

Утверждаю
Начальник Департамента
НИиР

_____ Н.Р. Белашенков

" ____ " _____ 202__ г.

**ЗАКЛЮЧЕНИЕ
о возможности открытого опубликования**

материалов статьи/доклада «Двухэтапный подход для генерации рекламы на естественном языке на основе архитектуры transformer» автор Неужин Егор Евгеньевич
(Ф.И.О. автора (ов), вид, наименование материалов, подлежащих экспертизе)

Экспертная комиссия федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет ИТМО» в составе: руководитель-эксперт – Птицын А.В., к.т.н., доц. ФБИТ
эксперт - Бухановский А.В., д.т.н., директор МФ ТИИТ
провела экспертизу материалов (протокол № __ от «__» _____ 202__ г.), на предмет отсутствия (наличия) в них сведений, составляющих государственную тайну, и возможности (невозможности) их открытого опубликования.

Руководствуясь Законом Российской Федерации «О государственной тайне», Перечнем сведений, отнесенных к государственной тайне, утвержденным Указом Президента Российской Федерации от 30 ноября 1995 г. № 1203, а также Перечнем сведений, подлежащих засекречиванию Минобрнауки России, утвержденным Приказом № 36 от 10 ноября 2014 г., комиссия установила:

сведения, содержащиеся в рассматриваемых материалах «Двухэтапный подход для генерации рекламы на естественном языке на основе архитектуры transformer»
(указываются сведения, содержащиеся в материалах)

не подпадают под действие Перечня сведений, составляющих государственную тайну (статья 5 Закона Российской Федерации «О государственной тайне»), не относятся к Перечню сведений, отнесенных к государственной тайне, утвержденному Указом Президента Российской Федерации от 30 ноября 1995 г. № 1203, не подлежат засекречиванию Перечнем сведений Минобрнауки России, утвержденным Приказом № 36 от 10 ноября 2014 г., и данные материалы могут быть открыто опубликованы.

Руководитель-эксперт

_____ (подпись)

/ Птицын А.В. /
(фамилия, инициалы)

Эксперт

_____ (подпись)

/ Бухановский А.В. /
(фамилия, инициалы)

Члены комиссии:

Секретарь

_____ (подпись)

/ Коваль Е.Н. /
(фамилия, инициалы)

Представитель ОИС и НТИ

_____ (подпись)

/ _____ /
(фамилия, инициалы)

УДК 004.896

ДВУХЭТАПНЫЙ ПОДХОД ДЛЯ ГЕНЕРАЦИИ РЕКЛАМЫ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ НА ОСНОВЕ АРХИТЕКТУРЫ TRANSFORMER

Невежин Е.Е. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – к.т.н., доцент ИДУ, старший научный сотрудник НЦКР Бутаков Николай Алексеевич (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

В данной работе описывается механизм генерации рекламных баннеров на основе двухступенчатого подхода на основе нейросетевой архитектуры transformer.

Введение. В настоящее время цифровая реклама является неотъемлемой частью нашей жизни. Он сопровождает нас на сайтах, в социальных сетях и мессенджерах и может рассматриваться как определенный тип связи от заинтересованного лица с потенциальным клиентом. Часто это представляется в виде баннера - небольшого текстового сообщения со связанной картинкой, которое приводит к целевой веб-странице, когда мы нажимаем на нее.

Доставка рекламы пользователям обеспечивается механизмами рекомендаций, облегчающими распространение баннеров среди пользователей. Современные движки персонализированы и пытаются ориентировать пользователей на рекламу, касающуюся их интересов и желаний. Исследовательские и промышленные сообщества вложили много усилий, чтобы эффективно решить эту часть работы. Но после того, как был найден нужный пользователь, его нужно поймать, показав соответствующую текстовую и визуальную информацию, в нашем случае, в виде баннера.

В настоящее время баннеры создаются людьми, и автоматизация этой задачи может принести пользу цифровой рекламе, подняв рутину генерации текста и даже усилив персонализацию с помощью баннеров, созданных специально для сайтов, отдельных веб-страниц и даже отдельных элементов.

Основная часть. Общая идея генерации рекламных баннеров построена на следующих шагах:

1. Собрать датасет, содержащий пары текст сайта и текст баннера
2. Произвести очистку, в том числе и по признаку соотносимости теста баннера тексту сайта
3. Обучить базовую модель на выборке
4. Использую тематические моделирование, кластеризировать существующие пары сайт-баннер в категории
5. Дообучить сеть на каждой категории
6. Обучить сеть для каждого класса с нуля

В настоящее время использование смарт-баннеров стало тенденцией для создания рекламы. Эти подходы предполагают некоторую степень адаптации для конкретного пользователя, который просматривает сайт или канал и генерирует изображения с текстом в соответствии с заданными шаблонами. Использование этого подхода вместе с нашей моделью может помочь автоматизировать и ускорить процесс. Таким образом, полученная модель дает неплохое качество генерации суммаризованного представления баннера.

Выводы. В данной работе мы представили подход, позволяющий генерировать рекламные баннеры из текстов веб-страниц. Проведенное сравнительное исследование показывает, что методы на основе трансформаторов работают лучше, чем другие, в то время как качество получаемых баннеров показывает, что они могут быть сгенерированы эффективно с

использованием предложенных методов на основе трансформаторов. Высокая степень неоднородности данных, обнаруженная в собранном наборе данных, помогает модели лучше обобщать, но снижает качество кластеров определенных доменов. Предложенный в работе двухступенчатый трансформаторный подход помогает решить эту проблему и может повысить эффективность до 5% в целом и до 10% для отдельных кластеров.

Невежин Е.Е. (автор)



Бутаков Н. А. (научный руководитель)