УДК 004.8:004.81

КОГНИТИВНЫЕ ДЕТЕРМИНАНТЫ УЯЗВИМОСТИ К ЦЕЛЕВЫМ ФИШИНГОВЫМ АТАКАМ В УСЛОВИЯХ ПРИМЕНЕНИЯ ГЕНЕРАТИВНЫХ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ

Плаксеев Д.А. (ВКА), **Бондаренко В.С.** (ВКА), **Гудков А.С.** (АО "Крибрум")
Научный руководитель – кандидат технических наук, преподаватель Тельбух В. В.
(Военно-космическая академия имени А.Ф.Можайского)

Введение. Развитие генеративных нейросетевых моделей привело к качественному преобразованию угроз в сфере информационной безопасности. Социальная инженерия, основанная на эталонных методах обмана, сегодня использует инструменты ИИ для создания персонализированного и практически неотличимого от реальности контента - текста, голоса, видео [1]. Крупные инциденты последних лет, включая хищение десятков миллионов долларов с помощью deepfake-видео и имитации голоса руководителей, демонстрируют, что технические барьеры перестают быть устойчивым препятствием для злоумышленников [2]. Согласно данным первого квартала 2025 года, фишинговые атаки нацелены на ключевые экономические сектора: государственные учреждения (22%), промышленность (16%) и финансовый сектор, где в 67% случаев негативным последствием становится утечка конфиденциальных данных. Общей целью всех инцидентов остается человек, чьи когнитивные особенности восприятия и принятия решений становятся основным вектором атаки [3]. Цель настоящей работы – системный анализ когнитивных детерминант уязвимости к целевым фишинговым атакам и построение комплексной модели защиты.

Основная часть.

Эволюция фишинговых атак в условиях применения генеративного ИИ

Традиционная модель целевых фишинговых атак описывает процесс воздействия на человека путем эксплуатации когнитивных искажений, организационной иерархии, социальных отношений, поведенческих привычек и технического окружения [4]. С появлением генеративного ИИ произошёл сдвиг от массовых рассылок к масштабируемой персонализации. Ключевые изменения парадигмы включают: лингвистический перфекционизм (точное воспроизведение речевых паттернов жертвы), мультимодальность (синтез голоса и deepfake-видео, эксплуатирующие эволюционное доверие к аудиовизуальным каналам) и автоматизацию разведывательной деятельности (сканирование открытых источников для выявления социальных связей и психологических триггеров).

Ключевые когнитивные детерминанты уязвимости.

Успех атак обусловлен не столько технологической сложностью, сколько целенаправленной эксплуатацией фундаментальных механизмов мышления [5]. Можно выделить пять основных детерминант:

1. **Слепое доверие к авторитету и социальному доказательству.** Генеративный ИИ создаёт мультимодальные атаки (голос руководителя, письмо, стилизованное под командное обсуждение), при которых одновременное подтверждение по нескольким каналам блокирует критическую оценку.
2. **Эффект срочности и дефицита.** Использование контекстуально точных предлогов для срочных действий переключает мышление с аналитического на импульсивное, делая угрозу неотличимой от рутинной задачи [6].
3. **Эвристика правдоподобия и когнитивная легкость.** Безупречные тексты, лишённые лингвистических аномалий, не встречают «когнитивных зацепок», и сообщение автоматически принимается за достоверное.

4. **Эвристика аффекта и доверия к знакомому.** Мимикрия под внутреннюю корпоративную среду (структура уведомлений, стиль писем, упоминание реальных проектов) создаёт ощущение «своего», подавляющее бдительность [7].

5. **Когнитивный разрыв в оценке возможностей ИИ.** Недооценка технологий генеративного ИИ приводит к тому, что жертва не активирует механизмы проверки, не допуская мысли о возможности синтеза голоса или видео.

Необходимость новой парадигмы защиты.

Традиционные методы, ориентированные на преодоление «человеческого фактора», неэффективны против атак, нацеленных на когнитивную сферу. В ответ на этот вызов предлагается концепция проактивной когнитивной защиты, построенная на принципах анализа данных, поведенческой психологии и машинного обучения. Примером реализации может служить архитектура, включающая модули профилирования угроз (создание «эталона нормы» организации), детекции аномалий (стилометрия, аудио-форензика, детекция дипфейков), когнитивного ассистента (помощь в преодолении искажений в момент принятия решения) и цифрового двойника безопасности (моделирование атак для адаптивного усиления защиты).

Выводы. Современная кибербезопасность переживает фундаментальный перелом: главная уязвимость сместилась из области программного кода в архитектуру человеческой психики. Угрозы, основанные на социальной инженерии с применением генеративного ИИ, стали новой стратегической реальностью. Сохранение традиционных подходов к защите равносильно стратегическому поражению. Необходим пересмотр принципов безопасности в сторону создания проактивных когнитивных систем, способных адаптироваться к эволюции угроз и усиливать способность человека принимать верные решения в условиях целенаправленной манипуляции.

Список использованных источников:

1. **Лепский, В. Е.** Когнитивные искажения в информационных войнах / В. Е. Лепский // Информационные войны. — 2023. — № 4 (64). — С. 44—52.

2. **Солдатова, Г. У.** Цифровая социализация и кибербезопасность: психологические аспекты / Г. У. Солдатова, Е. И. Рассказова // Психологический журнал. — 2024. — Т. 45, № 2. — С. 15—27.

3. **Белинская, Е. П.** Информационная социализация в цифровую эпоху: риски и возможности / Е. П. Белинская // Психологический журнал. — 2022. — Т. 43, № 3. — С. 105—115.

4. **Войскунский, А. Е.** Психология и интернет: монография / А. Е. Войскунский. — М.: Акрополь, 2021. — 340 с.

5. Зинченко, Ю. П. Психология безопасности в цифровой среде / Ю. П. Зинченко // Национальный психологический журнал. — 2023. — № 4 (48). — С. 85—96.

6. **Мартынова, Е. В**. Эффект срочности в фишинговых атаках: когнитивные механизмы / Е. В. Мартынова // Прикладная юридическая психология. — 2024. — № 3 (60). — С. 32—41.

7. **Нестик, Т. А.** Социально-психологические аспекты доверия в цифровой среде / Т. А. Нестик // Институт психологии РАН. Социальная и экономическая психология. — 2022. — Т. 7, № 4 (28). — С. 6—30.