！**