ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2022, том 508, с. 102–103

ПЕРЕДОВЫЕ ИССЛЕДОВАНИЯ В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И МАШИННОГО ОБУЧЕНИЯ

УДК 004.8

AI-РЕЦЕНЗИРОВАНИЕ ПОЛИГРАФНЫХ СКРИНИНГОВ

© 2022 г. Д. В. Асонов[1], М. А. Крылов[2,*]

Представлено академиком РАН А.Л. Семеновым
Поступило 28.10.2022 г.
После доработки 28.10.2022 г.
Принято к публикации 01.11.2022 г.

Представлен короткий обзор и результаты научного исследования возможности AI-рецензирования полиграфных скринингов. Результаты исследования будут применяться на практике для укрепления внутренней безопасности в ПАО Сбербанк в конце 2022 г. Полностью исследования будут представлены в публикации [1].

Ключевые слова: детекция лжи, полиграф, скрининг, рецензирование, AI

DOI: 10.31857/S2686954322070025

Защита денежных средств и данных клиентов традиционно является одним из столпов банковской культуры и репутации. В качестве одного из инструментов защиты клиентов банки используют полиграфные скрининги (ПС). Финансовая отрасль – не единственная, использующая ПС. Такие критичные с т.з. возможных последствий от внутреннего мошенничества отрасли как авиация, промышленность, правоохранительные структуры, государственные органы во всем мире также используют ПС.

Тема детекции лжи (выявление скрываемой информации) все больше интересует общество, что косвенно видно по количеству научно-популярных “лонгридов” [2–9]. На фоне увеличения значимости корпоративной безопасности и культуры нетерпимости к внутреннему мошенничеству, интенсивность научных исследований в области детекции лжи растет по всему миру, последние 2–3 года публикуется около тысячи научных статей в год. Исследования по теме детекции лжи являются ультра-мультидисциплинарными, ведутся учеными в таких областях, как психофизиология, нейронауки, безопасность, юриспруденция, компьютерные науки и AI, и т.д.

Возможно, самый острый вопрос в научных сообществах по теме, который до сих пор не разрешен уже на протяжении нескольких десятилетий: точность полиграфа, в частности, как ее считать и какова природа ошибок в выводах (ложь выявлена/не выявлена). В докладе мы расскажем, как мы частично ответили на этот вопрос, решив конкретную и практичную научную задачу.

В рамках минимизации рисков внутреннего мошенничества на рисковых направлениях деятельности Банка подразделение внутренней безопасности проводит скрининговые проверки на полиграфе кандидатов на трудоустройство и действующих сотрудников, которые проводятся только с их согласия и в полном соответствии с законодательством. Ежегодно проводится около 6 тысяч таких скринингов. Если предположить, что порядка 5% скринингов имеют ошибочные выводы специалистов-полиграфологов, существуют риски неправильной оценки благонадежности порядка 300 сотрудников. Чтобы избежать подобных ошибок, внутренняя безопасность проводит рецензирование (запрос второго мнения) по неоднозначным результатам скринингов. Это увеличивает расходы на полиграфные скрининги. Мы задумались над тем, как можно запрашивать второе мнение не только по неоднозначным результатам, а по всем без исключения исследованиям, и одновременно снизить расходы и уменьшить вероятность противоречивых результатов скринингов. Так началось данное исследование.

Ошибки полиграфа часто связывают с недостатками конкретного полиграфического метода и вероятностной природой выводов. Сотни научных работ изучили недостатки различных методов. Мы же рассмотрели вид ошибки, исследования которого ранее не опубликовались, и связанного с неоднозначностью принятия решения специалистом, вне зависимости чем она была вызвана: применяемой методикой или внешними факторами.

[1] *ПАО Сбербанк, Блок “Технологии”, Управление исследований и инноваций, Москва, Россия*
[2] *ПАО Сбербанк, Управление внутрибанковской безопасности, Москва, Россия*
*****E-mail: makrylov@sberbank.ru*

С помощью AI мы построили, насколько нам известно – первый в мире, прототип второго мнения для выводов полиграфологов, и пропилотировали его на исторических данных (записях полиграфных скринигов и выводах полиграфологов по ним) [1]. Результаты пилота подтвердили практическую пользу от применения модели для Внутренней Безопасности еще до окончания научного исследования и позволят значительно снизить привлекаемые человеческие ресурсы, временные и финансовые ресурсы на проведение ручного рецензирования. На основе прототипа сейчас завершается внедрение MVP, и в конце Q4'22 MVP начнет вносить дополнительный вклад в укрепление внутренней безопасности в Банке.

Исследование принесло также широкий набор побочных, значимых результатов:

i. Мы первые замерили качество моделей не только на всем датасете (который включает множество риск-факторов), а также на каждом риск-факторе отдельно. Это позволило нам выдвинуть гипотезу о том, что люди реагируют по-разному на собственную ложь при ответе на вопросы по разным темам. Косвенно, мы смогли получить количественную оценку качества вопросов, которые содержательно наполняют проверку отдельной темы. Чем сложнее модели натренироваться – тем менее четко сформулированы вопросы.

ii. Результаты пилота показали, что модели не только находят противоречия в выводах полиграфологов, но и последующий их анализ позволяет выявлять классы ранее неизвестных системных и процессных ошибок.

iii. Пилот показал возможность определять сверхсложные случаи, характеризующиеся тем, что испытуемый применял методы противодействия полиграфной проверке.

iv. Как и во многих других областях исследований в области AI, в детекции лжи практически нет выверенных golden standard датасетов. Запуск MVP и ручная перепроверка скрингов, где мнения модели и полиграфолога разошлись, инициирует быстрое накопление самого большого в мире golden standard по теме.

v. Мы придумали и реализуем возможность для ученых по всему миру далее раздвигать границы познания в этой области и проводить собственные эксперименты на обезличенных данных 2100+ проверок, при этом соблюдая законодательство РФ по трансграничной передаче персональных данных. Самый большой датасет реальных полиграфных проверок, который был доступен ученым, ранее имел размер 149 проверок, и ученые (Carnegie Mellon University) не могли им делиться [10].

СПИСОК ЛИТЕРАТУРЫ

1. *Dmitri Asonov, Maksim Krylov, Vladimir Omelyusik, Anastasiya Ryabikina, Evgeny Litvinov, Maksim Mitrofanov, Maksim Mikhailov*. Albert Efimov. Building a Second-Opinion Tool for Classical Polygraph. 2022, Under review at Scientific Reports, Nature.
2. Truth vs Lies. Special issue of Scientific American, 2022.
3. True story? Lie detection systems go high-tech. BBC, 2022.
4. Lie detectors have been unreliable for over a century but more Britons are being subjected to polygraph tests. iNews.co.uk, 2022.
5. Will Your Cheatin' Heart Tell on You? As Americans Lose Trust in Each Other, They're Turning to Tech to Detect Lies. TheInformation.com, 2022.
6. Lie detectors have always been suspect. AI has made the problem worse. MIT Technology Review, 2020.
7. AI lie detector developed for airport security. Financial Times, 2019.
8. The race to create a perfect lie detector – and the dangers of succeeding. The Guardian, 2019.
9. A New AI That Detects "Deception" May Bring an End to Lying as We Know It. Futurism.com, 2018.
10. *Aleksandra Slavkovic*. Evaluating Polygraph Data. Carnegie Mellon University, 2002.