

# Building NLP applications with Transformers



Julien Simon

Chief Evangelist, Hugging Face

[julsimon@huggingface.co](mailto:julsimon@huggingface.co)

# Deep Learning 1.0

Neural networks

Expert tools



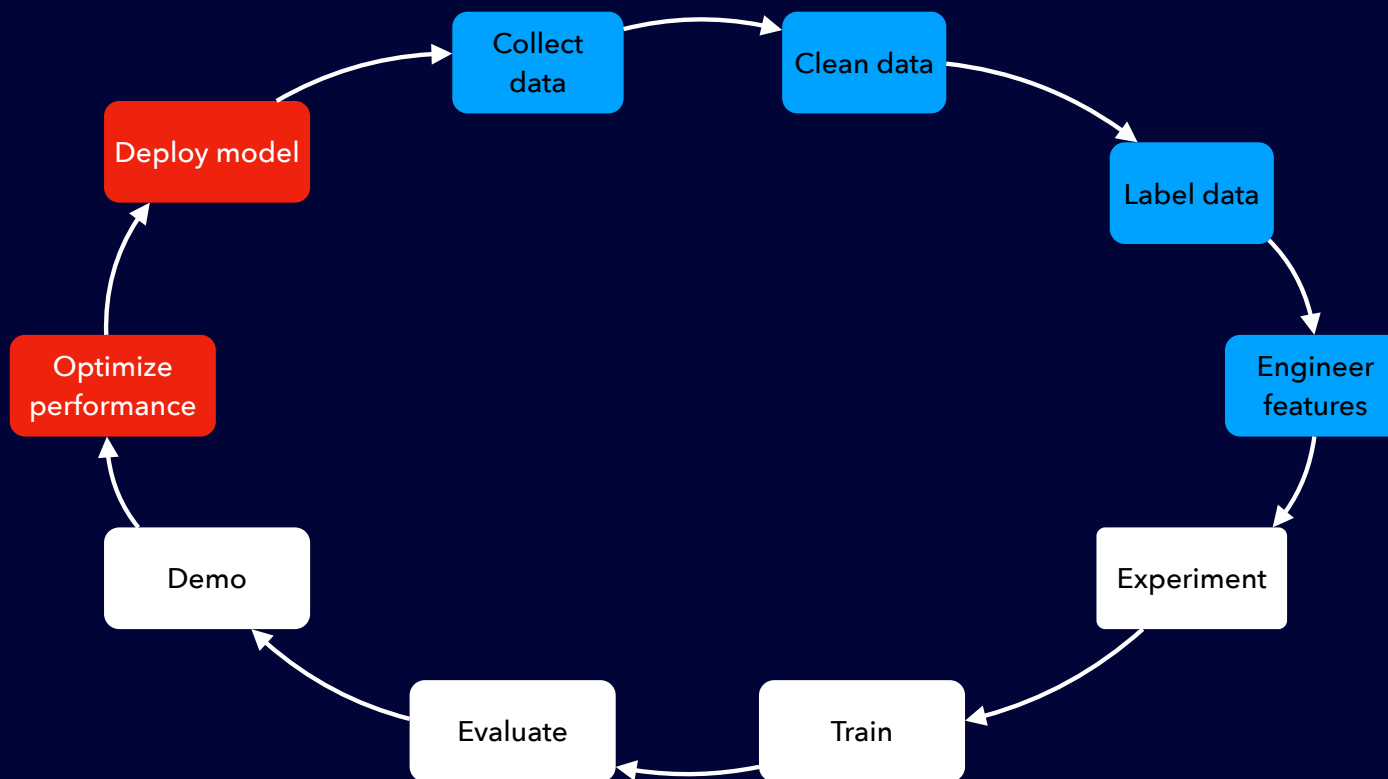
A few open datasets

Free images from pngset.com

GPUs



# A typical project (pretending waterfall is agile)



# Deep Learning 1.0: how it's going

87% of data science projects never make it into production

<https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production>



Only 25% of companies report widespread adoption

<https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html>



# Deep Learning 2.0

Transformers

Developer  
tools



Transfer  
Learning

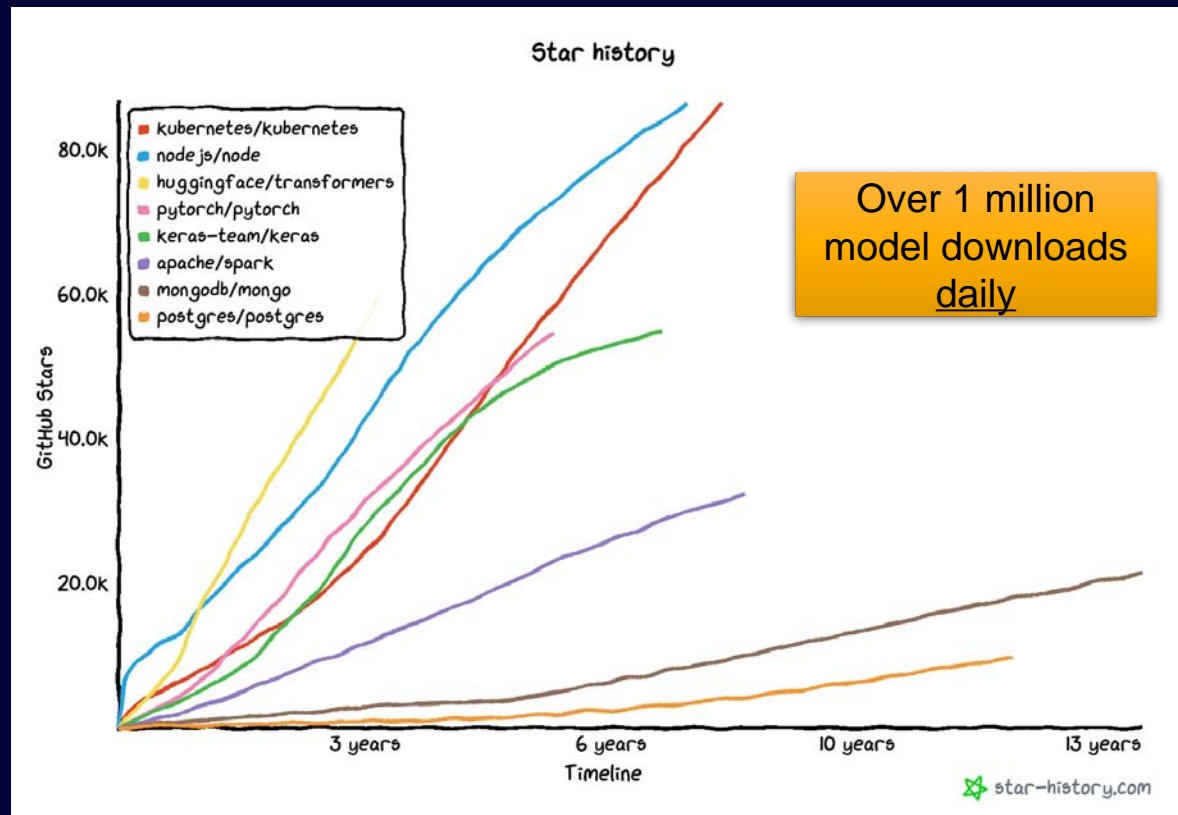
ML Hardware





# Transformers: one of the fastest-growing open source projects

<https://github.com/huggingface/transformers/>



*"Transformers are emerging as a general-purpose architecture for ML"*

<https://www.stateof.ai/>

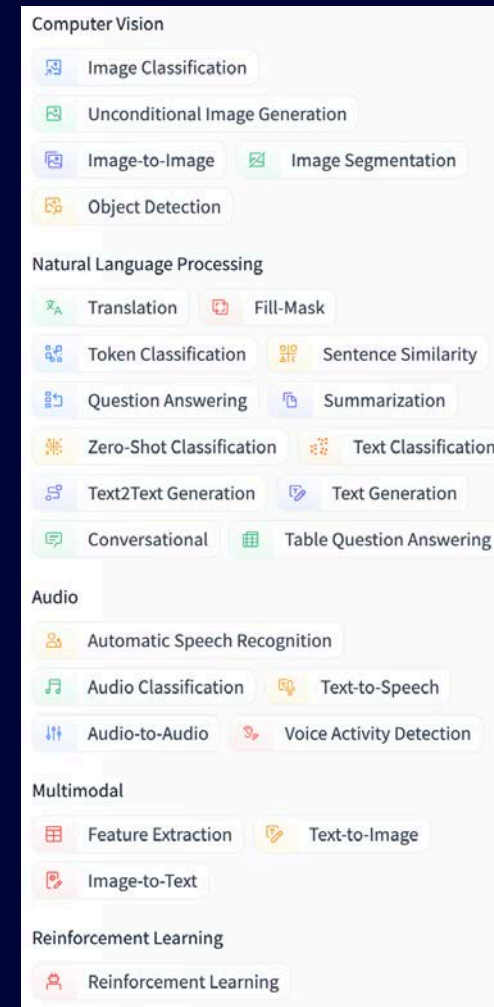
RNN and CNN usage down,  
Transformers usage up

<https://www.kaggle.com/kaggle-survey-2021>



# Transfer Learning

- Identify the **task type** for your business problem
- Pick and test a **pre-trained** model
  - No need to prepare a large dataset
  - Only a couple of lines of code
- Optionally, **fine-tune** the model on your data
  - Much less data is required
  - No need to train for long periods of time
  - Less than 50 lines of code



# Example: Translation + Part of Speech Tagging



```
from transformers import pipeline

translator = pipeline("text2text-generation", model="Helsinki-NLP/opus-mt-en-mul")
response = translator(">>hun<< In May 2022, Julien took his first trip to Hungary and loved it!")
text = response[0]["generated_text"]

classifier = pipeline("token-classification", model="novakat/nerkor-cars-onpp-hubert")
classifier(text)
```

2022. májusban **DATE** Julien **PER** az első **ORDINAL** utazása **Magyarországba** **GPE** szerette!





# Demo: pretrained models

## Multilingual voice queries on financial documents

- Speech-to-text in 21 languages (Facebook wav2vec2 300M)
- Semantic search on SEC filings (Sentence Transformers)

<https://huggingface.co/spaces/juliensimon/voice-queries>

<https://www.youtube.com/watch?v=YPme-gR0f80>



# Machine Learning Hardware

- A new generation of chips specially designed for **ML**
  - Faster training increases **agility** and **productivity**
  - Faster inference decreases **latency** and increases **throughput**
  - Get more work done with **less infrastructure** and at **lower cost**
- Hugging Face is partnering with ML hardware innovators
  - Training: Habana Labs, Graphcore,
  - Inference: Intel, Qualcomm, AWS Inferentia
  - Minimal code changes thanks to <https://github.com/huggingface/optimum>
  - Getting started posts at <https://huggingface.co/blog>



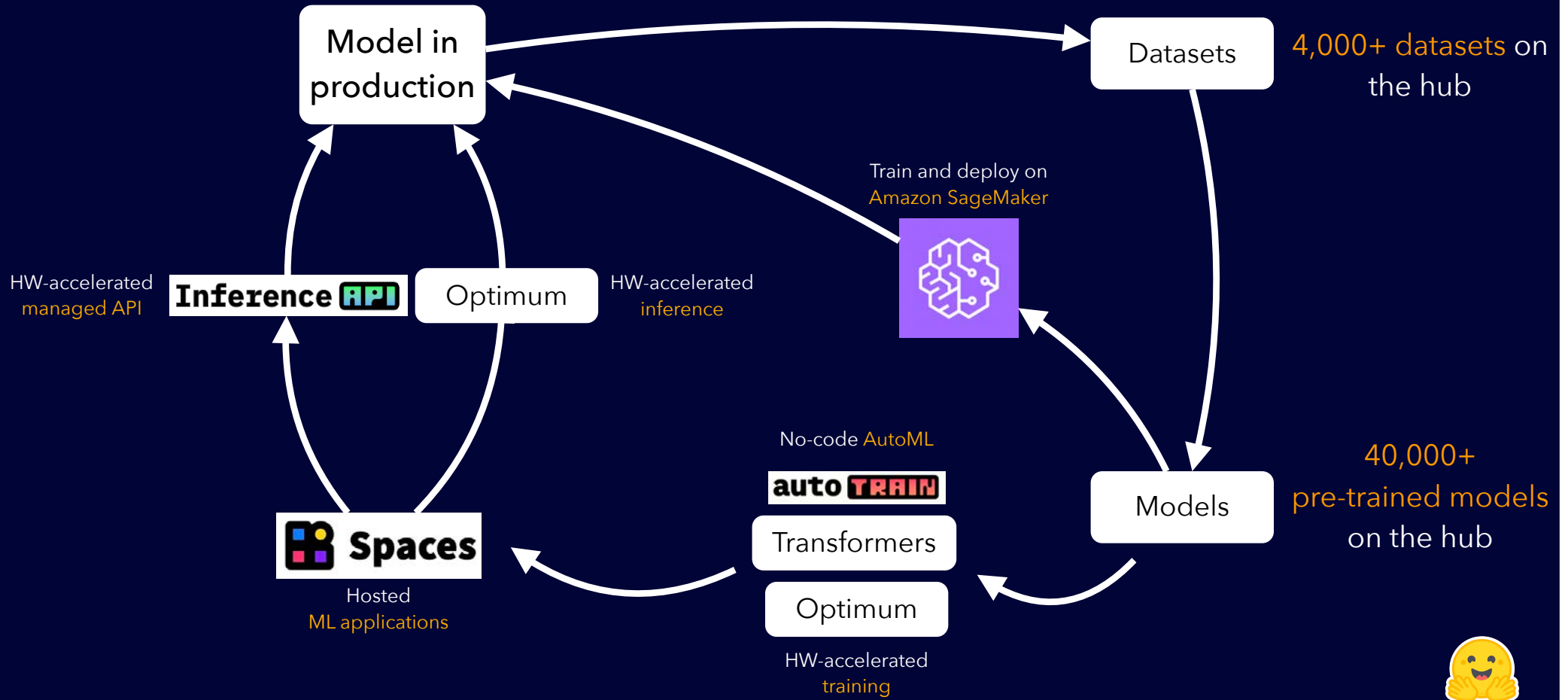
**GRAPHCORE**



**Qualcomm**



# Developer tools



 Demo: from the hub to AWS and back

Train and deploy a Hugging Face model  
on Amazon SageMaker

<https://huggingface.co/juliensimon/reviews-sentiment-analysis>



# Key Takeaways

- ML is complicated because we love to make it complicated
- Make sure to focus on the right things
  1. Find an **pre-trained model** that fits your business use case
  2. Identify a **business KPI** that shows success
  3. Measure the model on **real-life data**
  4. Good enough? Done!
  5. Need a bit more accuracy? **Fine-tune** on your data
  6. Optimize **prediction latency** and deploy in production
  7. Move to the next project
- Tools, platforms, and infrastructure are here: **no need to reinvent them**



# Getting started with Hugging Face

- Join our community  
<https://huggingface.co>
- New to Transformers?  
<https://huggingface.co/course>  
<https://discuss.huggingface.co>
- Need help? Ask about our Expert Acceleration Program (EAP)  
<https://huggingface.co/support>
- Need more privacy and compliance? Ask about a private hub deployment  
<https://huggingface.co/platform>



```
response = translator(">>your_language<< Thank you very much!")  
response[0]['generated_text']
```

[julsimon@huggingface.co](mailto:julsimon@huggingface.co)

@julsimon

<https://youtube.com/juliensimonfr/>

<https://julsimon.medium.com/>

