

Мультиязычие и культурный код в эпоху больших языковых моделей



Сергей Марков

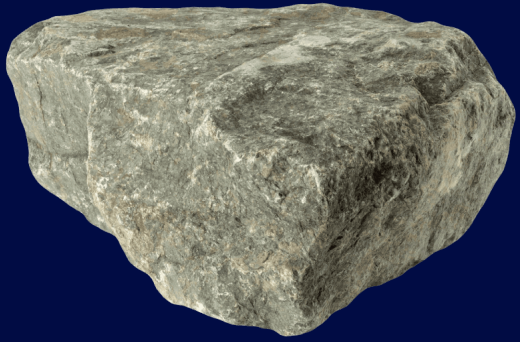
Директор по развитию технологий ИИ, ПАО «Сбербанк»

Markov.S.S@sberbank.ru

[@oulenspiegel_channel](#)

Революция в NLP и её плоды для смежных отраслей





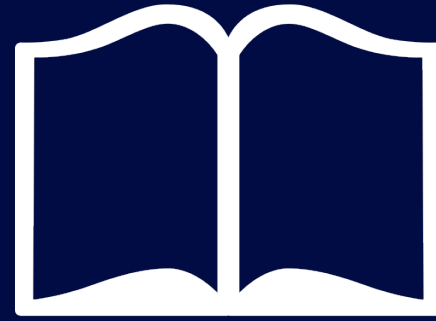
Камень



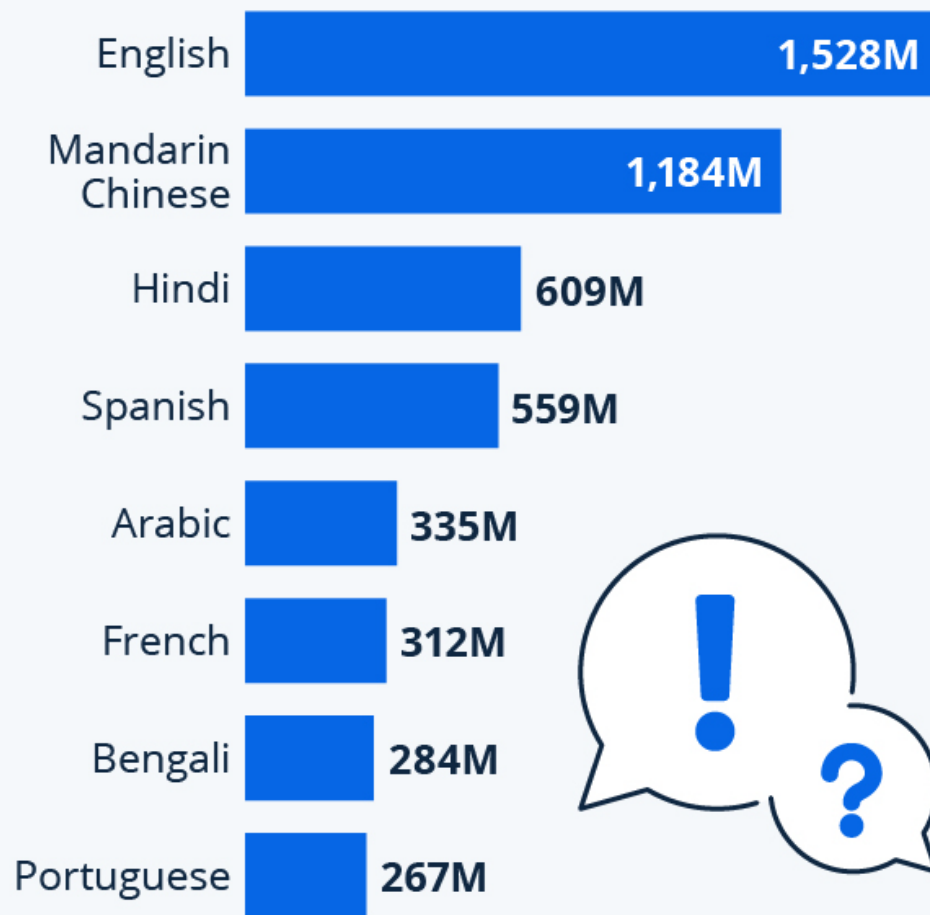
Бронза



Железо



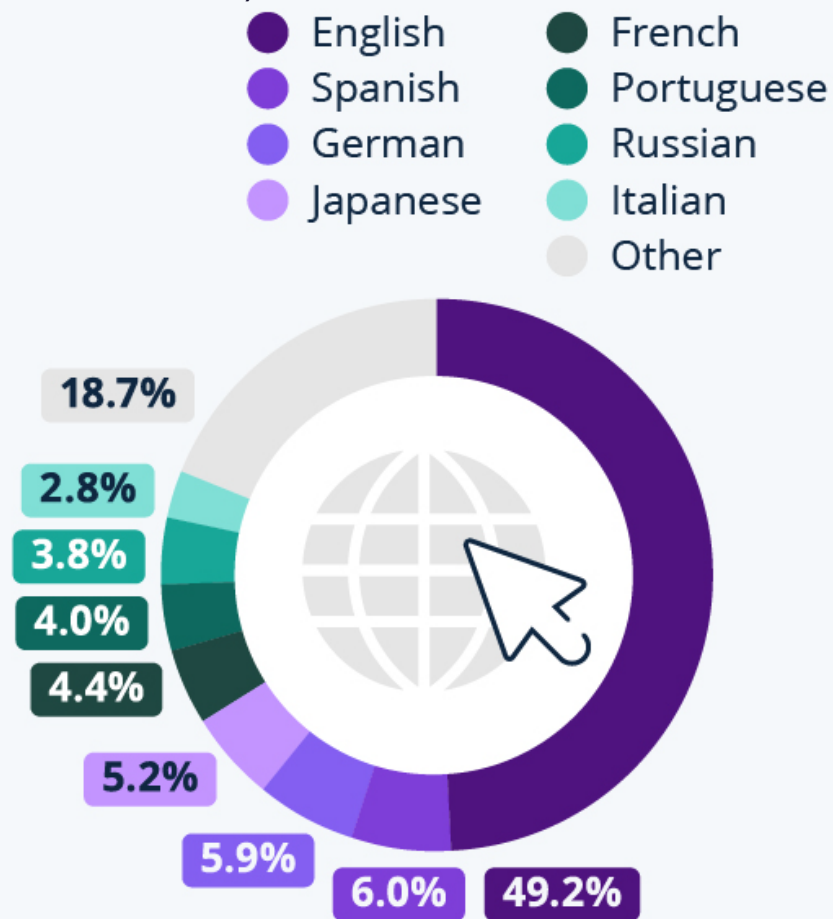
Самые распространённые языки в мире (число говорящих, для которых язык родной)



As of 2025

Sources: DataReportal, Ethnologue

Самые распространённые языки в интернете (доля сайтов)



LLM для многоязычия и многоязычие для LLM

- ❑ Сохранение разнообразия языков. В мире существует более 7000 языков, каждый из которых несет культурное и лингвистическое наследие.
- ❑ Поддержка малоресурсных языков. Многие из языков находятся на грани исчезновения, количество текстовых материалов на них ограничено.
- ❑ Улучшение качества перевода. Многоязычные модели могут значительно повысить качество машинного перевода, а добавление многоязычных данных в обучающие корпуса LLM расширяют их языковые способности.
- ❑ Развитие «интеллекта» моделей благодаря расширению репертуара рассуждений за счёт уникальных элементов культуры, представленных в различных языковых традициях.
- ❑ Этическая составляющая. Разработка многоязычных моделей способствует более справедливому технологическому развитию, уменьшая доминанцию нескольких самых популярных мировых языков.

Чем можно помочь

- Сбор моноязычных корпусов. Книги, статьи, документы и т. д.
 - Для обучения и дообучения моделей под конкретный язык.
 - Для обучения многоязычных моделей.
 - Для разработки NLP-инструментов — парсеры, лемматизаторы, токенизаторы, классификаторы, корректоры ошибок.
- Подготовка параллельных корпусов. Пары предложений.
 - В основном используются для машинного перевода.
 - Для более точного понимания моделями этих языков.
 - Для пословного и пофразного выравнивания. Для составления переводных словарей, словарей синонимов и т. д.
- Корпуса на родственных языках также помогают улучшать качество моделей для конкретного языка.

Многоязычные проекты



Мой сайт и книга



**Спасибо за
внимание!**