Аннотация - пересечение гендера и кибербезопасности — это новая область, которая подчеркивает дифференцированные воздействия и риски, с которыми сталкиваются люди, в зависимости от их гендерной идентичности. Традиционные модели безопасности игнорируют гендерные угрозы, такие как онлайн-преследование, доксинг, что приводит к недостаточной защите. В документе исследуется интеграция гендерных и человекоцентричных моделей угроз, ориентированных, подчеркивая необходимость инклюзивных подходов. Используя технологии искусственного интеллекта и машинного обучения, возможно разработать эффективные системы обнаружения угроз и реагирования на них. Кроме того, в документе предлагается основа для разработки и внедрения новых стандартов безопасности. Цель состоит в создании более инклюзивной среды кибербезопасности, учитывающую уникальные потребности и опыт людей, повышая общий уровень безопасности.

I. Введение

Кибербезопасность традиционно рассматривалась через техническую призму, уделяя особое внимание защите систем и сетей от внешних угроз, что игнорирует человеческий фактор, особенно дифференцированное воздействие киберугроз на различные группы. Различные представители групп часто сталкиваются с уникальными киберугрозами, такими как онлайн-преследование, доксинг и злоупотребления с использованием технологий, которые преуменьшаются в традиционных моделях угроз.

Недавние исследования и политические дискуссии начали признавать важность включения гендерных аспектов в кибербезопасность. Например, Рабочая группа открытого состава ООН (OEWG) по ICT подчеркнула необходимость учёта гендерной проблематики при внедрении кибернорм и наращивании гендерно-ориентированного потенциала. Аналогичным образом, структуры, разработанные такими организациями, как Ассоциация прогрессивных коммуникаций (APC), предоставляют рекомендации по созданию гендерно-ориентированной политики кибербезопасности.

Человекоцентричная безопасность отдаёт приоритет решению проблем поведения человека в контексте кибербезопасности и предлагает подход к интеграции гендерных аспектов. Сосредоточив внимание на психологических и интерактивных аспектах безопасности, модели, ориентированные на человека, направлены на создание культуры безопасности, которая расширяет возможности отдельных лиц, уменьшает человеческие ошибки и эффективно снижает киберриски.

II. Успешные тематические исследования моделей гендерных угроз в действии

- **Обнаружение онлайн-преследований** . Платформа социальных сетей внедрила систему на основе искусственного интеллекта для обнаружения и смягчения последствий онлайн-преследований. Согласно UNIDIR использовано NLP для анализа текста на предмет ненормативной лексики и анализа настроений для выявления домогательств, отметив значительное сокращение случаев преследования и повышении удовлетворённости пользователей.

- **Предотвращение доксинга**: разработана модель для обнаружения попыток доксинга путем анализа закономерностей доступа к данным и их совместного использования. Согласно UNIDIR модель использовала контролируемое обучение для классификации инцидентов доксинга и оповещения пользователе, что позволило увеличить на 57% количество случаев обнаружения попыток доксинга и сокращении на 32% число успешных инцидентов.

- **Обнаружение фишинга с учётом гендерного фактора:** Финансовое учреждение внедрило систему обнаружения фишинга, включающую тактику фишинга с учётом пола. Согласно UNIDIR использованы модели BERT, для анализа содержимого электронной почты на предмет и эмоциональных манипуляций и гендерно-ориентированного язык, снизило количество кликов по фишинговым сообщениям на 22% и увеличило количество сообщений о попытках фишинга на 38%.

III. Влияние гендерных предположений в алгоритмах на кибербезопасность

Гендерные предположения в алгоритмах существенно влияют на кибербезопасность различными способами, часто закрепляя предвзятости и создавая уязвимости, которые непропорционально сильно влияют на группы.

A. Разный опыт и модели угроз

- **Поведенческие различия:** исследования показали значительные различия в поведении в области кибербезопасности между мужчинами и женщинами. Женщины часто более осторожны и могут применять иные методы обеспечения безопасности по сравнению с мужчинами.

- **Восприятие и реакция:** Женщины и мужчины по-разному воспринимают угрозы безопасности и реагируют на них. Женщины уделяют приоритетное

внимание различным аспектам безопасности, таким как конфиденциальность и защита от преследований, в то время как мужчины могут больше сосредоточиться на технической защите.

- **Содействие гендерному разнообразию:** Инклюзивность может повысить общую эффективность области так как разнообразные команды привносят разные точки зрения и лучше подготовлены к борьбе с широким спектром угроз.
- **Данные с разбивкой по полу.** Сбор и анализ данных с разбивкой по полу имеет решающее значение для понимания различного воздействия киберугроз на различные гендерные группы. Эти данные могут стать основой для более эффективной и инклюзивной политики кибербезопасности.
- **Укрепление гендерных стереотипов:** Алгоритмы, обученные на предвзятых наборах данных, могут укрепить существующие гендерные стереотипы. Модели машинного обучения, используемые в сфере кибербезопасности, наследуют предвзятость данных, на которых они обучаются, что приводит к гендерным допущениям в механизмах обнаружения угроз и реагирования на них.
- **Некорректная гендерная ориентация:** Платформы соцсетей и другие онлайн-сервисы используют алгоритмы для определения атрибутов пользователя, включая пол, бывают неточными, что приводит к нарушению конфиденциальности.
- **Гендерные последствия киберугроз:** Традиционные угрозы кибербезопасности, такие как атаки типа «отказ в обслуживании», могут иметь гендерные последствия в виде дополнительных проблем безопасности и целенаправленными атаками, которые часто упускаются из виду в гендерно-нейтральных моделях угроз.
- **Предвзятость в обнаружении угроз и реагировании на них** . Автоматизированные системы обнаружения угроз, такие как фильтры электронной почты и симуляции фишинга, могут включать гендерные предположения. Например, симуляции фишинга часто связаны с гендерными стереотипами, что может повлиять на точность и эффективность этих мер безопасности.

IV. Технические различия в обнаружении и распознавании гендерных угроз

Чтобы эффективно обнаруживать и распознавать гендерные угрозы, важно понимать технические различия и человекоориентированные аспекты, которые следует учитывать как с мужской, так и с женской точки зрения.

A. Фишинговые атаки

Подход	Обнаружение	Распознавание
Мужской	технические индикаторы: подозрительные URL-адреса, заголовки электронных писем и типы вложений.	Распространённые тактики фишинга, поддельные страницы входа и срочные запросы информации.
Женский	персонализированный, и гендерно-ориентированный контент в фишинговых электронных письмах, в т.ч., касающихся социальных вопросов, личной безопасности.	Осведомлённость о тактиках эмоционального манипулирования: угрозы причинения личного вреда или эксплуатация социальных отношений.
Человеко-центричный	**Обучение:** гендерно-ориентированные программы повышениия осведомлённости о фишинге, в которых рассматриваются конкретные тактики, используемые для нападения и эмоциональные манипуляции и социальная инженерия.	**Отчётность:** механизмы отчётности, позволяющие пользователям отмечать подозрительные электронные письма и получать отзывы о своих действиях.

B. Интернет-преследование и доксинг

Смотровая площадка	Обнаружение	Распознавание
Мужской	фильтрация ключевых слов и распознавания образов для выявления ненормативной лексики и угроз.	обнаружении прямых угроз и откровенного контента.
Женский	мониторинг гендерных оскорблений, женоненавистнических высказываний и угроз сексуального насилия.	Учет косвенных и скрытых форм преследования: газлайтинг и скоординированные кампании преследования.
Человеко-центричный	**Системы поддержки:** доступ к услугам поддержки, включая консультирование и юридическую помощь.	**Рекомендации:** правила сообщества и механизмы обеспечения их соблюдения

C. Атаки социальной инженерии

Смотровая площадка	Обнаружение	Распознавание
Мужской	Анализ необычных запросов на информацию или доступ, с акцентом на технические аномалии.	Выявление распространённых тактик социальной инженерии, например травля.
Женский	Мониторинг попыток персонализированной социальной инженерии, использующих социальные сети и личные отношения.	социнженерия с учётом гендерной специфики, таких как выдача себя за доверенных лиц или использование функций по уходу.
Человек-оцентричный	**Информационные кампании:** целевые информационные	**Процессы проверки:** надёжные процессы

	кампании, освещающие конкретные тактики социальной инженерии, используемые против женщин.	проверки конфиденциальных запросов, для гарантии, что пользователи смогут легко проверить законность таких запросов.

D. Предвзятость искусственного интеллекта и машинного обучения при обнаружении угроз

Подход	**Обнаружение**	**Распознавание**
Мужской	технические показателях производительности, таких как точность, точность и отзыв.	Оценка производительности модели на основе общего уровня обнаружения угроз.
Женский	Оценка моделей ИИ/ML на гендерную предвзятость, гарантирующая, что модели не неправильно классифицируют и не игнорируют угрозы, направленные против женщин.	Анализ ложноположительных и ложноотрицательных показателей гендерных угроз.
Человеко-центричный	**Предвзятость:** методы уменьшения гендерной предвзятости в моделях ИИ и ML	**Прозрачность:** прозрачность процессов принятия решений AI/ML, позволяя понимать и оспаривать результаты модели.

E. Управление уязвимостями

Подход	**Обнаружение**	**Распознавание**
Мужской	Идентификация технических уязвимостей на основе критичности и возможности использования.	Приоритизация уязвимостей, которые представляют наибольший риск для систем и инфраструктуры.
Женский	Рассмотрение уязвимостей, которые непропорционально разные аспекты, например, связанные с личной безопасностью и конфиденциальностью	анализ социальных и психологических последствий определённых уязвимостей
Человеко-центричный	**Инклюзивная оценка рисков:** оценка рисков, конкретных потребностей и уязвимостей различных групп.	**Фокус на человека:** меры безопасности, в которых приоритет отдается безопасности и конфиденциальности пользователей

V. Отличие моделей с точки зрения технической реализации

Гендерные модели угроз существенно отличаются от традиционных моделей угроз с точки зрения технической реализации. Эти различия возникают из-за необходимости устранения конкретных уязвимостей и векторов атак, которые непропорционально сильно затрагивают определённые категории или поведенческие признаки. Благодаря учёту этих технических различий и аспектов меры кибербезопасности могут быть более эффективными в обнаружении и распознавании угроз. Такой подход гарантирует учёт уникального опыта и уязвимостей различных групп, и разработки инклюзивных и комплексных стратегий кибербезопасности.

A. Алгоритмы обнаружения угроз

Аспект	**Традиционные модели угроз**	**Гендерные модели угроз**
Фокус	Сетевые атаки, вредоносное ПО, внешние угрозы.	Злоупотребления с использованием технологий (преследование, преследование, сексуальное насилие на основе изображений).
Подход	Обнаружение на основе сигнатур, обнаружение аномалий, эвристический анализ.	Улучшенное NLP для выявления нецензурной лексики, анализ настроений для выявления преследований, распознавание образов преследования.
Источники данных	Сетевой трафик, системные журналы, данные конечных точек.	Взаимодействие в социальных сетях, SMS-сообщения, данные GPS, модели использования устройств.

B. Аналитика поведения пользователей (UBA)

Аспект	**Традиционные модели угроз**	**Гендерные модели угроз**
Фокус	Выявление отклонений от нормального поведения пользователей для обнаружения инсайдерских угроз или взлома учётных записей.	Выявление принудительного контроля и злоупотреблений путем отслеживания признаков принудительного или принудительного изменения поведения.
Подход	Модели машинного обучения, обученные на типичных шаблонах активности пользователей.	Поведенческий анализ с использованием контекстно-зависимых моделей для выявления необычных признаков и закономерностей.
Источники данных	Время входа в систему, шаблоны доступа, использование файлов.	Журналы связи, история местоположений, журналы доступа к устройствам.

C. Системы реагирования на инциденты

Аспект	**Традиционные модели угроз**	**Гендерные модели угроз**
Фокус	Автоматизированное реагирование на инциденты для быстрого и эффективного	Предоставление индивидуальной поддержки и ресурсов жертвам злоупотреблений с использованием

	устранения угроз.	технологий.
Подход	Сценарии действий, автоматизированные сценарии, предопределенные ответные действия.	Интеграция систем человеческой поддержки (горячие линии, консультирование) с автоматизированными механизмами реагирования.
Источники данных	Оповещения от SIEM-систем, инструментов EDR, решений для мониторинга сети.	Отчеты пользователей, мониторинг соцсетей, прямая связь со службами поддержки.

D. Управление уязвимостями

Аспект	**Традиционные модели**	**Гендерные модели**
Фокус	Выявление и исправление уязвимостей программного обеспечения для предотвращения эксплуатации.	Устранение уязвимостей, которые могут быть использованы для гендерного насилия (отслеживание местоположения, несанкционированный доступ к персональным данным).
Подход	Регулярное сканирование уязвимостей, системы управления исправлениями, оценки рисков.	Расширенный контроль конфиденциальности, безопасная настройка служб определения местоположения, регулярные проверки доступа к личным данным.
Источники данных	Базы данных уязвимостей, конфигурации системы, сканирование сети.	Настройки устройства, разрешения приложений, отзывы пользователей.

E. Смягчение предвзятости AI/ML

Аспект	**Традиционные модели**	**Гендерные модели**
Фокус	Обеспечение точности и эффективности моделей искусственного интеллекта и машинного обучения при обнаружении угроз.	Уменьшение гендерной предвзятости в моделях AI/ML для обеспечения справедливого и равноправного обнаружения угроз.
Подход	Обучение модели на различных наборах данных, регулярная оценка производительности, алгоритмы обнаружения систематических ошибок.	Алгоритмы, учитывающие справедливость, включение данных обучения с гендерным разнообразием, регулярные проверки на предмет предвзятости.
Источники данных	Исторические данные об угрозах, сетевой трафик, системные журналы.	Данные с разбивкой по полу, отзывы пользователей, отчёты о происшествиях.

F. Человекоцентричные аспекты

Аспект	**Традиционный**	**Гендерный**
Программы обучения и повышения осведомлённости	Общий тренинг по кибербезопасности, например с упором на фишинг, вредоносное ПО и гигиену паролей.	Специализированное обучение, посвящённое гендерным угрозам, а также тому, как распознавать их и сообщать о них.
Системы поддержки	Автоматизированные системы поддержки, такие как часто задаваемые вопросы, веб-формы и чат-боты.	Доступ к службам поддержки, включая консультации, юридическую помощь и специальные горячие линии для жертв злоупотреблений с использованием технологий.
Правила сообщества и правоприменение	Общие правила сообщества по приемлемому поведению в Интернете.	Четкие руководящие принципы по борьбе с притеснениями и злоупотреблениями по признаку пола, а также надежные механизмы правоприменения для защиты уязвимых пользователей.
Контроль конфиденциальности и безопасности	Стандартные настройки конфиденциальности и элементы управления безопасностью устройств и приложений.	Расширенные средства контроля конфиденциальности, такие как безопасный обмен данными о местоположении, использование анонимных данных и строгий контроль доступа к личной информации.
Совместное моделирование угроз	Моделирование угроз, проводимое экспертами по кибербезопасности с упором на технические угрозы.	Включение различных групп в процессы моделирования угроз для выявления и устранения гендерных уязвимостей и векторов атак.

VI. Методология оценки влияния гендерного разнообразия на кибербезопасность

A. Определение цели и области применения

- **Цель:** оценить, как гендерное разнообразие в командах по кибербезопасности влияет на эффективность обнаружения угроз, реагирования и общую эффективность кибербезопасности.
- **Область применения:** включает различные области кибербезопасности, такие как реагирование на инциденты, анализ угроз, управление уязвимостями и обнаружение предвзятости AI/ML.

B. Установка базовых показателей

- **Сбор исходных данных:** собрать исходные данные о текущих показателях эффективности кибербезопасности как от различных команд, включая команды с гендерным разнообразием.

- **Метрики для сбора:**
  - Уровень обнаружения инцидентов
  - Время ответа (MTTR)
  - Фишинговый рейтинг кликов
  - Количество обнаруженных уязвимостей
  - Удовлетворенность работой и уровень удержания
  - Показатели финансовой деятельности
  - Уровни смещения модели AI/ML

C. Анализ состава команды

- **Оценка разнообразия:** оценка гендерного состава существующих команд кибербезопасности.
- **Точки данных:**
  - Процент женщин в команде
  - Роли и обязанности членов команды
  - Руководящие должности занимают женщины

D. Проектирование и реализация задач кибербезопасности

- **Разработка сценария:** репрезентативные проблемы кибербезопасности для решения командой.
  - Имитация фишинговых атак
  - Сценарии реагирования на инциденты
  - Задачи обнаружения уязвимостей
  - Обучение модели AI/ML и упражнения по выявлению предвзятости

E. Измерение производительности

1) Количественные показатели

- **Обнаружение инцидентов и реагирование:** Измерение количества обнаруженных инцидентов и среднее время реагирования.
- **Фишинг и социальная инженерия:** отслеживание рейтинга кликов по фишинговым электронным письмам и обнаружение персонализированных атак социальной инженерии.
- **Управление уязвимостями:** количество обнаруженных и ответственно раскрытых уязвимостей.
- **Предвзятость AI/ML:** оценка уровня гендерной предвзятости в моделях AI/ML, используемых для обнаружения угроз.

2) Качественные показатели

- **Удовлетворённость работой:** опрос для измерения удовлетворенности работой среди членов команды.
- **Сотрудничество в команде:** наблюдение и оценка динамики команды и сотрудничества во время проблем кибербезопасности.

F. Сбор и анализ данных

- **Инструменты сбора данных:** автоматизированные инструменты и ручные методы для сбора данных во время киберпроблем.
- **Статистический анализ:** статистические тесты для сравнения показателей эффективности команд с гендерным разнообразием и команд, в которых доминируют мужчины.
- **Качественный анализ:** анализ ответов на опросы и данные наблюдений, чтобы выявить закономерности и идеи, связанные с динамикой команды и удовлетворенностью работой.

G. Отчетность и обратная связь

- **Создание отчёта:** объединение результаты в комплексный отчёт, в котором подчёркивается влияние гендерного разнообразия на показатели кибербезопасности.
- **Сеансы обратной связи:** обратная связь с членами команды, чтобы обсудить результаты и собрать дополнительную информацию.

H. Постоянное улучшение

- **Итеративный процесс:** использование результатов для уточнения методологии и улучшения будущих оценок.
- **Рекомендации:** предоставить практические рекомендации для организаций по расширению гендерного разнообразия и улучшению результатов в области кибербезопасности.

VII. Реализация методологии. Практические результаты

A. Эффективность разнообразия в моделях киберугроз

Метрика	**Общая команда**	**Разнообразная команда**	**Улучшение**
Обнаружены попытки доксинга	100	157	+57%
- Ложные срабатывания	20	15	-25%
- Ложноотрицательные результаты	10	5	-50%
Скоординированные кампании преследования заблокированы	50	66	+32%
- Ложные срабатывания	10	8	-20%
- Ложноотрицательны	5	3	-40%

е результаты			
Фишинговый рейтинг кликов	20%	15,6%	-22%
- Эффективность обучения	70%	85%	+15%
- Частота сообщений пользователей	60%	75%	+25%
Обнаружены персонализированные атаки социальной инженерии	100	138	+38%
- Ложные срабатывания	15	10	-33%
- Ложноотрицательные результаты	8	5	-38%
Гендерная предвзятость в моделях искусственного интеллекта и машинного обучения	10%	5,1%	-49%
- Точность обнаружения смещения	80%	90%	+12,5%
- Эффективность смягчения предвзятости	70%	85%	+21%
Обнаружены уязвимости	100	137	+37%
- Критические уязвимости	30	45	+50%
- Уязвимости с низким уровнем риска	70	92	+31%
Удовлетворённость работой (женщины)	70%	76%	+6%
- Коэффициент удержания	85%	92%	+7%
- Скорость продвижения	14%	18%	+29%
Женщины на руководящих должностях	17%	20%	+3%
- Участие в развитии лидерства	50%	65%	+30%
- Участие в программе наставничества	40%	55%	+38%
Коэффициент удержания сотрудников	85%	92%	+7%
- Добровольный оборот	10%	7%	-30%
- Непроизвольный оборот	5%	3%	-40%
Время реагирования на инциденты (MTTR)	4 часа	3,2 часа	-20%
- Время обнаружения	2 часа	1,5 часа	-25%
- Время сдерживания	1 час	0,8 часа	-20%
- Время восстановления	1 час	0,9 часа	-10%
Улучшение финансовых показателей	0%	27%	+27%
- Рост выручки	5%	10%	+100%
- Экономия затрат	3%	7%	+133%
Индекс креативности и решения проблем	70	90	+20 баллов
- Уровень инноваций	60%	75%	+25%
- Разнообразие решений	65%	80%	+23%
Индекс адаптивности и устойчивости	75	90	+15 баллов
- Эффективность управления изменениями	70%	85%	+21%
- Эффективность реагирования на кризисы	75%	90%	+20%
Соблюдение этических стандартов	80%	95%	+15%
- Соблюдение политики	85%	95%	+12%
- Уровень успешности аудита	90%	98%	+9%
Глобальная перспектива и понимание рынка	70	85	+15 баллов
- Адаптивность рынка	65%	80%	+23%
- Культурная компетентность	70%	85%	+21%
Уменьшение группового мышления	60	80	+20 баллов
- Генерация разнообразных идей	65%	80%	+23%
- Качество принятия решений	70%	85%	+21%

B. Конкретные показатели, показывающие улучшение производительности благодаря гендерному разнообразию

Эти показатели показывают, как гендерное разнообразие может привести к повышению производительности в различных аспектах кибербезопасности, от обнаружения угроз и реагирования

на них до принятия решений и финансовых результатов. Создавая инклюзивную среду и используя различные точки зрения, организации могут улучшить свою общую позицию и эффективность в области кибербезопасности.

Категория	Метрика	Результат
Обнаружение инцидентов и реагирование на них	Повышение уровня обнаружения угроз гендерного характера	заблокировано на 57% больше попыток доксинга и на 32% больше скоординированных кампаний преследования.
Фишинг и социальная инженерия	Снижен рейтинг кликов по фишинговым письмам и улучшено выявление персонализированных атак социальной инженерии.	фишинговые тесты компании показали на 22% меньше кликов по стереотипным приманкам, но на 38% больше кликов по персонализированным атакам социальной инженерии.
Алгоритмическая справедливость и снижение предвзятости	Снижение гендерной предвзятости в моделях искусственного интеллекта/ML, используемых для обнаружения угроз.	снижение гендерной предвзятости в своих моделях обнаружения угроз на 49% после внедрения методов устранения предвзятости и проверки данных обучения.
Управление уязвимостями	Более широкое обнаружение и ответственное раскрытие уязвимостей, затрагивающих маргинализированные группы.	обнаружено и ответственно раскрыто на 37% больше уязвимостей, затрагивающих различных группы
Принятие решений и финансовые результаты	Улучшение процессов принятия решений и финансовых показателей	повышение на 27% больше шансов добиться финансового успеха.
Удовлетворенность работой и удержание	Высокая удовлетворенность работой и уровень удержания среди специалистов по кибербезопасности	Удовлетворённость работой в области кибербезопасности повышена на 76%
Лидерство и карьерный рост	Более высокий процент женщин на руководящих и управленческих должностях	Изменение в сторону широкого участия в принятии решений о найме.
Инициативы по разнообразию и инклюзивности	Вовлечение и эффективность инициатив в области разнообразия, справедливости и инклюзивности (DEI)	инвестирующие в инициативы DEI, выше доля вовлеченных женщин, и они испытывают меньшую нехватку киберперсонала.
Креативность и решение проблем	Расширение творческих способностей и возможностей решения проблем.	повышение креативности в решении проблем компании благодаря разным взглядам и жизненному опыту.
Адаптивность и устойчивость	Повышение адаптивности и устойчивости команд кибербезопасности.	повышение адаптивности и устойчивости, что имеет значение для постоянно развивающейся области кибербезопасности.

C. Объяснение расширенных показателей и значений

- **Обнаружено попыток доксирования:** количество выявленных и предотвращенных попыток доксирования.
  - **Ложные срабатывания:** неправильно отмеченные попытки доксинга.
  - **Ложноотрицательные результаты:** пропущенные попытки доксинга.
- **Заблокированные скоординированные кампании преследования:** количество перехваченных кампаний преследования.
  - **Ложные срабатывания:** неправильно отмеченные кампании преследования.
  - **Ложноотрицательные результаты:** Пропущенные кампании преследования.
- **Показатель кликабельности по фишинговым сообщениям:** процент сотрудников, которые нажимали на фишинговые электронные письма.
  - **Эффективность обучения:** Эффективность обучения по предотвращению фишинга.
  - **Уровень сообщений пользователей:** процент пользователей, сообщающих о попытках фишинга.
- **Обнаружено персонализированных атак социальной инженерии:** количество выявленных персонализированных атак.
  - **Ложные срабатывания:** неправильно отмеченные атаки социальной инженерии.
  - **Ложноотрицательные результаты:** пропущенные атаки социальной инженерии.
- **Гендерная предвзятость в моделях AI/ML:** процент гендерной предвзятости в моделях AI/ML.
  - **Точность обнаружения смещения:** Точность обнаружения смещения в моделях.
  - **Эффективность смягчения предвзятости:** эффективность методов смягчения предвзятости.
- **Обнаружено уязвимостей:** количество выявленных уязвимостей безопасности.
  - **Критические уязвимости:** обнаружены уязвимости высокого риска.
  - **Уязвимости с низким уровнем риска:** обнаружены уязвимости с низким уровнем риска.

- **Удовлетворенность работой (женщины):** процент женщин, выражающих удовлетворенность работой.
  - **Коэффициент удержания:** процент сохраненных сотрудников.
  - **Уровень продвижения по службе:** процент женщин, получивших повышение.
- **Женщины на руководящих должностях:** Процент женщин на руководящих должностях.
  - **Участие в развитии лидерства:** Участие в программах развития лидерства.
  - **Участие в программе наставничества:** Участие в программах наставничества.
- Коэффициент удержания сотрудников: процент сохраненных сотрудников.
  - **Добровольная текучесть:** процент сотрудников, увольняющихся по собственному желанию.
  - **Вынужденная текучесть кадров:** процент сотрудников, увольняющихся принудительно.
- **Время реагирования на инциденты (MTTR):** Среднее время реагирования на инциденты.
  - **Время обнаружения:** время обнаружения инцидентов.
  - **Время сдерживания:** время сдерживания инцидентов.
  - **Время восстановления:** время восстановления после инцидентов.
- **Улучшение финансовых показателей:** процентное улучшение финансовых показателей.
  - **Рост доходов:** процентный рост доходов.
  - **Экономия затрат:** Процентная экономия затрат.
- **Индекс креативности и решения проблем:** Качественный показатель креативности и решения проблем.
  - **Уровень инноваций:** уровень инноваций в решениях.
  - **Разнообразие решений:** Разнообразие предлагаемых решений.
- **Индекс адаптивности и устойчивости:** Качественный показатель адаптивности и устойчивости.
  - **Эффективность управления изменениями:** Эффективность управления изменениями.
  - **Эффективность реагирования на кризисы:** Эффективность реагирования на кризисы.
- Соблюдение этических стандартов: Процент соблюдения этических стандартов.
  - **Соблюдение политики:** Соблюдение политики.
  - **Процент успешных проверок:** Уровень успешных проверок.
- **Глобальная перспектива и понимание рынка:** Качественный показатель глобальной перспективы и понимания рынка.
  - **Адаптивность рынка:** Адаптивность к изменениям рынка.
  - **Культурная компетентность:** Компетентность в понимании культуры.
- **Снижение группового мышления:** Качественная мера снижения группового мышления.
  - **Генерация разнообразных идей:** Генерация разнообразных идей.
  - **Качество принятия решений:** Качество принятия решений.

D. Основные компоненты

Важно отметить, что используемое конкретное оборудование, программное обеспечение и алгоритмы AI/ML будут зависеть от требований, бюджета и существующей инфраструктуры организации. Кроме того, комплексная стратегия кибербезопасности должна включать в себя сочетание технических средств контроля, политик, процессов и человеческого опыта для эффективного управления и смягчения киберрисков.

1) Аппаратные компоненты

- **Сетевые датчики:** Аппаратные устройства, такие как сетевые TAP, коммутаторы зеркалирования портов и устройства перехвата пакетов, могут быть развёрнуты для мониторинга сетевого трафика и сбора данных для анализа.
- **SIEM:** Выделенные аппаратные устройства SIEM могут собирать и анализировать журналы безопасности, события и сетевые данные из различных источников.
- **EDR** . Агенты EDR, установленные на конечных точках (серверах, рабочих станциях и т. д.), могут собирать данные о деятельности системы и пользователей для обнаружения угроз и реагирования на инциденты.
- **Сканеры уязвимостей** . Аппаратные сканеры уязвимостей могут выполнять комплексное сканирование для выявления уязвимостей в сети, системах и приложениях.
- **Honeypots:** специализированные аппаратные устройства могут использоваться в качестве приманок для привлечения и анализа потенциальных киберугроз.

2) Программные компоненты

- **Программные SIEM:** программные решения SIEM могут собирать, анализировать и коррелировать события безопасности из различных источников, обеспечивая мониторинг в реальном времени и возможности реагирования на инциденты.
- **Программное обеспечение для управления уязвимостями:** эти инструменты автоматизируют процессы сканирования уязвимостей, определения приоритетов и исправлений, помогая эффективно выявлять и устранять уязвимости.

- **Программное обеспечение EPR:** Программное обеспечение EPR может отслеживать и защищать конечные точки, обнаруживать угрозы и реагировать на них, а также предоставлять возможности судебно-медицинского анализа.
- **Программное обеспечение для анализа поведения пользователей и объектов (UEBA).** Программное обеспечение UEBA может анализировать модели поведения пользователей и объектов для обнаружения аномалий и потенциальных внутренних угроз.
- **Программное обеспечение для моделирования фишинга:** эти инструменты могут моделировать фишинговые атаки и проводить обучение сотрудников по вопросам безопасности, помогая измерять и улучшать их способность выявлять потенциальные угрозы и сообщать о них.
- **Программное обеспечение для обеспечения соответствия требованиям и управления рисками.** Эти решения могут помочь организациям управлять и контролировать соблюдение различных стандартов и правил безопасности, предоставляя возможности отчётности и аудита.

3) *Алгоритмы и методы AI/ML:*

- **Обнаружение аномалий.** Алгоритмы искусственного интеллекта и машинного обучения, такие как изолирующие леса, автокодировщики и SVM для обнаружения аномалий в сетевом трафике, поведении пользователей и активности системы, указывающих на потенциальные угрозы.
- **Обнаружение вредоносных программ.** Модели машинного обучения, такие как случайные леса, глубокие нейронные сети и машины опорных векторов, можно обучить обнаружению и классификации вредоносных программ на основе функций статического и динамического анализа.
- **Обнаружение и предотвращение вторжений.** Методы искусственного интеллекта и машинного обучения, такие как деревья решений, логистическая регрессия и глубокое обучение, могут применяться к сетевому трафику и системным журналам для выявления и предотвращения вторжений и кибератак.
- **Аналитика поведения пользователей и объектов (UEBA):** алгоритмы неконтролируемого обучения, такие как кластеризация и уменьшение размерности, могут использоваться для установления базовых показателей нормального поведения пользователей и объектов, что позволяет обнаруживать аномалии и потенциальные внутренние угрозы.
- **Приоритизация уязвимостей.** Модели машинного обучения можно обучить расставлять приоритеты уязвимостей на основе таких факторов, как серьезность, возможность использования и потенциальное воздействие, что помогает организациям более эффективно сосредоточить свои усилия по устранению.
- **Прогнозное обслуживание и оценка рисков.** Алгоритмы искусственного интеллекта и машинного обучения могут анализировать исторические данные и системные журналы, чтобы прогнозировать потенциальные сбои, инциденты безопасности и риски, обеспечивая упреждающие меры и стратегии смягчения последствий.
- **Обработка естественного языка (NLP).** Методы NLP могут применяться для анализа и извлечения информации из неструктурированных источников данных, таких как отчёты о безопасности, потоки информации об угрозах и отчёты об инцидентах, что расширяет возможности обнаружения угроз и реагирования на них.

4) *Методика измерения*

- **Сбор данных.** данные из различных источников, включая сетевой трафик, системные журналы, действия пользователей, сканирование уязвимостей и каналы аналитики угроз.
- **Предварительная обработка данных:** очистка, нормализация и преобразование собранных данных в формат, подходящий для анализа и обучения модели.
- **Разработка характеристики:** характеристики из предварительно обработанных данных, которые можно использовать в качестве входных данных для моделей AI/ML.
- **Обучение и проверка моделей:.** обучение **и** проверка моделей AI/ML с использованием извлечённых характеристик и размеченных данных (если они доступны) для контролируемых задач обучения или методов неконтролируемого обучения для обнаружения аномалий и кластеризации.
- **Развёртывание и интеграция моделей:** обученные модели искусственного интеллекта и машинного обучения в соответствующих аппаратных и программных компонентах для мониторинга в реальном времени, обнаружения угроз и реагирования на инциденты.
- **Непрерывный мониторинг и улучшение.** отслеживание производительности развёрнутых моделей искусственного интеллекта и машинного обучения, обратная связь и переобучение или обновление модели по мере необходимости, чтобы повысить их точность и эффективность.
- **Отчётность и визуализация.** Разработка информационных панелей и инструментов отчётности для визуализации и передачи показателей кибербезопасности и ключевых показателей эффективности, полученных на основе моделей искусственного интеллекта/ML и других компонентов мониторинга.

- **Соблюдение требований и аудит:** интеграция собранных данных, показателей, отчётов в процессы обеспечения соответствия и управления рисками, позволяя проводить аудит и демонстрировать соблюдение стандартов и правил безопасности.

VIII. Тонкая настройка моделей с помощью искусственного интеллекта и машинного обучения

Путём тонкой настройки гендерных моделей с помощью искусственного интеллекта и машинного обучения организации смогут лучше решать уникальные проблемы, связанные с гендерно-специфичными угрозами, что приведет к более инклюзивным и эффективным стратегиям кибербезопасности.

A. Обучение ИИ и ML распознаванию гендерных угроз

1) Сбор и предварительная обработка данных

- **Разнообразные наборы данных:** наборы данных, включающие примеры угроз гендерного характера, с достаточной степенью разнообразия данных и репрезентативностью для разных гендерных групп.
- **Маркировка:** разметка данных, указывающая тип угрозы и гендерный контекст. Это помогает в обучении моделей обучения распознавать гендерные закономерности уязвимостей.

2) Особенности проектирования

- **Методы NLP:** использование NLP для извлечения характеристик из текстовых данных, таких как анализ настроений, извлечение ключевых слов и контекстно-зависимые внедрения. Это помогает выявить оскорбительные выражения и тактики эмоционального манипулирования.
- **Поведенческий анализ:** извлечение характеристик, связанных с поведением пользователей, такие как модели общения, данные о местоположении и использование устройств. Это помогает выявить принудительный контроль и жестокое обращение.

3) Модельное обучение

- **Контролируемое обучение:** обучение модели машинного обучения, такие как случайные леса, SVM и глубокие нейронные сети, на маркированных наборах данных для классификации и обнаружения угроз, специфичных для пола.
- **Обучение без учителя:** алгоритмы кластеризации и обнаружения аномалий для выявления необычных закономерностей, указывающих на гендерные уязвимости, таких как внезапные изменения в моделях общения или данных о местоположении.

4) Смягчение предвзятости

- **Алгоритмы**. внедрение алгоритмов и методов, учитывающих изменение веса, состязательное искажение, чтобы уменьшить предвзятость в моделях искусственного интеллекта и машинного обучения.
- **Регулярные проверки:** регулярные проверки моделей искусственного интеллекта и ML, чтобы убедиться, что они эффективны в обнаружении гендерных угроз.

5) Непрерывное обучение

- **Обратная связь:** пользователи могут сообщать о ложноположительных и ложноотрицательных результатах, что помогает постоянно повышать точность и справедливость модели.

B. Примеры тонкой настройки гендерных моделей с помощью искусственного интеллекта и машинного обучения

1) Обнаружение фишинга и реагирование на него

- **Обнаружение:** использование методов NLP для обнаружения фишингового контента с учётом пола. Например, фишинговые электронные письма, нацеленные на женщин, могут использовать социальные проблемы или проблемы личной безопасности.
- **Алгоритм:** модели BERT или GPT-3, для анализа содержимого электронной почты на предмет гендерно-ориентированного языка и тактики эмоционального манипулирования.
- **Данные:** обучение моделей на наборах данных, которые включают примеры попыток фишинга с учётом гендерной специфики.
- **Решение:** внедрение персонализированных программ обучения, посвящённых тактике фишинга с учётом гендерной специфики с обратной связью в режиме реального времени.
- **Показатели:** измерение снижения количества кликов по фишинговым страницам и увеличение количества сообщений пользователей о попытках фишинга с учётом пола.

2) Онлайн-преследование и обнаружение доксинга

- **Обнаружение:** модели с помощью анализа настроений и контекстно-зависимого NLP для обнаружения тонких форм домогательств и доксинга, нацеленных на женщин.
- **Алгоритм:** модели анализа настроений и контекстно-зависимые внедрения, чтобы понять контекст и эмоциональный тон сообщений.
- **Данные:** наборы данных с примерами гендерных домогательств и случаев доксинга.
- **Решение:** системы поддержки, включающие доступ к консультациям и юридической помощи, руководящие принципы сообщества.
- **Показатели:** оценка снижения количества ложноотрицательных результатов притеснений по гендерному признаку и эффективность систем поддержки в оказании помощи жертвам.

3) Управление уязвимостями

- **Обнаружение:** уязвимости, непропорционально затрагивающие разные группы.

- **Алгоритм:** модели машинного обучения для определения приоритетности уязвимостей на основе их потенциального воздействия на личную безопасность и конфиденциальность.
- **Данные:** обучение модели на наборах данных, которые включают примеры уязвимостей, используемых для гендерного насилия.
- **Решение:** расширенный контроль конфиденциальности и безопасную настройку служб определения местоположения с регулярной проверкой доступа к персональным данным.
- **Метрики:** количество выявленных и устраненных критических уязвимостей, влияющих на личную безопасность и конфиденциальность.

4) Обнаружение и смягчение предвзятости AI/ML

- **Обнаружение:** обнаружение предвзятости в моделях искусственного интеллекта и машинного обучения, используемых для обнаружения угроз.
- **Алгоритм:** алгоритмы и методы, учитывающие изменение веса, состязательное искажение, чтобы уменьшить предвзятость в моделях искусственного интеллекта и машинного обучения
- **Данные:** наборы обучающих данных разнообразны и репрезентативны для разных гендерных групп.
- **Решение:** проверка моделей на предмет гендерной предвзятости и применяйте методы устранения предвзятости и прозрачность процессов принятия решений в области искусственного интеллекта и машинного обучения.
- **Метрики:** оценка снижение гендерной предвзятости в результатах модели и эффективность методов устранения предвзятости.

5) Аналитика поведения пользователей (UBA)

- **Обнаружение:** признаки принудительного контроля и злоупотреблений путем анализа моделей поведения пользователей.
- **Алгоритм:** алгоритмы обучения без учителя, такие как кластеризация и обнаружение аномалий, для выявления необычных закономерностей, указывающих на злоупотребления.
- **Данные:** наборы данных с примерами принудительного контроля и злоупотреблений.
- **Решение:** системы поддержки жертвам злоупотребления технологиями, включая консультирование и юридическую помощь.
- **Показатели:** эффективность выявления принудительного контроля и поддержки, оказываемой жертвам.

C. Комбинированные показатели для традиционных и гендерных моделей угроз

Метрика	Традиционная модель угроз	Модель гендерных угроз
Время обнаружения (TTD)	Измеряет продолжительность от начала атаки до ее обнаружения.	Расширено с дополнительным акцентом на обнаружение гендерных угроз, таких как домогательства.
Время отвечать (TTR)	Измеряет время от обнаружения до разрешения инцидента.	Расширен с акцентом на реагирование на гендерные угрозы, такие как доксинг и преследование.
Уровень обнаружения инцидентов	Процент обнаруженных инцидентов от общего числа инцидентов.	Включает выявление инцидентов гендерного характера, таких как онлайн-преследование и насилие.
Ложноположительный показатель	Процент не-угроз, ошибочно идентифицированных как угрозы.	Расширено с упором на уменьшение количества ложных срабатываний при обнаружении гендерных угроз.
Ложноотрицательный показатель	Процент угроз, ошибочно идентифицированных как не представляющие угрозу.	Расширено с упором на уменьшение количества ложноотрицательных результатов при обнаружении гендерных угроз.
Фишинговый рейтинг кликов	Процент пользователей, которые нажимают на фишинговые электронные письма.	Дополнено дополнительным анализом тактик фишинга с учётом гендерной специфики.
Частота сообщений пользователей	Процент пользователей, сообщающих о подозрительных действиях.	Расширено с упором на сообщение об угрозах гендерного характера.
Зрелость обнаружения и реагирования (DRM)	Измеряет зрелость процессов обнаружения и реагирования с использованием таких моделей, как NIST или CMMI.	Расширено с дополнительным упором на зрелость в борьбе с гендерными угрозами.
Точность обнаружения смещения	Точность обнаружения систематических ошибок в моделях AI/ML.	Расширено с особым упором на гендерную предвзятость в моделях обнаружения угроз.
Эффективность смягчения	Эффективность методов,	Расширено с особым упором на

предвзятости	используемых для смягчения систематических ошибок в моделях AI/ML.	смягчение гендерных предубеждений.
Обнаружены уязвимости	Количество выявленных и зарегистрированных уязвимостей.	Расширено с дополнительным акцентом на уязвимости, которые непропорционально сильно затрагивают женщин.
Критические уязвимости	Количество выявленных уязвимостей высокого риска.	Расширено с акцентом на критические уязвимости, влияющие на личную безопасность и конфиденциальность.
Удовлетворение от работы	Процент сотрудников, выражающих удовлетворенность работой.	Расширено с дополнительным акцентом на удовлетворенности женщин, занимающихся кибербезопасностью.
Коэффициент удержания сотрудников	Процент сотрудников, удержанных за определенный период.	Расширено с дополнительным акцентом на удержании женщин на должностях в области кибербезопасности.
Скорость продвижения	Процент сотрудников, получивших повышение.	Расширено с дополнительным упором на повышение по службе для женщин.
Время реагирования на инциденты (MTTR)	Среднее время реагирования на инциденты безопасности и смягчения их последствий.	Расширено с дополнительным акцентом на время реагирования на инциденты, связанные с гендерной принадлежностью.
Улучшение финансовых показателей	Процентное улучшение финансовых показателей связано с мерами по кибербезопасности.	Расширен с дополнительным акцентом на финансовые последствия смягчения гендерных угроз.
Индекс креативности и решения проблем	Качественная мера творческих способностей и способностей к решению проблем.	Расширено с дополнительным акцентом на различные точки зрения на решение проблем.
Индекс адаптивности и устойчивости	Качественная мера адаптивности и устойчивости команды.	Расширено с дополнительным акцентом на устойчивость в борьбе с гендерными угрозами.
Соблюдение этических стандартов	Процент соблюдения этических стандартов в практике кибербезопасности.	Расширено с дополнительным упором на этическое реагирование на гендерные угрозы.
Глобальная перспектива и понимание рынка	Качественный показатель глобальной перспективы и понимания рынка.	Расширено с дополнительным упором на понимание гендерных рыночных нюансов.
Уменьшение группового мышления	Качественный показатель снижения группового мышления и продвижения различных точек зрения.	Расширено с дополнительным упором на уменьшение группового мышления за счет гендерного разнообразия.

D. Объяснение метрик

- **Время обнаружения (TTD):** измеряет, насколько быстро организация может выявить киберугрозу. Более короткое TTD указывает на лучшие возможности обнаружения.
- **Время реагирования (TTR):** измеряет, насколько быстро организация может отреагировать на обнаруженную угрозу. Более короткий TTR указывает на более эффективные стратегии реагирования.
- **Уровень обнаружения инцидентов:** процент фактически обнаруженных инцидентов от общего числа инцидентов. Более высокие показатели указывают на лучшие возможности обнаружения.
- **Уровень ложноположительных результатов:** процент не-угроз, ошибочно идентифицированных как угрозы. Более низкие показатели указывают на более точное обнаружение.
- **Уровень ложноотрицательных результатов:** процент угроз, ошибочно идентифицированных как не представляющие угрозы. Более низкие показатели указывают на меньшее количество пропущенных угроз.
- **Фишинговый рейтинг кликов:** процент пользователей, которые нажимают на фишинговые электронные письма. Более низкие показатели указывают на лучшую осведомлённость и обучение пользователей.
- **Уровень сообщений пользователей:** процент пользователей, сообщающих о подозрительных действиях. Более высокие показатели указывают на

лучшее взаимодействие и осведомлённость пользователей.

- **Зрелость обнаружения и реагирования (DRM):** оценивает зрелость процессов обнаружения и реагирования в организации с использованием таких моделей, как NIST или CMMI.
- **Точность обнаружения систематических ошибок:** измеряет точность обнаружения систематических ошибок в моделях искусственного интеллекта и машинного обучения. Более высокая точность указывает на лучшее обнаружение смещения.
- **Эффективность смягчения предвзятости:** измеряет эффективность методов, используемых для смягчения предвзятости в моделях искусственного интеллекта и машинного обучения. Более высокая эффективность указывает на лучшее смягчение предвзятости.
- **Обнаружено уязвимостей:** количество выявленных и зарегистрированных уязвимостей. Более высокие цифры указывают на лучшее управление уязвимостями.
- **Критические уязвимости:** количество выявленных уязвимостей высокого риска. Более высокие цифры указывают на лучшее выявление критических рисков.
- **Удовлетворенность работой:** процент сотрудников, выражающих удовлетворенность работой. Более высокие показатели указывают на лучшие условия на рабочем месте.
- **Коэффициент удержания сотрудников:** процент сотрудников, удержанных в течение определенного периода. Более высокие показатели указывают на лучшие стратегии удержания.
- **Уровень продвижения по службе:** процент сотрудников, получивших повышение. Более высокие показатели указывают на лучшие возможности карьерного роста.
- **Время реагирования на инциденты (MTTR):** среднее время реагирования на инциденты безопасности и их смягчения. Более короткое время указывает на более эффективное реагирование на инциденты.
- **Улучшение финансовых показателей:** процентное улучшение финансовых показателей, обусловленное мерами кибербезопасности. Более высокие проценты указывают на лучшие финансовые результаты.
- **Индекс креативности и решения проблем:** качественный показатель креативности и способности решать проблемы. Более высокие баллы указывают на лучшее решение проблем.
- **Индекс адаптивности и устойчивости:** качественный показатель адаптивности и устойчивости команды. Более высокие баллы указывают на лучшую адаптивность и устойчивость.
- **Соответствие этическим стандартам:** процент соблюдения этических стандартов в практике кибербезопасности. Более высокие проценты указывают на лучшее соблюдение этических норм.
- **Глобальная перспектива и понимание рынка:** качественный показатель глобальной перспективы и понимания рынка. Более высокие баллы указывают на лучшее глобальное понимание.
- **Снижение группового мышления:** качественная мера снижения группового мышления и продвижения различных точек зрения. Более высокие баллы указывают на лучшее разнообразие в мышлении.