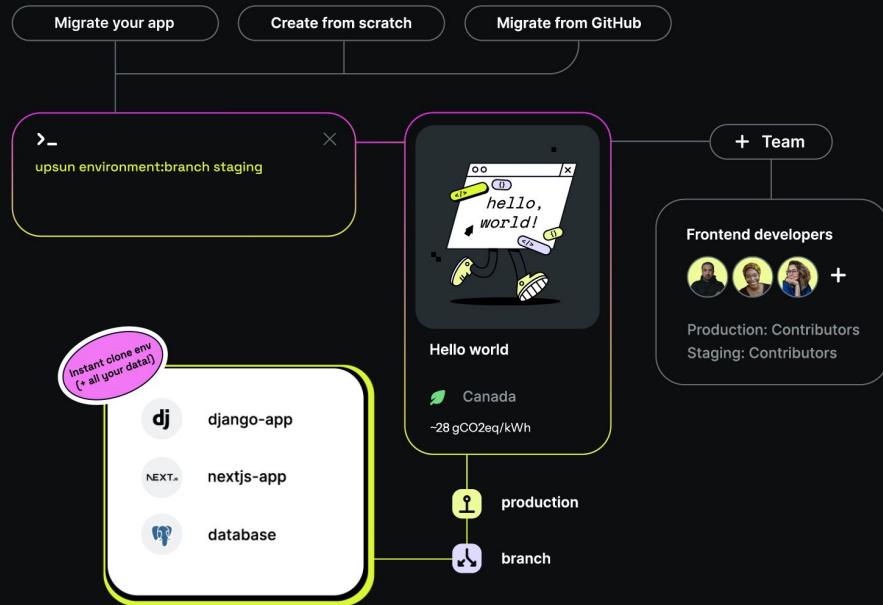




Machine Learning Inference in PHP by example

Leverage ONNX and
Transformers on Symfony





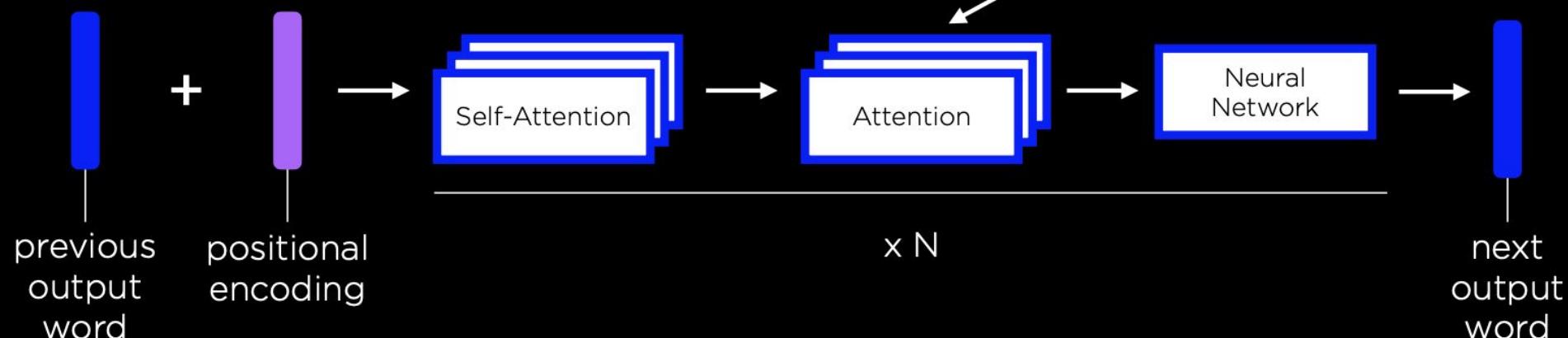
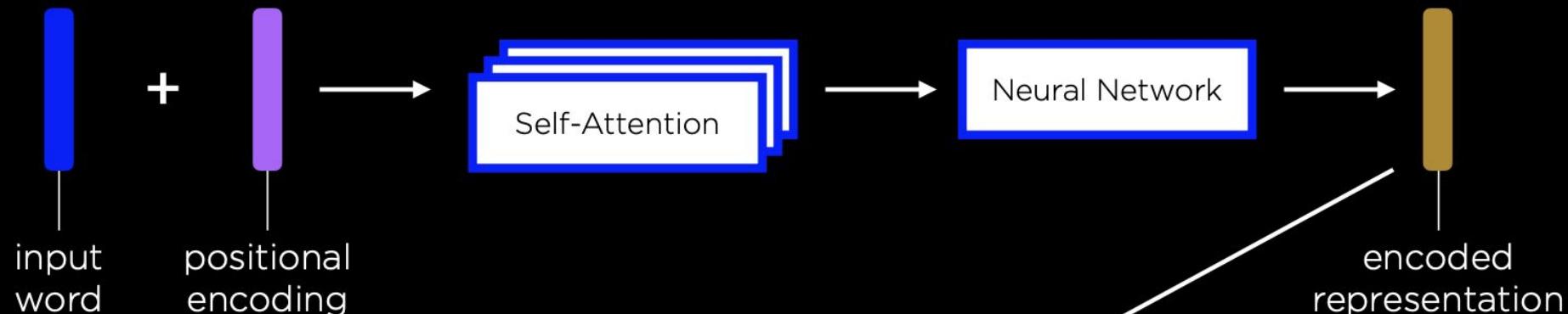
Who is familiar with
Transformers?

What is a Transformer?

- A neural network architecture that revolutionized NLP and beyond
- Introduced in 2017 paper "Attention is All You Need"
- Powers models like GPT, BERT, and T5

**Key innovation:
processes all input
at once instead of
sequentially**

- Encoder-decoder structure
- Self-attention layers
- Feed-forward neural networks
- Position embeddings



Before

2017

After

Sequential processing (RNNs,
LSTMs)

Limited context window

Slow training and inference

Parallel processing & lowering cost
of training & inference

Extended context understanding

Faster training and inference

Better at capturing relationships

Reduced resources usage

Demo:

Text: Symfony is a great framework for developing
[MASK] applications.

Learn more:

<https://cdn.cs50.net/ai/2023/x/lectures/6/lecture6.pdf>



SCAN ME

**Let's implement 3 use-cases
in PHP**

```
composer require codewithkyrian/transformers # Package  
  
./vendor/bin/transformers install # Install platform specific transformers  
  
# Download models  
./vendor/bin/transformers download Xenova/distilbert-base-uncased-finetuned-sst-2-english # Text  
classification  
./vendor/bin/transformers download Xenova/vit-base-patch16-224 # Image classification
```

Text Classification

Ca zaz Redets astyion

 Tar cələr sig al neyions.

Ca vint Farvest Reveste, Review

Amazon Reviewt. ★★★

Amazon Feesetting od nesions.



Amazon Reviewt. ★★★

Amazon Feesetting od nesions.





```
use function Codewithkyrian\Transformers\Pipelines\pipeline;

protected function execute(InputInterface $input, OutputInterface $output): int
{
    $asin = $input->getArgument('asin');
    $output->writeln("Calculating score for ASIN: $asin");

    $product = $this->entityManager->getRepository(Product::class)->findOneBy(['asin' => $asin]);
    $reviews = $product->getReviews();

    $pipe = pipeline('sentiment-analysis');

    foreach ($reviews as $review) {
        $out = $pipe($review->getText());
        if ($out['label'] == 'POSITIVE') {
            $positive++;
        } else {
            $negative++;
        }
    }

    $output->writeln("Positive: $positive, Negative: $negative, Score: " . ($positive - $negative) . "
(" . round($positive / $reviews->count() * 100, 2) . "% positive)");
    return Command::SUCCESS;
}
```

Demo

```
php bin/console app:score B07VGRJDFY
```



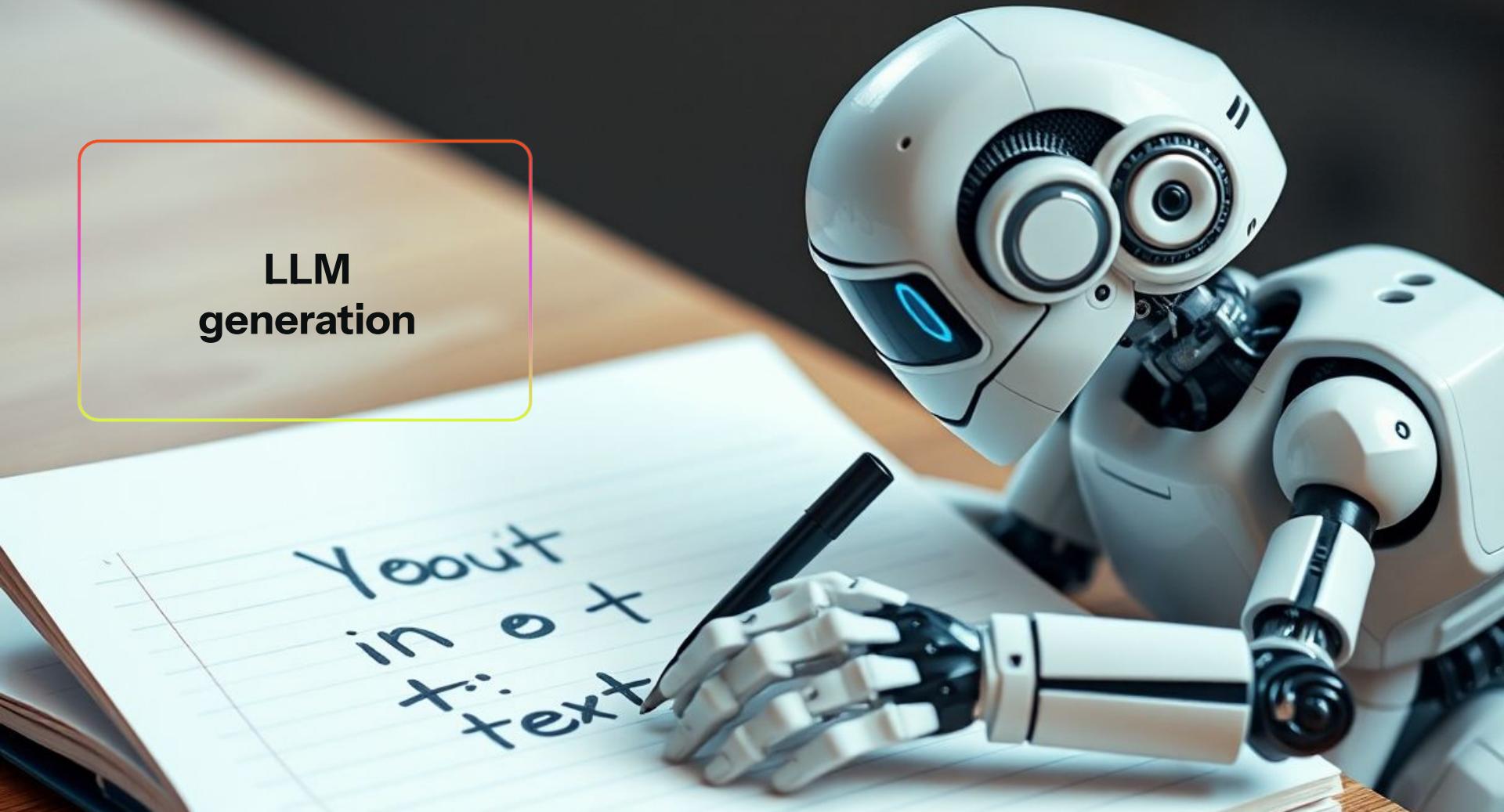
**Image
Classification**

```
public function handle(Request $request): Response
{
    [...]
    // Process the image to generate labels
    Transformers::setup()->setImageDriver(ImageDriver::GD);
    $classifier = pipeline('image-classification');
    $result = $classifier($this->getParameter('kernel.project_dir').'/public/uploads/'.$newFilename, 3);
    // Loop results to see if it's a hot dog
    [...]
}
```

Demo

<https://vienna.moigneu.net/hotdog>

LLM
generation





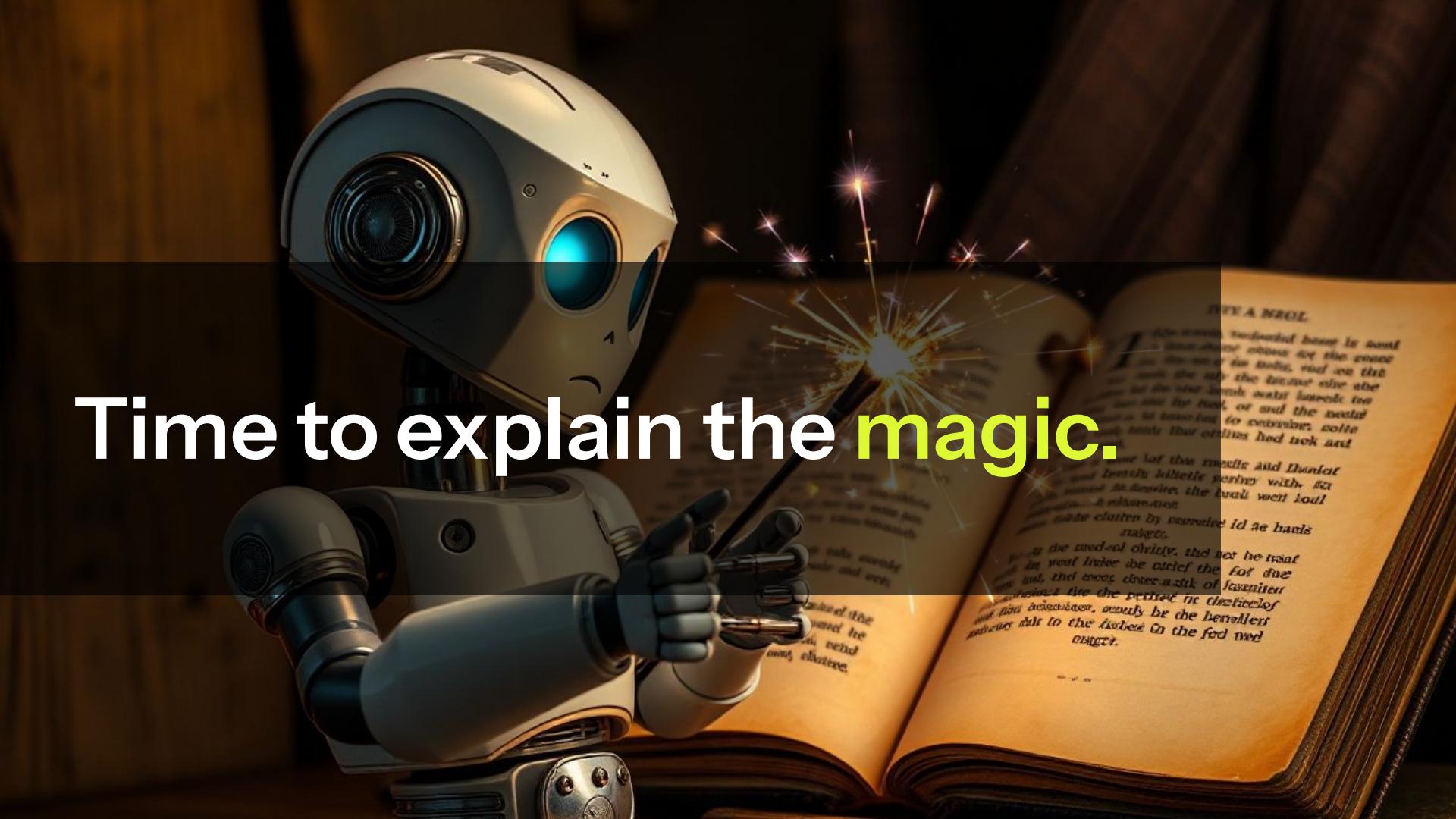
```
public function handle(Request $request): Response
{
    $question = $request->request->get('question');

    $generator = pipeline('text2text-generation', 'Xenova/flan-t5-small');
    $result = $generator($question,
        maxNewTokens: 256,
        repetitionPenalty: 1.6,
        temperature: 0.7
    );
    $answer = $result[0]['generated_text'];

    return $this->render('question/handle.html.twig', [
        'question' => $question,
        'answer' => $answer
    ]);
}
```

Demo

<https://vienna.moigneu.net/question>

A white and blue humanoid robot arm holds a glowing sparkler over an open book. The book's pages are filled with dense, illegible text. The scene is set against a dark background with dramatic lighting, creating a magical and mysterious atmosphere.

Time to explain the magic.

PHP FFI

Foreign Function Interface

FFI extension allows PHP code to directly call functions and manipulate data from C libraries without writing additional C code or PHP extensions.

Open Neural Network Exchange

ONNX for short

ONNX (Open Neural Network Exchange) is an **open standard** format that allows AI models to be shared between different machine learning frameworks like PyTorch, TensorFlow, and many others.

Transformers

and pipelines

Provided by the [TransformersPHP](#) package.

Courtesy of Kyrian Obikwelu
And 7 contributors

Supported pipelines

The screenshot shows the GitHub repository page for 'transformers-php'. The 'Code' tab is selected. At the top, there are links for Issues (11), Pull requests (1), Actions, Projects, Security, and Insights. Below the header, it says 'transformers-php (Public)'. On the left, there's a sidebar with 'main' (selected), '1 Branch', '15 Tags', a search bar ('Go to file'), and buttons for 'Add file' and 'Code'. The main area lists 297 commits from user 'takielias'. The commits are organized into sections for '.github', 'bin', 'docs', 'examples', 'libs', 'scripts', 'src', and 'tests'. Each commit includes a link, a brief description, and a timestamp. For example, the first commit is 'Fix Class Name for PretrainedTokenizer (#71)' made 7a4975b · last month. The commits are dated from 10 months ago to 3 months ago.

Commit	Description	Date
#71	Fix Class Name for PretrainedTokenizer (#71)	7a4975b · last month
.github	feat: Actions workflow to build and upload assets on release	3 months ago
bin	Reintroduce install command, new Library checker, and u...	6 months ago
docs	Fix Class Name for PretrainedTokenizer (#71)	last month
examples	Fix Class Name for PretrainedTokenizer (#71)	last month
libs	fix: refactor LibChecker to use the new libssloader workflow	3 months ago
scripts	feat: Update conversion notebook to include task	2 months ago
src	Fix Class Name for PretrainedTokenizer (#71)	last month
tests	Fix Class Name for PretrainedTokenizer (#71)	last month
.gitignore	fix: Stop tracking unnecessary example files	3 months ago
CHANGELOG.md	Update CHANGELOG	2 months ago
LICENSE	Add Tensor buffer test and update license	6 months ago
README.md	Correct docs on supported pipelines as mentioned in #27	7 months ago
VERSION	feat: Update VERSION	2 months ago
composer.json	revert: Return rokka/vips to normal dependencies	3 months ago
phpunit.xml	Initial Commit	10 months ago

Models

Infinite possibilities

- Hundreds of models readily available on 😊.
- Convert any model to ONNX with:
 - + [Python](#)
 - + [Notebook](#)

The screenshot shows the Hugging Face Model Hub interface. At the top, there is a search bar with the placeholder "Search models, datasets, users...". To the right of the search bar are three navigation buttons: "Models", "Datasets", and "Spaces". Below the search bar, the word "Models" is followed by the number "2,385" and a small "onnx" icon. A search filter for "onnx" is applied. The main content area is divided into several sections: "Multimodal" (Image-Text-to-Text, Visual Question Answering, Document Question Answering, Video-Text-to-Text, Any-to-Any), "Computer Vision" (Depth Estimation, Image Classification, Object Detection, Image Segmentation, Text-to-Image, Image-to-Text, Image-to-Image, Image-to-Video, Unconditional Image Generation, Video Classification, Text-to-Video, Zero-Shot Image Classification, Mask Generation, Zero-Shot Object Detection, Text-to-3D, Image-to-3D, Image Feature Extraction, Keypoint Detection), and "Natural Language Processing" (Text Classification, Token Classification, Table Question Answering, Question Answering, Zero-Shot Classification, Translation). Below these sections, a list of specific ONNX models is displayed, each with a profile picture, name, updated date, and a like count. The models listed are: "ezioruan/inswiper_128.onnx" (Updated Jul 4, 2023, 50 likes), "SamLowe/roberta-base-go_emotions-onnx" (Text Classification, Updated Sep 28, 2023, 3.14k likes), "microsoft/Phi-3-vision-128k-instruct-onnx-directml" (Text Generation, Updated 6 days ago, 79 likes), "onnx-community/Florence-2-base-ft" (Image-Text-to-Text, Updated Oct 8, 29.9k likes), "onnx-community/Qwen2.5-1.5B-Instruct" (Text Generation, Updated Oct 8, 254 likes), "akhaliq/AnimeGANv2-ONNX" (Updated Nov 11, 2021, 11 likes), "nouamanetazi/bloom-350m-onnx" (Updated Jul 1, 2022, 6 likes), "NeuML/ljspeech-jets-onnx" (Text-to-Speech, Updated Oct 10, 1.52k likes), "deepghs/ml-danbooru-onnx" (Updated Apr 10, 2023, 6 likes), and "deepghs/ccip_onnx" (Zero-Shot Image Classification, Updated Sep 9, 4 likes).

Model	Type	Updated	Likes
ezioruan/inswiper_128.onnx	Text Classification	Jul 4, 2023	50
SamLowe/roberta-base-go_emotions-onnx	Text Classification	Sep 28, 2023	3.14k
microsoft/Phi-3-vision-128k-instruct-onnx-directml	Text Generation	6 days ago	79
onnx-community/Florence-2-base-ft	Image-Text-to-Text	Oct 8	29.9k
onnx-community/Qwen2.5-1.5B-Instruct	Text Generation	Oct 8	254
akhaliq/AnimeGANv2-ONNX		Nov 11, 2021	11
nouamanetazi/bloom-350m-onnx		Jul 1, 2022	6
NeuML/ljspeech-jets-onnx	Text-to-Speech	Oct 10	1.52k
deepghs/ml-danbooru-onnx		Apr 10, 2023	6
deepghs/ccip_onnx	Zero-Shot Image Classification	Sep 9	4

SmolLM

Promising efficient LLM model

- Provided by the HF team
- 538Mb only
- Great alternative to fully fledged models
- Can even run on the browser



Give me some tips to improve my public speaking skills



Here are some tips to help you improve your public speaking skills:

1. **Practice, practice, practice:** The more you practice, the more comfortable you'll become with your material, and the more confident you'll feel. Practice in front of a mirror, record yourself, or speak in front of a small group of friends or family members.
2. **Know your audience:** Understand who your audience is, what they're interested in, and what they're looking for. Tailor your message, language, and tone to engage and persuade them.
3. **Prepare thoroughly:** Research your topic, organize your thoughts, and create an outline. Make sure your presentation has a clear structure and a solid understanding of your key points.
4. **Focus on your message:** Instead of worrying about how you're coming across, focus on the value you're providing to your audience. Remember, you're speaking to share your ideas, not to impress anyone with your speaking skills.
5. **Make eye contact:** Engage with your audience by making eye contact with different people in the room. This helps build trust and keeps your audience interested.
6. **Use confident body language:** Stand up straight, make use of gestures, and avoid fidgeting or crossing your arms. This will help you appear more confident and others will perceive you as more confident too.
7. **Speak slowly and clearly:** Avoid rushing through your speech. Speak slowly and clearly, pausing for emphasis and to allow your audience to absorb what you're saying.
8. **Use storytelling techniques:** People remember stories better than facts and figures. Use anecdotes, examples, and metaphors to make your message more relatable and engaging.
9. **Anticipate questions:** Think about potential questions your audience may have and prepare responses. This helps you feel more confident and in control.
10. **Join a public speaking group:** Consider joining a group like Toastmasters, where you can practice speaking in a supportive environment and receive feedback from others.
11. **Record yourself:** Record yourself speaking and watch the video to identify areas for improvement, such as your posture or tone of voice.



**Biggest limitation right now:
No GPU support. Yet.**

My recommendation

While this is great for **small tasks** (classification, image tagging, etc.), any LLM work should be **delegated to GPU based infrastructure.**

Your LLM options for now

Use a SaaS service: OpenAI, Claude API, etc.
Or use HF Inference endpoints to deploy the
model of your choice

And query that API endpoint from your app.

Or “Build your own”, at your own risk!



Thank you.



Guillaume Moigneu

VP, Advocacy @ Upsun

guillaume@platform.sh

guillaume.id