

**Казарян А.С.**

студент

kazarian.info@gmail.com

**Степаненко А.И.**

студент

leigst@gmail.com

**Дукага Кобонго Г.Д.**

студент

doukaga828@gmail.com

Научный руководитель: **Курбакова М.А.**

к.ф.н., доцент

kurbakovamarina@gmail.com

Московский Политехнический Университет,

Россия, Москва

## **МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОЦЕНКИ ГРАНИЦ БЕЗОПАСНОЙ АВТОМАТИЗАЦИИ**

*Аннотация.* В статье предлагается концептуальная модель оценки границ безопасной автоматизации человеко-машинных систем. Введены и формализованы коэффициент когнитивного делегирования и порог компетентностной необратимости. В отличие от существующих метрик, подход оценивает степень отстранения оператора от контура управления. Сформулированы ключевые практические следствия.

*Ключевые слова:* порог компетентностной необратимости; коэффициент когнитивного делегирования; дополненный интеллект; когнитивная деградация; функциональная безопасность; математическая модель; критическая инфраструктура.

### **Введение**

Современные человеко-машинные системы критической инфраструктуры эволюционируют в направлении всё более глубокой автоматизации. Оператор всё чаще выступает пассивным наблюдателем, что повышает эффективность штатных режимов, но создаёт новый класс рисков — постепенную, незаметную утрату способности к самостоятельным действиям в нештатных ситуациях. Данный парадокс, известный как «Иронии автоматизации» («Ironies of Automation», Л. Бейнбридж) [1], активно изучается в авиации, где нормативные требования [2, 3] как в России, так и за рубежом прямо предписывают регулярное прохождение пилотами тренировок на тренажёрах, включая обязательную отработку действий при имитации отказов двигателя и других систем. Однако за пределами авиации проблема когнитивной деградации оператора при избыточном делегировании решений искусственному интеллекту остаётся концептуально неосмысленной и лишённой формального аппарата количественной оценки.

Существующие метрики, такие как уровень автоматизации, степень автономности, фиксируют лишь долю операций, выполняемых машиной, но

не характеризуют состояние человека. В настоящей работе предлагается сместить аналитический фокус на степень фактического отстранения оператора от контура принятия решений.

Новизна работы заключается во введении в научный оборот понятий коэффициента когнитивного делегирования и порога компетентностной необратимости, которые дополняют существующие нормативные рамки [4, 5] в области функциональной безопасности и управления рисками отсутствовавшим в них ранее измерением когнитивной устойчивости оператора.

Цель статьи — введение и математическая формализация двух взаимосвязанных понятий: коэффициента когнитивного делегирования (ККД) и порога компетентностной необратимости (ПКН), образующих концептуальный каркас для оценки границ безопасной автоматизации.

### **Коэффициент когнитивного делегирования и порог компетентностной необратимости**

Коэффициент когнитивного делегирования (ККД) определяется как доля профессиональных решений, принимаемых автоматизированной системой без осознанного контроля и верификации со стороны оператора:

$$K_{cog} = \frac{N_{deleg}}{N_{total}} \quad (1)$$

где  $N_{deleg}$  — число делегированных решений,  $N_{total}$  — общее число решений в задаче.  $K_{cog} \in [0,1]$ , причём  $K_{cog} = 0$  соответствует полностью ручному управлению,  $K_{cog} = 1$  — полной автоматизации без контроля.

Под решением в контексте данной модели понимается элементарный акт профессионального суждения: интерпретация показания прибора, сравнение параметра с уставкой, оценка тренда, выбор режима работы, подтверждение или отклонение предложенного системой действия. Совокупность таких актов образует когнитивный контур оператора, целостность которого и является предметом оценки. В практическом плане вычисление ККД для конкретной человеко-машинной системы требует экспертного анализа рабочего процесса: выделения дискретных актов профессионального суждения и классификации каждого из них как делегированного либо контролируемого. В рамках настоящей теоретической работы формула (1) рассматривается как концептуальный каркас, а не как готовая инженерная методика.

Принципиальное отличие ККД от классического уровня автоматизации иллюстрируется следующим примером. Пусть дана система, в которой 90% операций выполняется автоматически, однако оператор осуществляет осознанный контроль и верификацию каждого автоматического действия. Показатель уровня автоматизации такой системы составляет 0,9 (при максимально возможном значении 1). С точки зрения ККД значение может быть близким к нулю, поскольку делегирования когнитивной функции не

происходит — оператор сохраняет активную позицию в контуре принятия решений.

Порог компетентностной необратимости (ПКН) — критическое значение  $K_{crit} \in (0,1)$ , при длительном превышении которого способность оператора к самостоятельному восстановлению контроля в нештатной ситуации становится практически невозможной без повторного обучения. Эмпирическим основанием для введения данного понятия служат многочисленные исследования [6, 7] в области человеческого фактора, фиксирующие явление утраты оператором способности понимать происходящее и предвидеть развитие событий и навыков ручного управления при длительной работе в режиме пассивного наблюдения за автоматикой. Здесь термин «необратимость» употребляется в инженерно-психологическом смысле: если время, необходимое для актуализации утраченной компетенции, превышает характерное время развития аварии, восстановление контроля невозможно.

Формально существование порога выражается через функцию восстановления компетенции  $R(t)$ :

$$\exists K_{crit} \in (0,1): \forall K_{cog} > K_{crit} \Rightarrow R(t) \rightarrow 0 \text{ при } t \rightarrow \infty \quad (2)$$

Иными словами, чем дольше оператор работает с превышением критического порога делегирования, тем меньше у него шансов восстановить контроль в нештатной ситуации.

Численное значение  $K_{crit}$  не универсально и требует эмпирического определения для каждого класса систем.

Для наглядного представления взаимосвязи между ККД, ПКН и состоянием оператора вводится градация режимов взаимодействия (таблица 1). Значение  $K_{cog} = 0,33$  принято в качестве иллюстративной границы и не претендует на нормативность, его эмпирическое обоснование требует отдельных исследований.

Таблица 1

**Диапазоны ККД и режимы взаимодействия**

| <b>Диапазон ККД</b>            | <b>Режим взаимодействия</b> | <b>Характеристика состояния оператора</b>                 |
|--------------------------------|-----------------------------|-----------------------------------------------------------|
| $0 \leq K_{cog} < 0,33$        | Активное управление         | Навык регулярно тренируется, риск деградации минимален    |
| $0,33 \leq K_{cog} < K_{crit}$ | Дополненный интеллект [8]   | Автоматизация разгружает, но контроль сохраняется         |
| $K_{cog} \geq K_{crit}$        | Замещающий интеллект        | Превышен ПКН, навык утрачивается функционально необратимо |

Термин «замещающий интеллект» используется в данной работе как авторский антоним к понятию дополненного интеллекта и обозначает

режим, при котором искусственный интеллект не усиливает, а вытесняет человека из контура управления.

### **Обсуждение: границы применимости и практические следствия**

Предложенная модель применима к профессиям с высокой ценой ошибки и невозможностью отложенной коррекции (операторы АСУ ТП, диспетчеры энергосистем, пилоты). Для творческих и офисных специальностей, допускающих итеративную работу, понятия ККД и ПКН в целом не применимы. Модель не отрицает пользу автоматизации, но указывает на необходимость контроля за степенью делегирования.

Из факта существования ПКН вытекают практические следствия: риск когнитивной деградации растёт нелинейно при приближении к  $K_{crit}$ ; необходимы периодические принудительные снижения ККД (тренировки, симуляции); опасность представляет постепенное, незаметное наращивание ККД.

Нелинейность риска означает, что увеличение ККД с 0,2 до 0,3 и с 0,8 до 0,9 не эквивалентны по последствиям. В первом случае оператор остаётся в зоне активного управления либо дополненного интеллекта, тогда как во втором — система может незаметно пересечь критический порог. Данное свойство требует особого внимания при проектировании и эксплуатации: мониторинг ККД должен сопровождаться не просто констатацией роста показателя, а анализом его близости к предполагаемому  $K_{crit}$ .

Из изложенного вытекают следующие практические рекомендации для разработчиков человеко-машинных интерфейсов и специалистов по безопасности критической инфраструктуры:

1. Внедрение периодических симуляций отказов автоматики по аналогии с авиационными тренажёрами для принудительного снижения ККД и поддержания навыка оператора.
2. Аудит логов человеко-машинного взаимодействия для вычисления фактического ККД и выявления докритических состояний.
3. Проектирование интерфейсов с задачами, требующими осознанного вмешательства оператора даже при наличии автоматического решения.
4. Разработка отраслевых стандартов на предельно допустимый ККД для различных классов критических систем.

### **Заключение**

В рамках настоящей работы предпринята попытка концептуального осмысления и математической формализации проблемы когнитивной

деградации оператора при избыточном делегировании решений автоматизированным системам. Основные результаты сводятся к следующему:

1. Введены и формализованы коэффициент когнитивного делегирования (ККД) и порог компетентностной необратимости (ПКН).
2. Показано, что модель применима к критической инфраструктуре и не универсальна.
3. Определены ключевые практические следствия, задающие направления для будущих исследований: эмпирическое определение  $K_{crit}$ , разработка методик расчёта ККД, экспериментальная проверка эффективности тренировочных интервенций.

Дальнейшее развитие предложенной модели связано с эмпирической верификацией численных значений  $K_{crit}$  для конкретных классов человеко-машинных систем, а также с разработкой инженерных методик расчёта ККД. Представляется, что своевременное внедрение метрик когнитивного делегирования в практику управления безопасностью позволит предотвратить утрату оператором контроля при дальнейшем росте автоматизации в условиях экспансии технологий искусственного интеллекта. Отдельный интерес представляет применение предложенной модели к системам автономного вождения, где полное отстранение человека является проектной целью, а не побочным эффектом.

### Благодарности

Авторы выражают признательность Налётову М.В., аналитику данных (Яндекс), за ценные замечания и предложенные направления развития темы.

### Список литературы

1. Bainbridge L. Ironies of Automation // Automatica. 1983. Vol. 19, No. 6. P. 775–779.
2. Приказ Минтранса России от 12.09.2008 N 147 (ред. от 02.02.2024) "Об утверждении Федеральных авиационных правил "Требования к членам экипажа воздушных судов, специалистам по техническому обслуживанию воздушных судов и сотрудникам по обеспечению полетов (полетным диспетчерам) гражданской авиации" (Зарегистрировано в Минюсте России 20.11.2008 N 12701) URL: [https://www.consultant.ru/document/cons\\_doc\\_LAW\\_82272/](https://www.consultant.ru/document/cons_doc_LAW_82272/) (дата обращения: 11.04.2026).
3. 14 CFR Part 121 — Operating Requirements: Domestic, Flag, and Supplemental Operations. Subpart N — Training Program. URL: <https://www.ecfr.gov/current/title-14/chapter-I/subchapter-G/part-121/subpart-N> (дата обращения: 11.04.2026).
4. IEC 61508-1:2010. Functional safety of electrical/electronic/programmable electronic safety-related systems — Part 1: General requirements. Geneva: International Electrotechnical Commission, 2010.

5. ISO 31000:2018. Risk management — Guidelines. Geneva: International Organization for Standardization, 2018.
6. Endsley M.R. Toward a Theory of Situation Awareness in Dynamic Systems // Human Factors. 1995. Vol. 37, No. 1. P. 32–64. DOI: 10.1518/001872095779049543.
7. Parasuraman R., Manzey D.H. Complacency and Bias in Human Use of Automation: An Attentional Integration // Human Factors. 2010. Vol. 52, No. 3. P. 381–410. DOI: 10.1177/0018720810376055.
8. Kazarian A.S. Augmented Intelligence: Research on the Effectiveness of Collaboration between Humans and Artificial Intelligence // Материалы конференции. 2025. P. 24–26. EDN VNVWBH.