

Методологические аспекты информационных и математических технологий

УДК 005.334:004.94

DOI:10.25729/ESI.2026.43.3.001

Риск как имманентное свойство эргатических систем в эпоху цифровых двойников: от «цифрового Левиафана» к парадигме симбиотического управления

Бочков Александр Владимирович

АО «НИИАС», Россия, Москва, a.bochkov@gmail.com

Аннотация. Современный переход к «управлению, основанному на данных», опирающемуся на создание полномасштабных цифровых двойников сложных социально-экономических систем, сталкивается с фундаментальным вызовом. В статье доказывается, что оптимизационные задачи, решаемые для таких цифровых моделей, являются некорректными по отношению к оригиналу, поскольку ключевые объекты управления – эргатические системы – принципиально неформализуемы из-за наличия человеческого компонента. Риск в этой парадигме перестает быть внешней угрозой, становясь имманентным свойством самой системы управления, порождаемым неизбежным и нарастающим расхождением траекторий виртуальной модели и реального эргатического процесса. Стратегии, оптимальные для «цифрового Левиафана» (тотальной симуляции), ведут к дестабилизации оригинала, эрозии человеческого компонента и выходу системы за границы управляемости. Автор обосновывает необходимость смены парадигмы: от поиска точечного оптимума для неполной модели к синтезу робастных и адаптивных стратегий, явно минимизирующих функционал риска. Вводится и противопоставляется два концепта: «цифровой Левиафан» (искушение управлять симуляцией) и «формализм риска» (дисциплина постоянного сравнения карты и территории). Ключевым элементом новой парадигмы «симбиотического управления» является формальное включение человеческого оператора, как корректирующего контура в итеративный диалог между моделью и реальностью. Целью управления становится не абсолютный оптимум в виртуальном пространстве, а динамическая адаптивная эффективность в реальном мире, достигаемая через постоянное управление имманентным риском расхождения траекторий. Предложенная архитектура симбиотического управления, основанная на принципах байесовского обновления и минимакса, позволяет не просто минимизировать риск, но и активно управлять эпистемической неопределенностью, превращая человеческую интуицию в структурированный источник коррекции модели. В работе показано, что такой подход обеспечивает устойчивость сложных социотехнических систем в условиях принципиальной неполноты знания, предлагая путь от иллюзии абсолютного контроля к мудрости совместной навигации.

Ключевые слова: цифровой двойник, эргатическая система, имманентный риск, оптимизационная задача, некорректность модели, симбиотическое управление, адаптивная стратегия, человеческий фактор, эпистемическая неопределенность

Цитирование: Бочков А.В. Риск как имманентное свойство эргатических систем в эпоху цифровых двойников: от «цифрового Левиафана» к парадигме симбиотического управления / А.В. Бочков // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 5-21. – DOI:10.25729/ESI.2026.43.3.001.

Введение. Современное управление все чаще строится на использовании «цифровых двойников» – виртуальных моделей физических объектов или процессов, обновляемых в реальном времени. Их применение формально сводится к задаче оптимизации заданного функционала (эффективности, затрат) на траекториях модельного процесса. Однако ключевые объекты управления – предприятия, города, транспортные системы – относятся к классу эргатических систем, чьи цели (устойчивость, легитимность, адаптивность) шире операционной эффективности. Полное пространство их состояний, особенно из-за человеческого фактора, принципиально неформализуемо. Любая цифровая модель оперирует лишь упрощенным подпространством, а её целевой функционал отражает лишь часть истинных, часто противоречивых целей оригинала. Это порождает фундаментальный парадокс: оптимизация идеализированной абстракции не гарантирует улучшения реальности.

Следствием становится онтологизация риска: риск превращается из внешней угрозы во внутреннее свойство архитектуры управления, вытекающее из несоответствия модели и реальности [1]. Цель статьи – проанализировать природу этого имманентного риска, описать деструктивную динамику, порождаемую «цифровым Левиафаном», и обосновать контуры новой парадигмы: формализма риска и симбиотического управления, где риск становится ключевым управляемым функционалом, а человек – главным корректирующим контуром.

1. «Цифровой Левиафан»: Некорректность оптимизационной задачи и генезис имманентного риска. Современный этап технологического развития, характеризуемый взрывным ростом вычислительных мощностей и доступности данных, породил новую управленческую утопию – возможность построения полномасштабного цифрового двойника сложной социально-экономической или социотехнической системы. Этот двойник, будучи насыщенным историческими и текущими данными, должен был стать идеальным полигоном для экспериментов, точным инструментом прогнозирования и, что самое главное, основой для принятия оптимальных управленческих решений. Формальная постановка задачи в рамках этой парадигмы выглядит безупречно с точки зрения математической теории управления и исследования операций. Пусть реальная управляемая система существует в некотором фазовом пространстве состояний $\mathcal{X}_R$. Управляющее воздействие $u(t)$ выбирается из допустимого множества $\mathcal{U}$. Динамика системы описывается стохастическим дифференциальным уравнением или их системой:

$$dX_t = f(X_t, u_t, t)dt + \sigma(X_t, u_t, t)dW_t,$$

где $X_t \in \mathcal{X}_R$ – состояние системы в момент времени t , f – детерминированный дрейф, σ – волатильность, а W_t – винеровский процесс, моделирующий случайные возмущения. Цель управления формализуется через целевой функционал $J_R(u)$, который требуется максимизировать (например, суммарный доход) или минимизировать (например, совокупные издержки или риск). Часто он имеет вид математического ожидания от некоторого интегрального или терминального критерия:

$$J_R(u) = E \left[\int_{t_0}^T L(X_t, u_t, t)dt + \Phi(X_T) \right],$$

где L – функция текущих затрат/доходов, а Φ – терминальный выигрыш. Классическая задача стохастической оптимизации заключается в нахождении стратегии управления $u^*(t)$ (часто в форме обратной связи $u^*(X_t, t)$), которая доставляет оптимальное значение функционалу: $u^* = \arg \max_{u \in \mathcal{U}} J_R(u)$.

Казалось бы, при наличии достаточно точной модели функций f , σ и критерия J_R , современные методы динамического программирования, принципа максимума Понтрягина или обучения с подкреплением позволяют приблизиться к решению этой задачи.

Однако фундаментальная проблема, лежащая в основе всей этой конструкции, заключается в том, что для эргатических систем исходное предположение об изоморфизме $\mathcal{X}_M \cong \mathcal{X}_R$ и точном соответствии (f, σ, J) и (f_M, σ_M, J_M) принципиально невыполнимо. Эргатическая система, по определению, является целостностью, возникающей из неразрывного взаимодействия человеческого, технологического и средового компонентов. Человеческий компонент вносит в систему качество, которое можно назвать «рефлексивной неформализуемостью» [2]. Пространство состояний $\mathcal{X}_R$ такой системы включает в себя не только измеримые технические параметры (обороты, напряжение, объемы), но и ментальные состояния операторов, их интенции, уровень доверия к системе, неформальные договоренности внутри коллектива, культурные нормы, имплицитное знание [3], которое невозможно вербализовать и, тем более, представить в виде структурированных данных. Эти

аспекты не просто сложно измерить; они принадлежат к качественно иной онтологической категории, нежели данные, обрабатываемые компьютером. Следовательно, оператор Π не является изоморфизмом, а представляет собой жесткую редукцию, проецирующую многомерное, отчасти континуальное, отчасти дискретное, отчасти качественное пространство реальности на конечномерное, чисто количественное подпространство модели: $\mathcal{X}_M \subset \mathcal{X}_R$, причем $\dim \mathcal{X}_M \ll \dim \mathcal{X}_R$). Более того, это подпространство не является статичным: по мере изменения системы, ее целей и контекста, релевантное подпространство состояний может «плавать» и модель, построенная на исторических данных, эти изменения уловить не в состоянии.

Это приводит к первому и главному источнику имманентного риска: некорректности оптимизационной задачи в постановке Адамара [4]. Задача называется корректно поставленной, если ее решение существует, единственно и непрерывно зависит от исходных данных. В нашем случае «исходными данными» является сама модель, как отображение реальности. Некорректность возникает именно из-за разрыва между моделью и оригиналом.

Имманентный риск возникает из трех ключевых ограничений при моделировании эргатических систем:

Редукция целей: Оптимизация по узкому формальному критерию (прибыль, эффективность) ведет к деградации по другим жизненно важным, но неформализуемым целям (устойчивость, адаптивность).

Игнорирование рефлексивности: Модель предполагает статичное поведение людей, тогда как в реальности они наблюдают и адаптируются к управляющим воздействиям, меняя систему и обесценивая прогнозы.

Слепота к «черным лебедям»: Модели, обученные на исторических данных, не учитывают катастрофические события за пределами этой выборки. Оптимизированная под «спокойную» модель система становится хрупкой к непредсказуемым потрясениям.

Таким образом, генезис риска коренится в применении аппарата оптимизации к принципиально неформализуемой эргатической системе. «Цифровой Левиафан» – это автономная управляющая сущность, действующая на основе своей редуцированной модели. Парадокс в том, что чем совершеннее его модель и алгоритмы, тем увереннее он в своей «оптимальности» и тем глубже пропасть между симуляцией и реальностью.

Риск при этом не уменьшается, а, напротив, маскируется и накапливается. Он становится не внешним параметром, который можно оценить и застраховать, а внутренним свойством архитектуры управления, его скрытой пружинкой, которая рано или поздно разожмет свою разрушительную силу, когда накопленное расхождение ρ_T превысит критический порог и система выйдет из области, где модель хоть как-то адекватна.

2. Деструктивная динамика «цифрового Левиафана»: механизмы накопления риска и потеря управляемости. Фундаментальная некорректность оптимизационной задачи для эргатической системы, рассмотренная в предыдущей главе, не является статическим, пассивным свойством. Она запускает активный, самоусиливающийся процесс деградации, в котором имманентный риск ρ_T не просто существует как константа, а монотонно нарастает во времени, приводя к качественным изменениям в состоянии системы.

Первичным и самым прямым следствием является **дестабилизация и эволюция в неописанные области пространства состояний**. Стратегия u_M^* вычислена, как аргумент максимума функционала $J_M(u)$ на траекториях модельного процесса X_t^M , подчиняющегося уравнению $dX_t^M = f_M(X_t^M, u_t, t)dt + \sigma_M dW_t$. Геометрически, в пространстве $\mathcal{X}_M$ эта стратегия «заставляет» модельную траекторию двигаться в область, где значение J_M максимально. Однако, из-за редукции $\Pi: \mathcal{X}_R \rightarrow \mathcal{X}_M$, эта область в модельном пространстве соответствует не

компактной, четко определенной области в реальном пространстве $\mathcal{X}_R$, а скорее конусу или многообразию высокой размерности. Точнее, прообраз $\Pi^{-1}(\mathcal{A}_M)$ некоторой «оптимальной» области $\mathcal{A}_M \subset \mathcal{X}_M$ в $\mathcal{X}_R$ может быть невыпуклым, несвязным и содержать состояния, качественно различные с точки зрения реальной динамики f_R . Применяя u_M^* , мы насильственно направляем реальный процесс X_t^R в этот прообраз, но поскольку реальная динамика f_R отличается от f_M , траектория X_t^R начинает отклоняться от проекции идеальной модельной траектории. Важно, что это отклонение не является случайным блужданием; оно имеет систематический характер, определяемый разностью операторов дрейфа: $\Delta f(X, u) = f_R(X, u) - f_M(\Pi(X), u)$. Эта разность, которую можно назвать **систематической ошибкой модели**, действует как дополнительное «силовое поле», сдвигающее реальную траекторию относительно модельной. Формально, если записать динамику ошибки $e_t = \widetilde{X}_t^R - X_t^{M,*}$, где $\widetilde{X}_t^R = \Pi(X_t^R)$, то, пренебрегая для простоты стохастическими членами, получим:

$$\dot{e}_t \approx \frac{\partial f_M}{\partial x} \big|_{X_t^{M,*}} e_t + \Delta f(X_t^R, u_M^*).$$

Первое слагаемое описывает линейную неустойчивость модельной траектории, второе – постоянно действующее возмущение из-за неадекватности модели. Поскольку стратегия u_M^* часто выводит систему на границу допустимой области в $\mathcal{X}_M$ (так как оптимум часто лежит на границе), линейная часть $\frac{\partial f_M}{\partial x}$ может иметь положительные собственные значения, делая ошибку e_t растущей даже при малом Δf . Но главное – это сам член Δf . По мере движения в область $\mathcal{A}_M$, где модель особенно плохо описывает реальность (например, в область высокой нагрузки на человеческий компонент, которую модель считает просто увеличением «производительности»), величина $|\Delta f|$ возрастает. Это приводит к тому, что реальная система X_t^R быстро покидает «коридор» $\Pi^{-1}(\mathcal{A}_M)$, в котором модель хотя бы приблизительно адекватна, и выходит в области пространства $\mathcal{X}_R$, вообще не представленные в модели. В этих областях предсказательная сила модели падает до нуля, так как для состояний X^R , для которых $\Pi(X^R)$ далеко от X^M , модель не может дать даже грубой оценки динамики. Управляющие воздействия u_M^* , рассчитываемые на основе текущего модельного состояния X_t^M , становятся не просто неоптимальными, а откровенно вредными, так как основываются на фундаментально неверной «карте местности».

Этот процесс можно описать, как динамику расходящихся траекторий с положительной обратной связью. Введем показатель степени расхождения $d(t) = E[|e_t|^2]$. Его производная по времени будет включать в себя:

1. Член, пропорциональный самой $d(t)$ (из-за возможной неустойчивости).
2. Член, пропорциональный ковариации ошибки и систематического сдвига $\text{Cov}(e_t, \Delta f)$.
3. Член, связанный с ростом нормы Δf .

Поскольку Δf сам зависит от состояния, а состояние все дальше уходит от области адекватности модели, возникает нелинейная зависимость: $|\Delta f| \approx \kappa(d(t))$, где κ – монотонно возрастающая функция. Тогда уравнение для $d(t)$ принимает вид:

$$\frac{d}{dt} d(t) \geq \alpha d(t) + \beta \kappa(d(t)) - \gamma,$$

где $\alpha, \beta > 0$, а γ – возможный демпфирующий член.

При достаточно больших $d(t)$ и возрастающей κ (например, $\kappa(d) \sim d^\delta, \delta > 0$), решение этого неравенства имеет взрывной характер: $d(t) \rightarrow \infty$ за конечное время T^* . Это математическое выражение катастрофического расхождения, когда разрыв между моделью и реальностью становится настолько велик, что система фактически выходит из-под контроля, основанного на этой модели. Момент T^* – точка потери управляемости.

Второй ключевой механизм деструктивной динамики – «эрозия» человеческого компонента и рост системной диссипации. В рамках модели M человеческий фактор часто представлен в предельно редуцированном виде: как источник задержек, шума (аддитивного или параметрического) или как детерминированный преобразователь входных сигналов в действия по фиксированному, хотя, возможно, и стохастическому, правилу. Оптимизация J_M естественным образом направлена на минимизацию негативного влияния этого «шума» и максимизацию «пропускной способности» человеческого звена.

На практике это выражается в:

- стандартизации и алгоритмизации действий оператора, лишаящих его свободы манёвра и возможности применять имплицитное знание¹;
- интенсификации труда для увеличения скорости обработки информации или принятия решений, что ведет к когнитивной перегрузке;
- сокращению «избыточных» с точки зрения модели функций, таких, как взаимный контроль, неформальное общение, создание смыслов, которые не имеют прямого измеримого вклада в J_M .

Формально, если представить человеческий компонент как подсистему с внутренним состоянием $h_t \in \mathcal{H}$ (включающим мотивацию, усталость, компетенцию, уровень доверия), а его вклад в общую динамику системы – через оператор $\psi(h_t, x_t, u_t)$, то в модели происходит грубая аппроксимация: $\psi_M = \psi_{id} + \xi_t$, где ψ_{id} – идеализированный, детерминированный отклик, а ξ_t – гауссов шум. Стратегия u_M^* , оптимизируя систему в целом, будет стремиться снизить зависимость от ψ (как от ненадежного элемента) и минимизировать дисперсию $E[|\xi_t|^2]$. В реальности это приводит к тому, что состояние h_t деградирует по определенному закону. Можно смоделировать эту деградацию как динамический процесс:

$$\dot{h}_t = Ah_t + Bu_t^{human} + \eta_t,$$

где u_t^{human} – управляющие воздействия, направленные на человека со стороны системы (задачи, инструкции, стимулы), вычисленные в соответствии со стратегией u_M^* .

Матрица A описывает внутреннюю динамику человеческого состояния (например, усталость накапливается, мотивация снижается без подкрепления). Оптимальная стратегия u_M^* порождает последовательность $\{u_t^{human}\}$, которая максимизирует краткосрочный вклад в J_M , но игнорирует долгосрочную динамику h_t . Фактически, u_t^{human} становится эксплуататорским, истощающим ресурс h_t . Например, если h_t включает запас «внимания» или «психической энергии», то u_t^{human} постоянно требует высокого уровня этого ресурса, не предусматривая периодов для его восстановления. Со временем это приводит к переходу состояния h_t в область низких значений (выгорание, цинизм, потеря квалификации). В этой области реальный оператор $\psi(h_t, \cdot)$ перестает быть даже приблизительно похожим на модельный ψ_M : вместо того, чтобы быть источником малого шума, он становится источником крупных, систематических ошибок или пассивно-сопротивляющимся элементом. Это напрямую увеличивает систематическую ошибку модели Δf , так как теперь f_R зависит от $\psi(h_t)$, а f_M – от ψ_M . Возникает петля положительной обратной связи: применение u_M^* ухудшает h_t , ухудшение h_t увеличивает Δf , увеличение Δf ускоряет расхождение траекторий $d(t)$, что, в свою очередь, может приводить к еще более агрессивным и деструктивным с точки зрения

¹ Имплицитное знание – неосознаваемое, скрытое знание или навык, который человек приобретает без явных инструкций и не может легко вербализовать, но применяет на практике. Оно формируется через «имплицитное научение», когда мозг усваивает закономерности, например, при езде на велосипеде, изучении языка или восприятии дизайна, и проявляется как интуитивное понимание, а не как набор фактов.

человека управляющим воздействиям со стороны модели, стремящейся «вернуть систему на курс».

Понятие **диссипации** здесь приобретает глубокий смысл. Если в термодинамике диссипация – это рассеяние энергии, превращение упорядоченной формы энергии в неупорядоченную (тепло), то в контексте эргатической системы, управляемой «цифровым Левиафаном», диссипация проявляется как рассеяние **системного потенциала**, который включает в себя не только материально-энергетические ресурсы, но и социальный капитал, доверие, адаптивный резерв, креативность человеческого компонента. Формально, можно ввести функционал системного потенциала $\Psi(t) = \Psi_S(X_t^R, h_t)$, отражающий способность системы к выполнению своих истинных, многоаспектных целей J_R . В нормальных условиях часть ресурсов тратится на поддержание этого потенциала (обучение, коммуникация, создание доверия, поддержание избыточности). Стратегия, оптимизирующая узкий критерий J_M , воспринимает эти затраты как неэффективные и стремится их минимизировать, перенаправляя ресурсы на непосредственное увеличение J_M . Это ведет к монотонному снижению потенциала: $d\Psi/dt < 0$. Скорость снижения может описываться уравнением:

$$\frac{d\Psi}{dt} = -\lambda_J \cdot (J_M(t) - J_M^{target}) - \lambda_\Psi \cdot \Psi(t) + v(t),$$

где первый член отражает «перекачку» ресурсов из потенциала в текущий модельный выход (чем агрессивнее оптимизация, тем больше отток), второй – естественную деградацию потенциала при отсутствии инвестиций, а $v(t)$ – случайные или внешние пополнения.

Решение такого уравнения при $\lambda_J > 0$ приводит к экспоненциальному затуханию $\Psi(t)$. Система, лишенная потенциала, теряет устойчивость (resilience): ее способность поглощать возмущения и восстанавливаться падает. Малое отклонение, которое раньше было бы компенсировано, теперь вызывает лавинообразный сбой. Она теряет адаптивность: для обучения и изменения уже нет свободных ресурсов и мотивации. Наконец, падает легитимность: участники системы перестают верить в ее справедливость и осмысленность, что ведет к росту транзакционных издержек, оппортунизму и, в пределе, к отказу от сотрудничества. Все эти потери напрямую увеличивают имманентный риск ρ_T , так как делают поведение системы все более непредсказуемым и все менее соответствующим простой, детерминированной логике модели M .

Третий механизм – формирование порочных циклов [6] и режимов с обострением [7].

Описанные выше процессы не независимы; они переплетаются, усиливая друг друга.

Кульминацией этой деструктивной динамики является катастрофическое расхождение и потеря управляемости. Точку, где система окончательно выходит из-под контроля, можно определить несколькими способами.

Первый – достижение расхождением $d(t)$ некоторого критического порога d_{crit} , после которого ни одно управляющее воздействие из допустимого множества $\mathcal{U}$, рассчитанное на основе модели M , не может вернуть систему в область, где модель хотя бы приблизительно адекватна. Формально, это момент T^* , когда множество достижимости реальной системы $\mathcal{R}_{\mathcal{R}}(T^*)$ под действием стратегий, основанных на M , полностью отделено от прообраза области адекватности модели $\Pi^{-1}(\mathcal{A}_{adm})$.

Второй способ – достижение потенциалом $\Psi(t)$ критически низкого уровня Ψ_{crit} , после которого система теряет структурную целостность: начинается массовый выход человеческих агентов, отказы технологических компонентов из-за отсутствия обслуживания, распад кооперативных связей. В этот момент система перестает быть управляемой в принципе, так как перестает существовать как целостный объект управления.

Третий способ – возникновение в системе новых, эмерджентных режимов, которые не только не описаны в модели, но и прямо противоречат ее базовым принципам. Например, вместо кооперации возникает жесткая конкуренция за истощающиеся ресурсы, вместо решения задач – их имитация, вместо инноваций – регресс к примитивным, но энергосберегающим практикам. Эти эмерджентные режимы обладают собственной, часто хаотической, динамикой, которую первоначальная модель M и связанная с ней стратегия u_M^* абсолютно не способны ни предсказать, ни парировать.

Таким образом, «цифровой Левиафан», начав с решения, казалось бы, рациональной задачи оптимизации, запускает лавинообразный процесс саморазрушения эргатической системы. Имманентный риск ρ_T , изначально присутствующий как следствие редукции, не остается постоянным, а становится функцией времени $\rho_T(t)$, монотонно растущей, часто нелинейно или даже гиперболически. Движущей силой этого роста являются не внешние шоки, а внутренние противоречия, заложенные в самой попытке подчинить живую, рефлексивную, многомерную реальность логике упрощенной симуляции. Система не просто отклоняется от плана; она активно трансформируется под воздействием этого «плана», и трансформируется в сторону большей хрупкости, меньшей адаптивности и, в конечном счете, к состоянию, где понятия «оптимальности» и «управления» теряют всякий смысл. Как показано в работе [5], анализ подобных деструктивных сценариев требует перехода от статических оценок безопасности к динамическим моделям, учитывающим эволюцию системного потенциала и скорости накопления риска. Выход из этой ловушки требует не уточнения модели в рамках старой парадигмы, а радикальной смены самой цели управления – от максимизации узкого функционала в симуляции к постоянному управлению и минимизации имманентного риска расхождения. Как это может быть реализовано в рамках новой парадигмы симбиотического управления, будет рассмотрено далее.

3. Формализм риска: от декомпозиции неопределенности к синтезу робастных стратегий. Кульминацией критического анализа «цифрового Левиафана» является не отрицание моделирования как такового, но осознание необходимости принципиально иной теоретической основы для управления эргатическими системами. Эта основа, обозначенная как формализм риска, предлагает совершить эпистемологический поворот: вместо того чтобы рассматривать риск как побочный продукт неточной модели, принять его имманентность в качестве центрального объекта управления. В рамках этого подхода цель смещается с поиска условного оптимума $u_M^* = \arg \max J_M(u)$ для неполной модели M к синтезу стратегий, минимизирующих функционал, явным образом измеряющий расхождение между моделью и реальностью. Однако для конструктивного синтеза таких стратегий необходима более тонкая, чем в предыдущих главах, декомпозиция самой природы неопределенности, порождающей риск.

Первый шаг формализма риска заключается в переходе от рассмотрения единой модели M к работе с ансамблем моделей или, что более фундаментально, с вероятностной мерой на пространстве моделей. Это признает тот факт, что наше знание об эргатической системе R принципиально неполно и может быть представлено лишь в виде распределения над возможными её описаниями. Пусть Θ – пространство параметров, определяющих класс допустимых моделей. Вместо единственной модели M_{θ_0} мы рассматриваем апостериорное распределение $p(\theta|\mathcal{D})$, где $\mathcal{D}$ – исторические и текущие данные. Тогда прогноз поведения системы и оценка последствий управления должны учитывать всю эту неопределенность. Целевой функционал, который в парадигме Левиафана был детерминированным $J_M(u)$, теперь сам становится случайной величиной, зависящей от θ : $J(u, \theta)$. Однако ключевое отличие формализма риска состоит в том, что мы отказываемся от простого взятия математического

ожидания $E_\theta[J(u, \theta)]$ в качестве нового критерия. Такой подход (стандартный в байесовской оптимизации) по-прежнему стремится к некоему «усредненному» оптимуму, который может оказаться столь же деструктивным, если распределение $p(\theta)$ широко или смещено.

Вместо этого формализм риска предлагает работать непосредственно с функционалом риска $\rho(u)$, который количественно измеряет потенциальный ущерб от расхождения между предсказаниями ансамбля моделей и реальностью.

Один из способов построения такого функционала – это определение риска как условной дивергенции между прогнозным и эмпирическим распределениями траекторий. Пусть P_θ^u – распределение траектории системы на горизонте T при управлении u , предсказанное моделью с параметрами θ . Пусть Q^u – истинное (неизвестное) распределение реальных траекторий. Мы никогда не знаем Q^u , но мы можем наблюдать эмпирические реализации $\{\xi^{(i)}\}_{i=1}^N$, порожденные реальной системой под управлением u . Тогда риск можно определить как верхнюю границу некоторой дивергенции D между прогнозными распределениями из нашего ансамбля и эмпирическим распределением, взятым по всем моделям, согласующимся с данными в пределах заданной доверительной области. Например, используя дивергенцию Реньи [8] порядка α :

$$\rho_\alpha(u) = \sup_{\theta \in \Theta_\varepsilon} D_\alpha(P_\theta^u | \widehat{Q}_N^u),$$

где $\Theta_\varepsilon = \{\theta: d(P_\theta^u, \widehat{Q}_N^u) \leq \varepsilon\}$ – множество параметров, чьи прогнозы не слишком противоречат эмпирическим данным в смысле некоторой метрики d , а $\widehat{Q}_N^u$ – эмпирическое распределение.

Такой функционал измеряет наихудшее возможное расхождение среди всех моделей, которые ещё не отвергнуты данными. Минимизация $\rho_\alpha(u)$ по u приводит к стратегии, которая старается сделать систему такой, чтобы даже самые пессимистичные из ещё допустимых моделей предсказывали поведение, близкое к наблюдаемому. Это есть не что иное, как формализация принципа минимакса применительно к эпистемической неопределенности.

Однако подобный подход, будучи теоретически строгим, зачастую вычислительно неосуществим. Практической альтернативой является введение модели рассогласования $\Delta(x, u)$, которая явно оценивает ошибку прогноза основной модели M в точке состояния x при управлении u . Эта модель может быть параметрической, $\Delta_\phi(x, u)$, и обучаться на исторических ошибках прогнозирования:

$$\phi^* = \arg \min_{\phi \in \Sigma_k} |x_{k+1} - (f_M(x_k, u_k) + \Delta_\phi(x_k, u_k))|^2.$$

В этом случае общая динамика системы представляется как:

$$x_{k+1} = f_M(x_k, u_k) + \Delta_{\phi^*}(x_k, u_k) + w_k,$$

где w_k – остаточный шум.

Риск тогда можно связать с величиной $|\Delta_{\phi^*}(x, u)|$ или, более точно, с ковариацией её оценки. Если модель рассогласования также выдает оценку неопределенности $\Sigma_\Delta(x, u)$, то функционал риска может быть определен, как ожидаемая норма ошибки с учетом этой неопределенности:

$$\rho(u) = E \left[\int_0^T |\Delta_{\phi^*}(x_t, u_t)|_{\Sigma_\Delta^{-1}(x_t, u_t)}^2 dt \right],$$

где норма взвешена обратной ковариацией, что придает больший вес областям, где модель рассогласования сама ненадежна. Задача управления тогда формулируется как:

$$u_\rho^* = \arg \min_{u \in \mathcal{U}} (\rho(u) + \lambda \cdot J_M(u)),$$

где $J_M(u)$ – традиционный функционал эффективности, а λ – параметр, регулирующий компромисс между минимизацией риска и достижением цели.

Это уже классическая задача стохастической оптимального управления, но с принципиально иным составным критерием.

Второй ключевой элемент формализма риска – явное разделение эпистемической и алеаторной неопределенности. Алеаторная неопределенность – это принципиально неустранимый шум, присущий самой системе (например, колебания спроса, погодные условия). Эпистемическая неопределенность – это незнание, проистекающее из неполноты модели, и она может быть сокращена путем сбора данных и обучения. Формализм риска фокусируется на управлении именно эпистемической неопределенностью, так как она является источником имманентного риска расхождения. В терминах байесовских глубоких нейронных сетей или гауссовских процессов это соответствует разделению предсказательной дисперсии на две компоненты.

Пусть для данной точки (x, u) модель предсказывает распределение следующего состояния $p(x'|x, u)$. Его общую дисперсию $\text{Var}[x']$ можно декомпозировать (в приближении) как:

$$\text{Var}[x'] \approx \underbrace{E_{\theta}[\text{Var}[x'|\theta]]}_{\text{Алеаторная}} + \underbrace{\text{Var}_{\theta}[E[x'|\theta]]}_{\text{Эпистемическая}},$$

где усреднение ведется по апостериорному распределению параметров $p(\theta|\mathcal{D})$.

Функционал риска тогда естественно связать с эпистемической компонентой, так как она отражает наше незнание правильной динамики. Стратегия, минимизирующая риск, будет стремиться выбирать такие управления u , которые уменьшают эпистемическую дисперсию, то есть вести систему в области, где наши модели более надежны, или, что эквивалентно, активно исследовать те области, где неопределенность велика, чтобы сократить её. Это приводит к фундаментальному пересмотру логики управления: не гнаться за сиюминутной эффективностью в области, где модель уверена (но, возможно, ошибается), а балансировать между эксплуатацией текущего знания и исследованием для его улучшения.

Наиболее перспективным направлением синтеза робастных стратегий в рамках формализма риска представляется стохастическое модельно-прогнозирующее управление (Stochastic Model Predictive Control, SMPC) с явным учетом ограничений на риск. В классической постановке на каждом шаге решается конечно-горизонтная задача оптимизации, первое управление применяется, и процесс повторяется. В SMPC учитывается стохастичность модели, и ограничения часто задаются в вероятностной форме. Формализм риска предлагает заменить вероятностные ограничения на ограничения по риску.

Сложность решения такой задачи заключается в том, что даже оценка риска ρ_k для заданной последовательности управлений является нетривиальной вычислительной проблемой, требующей методов Монте-Карло или аппроксимаций. Однако развитие методов глубокого обучения с подкреплением с учетом неопределенности [9] (например, на основе байесовских нейросетей или ансамблей) открывает пути для приближенного, но практически реализуемого синтеза таких стратегий.

Отдельного рассмотрения заслуживает проблема включения человеческого компонента в формализм риска. В парадигме «цифрового Левиафана» человек – источник шума. В формализме риска он должен стать активным источником информации для сокращения эпистемической неопределенности и корректирующим механизмом. Математически это можно представить как расширение пространства состояний системы x_t до (x_t, b_t) , где b_t – состояние доверия или убеждений оператора относительно адекватности модели. Динамика b_t описывает, как оператор обновляет свои убеждения, наблюдая расхождения. Стратегия управления теперь должна зависеть и от (b_t) : $(u_t = \pi(x_t, b_t))$. Более того, в контуре управления появляется второе управляющее воздействие – корректирующий сигнал оператора v_t ,

который может напрямую модифицировать управление, предлагаемое моделью: $u_t^{applied} = u_t^{model} + v_t$, или даже модифицировать саму модель: $\theta_{t+1} = \theta_t + g(v_t, data)$. Задача тогда становится игрой двух агентов (модели и человека) с общей, но не полностью совпадающей целью. Общая цель – минимизация долгосрочного риска ρ . Но у модели есть доступ только к J_M и оценке ρ через свои алгоритмы, а у человека – к качественным, контекстуальным знаниям и интуиции, которые не формализованы. Синтез совместной стратегии может быть описан в рамках теории совместной оптимизации или иерархического управления, где политика высокого уровня (человек) вмешивается в политику низкого уровня (модель) при превышении порога локального риска, оцениваемого человеком на основе своего b_t .

Это естественным образом приводит к **концепции симбиотического управления**, являющейся практической реализацией формализма риска. Его ядро – итеративный диалог, который можно формализовать, как цикл Байесовского обновления, но не только параметров модели, а всей структуры взаимодействия. Такой итеративный процесс принципиально сужает эпистемическую неопределенность, но не за счет создания всё более сложной модели, а за счет динамического уточнения границ её применимости и обогащения её контекстом от человеческого оператора. Риск ρ_t становится не объектом слепого минимизации, а управляющей переменной, значение которой поддерживается в допустимых пределах за счет диалога. Это перекликается с идеями адаптивного управления и управления по возмущениям, но на более высоком, мета-уровне, где возмущением является собственное несовершенство модели.

4. Архитектура симбиотического управления: итеративный диалог как механизм сжатия эпистемической неопределенности. Формализм риска, устанавливающий новую философию управления, требует для своей практической реализации конкретной системной архитектуры. Такой архитектурой, способной обеспечить динамический баланс между алгоритмической эффективностью и человеческой интуицией, является **симбиотическое управление**. Его сущность заключается не в простом соседстве человека и машины в контуре управления, а в организации между ними интенсивного, структурированного итеративного диалога, где каждый акт коммуникации приводит к сужению эпистемической неопределенности и уменьшению имманентного риска. Остановимся более подробно на построении математического аппарата и принципов функционирования такого диалога, рассмотрению его динамики, устойчивости и пределов эффективности.

Центральным объектом симбиотической системы является не статичная модель, а динамический эпистемический контур, состояние которого в момент времени t описывается кортежем

$$E_t = (\mathcal{M}_t, B_t, C_t, \rho_t),$$

где $\mathcal{M}_t$ – текущая агностическая модель системы, представленная, например, параметрическим семейством $\{f_{\theta_t}, \Sigma_{\theta_t}\}$, где $\theta_t \in \Theta$; B_t – модель убеждений человеческого оператора относительно адекватности $\mathcal{M}_t$ (это может быть распределение над некоторым пространством гипотез $\mathcal{H}$ или просто вектор доверия $b_t \in [0,1]^K$, где каждая компонента отражает уверенность в определенном аспекте модели (например, точность прогноза в разных режимах)); C_t – контекстуальный слой, содержащий неформализуемую или слабо формализуемую информацию, доступную оператору: социальную атмосферу, неписанные правила, слухи, интуитивные предчувствия, этические соображения (формально C_t можно рассматривать как точку в многообразии высокоразмерных качественных данных); $\rho_t: \mathcal{U} \rightarrow R^+$ – текущий оператор риска, оценивающий ожидаемое расхождение для любого допустимого управления u , с учетом $\mathcal{M}_t$, B_t и, возможно, C_t .

Динамика всей эргатической системы теперь описывается не одним уравнением, а связанной системой уравнений для физического состояния $x_t \in \mathcal{X}_{\mathcal{R}}$ и эпистемического состояния E_t :

$$\begin{aligned} x_{t+1} &= F(x_t, u_t, w_t; \mathcal{M}_t) + \varepsilon_t, \\ E_{t+1} &= G(E_t, x_t, u_t, y_{t+1}, v_t; \Omega). \end{aligned}$$

Здесь F – реальная, неизвестная динамика, которая в нашей парадигме аппроксимируется моделью $\mathcal{M}_t$, но содержит неустранимую ошибку ε_t . Функция G – закон эпистемического обновления, описывающий, как взаимодействие с реальностью (наблюдение нового состояния x_{t+1} или измерений y_{t+1}) и коррекция оператора v_t изменяют наше знание и доверие. Ω – набор мета-параметров, определяющих «скорость обучения» и «степень доверия» в диалоге.

Сердцем симбиоза является процесс генерации и коррекции управляющего воздействия. Он реализуется как двухуровневая оптимизация. На нижнем уровне алгоритмический агент решает задачу, сформулированную в рамках формализма риска:

$$u_t^0 = \arg \min_{u \in \mathcal{U}} \{J_{\mathcal{M}_t}(u) + \lambda_t \cdot \rho_t(u)\},$$

где λ_t – адаптивный коэффициент, регулирующий баланс между эффективностью (первый член) и осторожностью (второй член).

Ключевое отличие – функционал риска ρ_t учитывает не только неопределённость модели $\mathcal{M}_t$, но и степени доверия к ней $\mathcal{B}_t$. Если доверие к прогнозам в какой-либо области низкое, ρ_t будет штрафовать управляющие воздействия, ведущие систему в эту область.

Полученное оптимальное управление u_t^0 вместе с аналитикой (прогноз, оценка неопределённости, уязвимости модели) представляется оператору в виде структурированного предложения $\mathcal{P}_t$. На верхнем уровне человеческий оператор, обладая состоянием $\mathcal{B}_t$ и контекстом $\mathcal{C}_t$, производит оценку предложения $\mathcal{P}_t$. Он выступает как байесовский процессор с собственной неформализуемой моделью системы. Сравнивая прогноз со своей мысленной симуляцией, оператор вычисляет субъективную оценку риска $\rho_t^{\widehat{human}}$ и выявляет контекстуальные факторы, снижающие надёжность модели. Его задача – не просто отклонить предложение, а сформулировать структурированную коррективу v_t из заданного пространства $\mathcal{V}$, минимизирующую коммуникационные потери. Это может быть:

- аддитивная или мультипликативная поправка к управлению;
- семантическая коррекция (изменение приоритета целей);
- контекстуальная аннотация (добавление в модель нового ограничения).

Оператор выбирает корректив v_t^* , минимизирующий его субъективную оценку общего риска:

$$v_t^* = \arg \min_{v \in \mathcal{V}} \rho_t^{\widehat{human}}(\Phi(u_t^0, v), \mathcal{C}_t),$$

где Φ – оператор применения коррективы к исходному управлению (например, $\Phi(u, v) = u + v$). Итоговое управление, применяемое к системе, есть $u_t = \Phi(u_t^0, v_t^*)$.

После применения u_t система переходит в состояние x_{t+1} . Ошибка прогноза служит сигналом для **эпистемического обновления** – обучения модели на основе реального исхода. Описанный цикл (предложение, коррекция, применение, обновление) представляет собой одну итерацию эпистемического диалога.

Эффективность симбиоза оценивается по скорости снижения среднего реализованного риска: $\overline{\rho}_T = \frac{1}{T} \sum_{t=0}^{T-1} \rho_t(u_t) \rightarrow \rho_{min}$ при $T \rightarrow \infty$, где ρ_{min} – неустранимый уровень неопределённости. Скорость сходимости зависит от обучаемости модели и качества человеческих корректив.

Математически процесс можно представить как двух-агентное обучение с подкреплением, где наградой служит отрицательный риск: $r_t = -\rho_t(u_t)$. Политика алгоритмического агента параметризована π_ϕ^A , политика человеческого агента – π_ξ^H , где ξ – его неформализуемый опыт. Оптимизация направлена на нахождение точки равновесия, возможно, седловой, в игре с общей, но неполностью наблюдаемой наградой. Сложность заключается в нестационарности: политика π_ξ^H меняется по мере изменения доверия B_t , а политика π_ϕ^A меняется по мере обучения модели.

Важнейшим вопросом является устойчивость симбиотического контура. Существует опасность возникновения патологических режимов. Для предотвращения этих режимов необходимы метаправила управления самим диалогом. Архитектурно симбиотическая система требует специальных интерфейсов, которые превращают неявное знание оператора в структурированные коррективы. Симбиотическое управление – это не статичная архитектура, а мета-архитектура для «выращивания» совместной познавательной системы. Её ядро – итеративный диалог – представляет собой алгоритм с внешним оракулом (человеком), где каждая итерация выполняет два действия: улучшает формальную модель за счет данных и формализует часть контекстуального знания за счет человеческих коррективов. Динамическим инвариантом этого процесса является имманентный риск ρ_t , который служит и целью, и мерой прогресса. В пределе, успешный симбиоз ведет не к созданию идеальной модели, а к установлению устойчивого режима ко-эволюции, где модель и оператор совместно адаптируются к меняющемуся миру, постоянно поддерживая риск в приемлемых границах. Это превращает управление из задачи оптимизации в задачу совместного познания и адаптации, что является единственно адекватным ответом на вызовы, порождаемые сложностью эргатических систем.

5. Практическая имплементация и вызовы симбиотического управления: от теории к инженерии социотехнических систем. Теоретический каркас симбиотического управления, построенный на принципах формализма риска и архитектуре итеративного диалога, должен быть переведён в плоскость практической реализации. Эта задача сопряжена с комплексом междисциплинарных вызовов, лежащих на стыке теории управления, машинного обучения, когнитивных наук, человеко-машинного взаимодействия (HCI) и организационного проектирования [10].

5.1. Математическое и вычислительное ядро: от байесовских ансамблей к обучающимся картам риска. Практическая реализация эпистемического контура требует выбора конкретных математических представлений для модели $\mathcal{M}_t$, оператора риска ρ_t и механизма обновления. Ключевой дилеммой является баланс между выразительностью (способностью улавливать сложные зависимости) и вычислительной эффективностью (возможностью работать в режиме реального времени).

Для представления агностической модели $\mathcal{M}_t$ наиболее перспективным представляется использование ансамблей моделей или байесовских глубоких сетей. В первом случае $\mathcal{M}_t = \{f_{\theta^{(i)}}\}_{i=1}^N$, где каждый $\theta^{(i)}$ обучается на бутстраппированной выборке данных или с разными начальными условиями. Прогноз и неопределенность выводятся из статистики по ансамблю: $\widehat{x}_{t+1} = \frac{1}{N} \sum_i f_{\theta^{(i)}}(x_t, u_t)$, а эпистемическая дисперсия оценивается как $\Sigma_t^{epist} = \text{Var}_i[f_{\theta^{(i)}}(x_t, u_t)]$. Такой подход, хотя и требует в N раз больше вычислений, прозрачен и хорошо калибруется. Байесовские нейросети, использующие приближенные методы вывода вроде вариационного вывода или MC-Dropout [11], позволяют получить оценку неопределенности в рамках одной сети, но могут страдать от недооценки эпистемической

неопределенности. Выбор между ними зависит от критичности времени отклика и доступных вычислительных ресурсов.

Оператор риска ρ_t должен агрегировать информацию о неопределенности по всей прогнозируемой траектории. Пусть для предложенного управления $u_{t:t+H-1}$ модель генерирует распределение траекторий $p(x_{t+1:t+H}|x_t, u_{t:t+H-1}, \mathcal{M}_t)$. Тогда интегральный риск можно определить как математическое ожидание некоторой функции потерь L_{risk} , зависящей от состояния и его неопределенности, вдоль траектории:

$$\rho_t(u_{t:t+H-1}) = E_{p(x_{t+1:t+H}|\dots)} \left[\sum_{k=1}^H \gamma^k L_{risk}(x_{t+k}, \Sigma_{t+k}^{epist}) \right].$$

Функция L_{risk} должна штрафовать как попадание в нежелательные области фазового пространства (например, близость к границам безопасной работы), так и высокую эпистемическую неопределенность. Один из вариантов – использовать стоимость, зависящую от дисперсии: $L_{risk}(x, \Sigma) = x^T Q x + \text{Tr}(S \Sigma)$, где матрица S определяет «цену незнания». Вторая компонента прямо стимулирует выбор управлений, ведущих в области с низкой эпистемической неопределенностью. Задача оптимального управления тогда становится задачей стохастического MPC с составной стоимостью, которая может решаться методами стохастической оптимизации (например, на основе сэмплирования траекторий – Model Predictive Path Integral, MPPI [12]) или аппроксимации динамического программирования с глубоким обучением [13].

Отдельная сложность – эффективное обновление модели $\mathcal{M}_t$ в реальном времени после получения новой пары данных (x_t, u_t, x_{t+1}) . Полный пересчёт байесовского апостериорного распределения непрактичен. Необходимы инкрементальные методы. Для ансамблей можно применять онлайн-обучение каждого члена ансамбля на потоке данных с небольшой скоростью обучения. Для байесовских сетей существуют методы онлайн-вариационного вывода, где вариационные параметры ϕ_t (например, средние и дисперсии весов) обновляются по градиенту логарифмического правдоподобия нового наблюдения с учётом априорного распределения, играющего роль регуляризатора, предотвращающего катастрофическую забывчивость:

$$\phi_{t+1} = \phi_t + \eta \nabla_{\phi} [\log p(x_{t+1}|x_t, u_t, \phi_t) - D_{KL}(q_{\phi_t}(\theta) || p(\theta))].$$

Человеческие коррективы v_t должны также интегрироваться в процесс обучения. Если корректив был семантическим (например, «учитывать фактор Z »), система должна попытаться найти латентную переменную z_t , такую, что добавление её в модель уменьшает ошибку прогноза. Это можно сформулировать как задачу обучения с частичным привлечением учителя: по данным $\{(x_\tau, u_\tau, x_{\tau+1})\}$ и меткам-подсказкам от оператора в моменты вмешательства, обучить модель $f_\theta(x, u, z)$ и энкодер $q_\psi(z|x, u, \text{hint})$, предсказывающий латентную переменную.

5.2. Моделирование и интеграция человеческого компонента: от когнитивных архитектур к структурированным интерфейсам. Самой сложной частью является формализация и интеграция человеческой составляющей. Оператор – не детерминированный алгоритм, его состояние $\mathcal{B}_t$ и контекст $\mathcal{C}_t$ динамичны и подвержены когнитивным искажениям. Главный инженерный вызов – проектирование интерфейса передачи контекстуального знания с минимальными потерями и искажениями. Традиционные интерфейсы (кнопки, графики) не годятся для выражения качественных нюансов. Необходимы интерфейсы нового поколения. Для предотвращения когнитивной перегрузки оператора система должна сама определять моменты, когда вмешательство критически необходимо. Это можно сделать, отслеживая «сигналы беспокойства»: быстрый рост

эпистемической дисперсии $\frac{d}{dt} \text{Tr}(\Sigma_t^{\text{epist}})$, приближение траектории к ранее неисследованным областям фазового пространства, или расхождение в прогнозах между разными членами ансамбля моделей. Только в такие моменты система должна запрашивать активное внимание оператора, в остальное время работая автономно.

5.3. Организационные и этические аспекты: распределение ответственности и обучение культуре симбиоза. Симбиотическая система не существует в вакууме – она встроена в организацию со своей иерархией, нормами и культурой. Ключевой вопрос: распределение ответственности за решения. Если управляющее воздействие u_t является результатом коррективы оператора v_t^* к рекомендации модели u_t^0 , то кто отвечает в случае аварии? Оператор, утвердивший решение? Разработчик модели, чья рекомендация была ошибочной? Создатель интерфейса, который не позволил оператору правильно выразить свои опасения? На практике необходимо юридическое и организационное закрепление принципа совместной ответственности с чётким протоколом анализа происшествий, который реконструирует эпистемическое состояние E_t на момент принятия решения.

Это требует ведения полного протокола диалога, включающего не только последовательность $(u_t^0, v_t^*, u_t, x_{t+1})$, но и метаданные: прогнозы с неопределённостью, визуализации, представленные оператору, его аннотации, оценку субъективного риска ρ_t^{human} . Такой протокол позволит впоследствии определить, была ли ошибка следствием «слепого пятна» модели, которую оператор не заметил, или же оператор проигнорировал явные предупреждения системы. Другим организационным вызовом является подготовка операторов. Они должны быть не просто пассивными наблюдателями, а активными «напарниками» искусственного интеллекта. Этические аспекты включают проблему смещения в симбиозе. Если оператор обладает неосознанными предубеждениями (например, систематически не доверяет рекомендациям в определённых ситуациях из-за стереотипов), его коррективы могут «обучить» модель этим же предубеждениям, закрепляя их в алгоритме. Необходимы механизмы обнаружения таких паттернов, например, статистический анализ корреляции между характеристиками ситуаций и частотой/направленностью человеческих вмешательств, с последующей калибровкой.

5.4. Области применения и границы парадигмы. Симбиотическое управление наиболее востребовано в областях с высокой степенью неопределённости, где формальные модели заведомо неполны, а цена ошибки велика. Однако существуют и границы применимости. Симбиотическое управление малоэффективно в двух крайних случаях.

Во-первых, в полностью формализуемых, детерминированных системах, где модель может быть исчерпывающе точной (например, управление роботом в контролируемой среде). Здесь человеческое вмешательство только замедлит работу.

Во-вторых, в системах, где человеческий компонент настолько доминирует, а его поведение настолько иррационально или непредсказуемо, что любая формализация бессмысленна (например, управление творческим коллективом в момент кризиса). Здесь модель не сможет построить даже приближённо адекватное представление.

Оптимальная область – это зона сложности, где существует устойчивая, но неполная регулярность, а человеческая интуиция служит мощным, но не абсолютным источником информации о латентных переменных.

5.5. Выводы: путь к инженерии доверия. Практическая реализация симбиотического управления – это не просто техническая задача создания алгоритмов и интерфейсов. Это комплексная задача инженерии доверия между человеком и машиной. Её успех зависит от создания устойчивого, самообучающегося контура. Ключевым метрическим показателем

успеха такой системы является не только итоговая эффективность J_R , но и динамика эпистемической неопределенности – скорость, с которой сокращается разрыв между моделью и реальностью, и уровень доверия в системе, измеряемый через стабильность и адекватность убеждений B_t . Таким образом, переход от парадигмы «цифрового Левиафана» к симбиотическому управлению представляет собой не просто смену алгоритмов, а глубокую трансформацию самой логики взаимодействия с технологией в сложном мире. Это путь от иллюзии абсолютного контроля к мудрости совместной навигации в условиях принципиальной неполноты знания, где риск перестаёт быть угрозой и становится общим компасом для человека и машины.

Заключение. Цифровизация управления – необратимый тренд, однако искушение построить «цифрового Левиафана», полагающегося на иллюзию абсолютного знания и тотальной оптимизации, является опасной ловушкой. Такая парадигма превращает риск из внешней угрозы во внутреннее, деструктивное свойство системы, неизбежно ведущее к дестабилизации, эрозии человеческого компонента и потере управляемости. Выход лежит в смене фундаментальной философии: мы должны отказаться от погони за призрачным оптимумом в виртуальном мире и честно признать принципиальную ограниченность наших моделей. Предлагаемый формализм риска предлагает сделать эту ограниченность предметом явного управления, а парадигма симбиотического управления выступает практическим механизмом его реализации, выстраивая непрерывный диалог между человеческой интуицией и машинным расчетом. Устойчивое будущее лежит через «приручение» цифровых двойников, превращая их из самодовлеющих симулякров в инструменты, служащие подлинным целям живой, адаптивной и человеко-центрированной эргатической системы. Это путь от иллюзии контроля к мудрости совместной навигации в условиях принципиальной неполноты знания, где риск становится не угрозой, а общим компасом.

Список источников

1. Бочков А.В. На пути к формализму риска. Размышления о риске и его природе / А.В. Бочков, В.В. Лесных. – Вологда: Инфра-Инженерия, 2024. – 296 с. – EDN: MXDJTV.
2. Гастев Ю.А. Гомоморфизмы и модели. Логико-алгебраические аспекты моделирования. – М.: Наука, 1975. – 152 с.
3. Пуговкина А.С. Имплицитные знания как аспект интуиции и практического интеллекта // Евразийский Союз Ученых, 2015. – № 3-7 (12). – URL: <https://cyberleninka.ru/article/n/implitsitnye-znaniya-kak-aspekt-intuitsii-i-prakticheskogo-intellekta> (дата обращения: 07.01.2026).
4. Журавлев В.Ф. Некорректные задачи механики // Вестник МГТУ им. Н.Э. Баумана. Серия «Приборостроение», 2017. – № 2 (113). – URL: <https://cyberleninka.ru/article/n/nekorrektnye-zadachi-mehaniki> (дата обращения: 07.01.2026). – DOI:10.18698/0236-3933-2017-2-77-85.
5. Бочков А.В. На пути к формализму риска. Безопасность и устойчивость функционирования умных экспансивных систем. – Вологда; М.: Инфра-Инженерия, 2025. – 348 с. – EDN: VQVVDB.
6. Belikov A. Age-related diseases as vicious cycles. Preprints, 2017, DOI: 10.20944/preprints201704.0102.v1.
7. Самарский А.А. Режимы с обострением в задачах для квазилинейных параболических уравнений / А.А. Самарский, В.А. Галактионов, С.П. Курдюмов, А.П. Михайлов. – М.: Наука, 1987. – 470 с.
8. Прелов В.В. f-дивергенция и склеивание вероятностных распределений // Проблемы передачи информации, 2021. – Т. 57. – № 1. – С. 64–80. – DOI:<https://doi.org/10.31857/S0555292321010034>.
9. Шереужев М.А. Модификация алгоритма глубокого обучения для распределения функций и задач между робототехническим комплексом и человеком в условиях неопределенности и переменности окружающей среды / М.А. Шереужев, У. Го, В.В. Серебряный // Известия КБНЦ РАН, 2024. – № 6. – URL: <https://cyberleninka.ru/article/n/modifikatsiya-algoritma-glubokogo-obucheniya-dlya-raspredeleniya-funktsiy-i-zadach-mezhdu-robototekhnicheskim-kompleksom-i> (дата обращения: 07.01.2026).
10. Кучкаров З.А. Методы концептуального анализа и синтеза в теоретическом исследовании и проектировании социально-экономических систем: учебное пособие: [курс лекций: для слушателей специальности «Концептуальное проектирование систем организационного управления»: в 2 т.]. – Т. 1: Методология концептуального анализа и синтеза; Методология концептуального проектирования систем организационного управления; Организационное консультирование. – М.: Концепт, 2008. – 262 с.

11. Son S., Seok J. Improving monte carlo dropout uncertainty estimation with stable output layers. *Neurocomputing*, 2025, vol. 661, p. 131927, DOI: 10.1016/j.neucom.2025.131927.
12. Kim T., Park G., Kwak K. et al. Smooth model predictive path integral control without smoothing. *IEEE Robotics and Automation Letters*, 2022, vol. 7, pp. 1-8, DOI: 10.1109/LRA.2022.3192800.
13. Панфилов И.А. Разработка математического и алгоритмического обеспечения интеллектуального проектирования аппаратно-программных комплексов обработки информации в распределенных высокопроизводительных системах / И.А. Панфилов, Н.А. Смирнов, В.А. Терсков // Сибирский аэрокосмический журнал, 2011. – № 5 (38). – URL: <https://cyberleninka.ru/article/n/razrabotka-matematicheskogo-i-algoritmicheskogo-obespecheniya-intellektualnogo-proektirovaniya-apparatno-programmnyh-kompleksov> (дата обращения: 07.01.2026).

Бочков Александр Владимирович. Доктор технических наук, ученый секретарь, АО «НИИАС» (Россия, Москва). Автор более 100 научных публикаций, в том числе 6-х монографий. Область научных интересов: системы жизнеобеспечения человека в экстремальных условиях окружающей среды, исследования нестационарных процессов, вопросы синтеза и анализа риска, безопасности критической инфраструктуры, системный анализ и исследование операций, экспертные методы. Scopus Author ID: 57188739597, ORCID: 0000-0003-0211-8320, ResearcherID: A-4101-2014, SPIN: 8277-6641, AuthorID: 660808 a.bochkov@gmail.com. 107078, Россия, Москва, Орликов переулок, д.5, стр.1.

UDC 005.334:004.94

DOI:10.25729/ESI.2026.43.3.001

Risk as an immanent property of ergatic systems in the era of digital twins: from the "digital Leviathan" to the paradigm of symbiotic management

Alexander V. Bochkov

JSC "NIAS", Moscow, Russia, a.bochkov@gmail.com

Abstract. The current transition to so-called "data-driven management," based on the creation of full-scale digital twins of complex socio-economic systems, faces a fundamental challenge. Managed systems are ergatic, meaning they include ultimately non-formalizable human, technological, and environmental components. This article demonstrates that optimization problems solved for such digital models are inherently ill-posed with respect to the original, as the human element introduces reflexive non-formalizability and implicit knowledge that cannot be captured by any formal structure. Risk in this paradigm ceases to be an external threat, becoming an immanent property of the management system itself, generated by the inevitable and growing discrepancy between the trajectories of the virtual model and the real ergatic process. Strategies optimal for the "Digital Leviathan" (total simulation) lead to the destabilization of the original, erosion of the human component, and the system exceeding the boundaries of controllability. The authors substantiate the need for a paradigm shift: from searching for a point optimum for an incomplete model to synthesizing robust and adaptive strategies that explicitly minimize the risk functional. A key element of the new "symbiotic management" paradigm is the formal inclusion of the human operator as a corrective loop in the iterative dialogue between the model and reality. The proposed architecture of symbiotic management, grounded in Bayesian updating and minimax principles, enables not merely risk minimization but active management of epistemic uncertainty, transforming human intuition into a structured source of model correction. This approach ensures the resilience of complex socio-technical systems under conditions of fundamental knowledge incompleteness, offering a path from the illusion of absolute control to the wisdom of collaborative navigation. The goal of management becomes not an absolute optimum in virtual space, but dynamic adaptive efficiency in the real world, achieved through constant management of the immanent risk of trajectory divergence. The article introduces and contrasts two concepts: the "Digital Leviathan" (the temptation to manage the simulation) and the "formalism of risk" (the discipline of constantly comparing the map and the territory).

Keywords: digital twin, ergatic system, immanent risk, optimization problem, model incorrectness, symbiotic management, adaptive strategy, human factor, epistemic uncertainty

References

1. Bochkov A.V., Lesnykh V.V. Na puti k formalizmu riska. Razmyshleniya o riske i ego prirode [On the path to formalism of risk. Reflections on risk and its nature]. Vologda, Infra-Inzheneriya Publ., 2024, 296 p. ISBN 978-5-9729-2146-1, EDN: MXDJTV.
2. Gastev Yu.A. Gomomorfizmy i modeli. Logiko-algebraicheskiye aspekty modelirovaniya [Homomorphisms and models. Logical-algebraic aspects of modeling]. Moscow, Nauka [Science] Publ., 1975, 152 p.
3. Pugovkina A.S. Implitsitnyye znaniya kak aspekt intuitsii i prakticheskogo intellekta [Implicit knowledge as an aspect of intuition and practical intelligence]. Evraziyskiy Soyuz Uchenykh [Eurasian Union of Scientists], 2015, no. 3-7 (12). Available at: <https://cyberleninka.ru/article/n/implitsitnye-znaniya-kak-aspekt-intuitsii-i-prakticheskogo-intellekta> (accessed: 01.07.2026).
4. Zhuravlev V.F. Nekorrektnyye zadachi mekhaniki [Ill-posed problems of mechanics]. Vestnik MGTU im. N.E. Baumana. Seriya "Priborostroyeniye" [Bulletin of Bauman MSTU. Series "Instrument Engineering"], 2017, no. 2 (113). Available at: <https://cyberleninka.ru/article/n/nekorrektnye-zadachi-mekhaniki> (accessed: 01.07.2026), DOI:10.18698/0236-3933-2017-2-77-85.
5. Bochkov A.V. Na puti k formalizmu riska. Bezopasnost' i ustoychivost' funktsionirovaniya umnykh ekspansivnykh sistem [On the path to formalism of risk. Safety and stability of smart expansive systems]. Vologda; Moscow, Infra-Inzheneriya Publ., 2025, 348 p. ISBN 978-5-9729-2373-1, EDN: VQVVDDB.
6. Belikov A. Age-related diseases as vicious cycles. Preprints, 2017, DOI: 10.20944/preprints201704.0102.v1.
7. Samarskii A.A., Galaktionov V.A., Kurdyumov S.P., Mikhailov A.P. Rezhimy s obostreniyem v zadachakh dlya kvazilineynykh parabolicheskikh uravneniy [Blow-up in problems for quasilinear parabolic equations]. Moscow, Nauka [Science] Publ., 1987, 470 p. (English translation: Samarskii A.A., Galaktionov V.A., Kurdyumov S.P., Mikhailov A.P. Blow-up in problems for quasilinear parabolic equations. Berlin; N.Y., Walter de Gruyter, 1995, 533 p.)
8. Prelov V.V. f-divergence and gluing of probability distributions. Problems of Information Transmission, 2021, vol. 57, no. 1, pp. 64–80, DOI: 10.31857/S0555292321010034.
9. Sherezuzhev M.A., Guo U., Serebrenny V.V. Modifikatsiya algoritma glubokogo obucheniya dlya raspredeleniya funktsiy i zadach mezhdru robototekhnicheskimi kompleksami i chelovekom v usloviyakh neopredelennosti i peremennosti okruzhayushchey sredy [Modification of a deep learning algorithm for distributing functions and tasks between a robotic complex and a human in conditions of uncertainty and variability of the environment]. Izvestiya KBNTs RAN [Proceedings of the Kabardino-Balkarian Scientific Center of RAS], 2024, no. 6. Available at: <https://cyberleninka.ru/article/n/modifikatsiya-algoritma-glubokogo-obucheniya-dlya-raspredeleniya-funktsiy-i-zadach-mezhdru-robototekhnicheskimi-kompleksami-i> (accessed: 01.07.2026).
10. Kuchkarov Z.A. Metody kontseptual'nogo analiza i sinteza v teoreticheskom issledovanii i proyektirovanii sotsial'no-ekonomicheskikh sistem: uchebnoye posobiye [Methods of conceptual analysis and synthesis in theoretical research and design of socio-economic systems: textbook]. Vol. 1. Moscow, Kontsept Publ., 2008, 262 p.
11. Son S., Seok J. Improving monte carlo dropout uncertainty estimation with stable output layers. Neurocomputing, 2025, vol. 661, p. 131927, DOI: 10.1016/j.neucom.2025.131927.
12. Kim T., Park G., Kwak K. et al. Smooth model predictive path integral control without smoothing. IEEE Robotics and Automation Letters, 2022, vol. 7, pp. 1-8, DOI: 10.1109/LRA.2022.3192800.
13. Panfilov I.A., Smirnov N.A., Terskov V.A. Razrabotka matematicheskogo i algoritmicheskogo obespecheniya intellektual'nogo proyektirovaniya apparatno-programmnykh kompleksov obrabotki informatsii v raspredelennykh vysokoproizvoditel'nykh sistemakh [Development of mathematical and algorithmic support for intelligent design of hardware-software complexes for information processing in distributed high-performance systems]. Sibirskiy aerokosmicheskii zhurnal [Siberian Aerospace Journal], 2011, no. 5 (38). Available at: <https://cyberleninka.ru/article/n/razrabotka-matematicheskogo-i-algoritmicheskogo-obespecheniya-intellektualnogo-proektirovaniya-apparatno-programmnykh-kompleksov> (accessed: 01.07.2026).

Bochkov Alexander Vladimirovich. Doctor of Technical Sciences, Scientific Secretary, JSC "NIIAS" (Moscow, Russia). Author of over 100 scientific publications, including 5 monographs. Research interests: human life support systems in extreme environmental conditions, studies of non-stationary processes, risk synthesis and analysis, critical infrastructure safety, systems analysis and operations research, expert methods. Scopus AuthorID: 57188739597, ORCID: 0000-0003-0211-8320, ResearcherID: A-4101-2014, SPIN: 8277-6641, AuthorID: 660808 a.bochkov@gmail.com. 107078, Russia, Moscow, Orlikov Pereulok, 5/1.

Статья поступила в редакцию 28.01.2026; одобрена после рецензирования 18.02.2026; принята к публикации 11.08.2026.

The article was submitted 01/28/2026; approved after reviewing 02/18/2026; accepted for publication 08/11/2026.