Компания предлагает новую систему для борьбы с назойливыми «переворотами битов».

Модели искусственного интеллекта сталкиваются со множеством проблем, одной из которых являются аппаратные сбои, такие как «перевороты битов» (Bit Flips). Во время таких сбоев значения данных в системе изменяются с нуля на единицу и могут приводить к ошибкам, что создаёт серьёзные риски для точности и надёжности работы ИИ-моделей.

Исследователи компании Meta* – это американская технологическая компания, которая владеет и управляет такими продуктами и сервисами, как Facebook, Instagram, WhatsApp и другими.

 Компания была основана в 2004 году Марком Цукербергом и его друзьями под названием TheFacebook, Inc., в 2005 году переименована в Facebook, Inc., а в 2021 году в Meta Platforms, Inc., чтобы «отразить свой фокус на создании метавселенной» — интегрированной среды, которая связывает все продукты и сервисы компании.

 * Компания Meta и её продукты признаны экстремистскими, их деятельность запрещена на территории РФ." data-html="true" data-original-title="Meta" >Meta * отмечают, что такие ошибки во время инференса или обслуживания ИИ могут приводить к неверным или ухудшенным результатам моделей, что в конечном итоге влияет на качество предоставляемых ИИ-услуг. Meta задокументировала частоту таких битовых ошибок в своей инфраструктуре, подчеркнув, что справляться с этими сбоями всегда было непросто. А сложность и разнообразие современных аппаратных ИИ-систем делают их ещё более уязвимыми к таким сбоям.

Для решения проблемы Meta предложила новый подход к измерению аппаратных сбоев, чтобы разработчики ИИ-систем могли лучше понимать и управлять рисками. Они ввели новый метрический показатель — «коэффициент уязвимости параметров» (PVF), который стандартизирует оценку уязвимости ИИ-моделей к повреждениям параметров.

PVF можно адаптировать к различным моделям аппаратных сбоев и настроить для разных моделей и задач. Кроме того, его можно использовать на этапе обучения для оценки воздействия повреждений параметров на способность модели к сходимости.

Исследователи Meta смоделировали инциденты «тихой порчи данных» (Silent Data Corruption), используя инструмент DLRM, применяемый для генерации персонализированных рекомендаций контента. В некоторых случаях они обнаружили, что четыре из тысячи инференсов были неверными из-за битовых ошибок.

Представленный компанией подход во многом основывается на «коэффициенте уязвимости архитектуры» (AVF), предложенном в прошлом году исследователями из Intel и Университета Мичигана. Тем не менее, разработка Meta открывает перспективы для значительного улучшения надёжности и эффективности современных ИИ-моделей.

Применение PVF позволит разработчикам более точно оценивать уязвимости своих систем и принимать обоснованные решения по их оптимизации. В будущем это может привести к созданию более устойчивых к сбоям ИИ-архитектур, повышению точности результатов и, в конечном итоге, к расширению возможностей применения искусственного интеллекта в критически важных областях, где надёжность имеет первостепенное значение.

* Компания Meta и её продукты признаны экстремистскими, их деятельность запрещена на территории РФ.