РусКрипто

XXVII НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ

Безопасность технологий с искусственным интеллектом: спекуляции, домыслы прогнозы

Лось Владимир Павлович, Президент АЗИ, ГНС РГГУ,
доктор военных наук, профессор.

Тышук Екатерина Дмитриевна, СНС РГГУ.

РусКрипто

РусКрипто

Спекуляции и домыслы строятся на ложных высказываниях разработчиков о возможностях генеративных моделей ИИ.

РусКрипто

Опасения вызывает бесконтрольное использование моделей ИИ в следующих областях:

- оперативно-служебная деятельность сотрудников органов безопасности и правопорядка;
- КИИ;
- подготовка кадров для специальных служб;
- ……..

ПРОГНОЗ

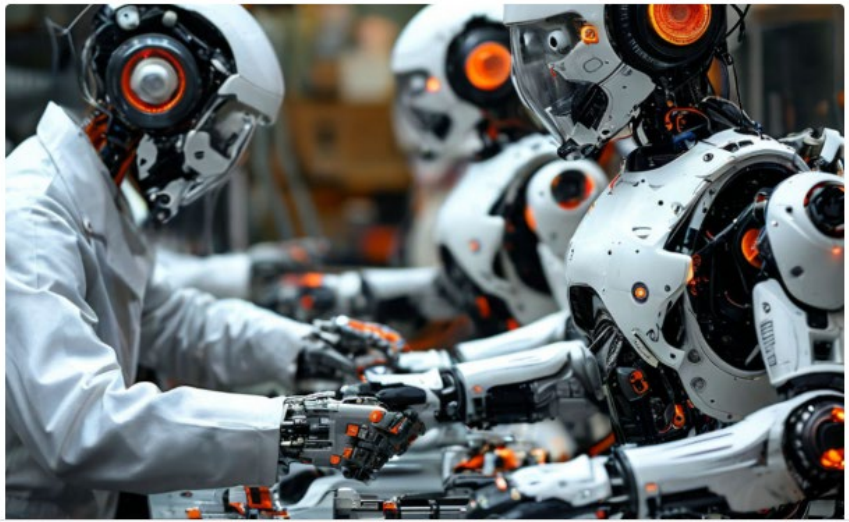
Рано или поздно встанет вопрос о проверке качества моделей ИИ, то есть об их сертификации

РусКрипто

Главная • ИТМО.ИСКИН • 17 марта 2025

В ИТМО разработали цифровой полигон для тестирования новых систем ИИ

«Полиокс» анализирует потенциал системы ИИ, прогнозирует ресурсную стоимость ее дообучения и объективно оценивает качество работы в экстремальных условиях



РусКрипто

В РГГУ разработан замысел создания лаборатории сертификации моделей ИИ

РусКрипто

Сертификация моделей ИИ – процесс установления их безопасности и применимости в данной предметной области.

На примере предметной области «Информационная безопасность» процедура сертификации включает следующие этапы:

1. Формирование верифицированной базы вопросов и ответов.
2. «Прогонка» всех вопросов базы через модель и получение ответов.
3. Сравнение полученных ответов с эталонными.
4. Подсчет процента неправильных ответов.
5. Вывод о возможности применимости модели.

Формирование верифицированной базы вопросов и ответов

РусКрипто

В качестве верифицированной базы вопросов и ответов для предметной области «Информационная безопасность» целесообразно использовать либо фонды оценочных средств по дисциплинам основных образовательных программ в области информационной безопасности, утвержденные ФУМО ВО ИБ, либо фонды оценочных средств для профессиональных экзаменов, разработанные НАРК.

РусКрипто

СПАСИБО
ЗА ВНИМАНИЕ