

Введение

Ask a question

Добро пожаловать на 😊 курс!



В этом курсе вы научитесь основам обработки естественного языка (NLP) с использованием библиотек от [Hugging Face](#). Экосистема состоит из: моделей (😬 [Transformers](#)), датасетов (😬 [Datasets](#)), вспомогательных библиотек (😬 [Accelerate](#), 😬 [Tokenizers](#)), а также репозитория [Hugging Face Hub](#). Это полностью бесплатно!

Чего ожидать от курса?

Краткое описание курса:

Introduction	Diving in	Advanced
Transformer models	The 😬 Datasets library	Building and sharing demos
Using 😬 Transformers	The 😬 Tokenizers library	Transformers can hear
Fine-tuning a pretrained model	Main NLP tasks	Transformers can see
Sharing models and tokenizers	How to ask for help	Optimizing for production

- Главы 1-4 содержат в себе введение в главные концепции библиотеки 😬 Transformers. К концу этой части курса вы будете знакомы с тем, как функционируют трансформеры, как применять модели из репозитория [Hugging Face Hub](#), как дообучить модели на собственных данных и опубликовать результаты на Hugging Face Hub!
- Главы 5-8 научат вас основам библиотек 😬 Datasets и 😬 Tokenizers (датасеты и токенизаторы); это необходимо для дальнейшего погружения в область обработки естественного языка. К концу этой части вы научитесь решать наиболее распространенные задачи в NLP самостоятельно!
- Главы 9-12 выходят за рамки NLP, в них описано, как можно применять трансформеры в задачах обработки речи и компьютерном зрении. Также вы узнаете, как создавать и демонстрировать свои модели, оптимизировать их для промышленного использования. После изучения этой части вы будете в силах применить 😬 Transformers к (почти) любой задаче машинного обучения!

Этот курс:

- Требует хорошего знания Python
- Будет лучше усвоен после ознакомления с курсом по глубокому обучению, например: [fast.ai's Practical Deep Learning for Coders](#) или одной из программ, подготовленных [DeepLearning.AI](#)
- Не предполагает предварительных знаний библиотек: [PyTorch](#) или [TensorFlow](#), однако знакомство с ними поможет вам в дальнейшем

После прохождения текущего курса мы рекомендуем ознакомиться со специализацией от DeepLearning.AI: [Natural Language Processing Specialization](#), которая покрывает широкий спектр традиционных моделей NLP: от наивного Байеса до LSTM-сетей!

Кто мы?

Об авторах:

Abubakar Abid окончил PhD в области прикладного машинного обучения в Стэнфордском университете. Во время PhD, он основал [Gradio](#) - свободная библиотека для Python, с помощью которой увидели свет свыше 600000 тысяч демоверсий моделей машинного обучения. Hugging Face приобрел Gradio, и теперь Abubakar работает с нами в качестве руководителя разработки машинного обучения.

Matthew Carrigan - ML-инженер в Hugging Face. Живет в Дублине, Ирландия, и ранее работал инженером по машинному обучению в Parse.ly, а до этого — научным сотрудником в Тринити-колледже в Дублине. Он не верит, что мы сможем реализовать теорию сильного искусственного интеллекта за счет масштабирования существующих архитектур, но все равно возлагает большие надежды на бессмертие роботов.

Lysandre Debut - ML-инженер в Hugging Face, работает над библиотекой 😬 Transformers с самых ранних этапов разработки. Его цель — сделать NLP доступным для всех, разработав инструменты с очень простым API.

Sylvain Gugger – инженер-исследователь в Hugging Face и один из ключевых участников разработки библиотеки 😬 Transformers. Ранее работал научным сотрудником в fast.ai и написал книгу [Deep Learning for Coders with fastai and PyTorch](#) в соавторстве с Jeremy Howard. Основное внимание в его исследованиях уделяется тому, чтобы сделать глубокое обучение более доступным путем разработки и улучшения методов, позволяющих моделям быстро обучаться при ограниченных ресурсах.

Dawood Khan - ML-инженер в Hugging Face. Dawood из Нью-Йорка, где он окончил Нью-Йоркский университет и получил степень бакалавра компьютерных наук. Проработав несколько лет iOS инженером, Dawood решил сменить работу и стал сооснователем Gradio. Позднее Hugging Face приобрел Gradio.

Merve Noyan - developer advocate в Hugging Face, работает над разработкой инструментов и созданием контента на их основе, чтобы машинное обучение более доступным.

Lucile Saulnier - ML-инженер в Hugging Face, разрабатывающая и поддерживающая использование инструментов с открытым исходным кодом. Она также активно участвует во многих исследовательских проектах в области NLP, таких как совместное обучение и BigScience.

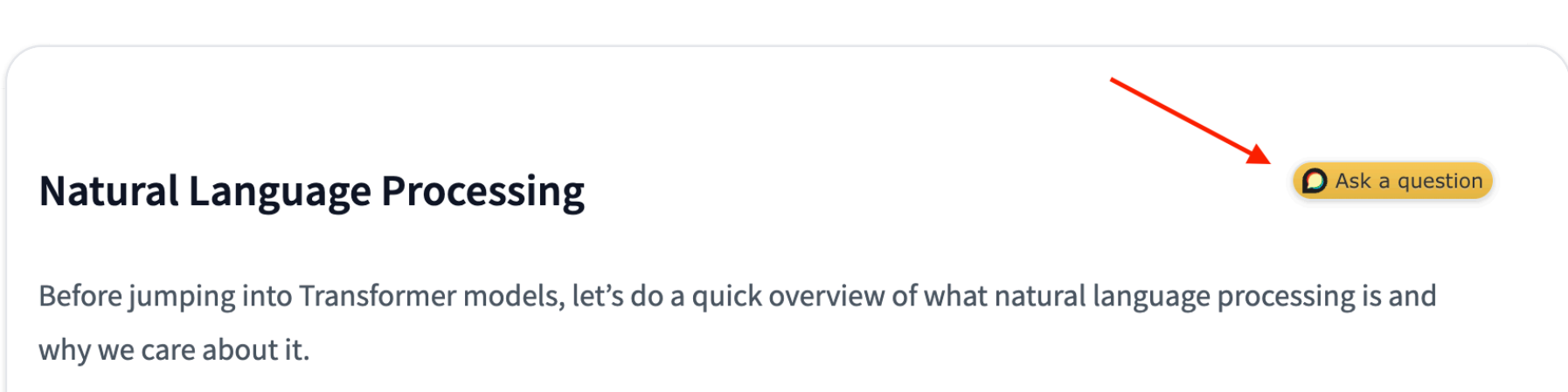
Lewis Tunstall - ML-инженер в Hugging Face, сосредоточен на разработке инструментов с открытым исходным кодом и обеспечении их доступности для более широкого сообщества. Соавтор будущей книги [O'Reilly book on Transformers](#).

Leandro von Werra - ML-инженер в команде, работающей над открытым исходным кодом Hugging Face и соавтор будущей будущей книги [O'Reilly book on Transformers](#). Обладает большим опытом реализации NLP-проектов в промышленности.

ЧАВО

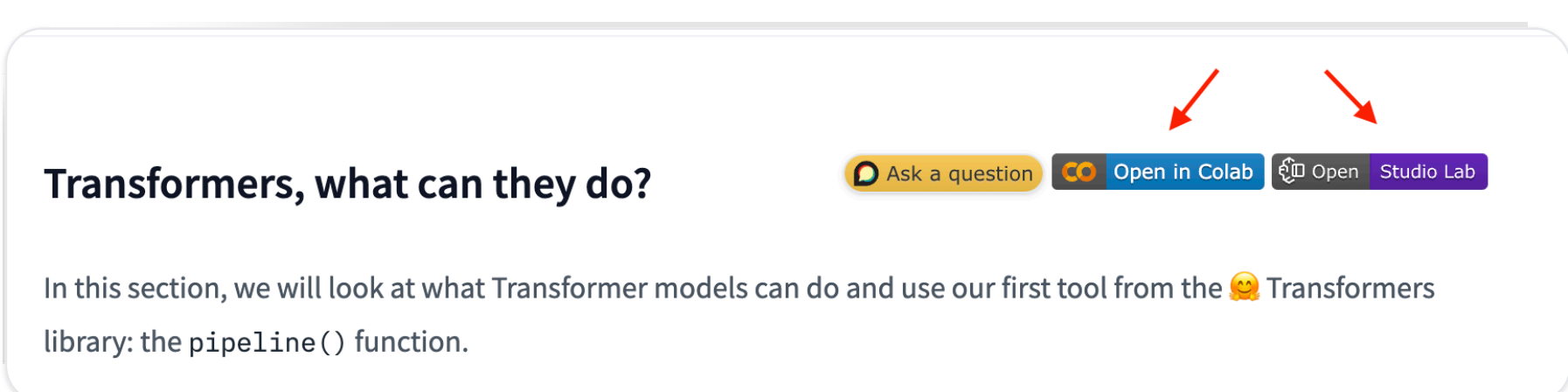
Мы собрали ответы на несколько часто задаваемых вопросов:

- Получу ли я сертификат после прохождения этого курса?** На данный момент у нас нет сертификации для этого курса. Мы работаем над получением сертификации для экосистемы Hugging Face. Следите за новостями!
- Сколько времени мне нужно будет потратить на прохождение этого курса?** Каждая глава этого курса рассчитана на неделю работы, то есть примерно 6-8 часов в неделю. Однако, вы можете проходить курс в любом удобном для вас ритме.
- Где я могу задать вопрос по материалам курса?** Если у вас возникли какие-либо вопросы по поводу любой части курса, просто нажмите на *Ask a question* наверху страницы, и вы будете автоматически перенаправлены в соответствующий раздел [форума Hugging Face](#) (форум на английском языке):



Обратите внимание, что на форуме также доступен список [идей для проектов](#), если вы хотите применить полученные знания на практике после прохождения курса.

- Где я могу посмотреть на код, используемый в этом курсе?** Внутри каждого раздела наверху страницы есть баннер, который позволит запустить код в Google Colab или Amazon SageMaker Studio Lab:



Блокноты Jupyter со всем кодом, используемом в материалах курса, доступны в репозитории [huggingface/notebooks](#). Если вы хотите сгенерировать их на своем компьютере, вы можете найти инструкцию в репозитории [course](#) на GitHub.

- Как я могу внести свой вклад в развитие курса?** Существует множество способов внести свой вклад в наш курс! Если вы найдете опечатку или баг, пожалуйста, откройте вопрос (issue) в репозитории [course](#). Если вы хотите помочь с переводом на ваш родной язык, вы можете найти инструкцию [здесь](#).
- Какие стандарты использовались при переводе?** Каждый перевод содержит глоссарий и файл TRANSLATING.txt, в которых описаны стандарты, используемые для перевода терминов и т.д. Вы можете посмотреть на пример для немецкого языка [здесь](#).
- Могу ли я использовать этот курс в своих целях?** Конечно! Этот курс распространяется по либеральной лицензии [Apache 2 license](#). Это означает, что вы должны упомянуть создателей этого курса, предоставить ссылку на лицензию и обозначить все изменения. Все это может быть сделано любым приемлемым способом, который, однако, не подразумевает, что правообладатель поддерживает вас или ваши действия по отношению этого курса. Если вы хотите процитировать этот курс, пожалуйста, используйте следующий BibTex:

```
@misc{huggingfacecourse,
  author = {Hugging Face},
  title = {The Hugging Face Course, 2022},
  howpublished = "\url{https://huggingface.co/course}",
  year = {2022},
  note = "[Online; accessed <today>]"
}
```

Поехали!

Вы готовы начать? В этой главе вы узнаете:

- Как использовать `pipeline()` для решения NLP-задач генерации и классификации текста
- Об архитектуре трансформеров
- Как различать архитектуры кодировщика, декодировщика и кодировщика-декодировщика и варианты их использования