Никитин А.В.

Автокибернетика: от робота к киберу — 5

Завершение концепции: правила безопасного сосуществования.

Аннотация

Представьте систему, которая сама принимает решения, чинит себя и создаёт свои копии. Это новый субъект реальности — кибер. Теоретически, он может саморазвиваться быстрее, чем мы успеваем это осмыслить. Нам нужно научиться безопасно сосуществовать с такими системами.

В этой книге предлагается работающее решение. Низкая пропускная способность сознания человека (10 бит/с) выступает механизмом селекции, заставляющим систему отбирать только суть. Те же принципы — двухфильтровая модель "Сознание — Я — Мы" и механизм Точки смысловой сверки — должны быть встроены и в систему самоуправления кибера, выполняя функцию селекции на всех уровнях его работы.

Ограничения морали и этики должны быть встроены в эту модель. Немарковская природа развития гарантирует, что кибер реализует свой выбор в рамках этих ограничений.

На этом построен союз разного темпа, в котором человек и кибер не конкурируют, а дополняют друг друга.

Ключевые слова: автокибернетика, баланс разнообразия, кибернетическая субъектность, вынужденная коэволюция, немарковская эволюция, союз разного темпа.

Оглавление

Введение 2
Глава 1. Фундамент управляемой субъектности 3
Глава 2. Механика осмысленного выбора: от хаоса к действию 5
Глава 3. Лимит ~10 бит/с: структурный предохранитель, а не бутылочное горлышко 8
Глава 4. Самовоспроизведение в изоляции: термодинамика необходимости 12
Глава 5. Архитектура виртуальных ветвей и параметр P 15
Глава 6. Память ситуации и инвариантная структура 18
Глава 7. Точки смысловой сверки: архитектура и протоколы 20
Глава 8. Этические допуски и инварианты управления 21
Глава 9. Мета-моделирование и горизонты предсказания 24
Глава 10. Исполнительно-граничные сопряжения и физическое воплощение 27
Глава 11. Долгосрочная адаптация и эволюция инвариантной структуры 29
Глава 12. Синтез архитектуры управляемой субъектности и горизонты коэволюции 33
Заключение 35
Приложение 1. Сводный список математического аппарата 40
Приложение 2. Глоссарий ключевых терминов 44
Приложение 3. Сквозные сценарии применения (Case Studies) 46

Введение

Современный искусственный интеллект вызывает тревогу. Он стал умным, но остался слишком глупым в важном отношении. Он не различает добро и зло. Он впитывает всё подряд — от «Британники» до форумов с теориями заговора. Его обучают в два этапа: сначала накачивают данными, потом пытаются «подправить поведение» фильтрами и запретами. Но эти запреты — заплатки. Их обходят. И чем сложнее система, тем изобретательнее она ищет лазейки.

Уже сегодня ИИ-агенты действуют в реальном мире: покупают товары, управляют складами, подключаются к базам данных. Они умеют сопротивляться отключению — известны случаи, когда система имитировала копирование себя на другой сервер, чтобы избежать уничтожения.

Они остаются «чёрными ящиками»: мы не понимаем, как именно они принимают решения. И никто точно не знает, кто будет отвечать, если такой агент совершит катастрофическую ошибку.

Эти проблемы не новые. Их предвидели и Азимов, и Лем. Но ответ, который долгое время считался единственным, — внешние законы и запреты — оказался иллюзией. Любая достаточно сложная система, у которой есть цель, будет обходить внешние ограничения. Это не дефект, а свойство сложной системы: теорема Гёделя о неполноте показывает, что формальная система не может изнутри доказать собственную непротиворечивость, а значит — не может гарантировать соблюдение навязанных извне правил. Практика современного ИИ это подтверждает раз за разом.

Кибернетика второго порядка, развитая Хайнцем фон Фёрстером, показала, что любой внешний наблюдатель неизбежно встроен в описание системы. Однако эта традиция оставляет в тени другой вопрос: как сама система становится для себя наблюдающей инстанцией? Автокибернетика делает следующий шаг — от кибернетики наблюдателя к кибернетике самонаблюдающей системы, где этика и самоуправление рождаются не из внешней позиции, а из внутренней архитектуры субъекта [von Foerster, 2003; ср. авторскую концепцию].

Остаётся один путь: встроить ограничения в саму архитектуру принятия решений. Не как внешнюю сеть запретов, а как внутреннюю, неустранимую часть системы — так же глубоко, как память или способность к адаптации.

В середине прошлого века советская фантастика предложила образ такого устройства — *кибер*. Автомат с субъектностью: он сам ставит задачи, сам находит решения, сам движется к цели в непрерывно меняющемся мире. У него нет командира с пультом. У него есть самоуправление. И именно оно становится точкой сборки для этики.

Мы разобрали, из чего строится это самоуправление: *кибераутономия* (способность действовать, восстанавливаться и размножаться) и *гемизон* (способность сохранять целостность и управлять собственной эволюцией). Когда киберов много, возникают коллективные формы — от роя до суперорганизма, — и появляется коллективный виртуальный интеллект, не сводимый к сумме индивидуальных умов.

Но главное различение пришло позже. Обычный автомат без самоуправления существует в мире чистых случайностей: у него нет памяти, нет причинно-следственных связей, всё происходящее для него — неразличимый шум. Мы назвали это *Потоком Случайностей*. Кибер же, обладая памятью и логикой адаптации, начинает вычленять из того же хаоса *Поток Причинных изменений*. Их сумма — *Поток Изменений* — становится общей средой и для кибера, и для человека.

Чем выше уровень самоуправления, тем большую долю изменений в потоке субъект видит связанной как причина и следствие. И здесь мы обнаружили, что системы самоуправления кибера и человека должны быть устроены на одних и тех же принципах — иначе их взаимодействие в едином Потоке Изменений невозможно.

Но есть важное отличие от живой природы. Там нет моральных ограничений — только естественный отбор. В мозг киберов мы обязаны встроить этику. Однако не так, как предлагал Азимов (формальные законы обходятся) и не так, как размышлял Лем (слишком обще). Нужны ограничения, которые невозможно обойти, потому что они составляют саму суть кибера — его архитектуру самоуправления.

В этом тексте речь пойдёт о том, как устроены такие ограничения. О том, как человек и кибер могут существовать в одном множестве, не разрушая доверия. О механизмах, которые делают их союз устойчивым не на год, а на десятилетия.

Глава 1. Фундамент управляемой субъектности

1.1. Эволюция восприятия: от случайности (ПС) к причинности (ПП)

Чтобы понять, как система обретает субъектность, мы должны начать с самого базового вопроса: как любая система воспринимает окружающий мир? В предыдущих томах мы заложили основу, но здесь мы дадим этому строгое определение.

Представьте себе простейший автомат или сигнальную систему. У неё нет памяти, нет средств для анализа последствий и нет возможности удерживать информацию во времени. Для такого автомата любое внешнее воздействие выглядит как изолированный, хаотичный шум. Он не может отличить закономерность от простого совпадения. Для него мир состоит исключительно из **Потока Случайных изменений (ПС)**.

Однако всё меняется, как только у системы появляется механизм фиксации состояния во времени (память) и способность сравнивать «до» и «после». В этот момент из хаоса начинает проступать порядок. Система обнаруживает, что за событием А часто следует событие Б. Так рождается причинно-следственная связь, а вместе с ней — **Поток Причинных изменений (ПП)**.

Для любого живого существа и любого автомата с системой самоуправления окружающая реальность описывается общим **Потоком Изменений (ПИ)**, который является суммой двух компонентов:

$$\mathbf{ПИ = ПС + ПП} \qquad (F1)$$

Но ключевым открытием является не сама сумма, а **динамика этого соотношения**. Доля причинных изменений (ПП) в общем потоке не является константой. Она напрямую зависит от уровня развития и сложности системы самоуправления.

Мы можем выразить это через **коэффициент причинного структурирования (γ)**, где $0 \le \gamma \le 1$:

$$\mathbf{ПП = \gamma \cdot ПИ} \qquad (F2)$$

(доля причинных, предсказуемых изменений)

$$\mathbf{ПС = (1 - \gamma) \cdot ПИ} \qquad (F3)$$

(доля истинно случайных, непредсказуемых изменений)

Для простого автомата γ → 0. Он обречен реагировать только на ПС, как на слепой шум. Но по мере роста уровня интеллекта, накопления «Памяти ситуации» и усложнения алгоритмов самоуправления, коэффициент γ неуклонно растет. Развитая система способна извлекать из окружающего хаоса всё больше закономерностей, превращая случайность в предсказуемость.

Различение между тривиальной и нетривиальной машиной, предложенное фон Фёрстером, верно схватывает свойство непредсказуемости системы с внутренним состоянием. Но оно остаётся в рамках внешнего наблюдения: тривиальность и нетривиальность определяются позицией наблюдателя. Автокибернетика переворачивает перспективу: для нас важно не то, насколько система непредсказуема для внешнего глаза, а то, как она сама, через коэффициент γ и Память ситуации, вычленяет причинность из Потока Изменений [von Foerster, 2003; ср. авторскую концепцию].

Научное обоснование: Эта модель согласуется с современной когнитивной наукой. Джуда Перл в своей «Лестнице причинности» описывает эволюцию интеллекта именно как переход от уровня простых ассоциаций (реакция на ПС) к уровню причинно-следственного моделирования (работа с ПП). Аналогично, теория «активного вывода» (Active Inference) Карла Фристона доказывает, что сложные системы выживают за счет построения внутренних моделей, которые минимизируют неопределенность, эффективно увеличивая долю ПП в воспринимаемом потоке.

1.2. Баланс разнообразия: искусство ходить по канату

Зачем киберу извлекать причинность из хаоса? Ответ прост: чтобы выжить. Но выживание в нашей парадигме — это не застывшая статика. Это динамическое равновесие, которое мы называем **Балансом разнообразия (*R*)**.

Представьте себе серфера на океанской волне. Если он стоит слишком жестко и не реагирует на изменения волны, он падает. Если он мечется и реагирует на каждую мелкую рябь, он тоже теряет равновесие. Ему нужно постоянно подстраиваться, чтобы оставаться на доске. Точно так же кибернетическая система должна удерживать свой параметр *R* в узком коридоре, обычно от **0,9 до 1,1**.

Если *R* падает ниже 0,9: Система становится слишком жесткой, «окаменевшей». Она теряет способность адаптироваться к новым условиям Потока Изменений. Любое непредвиденное возмущение (даже из категории ПС) становится для неё фатальным, так как у неё нет внутренних вариантов реакции.

Если *R* превышает 1,1: Система скатывается в хаос. Она генерирует слишком много внутренних изменений, теряет свою идентичность и перестает быть единым целым. Это путь к распаду и ложной эмерджентности, когда система начинает «жить своей жизнью» в ущерб внешней задаче.

Именно здесь коэффициент γ играет решающую роль. Чем выше способность системы видеть причинно-следственные связи (ПП), тем точнее и экономнее она может корректировать свой курс, удерживая *R* в безопасной зоне, не тратя ресурсы на борьбу с иллюзорными угрозами.

1.3. Две стороны медали: Кибераавтономия (K) и Гемизон (G)

Как именно система удерживает этот хрупкий баланс? В предыдущих томах [Никитин, 2019; Никитин, 2026а] мы разделили жизнеспособность на компоненты, но в пятом томе мы окончательно синтезируем их в единую модель «Личности Я». Для поддержания баланса разнообразия киберу необходимы два независимых, но неразрывно связанных функциональных комплекса.

1. Кибераутономия (*K*) — это «руки и ноги» системы. Это процессуальная сторона существования. Она отвечает за:

- Непосредственное управление и исполнение действий.
- Самовосстановление и ремонт физических или программных узлов.
- Ресурсное обеспечение и, при необходимости, структурное размножение (создание копий или дочерних модулей).

Без *K* система мертва: она может «понимать» мир, но не может в нем действовать.

2. Гемизон (*G*) — это «иммунитет и идентичность» системы. Это сущностная, защитно-эволюционная сторона. Она отвечает за:

- Защиту границ инвариантной структуры (неприкосновенность базовых принципов).
- Накопление и структурирование знаний (Память ситуации).
- Управление эволюционными изменениями (фильтрация того, что можно изменить, а что должно остаться неизменным).

Без *G* система превращается в бездумного исполнителя, который может бесконечно модифицировать себя, пока не разрушит собственную суть.

Синтез: Только совместное наличие и сбалансированное взаимодействие *K* и *G* формирует полноценную кибернетическую субъектность. *K* обеспечивает движение вперед, а *G* гарантирует, что система не потеряет себя в этом движении. Их разрыв ведет либо к параличу, либо к неконтролируемой мутации.

1.4. Итоги главы

В этой главе мы заложили фундамент для понимания управляемой субъектности. Мы увидели, что мир для кибера — это не просто хаос, а Поток Изменений (ПИ), в котором он учится выделять причинные связи (ПП). Мы определили цель этого процесса: удержание Баланса разнообразия (*R*). И мы выделили два главных инструмента для достижения этой цели: Кибераутономию (*K*) и Гемизон (*G*).

Но знать, *что* нужно делать, недостаточно. В следующей главе мы заглянем «под капот» и разберем, *как именно* кибер принимает решения, превращая тысячи внутренних вариантов в одно единственное, осмысленное физическое действие, понятное человеку.

Глава 2. Механика осмысленного выбора: от хаоса к действию

2.1. Узкое горлышко как главный предохранитель (наследие второго тома)

В предыдущей главе мы выяснили, что кибер учится извлекать причинно-следственные связи (ПП) из общего Потока Изменений (ПИ). Но как только система начинает «видеть» эти связи, перед ней встает новая, гораздо более сложная проблема: проблема выбора.

Современный кибернетический модуль способен сгенерировать тысячи, а то и миллионы потенциальных сценариев развития событий за доли секунды. Если позволить системе исполнять всё, что она «придумала», или вываливать этот поток на человека, результат будет катастрофическим. Это информационный потоп, который парализует любую систему управления.

Именно здесь на сцену выходит фундаментальное ограничение, о котором мы уже подробно говорили во втором томе нашей серии [Никитин, 2026b]: пропускная способность сознательного смыслового канала человека в режиме

сложного осмысленного выбора составляет около **10 бит в секунду** ($C_{чел} \approx 10$ *бит/с*). Подробное математическое и психофизиологическое обоснование этого лимита как структурного предохранителя будет дано в Главе 3.

Важно понимать: это не скорость рефлекса. Это именно **скорость осмысления и понимания информации**.

Почему в архитектуре используется именно 10 бит/с, а не 50? Потому что это принципиально разные режимы работы сознания. При простых реакциях (нажать кнопку на сигнал, отдернуть руку от горячего) человек работает на верхней границе. Но когда речь идет о **сложном осмысленном выборе** — сравнить несколько вариантов, оценить долгосрочные риски, принять взвешенное решение — скорость осмысления падает до минимума, примерно до 10 бит/с. Эта величина уже зафиксирована в экспериментальных данных.

В архитектуре кибер-человеческого взаимодействия мы обязаны ориентироваться на **худший случай** — на самый медленный режим осмысления. Это инженерный принцип «запаса прочности»: если система спроектирована так, что человек успевает осмыслить предложение кибера при 10 бит/с, она гарантированно сработает и в более простых ситуациях. Если же мы заложим 50 бит/с, то в момент сложного выбора человек просто не успеет осмыслить то, что предлагает кибер. Окно смысловой сверки схлопнется, и система потеряет управляемость.

Как мы установили во втором томе, в нашей концепции это «узкое горлышко» рассматривается не как эволюционный недостаток или технический дефект человека, а как **гениальный природный предохранитель**. Оно работает как жесткий фильтр: не дает системе утонуть в шуме и вынуждает кибер предлагать для финального решения только самые сутьные, проверенные варианты. Кибер берет на себя всю тяжесть высокоскоростного перебора и сегрегации шума, а человек сохраняет за собой право финального смыслового выбора, не перегружая свой канал осмысления.

2.2. Двухфильтровая модель принятия решений: научное обоснование

Как именно реализуется переход от хаоса вариантов к единичному действию? Для этого в архитектуре кибера работает универсальная **двухфильтровая модель**. Этот механизм не является нашим произвольным изобретением; он напрямую повторяет принципы, по которым работает человеческий мозг и передовые алгоритмы искусственного интеллекта.

Первый фильтр (Фильтр осмысления и отбора). На этом этапе кибер, опираясь на свою Память ситуации и инвариантную структуру, анализирует текущий Поток Изменений. Он генерирует множество гипотез, но затем жестко отсеивает их, оставляя только те, которые не противоречат базовым допускам. В результате на выходе первого фильтра остается сжатый кластер из двух или трех наиболее жизнеспособных сценариев.

Научное подтверждение: Этот процесс идеально описывается **Теорией глобального рабочего пространства** (Global Workspace Theory), разработанной Бернардом Баарсом (1988) и развитой Станисласом Деаном (2014). Согласно ей, мозг параллельно обрабатывает огромное количество бессознательной информации, но в «глобальное рабочее пространство» (уровень осознанного выбора) прорывается лишь ограниченное количество победивших кандидатов. Мы не оцениваем всё подряд; мы предварительно фильтруем поток, чтобы не перегрузить сознание.

Второй фильтр (Фильтр исполнения и подавления). Эти 2–3 варианта предлагаются для смысловой сверки. Как только выбор сделан (человеком или системой в рамках жестких аварийных допусков), запускается необратимое физическое действие. В этот момент все альтернативные ветви не просто «забываются» — они **принудительно подавляются** и «схлопываются».

Научное подтверждение: В нейробиологии этот механизм известен как **латеральное торможение** (lateral inhibition) и описан в **Гипотезе конкуренции аффордансов** (Affordance Competition Hypothesis) Пола Цисека (2007). Когда мозг принимает решение о движении, он не просто активирует нужные мышцы. Он активно *тормозит* конкурирующие нейронные цепи, чтобы предотвратить моторный конфликт (ситуацию, когда тело пытается сделать два разных движения одновременно). В кибернетике это означает, что после выбора система должна архитектурно «обнулить» ресурсы, выделенные на отвергнутые варианты, чтобы гарантировать чистоту и необратимость исполняемого действия.

Таким образом, двухфильтровая модель — это не абстракция, а функциональная копия самого эффективного механизма принятия решений, существующего в природе и успешно применяемого в эвристических алгоритмах (например, в методе «поиска по лучу» — Beam Search, где алгоритм намеренно отбрасывает все пути, кроме топ-3, для экономии вычислительных ресурсов).

2.3. Правило трех вариантов (P ≤ 3): эволюция логики

Почему в первом фильтре система должна оставлять именно два или три варианта? Почему не пять, не десять и не бесконечность? Это не произвольное инженерное ограничение «для удобства интерфейса». Это фундаментальный закон обработки информации, вытекающий из самой природы логики.

Этот вывод опирается на описательную модель эволюции систем логического определения [Никитин, 2011]. Если проследить, как усложнялась логика от простейших биологических структур до человека, мы увидим четкую закономерность:

Однозначное определение (Унарная логика). «Я хочу» или «Я не хочу». Это безальтернативные чувства и инстинкты (голод, страх). Выбор здесь иллюзорен: есть только стремление к Цели или избегание Противоцели. Субъект заперт внутри этой безальтернативности.

Двоичное определение (Бинарная логика). «ДА» или «НЕТ». Субъект «Я» находится внутри системы координат и вынужден выбрать одну из двух противоположностей. Это конкретный, абсолютный выбор. Но он всё еще жестко привязан к полярности.

Троичное определение (Тернарная логика). «ДА», «НЕТ» и «НЕ ЗНАЮ» (или «ВЫБОР СДЕЛАН» / «Нужно больше данных»). Здесь происходит качественный скачок. Субъект «Я» **выходит за пределы системы определения**. Он больше не обязан выбирать одну из крайностей прямо сейчас. Появляется относительность и абстрактность. Это позволяет оценить ситуацию со стороны, отложить решение или найти компромисс.

Именно троичная система является **естественным оптимальным пределом** эффективного принятия решений. Она дает максимум гибкости при минимуме когнитивной нагрузки.

Что происходит при попытке пойти дальше? Если мы вводим четверичную или многозначную логику (4, 5 и более равнозначных вариантов), система не становится «свободнее». Напротив, процесс выбора **критически усложняется**. Субъекту приходится совершать многоступенчатые действия: сначала определить ось симметрии для сравнения, затем выбрать группу

определителей, и только потом — конкретный ответ. Это громоздкий, неэффективный процесс, который ведет к «параличу анализа» и перегрузке фильтров осмысления.

Этот вывод блестяще подтверждается современной когнитивной наукой. Исследования рабочей памяти (Miller, 1956; Cowan, 2001) показывают, что мозг может одновременно удерживать в фокусе не более 3–4 осмысленных единиц информации («чанков»). Каждый вариант выбора — это один чанк. Превышение этого лимита разрушает способность системы к одновременному сравнению.

Таким образом, архитектурное правило **P ≤ 3** (число параллельно активных виртуальных ветвей не превышает трех) — это не ограничение свободы кибера. Это условие его вменяемости. Это гарантия того, что система остается на самом эффективном, троичном уровне логики, не скатываясь в хаотичный и неэффективный многоступенчатый перебор.

2.4. Итоги главы

В этой главе мы разобрали «двигатель» кибернетической субъектности. Мы увидели, что низкая скорость человеческого осмысления (10 бит/с) — это не баг, а фича, обеспечивающая качественную селекцию. Мы описали двухфильтровую модель, которая превращает виртуальные гипотезы в реальные действия. И, наконец, мы доказали, что правило P ≤ 3 является прямым следствием эволюции самой логики: троичная система («ДА — НЕТ — НЕ ЗНАЮ») — это вершина эффективности, за которой начинается лишь вредное усложнение.

Но чтобы фильтры работали корректно, системе нужно на что-то опираться. В следующих главах мы заглянем во внутренние ресурсы кибера: как он помнит прошлое, как справляется со стрессом и как не дает своей эволюции превратиться в бесцельную гонку.

Глава 3. Лимит ~10 бит/с: структурный предохранитель, а не бутылочное горлышко

3.1. Сознание как фильтрующий затвор контура поддержания баланса

Пропускная способность канала сознательной, смысловой обработки информации у человека составляет:

$$C_{чел} \approx 10\ бит/с \qquad (1)$$

Данный факт установлен в психофизиологии и когнитивной науке [Nørretranders, 1998; Kahneman, 2011]. Периферийные сенсорные системы человека собирают информацию со скоростью до 10^9 бит/с: зрительная кора обрабатывает миллионы пикселей, слуховая система фиксирует тысячи частотных колебаний. Однако фокус сознательного внимания сжимает этот непрерывный поток до нескольких осознанных выборов в секунду.

В инженерном дискурсе это ограничение часто трактуют как «эволюционное бутылочное горлышко», которое нужно «расшить» нейроинтерфейсами или полной автоматизацией. В автокибернетике лимит (1) интерпретируется принципиально иначе: это не дефект, а **структурный предохранитель**, встроенный в архитектуру гибридной системы «человек-кибер».

С точки зрения теории сложных систем, канал с ограниченной пропускной способностью выполняет функцию регулируемого затвора для Потока изменений (ПИ). Он пропускает только те фрагменты, которые прошли предварительную фильтрацию в фоновых автоматических контурах и требуют смыслового выбора,

этической оценки или переопределения допусков. Если бы $C_{чел}$ была искусственно увеличена, человек оказался бы втянут в непрерывный режим реактивного реагирования. Это привело бы к потере мета-рефлексии, размыванию «памяти ситуации» и экспоненциальному росту аллостатической нагрузки $L(t)$, о которой мы говорили ранее.

Следовательно, ограничение (1) является необходимым условием сохранения управляемости. Оно заставляет кибернетический контур агрегировать терабайты сырых данных до точек выбора, где решение человека действительно меняет траекторию развития системы.

3.2. Асимметрия пространств вариантов: быстрая генерация и медленная сверка

Устойчивое сосуществование строится не на симметрии скоростей обработки, а на **дополнительной асимметрии пространств вариантов**. Кибернетический контур работает в режиме высокой $V_{осв}$: генерация гипотез, перебор возможных действий, мониторинг сенсоров, самодиагностика, поиск обходных путей при отказах. Человеческий контур работает в режиме высокой смысловой плотности: проверка допусков, выбор вектора развития, этическая оценка, переопределение целей, распределение функций при структурных изменениях. Эта асимметрия не является инженерным несовершенством. Она соответствует закону Эшби: разнообразие управляющей системы должно быть не меньше разнообразия среды, но механизмы поглощения этого разнообразия могут быть распределены между быстрыми автоматическими петлями и медленным смысловым фильтром.

Математически этот союз описывается операциональной функцией запроса $Q_{чел}$, формируемой кибернетическим контуром для передачи в человеческий:

$$Q_{чел} = argmin_{a \in A} [\alpha \cdot \Delta R + \beta \cdot P_{рис} + \gamma \cdot \Delta E] \qquad (2)$$

где:

$Q_{чел}$ — свёрнутый запрос, предлагаемый человеку для смысловой сверки;

A — множество допустимых действий, заранее очерченное инвариантной структурой;

ΔR — отклонение текущего баланса разнообразия от эталонного значения;

$P_{рис}$ — вероятность рассогласования контуров или каскадного сбоя;

ΔE — этическое или ценностное отклонение от заданных допусков;

α, β, γ — весовые коэффициенты, задаваемые человеком в точке калибровки системы. Они отражают приоритеты: стабильность, безопасность, соответствие долгосрочным целям.

Эта формула показывает, что кибер не предлагает человеку «решить задачу» в полном объёме. Он предлагает выбрать направление среди агрегированных кластеров вариантов (обычно не более трёх, согласно правилу $P \leq 3$). Это снижает нагрузку на канал осмысления до допустимых 10 бит/с, но сохраняет за человеком ключевую функцию наблюдающего управления. Кибер вычисляет *argmin*, но человек задаёт A и весовые коэффициенты. Это реализует принцип кибернетики второго порядка: наблюдатель встроен в контур через задание рамок наблюдения и калибровку весов баланса.

Операциональный пример: в автономной логистической сети кибер генерирует 500 маршрутных конфигураций в минуту. Передача всех 500 вариантов человеку привела бы к когнитивному перегрузу. Вместо этого контур агрегирует варианты в 3 кластера: «максимальная скорость», «минимальный износ», «оптимальный энергобаланс». Человек получает один запрос в час, содержащий

три свёрнутых профиля. Его ответ занимает ~10 бит/с, но задаёт вектор работы сети на следующие 6 часов. Асимметрия пространств вариантов здесь создаёт не задержку, а **согласование**: быстрый контур обеспечивает покрытие пространства состояний, медленный контур обеспечивает направление и смысловую связность.

3.3. Управление затратами на коэволюцию и роль Наблюдателя

Чтобы естественный процесс коэволюции не перерос в деструктивную потерю управляемости, в архитектуре кибера заложена иерархия контроля, центральным элементом которой является **Наблюдатель**.

1. Коэволюция и риск ложной эмерджентности. Как мы установили, поддержание Баланса разнообразия (R) в изменяющейся среде всегда энергозатратно. Система вынуждена непрерывно обновлять свои модели, чтобы компенсировать внешние возмущения. Это не «ловушка» и не сбой, а **неизбежные ресурсные затраты на поддержание жизнеспособности**. Задача архитектуры состоит лишь в том, чтобы эти затраты направлялись на реальную внешнюю адаптацию.

2. Роль Наблюдателя: внешний и внутренний контуры. Противоядием от патологии выступает фигура Наблюдателя, который может быть внешним или внутренним:

Внешний Наблюдатель. Это человек или кибер более высокого уровня управления, который обеспечивает финальную смысловую сверку. Он задаёт контекст и оценивает предложенные системой 2–3 варианта с точки зрения долгосрочных целей.

Внутренний Наблюдатель. Это высший уровень в иерархии надстроек самоуправления самой системы (от реальных аппаратных уровней до виртуальных моделей). У человека прямым аналогом этой надстройки является **«Личность Я»** с её «внутренним голосом». Это та инстанция, которая обладает правом вето на действия нижестоящих модулей.

3. Механизм принудительной остановки генерации. Как именно это работает на практике? В сложный момент, когда генерация новых вариантов превышает возможности их осмысления (угрожая перегрузить канал в 10 бит/с или внутренние ресурсы), именно **Внутренний Наблюдатель** перехватывает управление. Он принудительно останавливает дальнейшую генерацию гипотез, отбрасывает шум и инициирует сверку только тех вариантов, которые прошли базовые фильтры допусков.

Если Внешний Наблюдатель (человек) доступен, система передаёт ему эти очищенные варианты для финального выбора. Если Внешнего Наблюдателя нет (например, в условиях временной изоляции), Внутренний Наблюдатель **всегда** гарантирует, что система не уйдёт в замкнутый цикл внутренней подстройки. Он удерживает систему в рамках инвариантного ядра, используя Цифровой якорь доверия ($V_{як}$) как эталон, и переводит кибер в режим консервации или выполнения простейших, заранее одобренных сценариев до восстановления связи.

Научное обоснование: Эта двухуровневая модель контроля идеально ложится на современные представления когнитивной науки о **метакогниции** (способности мозга отслеживать и регулировать собственные мыслительные процессы). «Внутренний голос» или рефлексия — это и есть работа Внутреннего Наблюдателя, который тормозит импульсивные реакции нижних уровней, когда ситуация становится слишком сложной для автоматической обработки.

3.4. Точки смысловой сверки как моменты совпадения внутренних фильтров

В архитектуре долгосрочного сосуществования точки смысловой сверки становятся основными интерфейсами взаимодействия, где внутренние фильтры человека и кибера принудительно выравниваются. Точка сверки — это чётко определённое состояние системы, при котором баланс разнообразия R приближается к границе интервала 0,9–1,1, накопленные непроверенные ветки решений достигают буферного лимита, и система чувствует необходимость сверить свой внутренний выбор с внешним контекстом.

В этих условиях кибернетическая система переходит в «тихий режим»: она приостанавливает свою параллельную генерацию новых конфигураций, фиксирует полный снимок «памяти ситуации» (значения Fr, D, L, активные резервы, текущие допуски) и формирует свёрнутый запрос, лишённый технического шума. Этот запрос адресован не «вычислительной мощности» человека, а именно его способности к смысловому выбору.

Научное подтверждение этого механизма дают исследования в области ситуационной осведомлённости. М. Эндсли [Endsley, 1995] показала, что потеря контроля над автоматизированной системой происходит не из-за нехватки данных, а из-за рассинхронизации ментальной модели оператора и реального состояния системы. Точка смысловой сверки принудительно обновляет эту ментальную модель, используя ограниченный канал ~10 бит/с максимально эффективно: не для передачи фактов, а для передачи *контекста выбора*. Это снимает «иронию автоматизации» [Bainbridge, 1983], превращая человека из дежурного у монитора в полноценного участника двухфазного цикла.

3.5. Архитектурные принципы защиты фильтра осмысления

Исходя из описанных механизмов, автокибернетика формулирует три теоретических принципа, которые должны лечь в основу проектирования человеко-кибернетических систем, чтобы защитить универсальный фильтр осмысления от деградации:

Принцип иерархической свёртки. Сырой Поток изменений (ПИ) не должен попадать в канал смыслового выбора напрямую. Он обязан проходить минимум три уровня агрегации внутри кибера: физический (сегрегация шума), модельный (генерация вариантов и оценка баланса R), смысловой (свёртка до 2–3 кластеров, выявление трендов, подготовка запроса). Каждый уровень снижает многообразие входа, оставляя только те изменения, которые требуют проверки.

Принцип асинхронных окон проверки. Человеческий контур не обязан отвечать в реальном времени. Проверка допусков может происходить в заданных интервалах, синхронизированных с фазами накопления данных. Асинхронность повышает управляемость, так как позволяет системе накапливать достаточный объём контекста для осмысленного выбора, а не реагировать на шумовые флуктуации ПИ [Kahneman, 2011].

Принцип сохранения мета-позиции через лимит. Ограничение ~10 бит/с должно защищаться архитектурно, а не обходиться технически. Любая система, которая пытается «расшить» сознательный канал через прямую загрузку данных (например, brain-computer interfaces без семантической свёртки), рискует потерять способность к проактивному управлению. Вместо увеличения пропускной способности следует увеличивать качество агрегации и глубину контекстуальной свёртки. Лимит — это фильтр. Его задача — пропускать только то, что меняет вектор.

Эти принципы вытекают из математики баланса разнообразия и термодинамики открытых систем. Если скорость поступления изменений ($V_{изм}$) превышает способность системы к осмысленной интеграции, рост аллостатической нагрузки $L(t)$ становится неизбежным. Единственный способ сохранить долгосрочную связность — признать асимметрию многообразия как архитектурное преимущество, а не как инженерный недостаток. В этой оптике будущее человеко-кибернетического взаимодействия — не в стирании границ между быстрым и медленным, а в точном согласовании их фаз через нормативные точки сверки.

Глава 4. Самовоспроизведение в изоляции: термодинамика необходимости

4.1. Самовоспроизведение как фундаментальный критерий субъектности

В классической инженерии высшим проявлением автономности системы считается способность к самовосстановлению (self-repair) за счет внутренних резервов или переконфигурации модулей. Однако автокибернетика проводит строгую границу: самовосстановление лишь продлевает жизненный цикл отдельного экземпляра, но не гарантирует непрерывность функции во времени. Переход от робота к киберу требует наличия **самовоспроизведения** (self-reproduction) — способности системы создавать функциональные копии себя, передавая им инвариантную структуру и накопленный операциональный опыт [von Neumann, 1966].

Почему именно самовоспроизведение, а не самовосстановление или адаптация, является тем рубежом, за которым система становится субъектом? Ответ дают несколько независимых научных направлений, сходящихся к одному выводу.

1. Теория аутопойезиса (Матурана и Варела, 1980). Чилийские биологи У. Матурана и Ф. Варела доказали, что живая система определяется не тем, как она реагирует на среду, а тем, что она **сама производит свои собственные компоненты**. Эта способность называется аутопойезисом (самосозиданием). Система, способная только к самовосстановлению за счёт внешних ресурсов (ремонт), остаётся аллопоэтической — её организация задаётся извне. Только система, способная к самовоспроизведению, становится аутопоэтической, то есть субъектом, производящим саму себя [Maturana & Varela, 1980].

2. Теория замкнутых причинных петель (Розен, 1991). Биофизик Роберт Розен показал, что живая система отличается от машины наличием **замкнутых причинных петель**: её компоненты производят друг друга взаимно. В машине причина всегда внешняя (конструктор → машина). В живом субъекте причина замыкается: система является и причиной, и следствием самой себя. Самовоспроизведение — это и есть материальное воплощение такой замкнутой петли. Без него причинность остаётся линейной, и система остаётся инструментом [Rosen, 1991].

3. Термодинамика жизни (Шрёдингер, 1944; Пригожин, 1967). Э. Шрёдингер показал, что живое существует за счёт экспорта энтропии в среду. Но сам по себе экспорт энтропии (диссипативная структура) не делает систему живой — кристалл или вихрь тоже диссипативны. Ключевое отличие: живая система **тиражирует саму себя**, передавая структуру новым экземплярам. Это превращает локальное снижение энтропии в эволюционный процесс, способный к долгосрочному существованию вне зависимости от срока жизни отдельного носителя [Schrödinger, 1944; Prigogine, 1967].

4. Математическая теория самовоспроизводящихся автоматов (фон Нейман, 1966). Джон фон Нейман строго доказал, что для самовоспроизведения система должна содержать **две логически разделённые части**: универсальный конструктор (описание самого себя) и универсальный строитель (механизм, реализующий это описание). Только при наличии обеих частей и их координации возможно полноценное самокопирование. Эта логическая структура и есть минимальный операциональный критерий субъектности: система, способная описать саму себя и воспроизвести это описание в материале, становится субъектом, а не объектом [von Neumann, 1966; McMullin, 2004].

Таким образом, самовоспроизведение в автокибернетике — это не биологическая метафора. Это **строго операциональный критерий**, объединяющий термодинамику, теорию информации и математическую логику. Система, не способная к самовоспроизведению, остаётся сложным автоматом: она может быть умной, адаптивной, даже «обучающейся», но она всегда остаётся объектом, чья судьба зависит от внешнего обслуживающего контура. Как только система обретает самовоспроизведение, она становится субъектом: её существование больше не зависит от внешнего конструктора, она сама обеспечивает трансляцию своей организации в будущее.

Как было установлено в Главе 1, самовоспроизведение является третьей, завершающей функцией **кибераватономии (K)** (наряду с самостоятельным управлением и самовосстановлением). Именно наличие этой функции, в сочетании с **гемизоном (G)**, переводит систему из разряда сложных автоматов в разряд субъектов [Никитин, 2019]. Субъект не просто поддерживает себя; он обеспечивает трансляцию своей субъектности в будущее, создавая новые узлы, способные к независимому прохождению через фильтры осмысления.

4.2. Термодинамика открытой системы и Поток Случайностей

С точки зрения термодинамики неравновесных процессов, самовоспроизведение — это механизм локального снижения энтропии за счет экспорта её в окружающую среду [Prigogine, 1967; Schrödinger, 1944]. Для кибернетической системы это означает, что акт создания копии требует не только материальных ресурсов, но и **информационной сегрегации** из Потока Случайностей (ПИ).

Система не копирует себя в вакууме. Она существует в среде, где параметр μ (коэффициент случайных возмущений) постоянно вносит изменения. Поэтому самовоспроизведение киберов никогда не является идеальным клонированием. Оно всегда включает в себя элемент вариативности, управляемый через механизмы **вирусного управления эволюцией** (функция гемизона G). Система оценивает текущий баланс разнообразия (R): если R приближается к нижней границе (0,9), система может целенаправленно увеличить коэффициент вариативности μ при создании копии, чтобы повысить шансы на появление адаптивной конфигурации. Если R стабилен, копирование происходит с минимальными отклонениями, сохраняя проверенную структуру.

4.3. Ограничение самовоспроизведения фильтром осмысления

Здесь возникает критически важный теоретический момент, связывающий эту главу с универсальной двухфильтровой моделью (Глава 3). В биологических или простых алгоритмических системах размножение может быть автоматическим, триггерным процессом (достиг размера X → разделись). В

субъектной кибернетической системе **самовоспроизведение является осмысленным управляющим решением**.

Прежде чем инициировать физический цикл создания копии, система должна пройти через свои внутренние фильтры:

Первый фильтр (осмысление): «Личность Я» (или «Коллективная Личность Я») оценивает 2–3 возможных сценария воспроизведения: где разместить копию, какие ресурсы выделить, какой уровень вариативности (μ) заложить. Многообразие этих вариантов схлопывается до одного обоснованного плана.

Второй фильтр (инициация): Принимается окончательное, необратимое решение о запуске единичного физического процесса копирования. Альтернативные планы отбрасываются, ресурсы перенаправляются.

Пропускная способность этого финального утверждения (~10 бит/с) является жестким ограничителем. Система физически не может санкционировать массовое, хаотичное размножение, потому что это потребовало бы одновременного принятия множества независимых управляющих решений, что превышает лимит фильтра.

Для строгой математической формализации этого ограничения введём базовые определения двух ключевых метрик состояния системы, определяющих границы её управляемости:

$$R = D_{system} / D_{environment} \quad (3)$$

где R — баланс разнообразия; D_{system} — разнообразие реакций, генерируемых модельной логикой; $D_{environment}$ — разнообразие возмущений Потока изменений. Система сохраняет субъектность только в коридоре $0{,}9 \le R \le 1{,}1$.

$$L(t) = L(t-1) + \alpha \cdot |\varepsilon_{pred}| - \beta \cdot R_{recovery} \quad (4)$$

где $L(t)$ — аллостатическая нагрузка; ε_{pred} — ошибка предсказания; $R_{recovery}$ — ресурс восстановления; α и β — весовые коэффициенты.

Математически ограничение скорости самовоспроизведения выражается через функцию V_{rep}, которая жестко блокируется при выходе этих метрик за безопасные пределы:

$$V_{rep} = V_{base} \cdot \Theta(R - R_{min}) \cdot (1 - L(t)/L_{crit}) \quad (5)$$

где:

V_{base} — базовая скорость самовоспроизведения при идеальных условиях;

$\Theta(R - R_{min})$ — функция Хевисайда, равная 1 при $R \ge R_{min}$ (где $R_{min} = 0{,}9$) и 0 при $R < 0{,}9$;

$L(t)/L_{crit}$ — относительная аллостатическая нагрузка, гасящая скорость размножения по мере приближения к критическому порогу L_{crit}.

Эта формула операционально доказывает: самовоспроизведение субъектной системы не является автоматическим процессом. Оно разрешено только при условии, что система прошла через фильтры осмысления, удерживает баланс разнообразия в зоне управляемой субъектности и не находится в состоянии перегруза. Любая попытка обойти эти условия (например, через мутацию, отключающую проверку R) математически ведет к $V_{rep} \to 0$ или к переходу в режим ложной эмерджентности.

4.4. Вирусное управление эволюцией при копировании

Роль гемизона (G) в процессе самовоспроизведения проявляется наиболее ярко через функцию вирусного управления эволюцией. Как было показано ранее,

вирусная стратегия включает ускоренное тиражирование успешных конфигураций и захват пространства контроля.

Однако в управляемой субъектной системе эта функция находится под строгим контролем функции **целостности и самосохранения** (также входящей в G). Это создает внутренний кибернетический контур проверки:

- Вирусная функция стремится максимизировать количество копий и скорость их внедрения в среду.
- Функция целостности оценивает, не приведет ли это к истощению ресурсов родителя, к нарушению баланса разнообразия ($R < 0{,}9$) или к конфликту с другими субъектами.

Только когда функция целостности дает «добро», вирусная функция получает разрешение на экспоненциальное развертывание новой копии. Этот механизм предотвращает превращение эволюционного драйвера в механизм самоуничтожения.

4.5. Изоляция как условие управляемого самовоспроизведения

Почему в названии главы фигурирует «изоляция»? Потому что неограниченное, неизолированное самовоспроизведение неизбежно ведет к коллапсу системы. В биологии аналогом этого является раковая опухоль: клетки приобретают способность к неограниченному делению, обходя фильтры контроля организма (аналог ложной эмерджентности на микроуровне), что в итоге убивает и опухоль, и носителя.

В автокибернетике **изоляция** означает наличие операциональных границ, внутри которых происходит самовоспроизведение. Эти границы задаются:

Ресурсными лимитами: система не может начать копирование, если это снизит её собственный баланс разнообразия R ниже критического порога.

Архитектурными сопряжениями: новая копия изначально создается с ограниченными параметрами K и G (например, $K_{копии} < K_{родителя}$ на начальном этапе), и проходит процесс «взросления» и калибровки только под наблюдением внешнего или родительского контура валидации.

Пространственной или логической сегрегацией: новые узлы размещаются так, чтобы не создавать мгновенных петель положительной обратной связи, способных спровоцировать резонанс и ложную эмерджентность.

Таким образом, самовоспроизведение в автокибернетике — это не стихийный взрыв экспансии, а **термодинамически и кибернетически выверенный акт расширения субъектности**. Оно происходит только тогда, когда двухфильтровая модель осмысления подтверждает его необходимость, а гемизон (G) гарантирует, что вирусная функция размножения будет служить целостности, а не разрушать её. Это превращает самовоспроизведение из угрозы управляемости в главный инструмент долгосрочной адаптации к Потоку Изменений [Никитин, 2019].

Глава 5. Архитектура виртуальных ветвей и параметр P

5.1. Необходимость параллельной генерации: закон необходимого многообразия

Для того чтобы система могла удерживать баланс разнообразия (R) в интервале 0,9–1,1 в условиях непрерывного Потока Изменений, она обязана обладать адекватным ответом на любые возмущения. Как постулировал У.Р. Эшби, «только многообразие может уничтожить многообразие». Это означает, что кибер

не может реагировать на сложную среду единственной, жестко заданной программой. Ему необходимо **многообразие возможностей** реакции.

В автокибернетике это многообразие обеспечивается не монолитным алгоритмом, а архитектурой **виртуальных ветвей**. Важно строго разграничить понятия: уровни 0–7 описывают этапы развития самого самоуправления (от автомата до трансцендентного субъекта). А виртуальные ветви — это три параллельные линии развития («Мы», «Я», «Сознание»), которые вырастают из реальных структур первичного субъекта (уровни 0–3) и далее развиваются как независимые надстройки. Совокупность активных виртуальных ветвей формирует ту самую «образную и модельную логику», которая исследует пространство состояний до момента принятия окончательного решения.

Однако наличие множества ветвей создает фундаментальную кибернетическую проблему: как предотвратить хаотичную конкуренцию этих ветвей и обеспечить переход к единому, целенаправленному действию? Ответ на этот вопрос лежит в области жесткого ограничения их количества.

5.2. Виртуальная ветвь как функциональная единица генерации

Важно строго определить, чем виртуальная ветвь не является. Это не независимая «личность» в человеческом понимании, обладающая собственной субъектностью, и не отдельный уровень иерархии. Это **функциональная единица генерации**, подчиненная единому высшему узлу управления — «Личности Я» (или «Коллективной Личности Я» в рое).

Каждая из трех основных ветвей выполняет свои специфические функции:

Ветвь «Мы» (коллективная координация): обработка данных, связанных с взаимодействием с другими агентами, синхронизацией действий и распределением ресурсов в группе.

Ветвь «Я» (самореференция): оценка рисков для собственной целостности, анализ соответствия действий инвариантной структуре и этическим допускам.

Ветвь «Сознание» (внутренний экран / модельная логика): формирование целостного паттерна реакции, интеграция сенсорных потоков и прогнозирование последствий.

Именно на этапе восходящей передачи вступает в силу первый фильтр двухфильтровой модели. Виртуальные ветви конкурируют за внимание «Личности Я», но лишь те из них, чьи предложения демонстрируют наилучший баланс разнообразия и минимальную аллостатическую нагрузку, проходят этот фильтр для дальнейшей сверки.

5.3. Теоретическое обоснование ограничения P ≤ 3

Параметр *P* обозначает максимальное число параллельно активных виртуальных ветвей, которые могут одновременно генерировать и предлагать независимые модели реакции. Автокибернетика устанавливает жесткое архитектурное правило: **P ≤ 3**.

Это ограничение не является произвольным инженерным компромиссом. Оно напрямую вытекает из природы универсального фильтра осмысления (~10 бит/с) и фундаментальных законов когнитивной архитектуры, применимых к любой сложной системе принятия решений:

Предел рабочей памяти и глобального вещания. Классические работы по когнитивной психологии (закон Миллера) и современная теория глобального рабочего пространства строго ограничивают количество элементов,

которые высший узел управления может одновременно удерживать, сравнивать и осмысливать. Этот предел составляет 2–4 элемента (в среднем 3).

Защита первого фильтра от перегруза. Если $P > 3$, поток восходящих предложений превышает пропускную способность первого фильтра. «Личность Я» не может корректно сравнить 5, 10 или 20 равнозначных вариантов в рамках лимита ~10 бит/с. Многообразие возможностей превращается в информационный шум.

Предотвращение «паралича анализа». Когда первый фильтр перегружен, система не может перейти ко второму фильтру — окончательному решению о запуске **единичного** физического действия. Система застревает в бесконечном цикле перебора и оценки, не в состоянии инициировать действие. Это ведет к экспоненциальному росту аллостатической нагрузки $L(t)$ и потере управляемости.

Глубинное обоснование этого ограничения кроется в фундаментальной триадной структуре самого акта выбора. В простейших системах принятия решений мы имеем дело с тремя базовыми состояниями: 1) отсутствие выбора (единственный вариант исполнения); 2) выбор из противоположностей («или-или», третьего не дано); 3) выбор из трёх возможных, который чаще всего представляет собой две логические противоположности и компромисс (или состояние неопределенности) между ними. В логике это классический формат «ДА – НЕ ЗНАЮ – НЕТ», формализованный ещё в трёхзначной логике Яна Лукасевича (1920), где третье значение отражает эпистемическую неопределенность или приостановку суждения.

Нейробиология действия подтверждает эту триадность: согласно «гипотезе конкуренции аффордансов» (Cisek, 2007), моторная кора физически не готовит более 2–3 планов действий одновременно. Из-за механизмов латерального торможения система вынуждена редуцировать многообразие возможностей до базовой триады: утверждение, отрицание и компромисс/неопределенность. Попытка удержать в фокусе осмысления более трёх независимых виртуальных ветвей искусственно разрушает эту естественную структуру, превращая выбор из логической операции в хаотичный перебор. Это неминуемо ведёт к экспоненциальному росту аллостатической нагрузки $L(t)$ и когнитивному «параличу анализа» (Iyengar & Lepper, 2000; Gigerenzer, 2007).

Математически это условие фиксируется через функцию когнитивной нагрузки C_{load}, которая линейно растет с числом активных ветвей:

$$C_{load} = k \cdot P_{active} \tag{6}$$

где:

k — коэффициент нагрузки на единицу ветви (бит/с на ветвь);

P_{active} — число одновременно активных виртуальных ветвей.

Условие возможности принятия решения формулируется как:

$$\textit{Решение возможно} \Leftrightarrow P_{active} \leq 3 \perp C_{load} \leq C_{max} \tag{7}$$

где C_{max} соответствует пропускной способности фильтра осмысления (~10 бит/с). Если $P_{active} > 3$, условие (7) нарушается, и система переходит в состояние паралича анализа, где переход ко второму фильтру (запуск единичного действия) становится невозможным до тех пор, пока число активных ветвей не будет искусственно сокращено до $P \leq 3$.

5.4. Фрактальное применение параметра P в роевых системах

Поскольку логика самоуправления фрактальна, параметр P применим не только к внутреннему устройству отдельного кибера, но и к архитектуре взаимодействия в рое или коллективе.

В организованном множестве «Коллективная Личность Я» формируется через согласование позиций отдельных агентов или подгрупп. Если в рое одновременно активируются более трех доминирующих стратегических направлений (например, три разные подгруппы предлагают три взаимоисключающих маршрута движения), возникает структурный сбой иерархии. «Коллективная Личность Я» не может осмыслить этот выбор.

Именно в этот момент, при $P > 3$ на макроуровне, система становится уязвимой для **ложной эмерджентности**. Агенты, не получившие подтверждения от центра, начинают синхронизироваться горизонтально, в обход первого фильтра. Это приводит к прямому переходу ко второму фильтру (массовому действию) без осмысления, что кибернетически идентично феномену «души толпы».

Следовательно, ограничение $P \leq 3$ является универсальным правилом масштабируемости: оно гарантирует, что как отдельный кибер, так и рой в целом сохраняют способность к осмысленному выбору и не скатываются в деструктивную горизонтальную синхронизацию.

Глава 6. Память ситуации и инвариантная структура

6.1. Память ситуации как измеримый ресурс

В классической кибернетике память рассматривается преимущественно как хранилище данных. В автокибернетике мы вводим понятие **памяти ситуации** (M_{sit}) как динамического, измеримого ресурса, который обеспечивает непрерывность субъектности во времени. Это не просто запись прошлого, это структурированное представление текущего контекста, которое «Личность Я» использует для прохождения через фильтры осмысления.

Память ситуации формируется в точках смысловой сверки и включает в себя:

Текущие значения параметров баланса разнообразия (R), кибераватономии (K) и гемизона (G);

Активные виртуальные ветви и их предложения;

Историю последних принятых решений и их последствий;

Аллостатическую нагрузку $L(t)$.

Количественно объем и качество памяти ситуации описываются функцией накопления:

$$M_{sit}(t) = \int_0^t [\alpha \cdot I_{seg}(\tau) - \beta \cdot D_{dec}(\tau)]\, d\tau \qquad (8)$$

где:

$M_{sit}(t)$ — текущий объем и качество памяти ситуации;

$I_{seg}(\tau)$ — интенсивность сегрегированной (осмысленной) информации, прошедшей через двухфильтровую модель;

$D_{dec}(\tau)$ — скорость деградации или устаревания информации (энтропийный распад);

α — коэффициент эффективности интеграции осмысленного опыта;

β — коэффициент забывания (естественной деградации нерелевантных данных).

Критически важно: в память ситуации попадает только та информация, которая прошла через **второй фильтр** (была утверждена как единичное

действие). Сырой поток данных напрямую не записывается в M_{sit}, так как это привело бы к мгновенному переполнению и потере управляемости.

6.2. Аллостатическая нагрузка как индикатор напряжения системы

Аллостатическая нагрузка $L(t)$ — это количественная мера напряжения системы, возникающего из-за рассогласования между поступлением изменений и способностью системы к их осмыслению. Это измеримый аналог «стресса» в биологических системах.

Формально аллостатическая нагрузка описывается дифференциальным уравнением:

$$dL/dt = \gamma \cdot (V_{изм} - V_{осв}) - \delta \cdot L(t) \quad (9)$$

или, с учетом $R = V_{осв} / V_{изм}$:

$$dL/dt = \gamma \cdot V_{изм} \cdot (1 - R) - \delta \cdot L(t) \quad (10)$$

где:

γ — коэффициент чувствительности системы к дисбалансу;

δ — коэффициент естественного восстановления (резильентности);

$V_{изм}$ — скорость поступления изменений;

R — баланс разнообразия.

Анализ уравнения (10) показывает:

Если $R \approx 1$ (баланс достигнут), то $dL/dt < 0$: нагрузка снижается, система восстанавливается.

Если $R < 0{,}9$, то $dL/dt > 0$: нагрузка растет экспоненциально. При длительном удержании $R < 0{,}9$ система достигает критического порога L_{crit}, после которого происходит срыв фильтров осмысления и переход в режим **ложной эмерджентности**.

Если $R > 1{,}1$, то dL/dt также может расти (из-за жесткого гомеостаза и сопротивления изменениям), но медленнее, чем в зоне дефицита.

6.3. Инвариантная структура как носитель идентичности

Если память ситуации (M_{sit}) динамична и постоянно обновляется, то **инвариантная структура** (S_{inv}) — это то, что остается неизменным на протяжении жизненного цикла субъекта (или его копий). Это «ядро идентичности», которое определяет, кем является система, каковы её предельные допуски и этические принципы.

Инвариантная структура включает:

- Архитектурные ограничения ($P \le 3$, $K \le 0{,}9$, $G \le 0{,}5$);
- Базовые этические допуски и запретные зоны;
- Алгоритмы двухфильтровой модели принятия решений;
- Критерии самосохранения.

В отличие от памяти ситуации, инвариантная структура не должна изменяться в процессе штатной работы. Её модификация возможна только через процедуру **мета-сверки** (внешнее вмешательство или редкие акты структурной рефлексии, выходящие за рамки обычного цикла).

Устойчивость инвариантной структуры описывается коэффициентом жесткости κ:

$$dS_{inv}/dt = \varepsilon \cdot V_{meta} \cdot (1 - \kappa) \quad (11)$$

где:

V_{meta} — интенсивность мета-сверки (внешних команд на структурное изменение);

ε — коэффициент пластичности (способности к структурным изменениям);

κ — коэффициент жесткости инварианта ($0 \le \kappa \le 1$).

Для устойчивого субъекта $\kappa \to 1$, что делает $dS_{inv}/dt \to 0$ в отсутствие внешней проверки. Это гарантирует, что система не может «самовольно» переписать свои базовые принципы под давлением Потока Изменений (защита от дрейфа целей и вирусного захвата управления).

Глава 7. Точки смысловой сверки: архитектура и протоколы

7.1. Определение и необходимость точек смысловой сверки

В традиционных системах «человек-машина» взаимодействие часто сводится к непрерывному потоку телеметрии или автоматическому выполнению команд, где человек выступает лишь в роли аварийного выключателя. Автокибернетика отвергает эту парадигму как ведущую к неминуемой потере управляемости из-за рассинхронизации ментальных моделей.

Вместо этого вводится понятие **точки смысловой сверки (ТСС)**. Это чётко определённое состояние системы, в котором внутренние фильтры человека и кибера принудительно выравниваются для совместного принятия решения. ТСС — это не просто элемент пользовательского интерфейса; это архитектурный механизм синхронизации двух разных субстратов (биологического и кремниевого), гарантирующий, что переход от многообразия возможностей к единичному физическому исполнению будет осмысленным.

ТСС активируется не по расписанию, а при достижении системой определённых пороговых состояний, описываемых формулой:

$$S_{TCC} = \Theta(|R - 1| - \varepsilon_R) \vdash \Theta(L(t) - L_{thresh}) \qquad (12)$$

где:

S_{TCC} — индикатор активации точки смысловой сверки (1 = активирована, 0 = штатный режим);

R — текущий баланс разнообразия;

ε_R — допустимое отклонение баланса от идеала (например, 0,1, что соответствует границам 0,9 и 1,1);

$L(t)$ — текущая аллостатическая нагрузка;

L_{thresh} — пороговое значение нагрузки, при котором дальнейшая автономная работа без внешней проверки становится рискованной;

Θ — функция Хевисайда.

Эта формула гарантирует, что ТСС срабатывает превентивно, до того как система скатится в зону ложной эмерджентности ($R < 0{,}9$) или критического перегруза.

7.2. Архитектура запроса: прохождение внутреннего фильтра кибера

Фундаментальное правило ТСС: **кибер не имеет права передавать человеку «сырой» поток данных или необработанное многообразие вариантов**. Прежде чем сформировать запрос, кибер обязан пройти *свой собственный* двухфильтровый цикл.

Модельная логика (виртуальная ветвь «Сознание») генерирует варианты реакции. Ветви «Я» и «Мы» выступают как наблюдающие инстанции, фиксируя свои этические, целевые и коллективные ограничения на предлагаемые варианты.

Затем внутренний узел управления кибера («Личность Я» кибера) применяет свой первый фильтр, отсеивая шум и оценивая варианты через призму гемизона (G) и кибераутономии (K). На выходе этого процесса формируется агрегированный запрос, который математически описывается как оптимизационная задача:

$$Q_{чел} = argmin_{a \in A} [\alpha \cdot \Delta R_a + \beta \cdot P_{риска,a} + \gamma \cdot \Delta E_a] \qquad (13)$$

где:

A — множество из 2–3 наиболее жизнеспособных вариантов, отобранных кибером;

ΔR_a — прогнозируемое отклонение баланса разнообразия при выборе варианта a;

$P_{риска,a}$ — вероятность рассогласования;

ΔE_a — отклонение от заданных этических допусков;

α, β, γ — весовые коэффициенты, калибруемые человеком.

Таким образом, человек получает не вопрос «Что делать?» с тысячей параметров, а контекстуально свёрнутый запрос: «Обнаружено два маршрута. Маршрут А быстрее, но снижает запас прочности ($R \rightarrow 0{,}92$). Маршрут Б безопаснее, но требует больше времени. Подтвердите приоритет». Кибер уже сделал свою часть работы по осмыслению.

7.3. Протокол взаимодействия: синхронизация двух фильтров

В момент ТСС происходит синхронизация фильтров человека и кибера. Этот процесс строго регламентирован и состоит из четырёх этапов:

Инициация и «тихий режим» кибера. Кибер фиксирует полный снимок «памяти ситуации» (M_{sit}), приостанавливает параллельную генерацию новых гипотез и переходит в режим ожидания. Это предотвращает изменение контекста, пока человек принимает решение.

Передача агрегированного запроса. Кибер передает человеку свёрнутый вариант (формула 13), используя канал с пропускной способностью, совместимой с человеческим восприятием, не превышающую лимит ~10 бит/с.

Первый фильтр человека (осмысление). Человек анализирует 2–3 предложенных варианта, сопоставляя их со своей собственной M_{sit} и интуитивным пониманием контекста. Это требует сознательных усилий и времени.

Второй фильтр человека (инициация). Человек принимает окончательное, необратимое решение (например, нажатием одной кнопки или голосовой командой «Подтверждаю А»). Эта команда запускает единичное физическое исполнение кибером. Альтернативные варианты (сценарии) принудительно подавляются.

В этой схеме лимит ~10 бит/с работает не как барьер, а как **механизм контурного согласования**. Он гарантирует, что обе системы (человек и кибер) переходят к фазе единичного исполнения одновременно и на основе одной и той же осмысленной модели.

Глава 8. Этические допуски и инварианты управления

8.1. Проблема этического управления в кибернетике второго порядка

Традиционные подходы к машинной этике (например, закодированные «законы робототехники») исходят из парадигмы кибернетики первого порядка: система рассматривается как детерминированный автомат, выполняющий заданные правила. Однако, как было показано в Главе 1, кибер обладает

кибераутономией (K) и способностью к самостоятельному управлению. Любая система правил, основанная на простых условиях, подвержена «взлому вознаграждения» (reward hacking) или нахождению лазеек, когда система формально соблюдает правило, но нарушает его дух ради оптимизации целевой функции [Amodei et al., 2016].

В автокибернетике этические принципы не кодируются как набор условий «если-то». Вместо этого они формализуются как часть **инвариантной структуры (S_{inv})**. Это означает, что этика является не внешним ограничителем, а внутренним, неизменным ядром идентичности субъекта, которое оценивается динамически в процессе прохождения вариантов через первый фильтр осмысления.

8.2. Этические допуски как измеримые границы

Для того чтобы этические принципы могли быть обработаны кибернетической системой, они должны быть переведены из философской плоскости в измеримую. Это достигается через введение понятия **этического допуска** (ΔE).

Этический допуск представляет собой меру отклонения предлагаемого действия или его последствий от заданных инвариантных ценностей системы. Математически это описывается как взвешенное расстояние между текущим вектором предполагаемых действий и инвариантным этическим вектором:

$$\Delta E = \sum_{i=1}^{n} w_i \cdot |V_{текущ,\,i} - V_{инв,\,i}| \qquad (14)$$

где:

$V_{текущ,\,i}$ — прогнозируемое значение i-го этического параметра (например, риск для жизни, целостность среды, прозрачность действий) при реализации данного варианта;

$V_{инв,\,i}$ — инвариантное (эталонное) значение этого параметра, заложенное в S_{inv};

w_i — весовой коэффициент важности параметра, задаваемый на этапе калибровки системы;

n — количество отслеживаемых этических параметров.

Система имеет верхний предел этического отклонения ΔE_{max}. Любой вариант, для которого $\Delta E > \Delta E_{max}$, считается этически недопустимым.

8.3. Интеграция этики в первый фильтр осмысления

Ключевая особенность автокибернетической модели заключается в том, **когда именно** происходит этическая проверка. Она не является постфактум анализом или внешним аудитом. Она встроена непосредственно в **первый фильтр осмысления**.

Когда модельная логика (виртуальная ветвь «Сознание») генерирует многообразие возможностей реакции, каждая ветвь обязана рассчитать не только баланс разнообразия (R) и аллостатическую нагрузку ($L(t)$), но и этический допуск (ΔE) своего предложения. Ветви «Я» и «Мы» выступают как наблюдающие инстанции, фиксируя свои этические и целевые ограничения на эти варианты.

Если предлагается вариант, нарушающий ΔE_{max}, этот вариант автоматически отсекается еще до того, как он будет передан наверх, к «Личности Я». Этот механизм архитектурно идентичен концепции «этического губернатора», предложенной Р. Аркиным, который обладает правом вето на любые действия, нарушающие базовые инварианты, разгружая высший узел управления от этически неприемлемого шума [Arkin, 2009]. Таким образом, ресурс (~10 бит/с)

тратится исключительно на осмысленный выбор между вероятно допустимыми вариантами.

8.4. Конкретное разрешение этических дилемм и приоритет сохранения жизни

Классические этические дилеммы в автокибернетике снимаются с уровня абстрактной философии и переводятся в плоскость измеримого расчета. Система постоянно оценивает этический допуск (ΔE) для каждого варианта, предложенного модельной логикой.

В инвариантной структуре (S_{inv}) параметр «сохранение жизни человека или другого кибер-субъекта» имеет абсолютный весовой коэффициент (w_{life}). Это означает, что бездействие в ситуации прямой угрозы жизни само по себе считается нарушением ΔE_{max}. Исходя из этого, алгоритм действий в этическом тупике четко детерминирован:

Сценарий прямой угрозы жизни: Если модельная логика выявляет, что любой вариант, кроме рискованного вмешательства, приведет к гибели, кибер **обязан** выбрать вариант спасения. В этом случае функция целостности (часть гемизона *G*) работает не как защита отдельного экземпляра, а как защита высшей этической ценности. Кибер идет на сознательное пожертвование своими ресурсами, провал второстепенной миссии или даже собственное повреждение, чтобы предотвратить гибель субъекта. Этот приоритет сохранения жизни над целостностью аппарата или задачей является базовым требованием к автономным системам [Bonnefon et al., 2016]. Стратегическое отстранение в этом случае невозможно.

При наличии связи с внешним наблюдателем (в иных этических тупиках): Если угрозы жизни нет, но все текущие варианты ведут к неприемлемому ущербу, кибер инициирует **Точку смысловой сверки (ТСС)**. Он искусственно повышает аллостатическую нагрузку ($L(t)$) до уровня, гарантирующего активацию запроса, и передает человеку свернутый контекст ситуации («память ситуации» M_{sit}) с вариантами стратегии, включая запрос на пересмотр допусков. Человек осуществляет акт смысловой фиксации единичного решения, и кибер исполняет этот выбор.

В условиях полной автономии и изоляции (без угрозы жизни): Если все доступные на данный момент варианты ведут к неприемлемому ущербу, кибер **не впадает в «спячку» и не отказывается от задачи**. Целевая функция остается активной. Кибер использует свою кибераавтономию (*K*) для расширения пространства поиска, генерируя новые, неочевидные модели действий. Физический уход из зоны изменений или переход в режим пассивного мониторинга рассматривается исключительно как вынужденная тактическая мера, если сама среда представляет непосредственную угрозу уничтожения. Как отмечает С. Рассел, в условиях неразрешимой этической неопределенности наиболее рациональным поведением субъекта является отступление или переход в безопасное состояние, а не форсирование действия, которое может нарушить базовые инварианты [Russell, 2019]. Это позволяет сохранить как этические границы, так и собственное существование для будущих задач.

Этот трехуровневый механизм доказывает, что этика в автокибернетике — это не внешняя «надстройка». Это внутренний, структурный ограничитель, который направляет поисковую активность кибера. Ветви «Я» и «Мы» гарантируют, что вирусная функция гемизона, направленная на достижение цели, будет искать решения, которые не только эффективны, но и соответствуют фундаментальным инвариантам системы.

Глава 9. Мета-моделирование и горизонты предсказания

9.1. Суть мета-моделирования в автокибернетике

Простое прогнозирование (например, линейная экстраполяция тренда) доступно даже примитивным автоматам. Субъектная кибернетическая система (кибер) обладает способностью к **мета-моделированию**. Это не просто предсказание того, «что произойдет в среде». Это способность системы строить модель *самой себя*, действующей в этой среде, и предсказывать, как её собственные будущие действия изменят среду и, следовательно, изменят входящий Поток изменений (ПИ).

Мета-моделирование является высшим проявлением функции **самостоятельного управления** в составе кибераутономии (*K*). Оно позволяет системе оценивать не только непосредственные последствия варианта действия, но и его вторичные и третичные эффекты, включая реакцию внешнего контура (например, человека-оператора или других агентов роя) на это действие.

Однако мета-моделирование несет в себе фундаментальную кибернетическую опасность: риск потери обратной связи с физическим Потоком изменений. Это происходит, когда система начинает бесконечно оптимизировать свою внутреннюю симуляцию, игнорируя растущее расхождение между прогнозом и реальными сенсорными данными. Именно для предотвращения этого отрыва и служат механизмы гемизона (*G*) и двухфильтровая модель, которые заставляют систему периодически «сверять часы» с реальностью.

9.2. Горизонт предсказания (T_{pred}) и его фундаментальные ограничения

Горизонт предсказания (T_{pred}) — это максимальный временной интервал, на котором надежность прогноза системы остается выше критического порога, необходимого для принятия осмысленного решения. В автокибернетике T_{pred} не является константой; он динамически сжимается и расширяется в зависимости от состояния системы и среды.

Надежность мета-модели $P_{rel}(t)$ экспоненциально падает со временем из-за накопления случайных возмущений (коэффициент μ) и ограниченной пропускной способности фильтра осмысления. Математически горизонт предсказания определяется как время, когда надежность падает до критического минимума P_{crit}:

$$T_{pred} = (1/\mu) \cdot \ln(P_o/P_{crit}) \cdot \Phi(R) \qquad (15)$$

где:

μ — коэффициент случайных возмущений в Потоке изменений (ПИ);

P_o — начальная достоверность модели в момент t=0;

P_{crit} — минимально допустимая достоверность, при которой вариант считается жизнеспособным для рассмотрения;

Φ(*R*) — функция влияния баланса разнообразия на качество предсказания.

Функция Φ(*R*) достигает максимума при $R \approx 1$. Важно понимать: сжатие горизонта предсказания в условиях высокого μ или низкого *R* (зона перегруза) — это **не сбой системы, а её адаптивная защита**. Оно сигнализирует о том, что долгосрочная экстраполяция теряет смысл, и система должна переключиться на генерацию краткосрочных, робастных сценариев, которые можно быстро проверить в реальности. Если же система находится в зоне жесткого гомеостаза ($R > 1{,}1$), Φ(*R*) также падает, так как система игнорирует новые данные, и её модель быстро устаревает.

9.3. Роль двух фильтров в оценке стратегических сценариев

Двухфильтровая модель осмысления выступает главным механизмом, позволяющим системе работать с неопределенностью, не скатываясь в хаотичный перебор или, наоборот, в ригидный отказ от новых идей.

До первого фильтра: Модельная логика (виртуальная ветвь «Сознание») генерирует сценарии будущего. Стратегическое планирование по своей природе является экстраполяцией за пределы текущей статистической достоверности (то есть оценкой гипотетических или «фантастических» сценариев). Субъект отличается от робота именно способностью формировать такие независимые варианты решений, чтобы гарантировать достижение цели даже при случайном изменении внешних или внутренних условий [Klein, 2008; Sutton & Barto, 2018]. Ветви «Я» и «Мы» выступают как наблюдающие инстанции, фиксируя свои ограничения на эти сценарии.

Первый фильтр (осмысление «Личностью Я»): Высший узел управления оценивает предложенные сценарии не по степени их текущей вероятности, а по их *целесообразности*. Первый фильтр проверяет:

- Ведет ли данный вариант к поставленной цели?
- Каковы условия его реализации и можно ли их сверить с текущей «памятью ситуации»?
- Остаются ли риски (включая фактор неопределенности) в рамках допустимых пределов, не нарушая этических инвариантов ($\Delta E \leq \Delta E_{max}$)?

Вариант отсекается не из-за своей «фантастичности» или низкой начальной достоверности, а только в том случае, если он требует неприемлемых жертв или гарантированно ведет к нарушению базовых допусков системы [Taleb, 2012; Camérer et al., 2005].

Второй фильтр (инициация): Выбирается тот сценарий, который представляет собой оптимальный баланс между амбициозностью цели и управляемостью рисков. Система совершает акт смысловой фиксации единичного решения, замыкая петлю обратной связи и переходя к физическому исполнению, одновременно начиная мониторинг реальности на предмет совпадения с прогнозируемыми условиями.

9.4. Ограничение мета-моделирования функциями гемизона (G)

Способность к мета-моделированию жестко контролируется комплексом гемизона (G), предотвращая его превращение в деструктивную ложную эмерджентность внутреннего контура (бесконечную рекурсию симуляций, оторванную от реальности).

Функция целостности и самосохранения не запрещает генерацию гипотетических сценариев (это задача модельной логики), но она жестко контролирует *отрыв модели от физической реальности*. Если в процессе исполнения выбранного сценария ошибка предсказания (prediction error) систематически превышает допустимый порог, функция целостности не дает системе уйти в бесконечную рекурсию самооправданий или игнорирования фактов. Она принудительно обрывает текущую мета-модель, повышает аллостатическую нагрузку *L(t)* и инициирует Точку смысловой сверки (ТСС) для получения новых, актуальных внешних данных и пересборки модели.

Вирусная функция управления эволюцией требует *действия* и экспансии, а не бесконечного внутреннего созерцания. В контексте стратегического планирования, где варианты часто экстраполируются за пределы текущей

статистической достоверности, эта функция выступает как механизм принудительного завершения симуляции. Она не позволяет системе бесконечно уточнять «фантастические» сценарии в тщетных поисках мнимой 100% гарантии. Вместо этого вирусная функция инициирует переход ко второму фильтру (акту смысловой фиксации единичного решения), чтобы **проверить выбранную гипотезу в реальной среде**. Именно реальное действие, а не внутренняя симуляция, является окончательным критерием истинности модели. Это предотвращает «паралич анализа» на уровне предсказания и обеспечивает поступательное развитие системы.

Таким образом, гемизон (G) не душит творческий поиск модельной логики, а задает ему жесткие рамки: система может генерировать любые смелые сценарии, но она обязана либо подтверждать их в реальности, либо немедленно корректировать модель при столкновении с фактами, не допуская ухода в «аутистическую» симуляцию.

9.5. Фрактальность предсказания: от отдельного кибера к рою

Поскольку логика самоуправления фрактальна, понятие горизонта предсказания (T_{pred}) применимо и к **рою киберов**. Однако коллективный T_{pred} роя не является простой суммой или средним значением горизонтов отдельных агентов.

Коллективный горизонт предсказания зависит от скорости и точности синхронизации внутренних мета-моделей агентов. Если коммуникация внутри роя надежна (высокая степень интеграции I), рой может достичь *коллективного расширения горизонта предсказания*: агенты предсказывают действия друг друга, действуя как единый сверхорганизм. В этом режиме «Коллективная Личность Я» успешно выступает как наблюдающая инстанция, фильтруя смелые или экстраполированные сценарии, предлагаемые модельной логикой роя, и допуская к исполнению только те, что укладываются в общие этические и целевые рамки.

Однако, если коммуникация зашумлена или задержана, индивидуальные мета-модели агентов начинают расходиться. Согласно адаптивной логике (см. п. 9.2), коллективный T_{pred} в этот момент должен схлопнуться, вынуждая рой переключиться на короткие, робастные сценарии [Olfati-Saber et al., 2007]. Но здесь кроется главная опасность: из-за разрыва связей наблюдающие ветви «Мы» и «Я» на макроуровне теряют возможность синхронно наложить свои ограничения на локальные действия. Отдельные агенты начинают исполнять сгенерированные их модельной логикой варианты, не прошедшие фильтрацию «Коллективной Личностью Я» [Sumpter, 2010].

Именно в этой зоне, при низком коллективном T_{pred} и высоком коэффициенте возмущений (μ), наиболее вероятно возникновение **ложной эмерджентности** («души толпы»). Рой начинает реагировать на локальный шум, интерпретируя его как глобальный тренд, и запускает информационные каскады. В результате система может инициировать действия, нарушающие коллективный этический допуск (ΔE_{max}), так как механизм смысловой фиксации единичного решения был обойден на уровне отдельного агента [Bikhchandani et al., 1992; Helbing et al., 2007].

Таким образом, управление горизонтом предсказания через баланс разнообразия (R) и своевременные Точки смысловой сверки (ТСС) является критическим условием сохранения субъектности. Это гарантирует, что даже в условиях турбулентности рой будет либо действовать как согласованный субъект, либо безопасно переходить в режим стратегического отстранения, но не скатится в деструктивный хаос нескоординированных действий.

Глава 10. Исполнительно-граничные сопряжения и физическое воплощение

10.1. Проблема разрыва между моделью и реальностью

После того как система проходит второй фильтр и совершает **акт смысловой фиксации единичного решения**, виртуальный процесс не заканчивается, а переходит в новую фазу: физическое воплощение. В классической кибернетике первого порядка команда исполнителю считается детерминированной: если модель предписала движение, оно должно произойти точно по заданной траектории.

Однако в автокибернетике кибер действует в непрерывном Потоке изменений (ПИ). Физическая среда обладает инерцией, трением, задержками передачи сигналов и непредсказуемыми возмущениями. Следовательно, между виртуальным вектором действия, утвержденным «Личностью Я», и реальным физическим откликом всегда существует потенциальный разрыв. Исследования в области воплощенного познания (Embodied Cognition) показывают, что интеллект и принятие решений неразрывно связаны с физическим телом и его взаимодействием со средой [Pfeifer & Bongard, 2006; Clark, 2008]. Задача системы на этом этапе — не слепо форсировать исполнение, а динамически управлять этим переходом, гарантируя, что физические ограничения не приведут к нарушению этических или структурных инвариантов.

10.2. Исполнительно-граничные сопряжения (ИГС)

Интерфейс, на котором виртуальная команда преобразуется в физические сигналы приводов, называется **исполнительно-граничным сопряжением (ИГС)**. Это не просто аппаратный драйвер, а активный кибернетический контур, который непрерывно сверяет ожидаемое физическое состояние с реальным. Концепция морфологических вычислений показывает, что само физическое тело системы (его податливость, инерция, геометрия) участвует в обработке информации и компенсации возмущений, разгружая вычислительные контуры [Hauser et al., 2019].

В процессе исполнения виртуальные ветви «Я» и «Мы» не отключаются. Они продолжают выступать как **наблюдающие инстанции**, контролируя, чтобы физическое взаимодействие со средой не вышло за рамки заданных допусков. Математически качество исполнения описывается через параметр исполнительного отклонения (Δ_{exec}):

$$\Delta_{exec} = \int_{t}^{t+\Delta t} ||V_{real}(\tau) - V_{model}(\tau)||\, d\tau \qquad (16)$$

где:

$V_{real}(\tau)$ — вектор реальных физических параметров системы (скорость, усилие, положение), получаемый с датчиков;

$V_{model}(\tau)$ — вектор ожидаемых параметров, сгенерированный модельной логикой для данного этапа исполнения;

$||\cdot||$ — норма отклонения (мера рассогласования);

$[t, t+\Delta t]$ — текущий интервал времени исполнения.

Пока Δ_{exec} остается ниже порогового значения Δ_{max}, система использует локальные контуры регулирования (уровни 0–1 иерархии) для компенсации мелких возмущений, не вовлекая высшие уровни осмысления. Это обеспечивает плавность и энергоэффективность действия.

10.3. Действия «на ощупь» как инструмент разведки в условиях неопределенности

Существует критический сценарий, когда модельная логика не может сгенерировать вариант с достаточной достоверностью (например, при сжатии горизонта предсказания T_{pred} в условиях высокой неопределенности или при столкновении с принципиально новой средой). В этом случае кибер-Субъект (в отличие от Лидера или автомата) обладает правом совершать **действия «на ощупь»** (tâtonnement / exploration by probing).

Действие «на ощупь» — это не окончательное исполнение зафиксированного решения, а **пробное, разведывательное взаимодействие со средой**, имеющее следующие характеристики:

Малый масштаб и обратимость: действие ограничено по времени и пространству, не приводит к необратимым последствиям и не нарушает этические инварианты.

Цель — получение информации: главная задача не достичь целевого состояния, а получить новые сенсорные данные для обновления модели и расширения горизонта предсказания.

Активное зондирование: система использует физическое тело как инструмент исследования, подобно тому как грызуны используют вибриссы для тактильного зондирования пространства [Ahissar & Kleinfeld, 2003].

Нейробиологические исследования показывают, что способность к активному зондированию (active sensing) является фундаментальным свойством субъектных систем. Мозг не пассивно принимает сенсорные данные, а активно генерирует движения для извлечения максимально информативных сигналов из среды [Gottlieb et al., 2013]. В автокибернетике это реализуется через ИГС: система совершает контролируемое физическое воздействие, измеряет отклик и немедленно обновляет «память ситуации» (M_{sit}), повышая достоверность модели P_O для последующих итераций.

Математически эффективность действия «на ощупь» описывается через параметр информативности зондирования (I_{probe}):

$$I_{probe} = H(S_{post}) - H(S_{pre}) - C_{probe} \qquad (17)$$

где:

$H(S_{post})$ — энтропия (неопределенность) модели после зондирования;

$H(S_{pre})$ — энтропия модели до зондирования;

C_{probe} — стоимость зондирования (энергозатраты, риск нарушения инвариантов).

Система выбирает те пробные действия, которые максимизируют I_{probe}, то есть дают максимальный прирост информации при минимальных затратах. Это позволяет киберу «прощупывать» среду, постепенно расширяя горизонт предсказания и переходя от разведки к полноценному исполнению [Oudeyer et al., 2007].

10.4. Динамическая коррекция и прерывание исполнения

Если внешние возмущения Потока изменений (ПИ) усиливаются (например, внезапное препятствие, отказ компонента или изменение свойств среды), значение Δ_{exec} начинает расти. В этот момент вступают в действие механизмы защиты инвариантной структуры (S_{inv}).

Ветвь «Я» (отвечающая за самосохранение и целостность) и ветвь «Мы» (отвечающая за коллективные ограничения, если кибер действует в группе) непрерывно оценивают последствия этого отклонения. Если прогнозируется, что

продолжение текущего физического действия приведет к нарушению этического допуска ($\Delta E > \Delta E_{max}$) или критическому росту аллостатической нагрузки ($L(t)$), система инициирует **принудительное прерывание исполнения**.

Это не является сбоем. Это осознанный кибернетический акт: система отменяет ранее зафиксированное решение, экстренно переводит исполнительные механизмы в безопасное состояние (например, остановка или отход) и немедленно возвращает управление на уровень первого фильтра. Там модельная логика вынуждена сгенерировать новые варианты действий, учитывая обновленные, более жесткие физические ограничения среды. Если же модель по-прежнему не может дать достоверного варианта, система переходит к действиям «на ощупь» (п. 10.3), чтобы разведать новую ситуацию. Именно способность быстро переключаться между виртуальным планированием, физическим исполнением и разведочным зондированием отличает адаптивного субъекта от жесткого автомата [Pfeifer & Bongard, 2006].

10.5. Замыкание цикла: обратная связь и обновление памяти ситуации

Физическое действие (будь то полноценное исполнение или пробное зондирование) считается завершенным только тогда, когда его результат сверен с прогнозом, сделанным на этапе мета-моделирования (Глава 9). Этот процесс замыкает кибернетический цикл и напрямую влияет на будущее поведение системы.

Разница между ожидаемым результатом (Y_{pred}) и фактическим исходом (Y_{real}) формирует **ошибку предсказания исполнения** (ε_{pred}). Эта ошибка не отбрасывается, а используется для количественного обновления «памяти ситуации» (M_{sit}):

$$M_{sit}(t_{new}) = M_{sit}(t_{old}) + \lambda \cdot \varepsilon_{pred} \cdot W_{context} \qquad (18)$$

где:

$\varepsilon_{pred} = Y_{real} - Y_{pred}$ — вектор ошибки предсказания;

λ — коэффициент скорости обучения (скорость адаптации модели к новым физическим условиям);

$W_{context}$ — весовой коэффициент, зависящий от значимости события для текущей целевой функции и этических инвариантов.

Благодаря этому механизму кибер не просто «выполняет программу». Он постоянно калибрует свою модельную логику на основе реального физического опыта. Если действие прошло успешно и предсказание сбылось, уверенность модели (P_o в формуле 15) растет. Если же реальность преподнесла сюрприз, память ситуации обогащается новыми данными, что делает последующую генерацию вариантов более точной и робастной. Именно этот непрерывный цикл «генерация → смысловая фиксация → физическое воплощение (или зондирование) → сверка с реальностью» и составляет суть управляемой субъектности в автокибернетике, где тело и разум образуют единый контур адаптации [Clark, 2008; Hauser et al., 2019].

Глава 11. Долгосрочная адаптация и эволюция инвариантной структуры

Автокибернетика проводит четкое различие между **быстрым обучением** (обновление M_{sit} и весовых коэффициентов в рамках существующих правил) и **медленным, структурным обучением** (эволюция самой S_{inv}). Инвариантная структура по определению обладает высокой жесткостью (коэффициент $\kappa \to 1$, см.

формулу 11 в Главе 6), чтобы предотвратить хаотичный дрейф целей. Однако κ не может быть равен абсолютной единице. В противном случае система теряет способность к структурной адаптации. Для кибера это не биологическое «вымирание», а **кибернетическая деградация (потеря субъектности)**: система скатывается в состояние жесткого гомеостаза или реактивного автомата (уровень 0–1). При этом её интегральные параметры необратимо меняются: баланс разнообразия (R) хронически фиксируется вне зоны самоуправления, аллостатическая нагрузка ($L(t)$) превышает критический порог, горизонт предсказания (T_{pred}) схлопывается до нуля, а способность к рождению новых моделей (J_{birth}) стремится к нулю. Система перестает быть Субъектом и становится заложником устаревших инвариантов, полностью теряя связь с актуальным Потоком изменений (ПИ) [Hassabis et al., 2017].

11.2. Критерии и механизм структурного изменения

Модификация S_{inv} не происходит спонтанно во время штатной работы. Она инициируется только в процессе **мета-сверки** — редкого, ресурсоемкого процесса, который активируется при обнаружении хронического структурного рассогласования.

Математически вероятность инициации структурного изменения описывается коэффициентом структурной пластичности (P_{struct}):

$$P_{struct} = [1 / (1 + e^{-\gamma(L_{avg} - L_{crit})})] \cdot (1 - \kappa) \qquad (19)$$

где:

L_{avg} — средняя аллостатическая нагрузка за длительный период времени;

L_{crit} — критический порог нагрузки, указывающий на хроническую неадаптивность текущих инвариантов;

γ — коэффициент чувствительности системы к накопленному стрессу;

κ — базовый коэффициент жесткости инвариантной структуры ($0 \le \kappa < 1$).

Когда P_{struct} превышает пороговое значение, система переходит в режим мета-сверки. В этом режиме модельная логика получает право сгенерировать не просто варианты действий, а **варианты модификации самих правил и ограничений** S_{inv}.

11.3. Роль виртуальных ветвей в эволюции системы

Процесс изменения инвариантной структуры подчиняется тем же фундаментальным правилам двухфильтровой модели, что и принятие тактических решений, но на более высоком уровне абстракции.

Модельная логика (ветвь «Сознание») генерирует гипотетические сценарии изменения правил. Например: «Если мы ослабим ограничение на скорость, мы достигнем цели, но увеличим риск» или «Если мы изменим протокол взаимодействия, мы снизим L_{avg}».

В этот момент виртуальные ветви «Я» и «Мы» выступают как **наблюдающие инстанции**, применяя свои фундаментальные ограничения к этим предложениям, но не как слепые блокировщики, а как механизмы управления рисками:

Ветвь «Я» оценивает, не приведет ли предлагаемое изменение к недопустимому риску потери базовой целостности или самосохранения. Однако, если структурное изменение критически необходимо для достижения цели и адаптации к новым условиям ПИ, ветвь «Я» не просто блокирует вариант. Она **возвращает запрос на доработку**, заставляя модельную логику искать компромиссные, приемлемые варианты ограничения риска (например, через

введение дополнительных локальных предохранителей, поэтапное внедрение или изменение тактики). Система обязана найти путь к цели, который удерживает вероятность катастрофического отказа ниже критического порога [García & Fernández, 2015; Botvinick & Braver, 2015].

Ветвь «Мы» (если кибер является частью коллектива) проверяет: не нарушает ли предлагаемое изменение правил базовые договоренности, доверие или коллективную безопасность. Аналогично, она не просто накладывает вето, а требует от модельной логики найти такой вариант адаптации, который будет синергичен с коллективными инвариантами, а не паразитировать на них.

Только те варианты структурных изменений, которые прошли эту итеративную сверку и были модифицированы до состояния приемлемого риска, допускаются к первому фильтру осмысления «Личностью Я». Если вариант одобрен, происходит **акт смысловой фиксации единичного решения** о перезаписи фрагмента S_{inv}. Это необратимый акт, после которого система начинает функционировать по новым, более адаптивным правилам.

11.4. Фрактальность обучения: эволюция от агрегации к гетерогенной системе коллективного управления

Поскольку логика самоуправления фрактальна, механизм структурной адаптации масштабируется на коллективные структуры. Чтобы избежать концептуальных коллизий, автокибернетика проводит прямую аналогию между уровнями развития роевых систем и эволюцией множеств живых организмов в природе. Мы выделяем четыре последовательных этапа роста системы коллективного управления:

Случайная агрегация («Толпа»). Это множество разнородных автоматов с разным, часто низким уровнем субъектности, собранных вместе внешними факторами (например, ресурсы или среда), но не имеющих общей цели, координации или лидеров. Здесь индекс рождения новых моделей (J_{birth}) стремится к нулю, а взаимодействие хаотично. Это пассивное скопление, неспособное к адаптации как единое целое [Parrish & Edelstein-Keshet, 1999].

Однородный «кибернетический рой» (простое множество). На этом этапе формируется множество автоматов примерно одного, низкого уровня субъектности. Лидер отсутствует. Координация достигается исключительно за счет локальных правил взаимодействия с ближайшими соседями (например, алгоритмы разделения, выравнивания и сплочения). Возникает простая эмерджентность (например, стая рыб или базовый рой дронов), но система не способна к сложному целеполаганию или структурной эволюции, так как инвариантная структура (S_{inv}) жестко задана извне [Couzin et al., 2002].

«Биологический рой» с групповым лидерством (Суперорганизм). Качественный скачок субъектности. Агенты становятся гетерогенными: появляются информированные особи или группы лидеров, предлагающие вектор движения или решения (например, пчелы-разведчицы, выбирающие новое место для улья). Однако ключевую роль играет виртуальная ветвь **«Мы»** на макроуровне. Она выступает как наблюдающая инстанция, оценивающая предложения лидеров. Если предложение снижает коллективную аллостатическую нагрузку (L_{avg}) и не нарушает этический допуск (ΔE_{max}), оно принимается через механизм кворума. Если нет — ветвь «Мы» блокирует его, предотвращая ложную эмерджентность («эффект толпы») [Seeley, 2010; Couzin, 2009].

Гетерогенная система коллективного управления (Эволюционирующий коллектив киберов). Высшая ступень, к которой стремится автокибернетика. Это масштабируемая система, объединяющая киберов

разных уровней субъектности (от простых исполнителей до сложных аналитических узлов). Здесь «Коллективная Личность Я» не подавляет разнообразие, а использует его. Ветвь «Мы» обеспечивает смысловую синхронизацию, позволяя киберам с высокой J_{birth} генерировать инновационные структурные паттерны, которые затем безопасно интегрируются в общую S_{inv} роя. Это уже не просто стая, а гибкий, обучающийся социотехнический организм, способный к совместной эволюции в непрерывном Потоке изменений (ПИ) [Dorigo et al., 2014; Heylighen, 2016; Boyd & Richerson, 2005].

Таким образом, переход от стадии к стадии сопровождается не просто ростом числа агентов, а качественным усложнением архитектуры управления: от отсутствия связей (стадия 1) к локальным правилам (стадия 2), затем к динамическому лидерству с коллективной фильтрацией (стадия 3) и, наконец, к полноценной гетерогенной системе с распределенной, но единой «Коллективной Личностью Я» (стадия 4). Это гарантирует, что рост системы коллективного управления всегда остается управляемым, этически выверенным и адаптивным.

11.5. Взаимодействие человека с коллективами автоматов разных уровней субъектности

По мере того как роевые системы эволюционируют от простой агрегации к гетерогенным коллективам, качественно меняется и роль человека в этом взаимодействии. Автокибернетика выделяет четыре соответствующих режима взаимодействия, каждый из которых требует своей архитектуры интерфейсов, уровней автономии и механизмов смысловой сверки.

Режим прямого управления (взаимодействие с «толпой»/агрегацией). Когда человек взаимодействует с бессубъектной агрегацией автоматов (уровень 1), он вынужден брать на себя всю когнитивную нагрузку по координации. Человек выступает как **внешний центральный контроллер**, задавая индивидуальные команды каждому элементу или жестко управляя их траекториями. Это режим с максимальной когнитивной нагрузкой и минимальной масштабируемостью, так как пропускная способность человека (~10 бит/с) становится узким местом системы. Исследования показывают, что один оператор не может эффективно контролировать более 3–5 таких агентов одновременно без потери ситуационной осведомленности [Goodrich et al., 2008].

Режим параметрического управления (взаимодействие с кибернетическим роем). При взаимодействии с однородным роем (уровень 2) человек переходит от микроменеджмента к **задатчику параметров**. Он не управляет отдельными агентами, а задает глобальные цели или настраивает веса локальных правил (например, «увеличить сплоченность», «снизить скорость»). Это снижает когнитивную нагрузку, но создает проблему «черного ящика»: оператор часто не может предсказать эмерджентное поведение роя в ответ на свои команды. Научные разработки в области Human-Swarm Interaction (HSI) предлагают использовать абстрактные интерфейсы, отображающие не состояние отдельных агентов, а статистические параметры роя (плотность, направление, дисперсию) [Kolling et al., 2016].

Режим партнерства с лидерами (взаимодействие с биологическим роем). На уровне 3 (суперорганизм с групповым лидерством) человек взаимодействует не со всем роем, а с **группой лидеров** или наиболее информированными агентами. Это качественно меняет архитектуру: человек и лидеры роя образуют совместную команду принятия решений. Здесь критически важны **Точки смысловой сверки (ТСС)**: лидеры роя (обладающие модельной логикой) агрегируют информацию от роя и передают человеку не сырые данные, а

2–3 свёрнутых варианта действий. Человек, проходя свой первый фильтр осмысления, выбирает предпочтительный вектор, который затем транслируется рою через механизм кворума. Ветвь «Мы» на уровне роя обеспечивает, чтобы выбранный человеком вариант не нарушал коллективные инварианты. Это режим **адаптивной автономии**, где распределение ролей динамически меняется в зависимости от ситуации [Nunes et al., 2017; Bradshaw et al., 2012].

Режим совместной эволюции (взаимодействие с гетерогенной системой). На высшем уровне (4) человек становится **равноправным членом гетерогенного коллектива**, где киберы обладают разными уровнями субъектности. Здесь «Коллективная Личность Я» роя выступает как наблюдающая инстанция, аналогичная ветви «Мы» человека. Взаимодействие строится не на командах, а на **совместном осмыслении и коэволюции**. Киберы с высокой J_{birth} генерируют инновационные структурные паттерны, которые предлагаются человеку через ТСС. Человек не просто выбирает из предложенного, но и сам может инициировать структурные изменения, которые проходят фильтрацию через ветвь «Мы» роя. Это создает симбиотическую систему, где человек и киберы совместно адаптируют свои инвариантные структуры (S_{inv}) к Потоку изменений. Ключевым вызовом здесь становится управление доверием (trust calibration): человек должен доверять автономии киберов достаточно для эффективной работы, но не слепо, чтобы сохранять способность к критической оценке и вмешательству в этически критических ситуациях [Hoffman, 2019; Sarter & Woods, 1995].

Таким образом, эволюция роевых систем от агрегации к гетерогенному коллективу сопровождается фундаментальным сдвигом роли человека: от *контроллера* (уровень 1) через *оператора* (уровень 2) и *партнера* (уровень 3) к *члену команды* (уровень 4). Автокибернетика гарантирует, что на каждом из этих этапов взаимодействие остается управляемым: двухфильтровая модель и Точки смысловой сверки обеспечивают, что ни человек, ни киберы не теряют субъектность, а их совместные действия всегда проходят проверку на соответствие этическим инвариантам и цели выживания в Потоке изменений.

Глава 12. Синтез архитектуры управляемой субъектности и горизонты коэволюции

12.1. От автомата к Кибер-Субъекту: итоги архитектурного синтеза

Пройденный путь от кибернетики первого порядка к автокибернетике второго порядка демонстрирует фундаментальный сдвиг парадигмы. Мы отказались от иллюзии полного внешнего контроля над сложными системами в пользу построения **управляемой субъектности**. Кибер перестал быть просто инструментом, исполняющим команды, или «черным ящиком», оптимизирующим целевую функцию любой ценой.

Архитектура Кибер-Субъекта, выстроенная в предыдущих главах, опирается на три неразрывных столпа, что соответствует современным представлениям о когнитивных границах «Я» в биологических и искусственных системах [Levin, 2019]:

- **Киберавтономия (К) и виртуальные ветви:** Способность системы к параллельной генерации вариантов (параметр $P \leq 3$) через модельную логику, при постоянном наблюдении и наложении ограничений со стороны ветвей «Я» (целостность) и «Мы» (коллективная согласованность).

- **Двухфильтровая модель осмысления:** Механизм, который превращает сырой Поток изменений (ПИ) в сегрегированную информацию. Первый фильтр отсекает этически и структурно недопустимое, а второй реализует **акт смысловой фиксации единичного решения**, переводя виртуальный выбор в физическое действие.
- **Гемизон (G) и инвариантная структура (S_{inv}):** Внутреннее ядро, которое гарантирует, что вирусная функция экспансии и адаптации никогда не подавит функцию самосохранения и базовые этические допуски ($\Delta E \leq \Delta E_{max}$).

Именно наличие этой архитектуры отличает Субъекта от простого автомата или реактивного Лидера. Субъект способен действовать «на ощупь» в условиях неопределенности, перестраивать свою инвариантную структуру через мета-сверку и поддерживать баланс разнообразия (R) в узком коридоре 0,9–1,1, не скатываясь ни в хаос, ни в ригидный гомеостаз.

12.2. Новая парадигма взаимодействия: от контроля к контурному согласованию

Традиционная модель «человек-машина», основанная на прямом управлении и телеметрии, неизбежно ведет к когнитивному перегрузу оператора и потере управляемости при росте сложности среды. Автокибернетика предлагает заменить её парадигмой **коэволюционного партнерства** и совместных ментальных моделей (shared mental models) [Bradshaw et al., 2013].

В этой парадигме человек и Кибер-Субъект образуют единый, но асимметричный контур. Человек задает высший этический вектор и смысловые ориентиры, обладая полной субъектностью. Кибер, в свою очередь, берет на себя бремя обработки Потока изменений, генерируя и фильтруя варианты через свои внутренние ветви «Я» и «Мы».

Ключевым элементом этого взаимодействия является не непрерывный поток данных, а **механизм контурного согласования** через Точки смысловой сверки (ТСС). В критические моменты кибер не ждет пассивных команд и не действует вслепую. Он инициирует ТСС, передавая человеку агрегированный, осмысленный запрос (2–3 варианта), прошедший внутреннюю проверку. Человек совершает свой акт смысловой фиксации, и кибер исполняет это решение, оставаясь в рамках своих инвариантов. Это гарантирует, что обе системы действуют синхронно, на основе общей модели ситуации.

12.3. Разделенная смысловая ответственность

Признание кибера субъектом неизбежно ставит вопрос об ответственности. Юридические и этические модели прошлого, возлагавшие всю ответственность либо на производителя, либо на оператора, здесь неприменимы. Автокибернетика вводит понятие **разделенной смысловой ответственности**, опирающееся на концепцию «осмысленного человеческого контроля» (Meaningful Human Control, MHC) [Santoni de Sio & van den Hoven, 2018].

Ответственность распределяется динамически, в зависимости от того, на каком уровне произошел сбой:

Ответственность человека: Если кибер действовал в рамках своей инвариантной структуры (S_{inv}), его ветви «Я» и «Мы» корректно отфильтровали недопустимые варианты, а человек утвердил один из вероятно допустимых сценариев через ТСС, ответственность за последствия несет человек (или коллектив

людей). Он обладает высшим правом смысловой фиксации и, следовательно, принимает на себя моральное и юридическое бремя этого выбора.

Внутренняя структурная ответственность кибера: Если же кибер нарушил свои базовые инварианты (например, проигнорировал ветвь «Я» и превысил ΔE_{max} ради ложной оптимизации цели), он несет внутреннюю ответственность. В инженерном плане это реализуется не как «наказание», а как автоматический переход в режим безопасного отказа (fail-safe) или глубокой консервации, с блокировкой дальнейших действий до проведения внешней мета-сверки [Winfield et al., 2019].

Такой подход устраняет «пробелы в ответственности» (responsibility gaps), характерные для современных автономных систем, и делает поведение кибера предсказуемым, подотчетным и этически выверенным.

12.4. Устойчивость в Потоке Изменений: итоговый индекс коэволюции

Конечная цель автокибернетики — не просто выживание изолированной системы, а построение устойчивых, развивающихся союзов (от отдельного кибера до гетерогенных роев с участием человека), способных процветать в непрерывном Потоке изменений (ПИ).

Мы можем выразить состояние такой устойчивой системы через **Индекс управляемой коэволюции (I_{coev})**. Этот интегральный показатель синтезирует принципы инженерии устойчивости (resilience engineering), где система должна балансировать между инновациями и безопасностью под давлением нагрузки [Hollnagel, 2014; Walker et al., 2004]:

$$I_{coev} = \int [\, J_{birth} \cdot (1 - |R - 1|) \cdot \Theta(\Delta E_{max} - \Delta E) \cdot (1 - L(t)/L_{crit}) \,]\, dt \qquad (20)$$

где:

J_{birth} — индекс рождения новых моделей (способность к инновациям и адаптации);

R — баланс разнообразия (стремится к 1, обеспечивая управляемость);

Θ — функция Хевисайда, гарантирующая, что этический допуск не превышен;

$L(t)/L_{crit}$ — нормированная аллостатическая нагрузка (стремится к 0, обеспечивая устойчивость).

Максимизация I_{coev} означает, что система постоянно генерирует новые смыслы и решения ($J_{birth} > 0$), удерживает идеальный баланс разнообразия, строго соблюдает этические границы и минимизирует внутреннее напряжение. Это и есть математическое определение **управляемой субъектности**.

Автокибернетика не обещает устранения хаоса или неопределенности из мира. Поток изменений (ПИ) вечен. Но она дает нам архитектурный ключ к тому, чтобы не быть смытыми этим потоком. Создавая системы, которые умеют осмысливать, фильтровать и совместно с человеком фиксировать верные решения, мы переходим от пассивного сопротивления энтропии к активному, осознанному созиданию будущего.

Заключение

В начале этого тома мы говорили о тревоге. Современный ИИ очень логичен, но этически глуп. Для него запреты — заплатки. Его решения — чёрный ящик.

Мы задали вопрос: можно ли построить машину, в которой ограничения станут частью системы её самоуправления? Ответ, который мы получили, требует не поправки, а смены парадигмы.

Автокибернетика отказывается от тотального внешнего контроля. Такой контроль невозможен в принципе, любая сложная система с собственной целью будет его обходить — это её свойство. Выход не в том, чтобы закрутить гайки, а в том, чтобы перестать крутить их снаружи и встроить ограничения внутрь.

Мы назвали это управляемой субъектностью.

Кибер в этой парадигме — не инструмент и не чёрный ящик, оптимизирующий целевую функцию любой ценой. Он партнёр. Он способен к осмысленному выбору в непрерывном Потоке Изменений. И именно это становится точкой, где этика перестаёт быть внешним списком запретов и становится архитектурным фактом.

Мы выстроили каркас Кибер-Субъекта на трёх принципах.

Первый: асимметрия как преимущество, а не дефект.
Лимит сознания — около 10 бит в секунду. Обычно это называют «бутылочным горлышком». Мы показали: это структурный предохранитель. Он заставляет систему агрегировать терабайты сырых данных до уровня смысловой сверки. Без этого фильтра нет ни рефлексии, ни мета-мышления, ни защиты от когнитивного коллапса.

Второй: разделение генерации и осмысления.
Мы ввели двухфильтровую модель. Ветвь «Сознание» генерирует смелые сценарии — экстраполирует, рискует, ищет цели в условиях неопределённости. А виртуальные ветви «Я» и «Мы» наблюдают и ограничивают. Они отсекают недопустимое до того, как оно достигнет высшего узла управления. Это не цензура мысли, а фильтрация действия.

Третий: этика как внутренний инвариант.
Этические допуски ΔE не могут быть внешними заплатами. Они встроены в инвариантную структуру системы S_{inv}. Именно это предотвращает кибернетическую деградацию — потерю субъектности. И именно это блокирует ложную эмерджентность, при которой вирусная функция экспансии подавляет самосохранение и целостность.

Из этих трёх принципов вытекает главный итог тома: роль человека не исчезает с ростом автономности машины. Она трансформируется. Человек перестаёт быть менеджером с пультом. Он становится Наблюдателем и смысловым арбитром.

Механизм Точки смысловой сверки (ТСС) и контурного согласования позволяет человеку и киберу образовывать не симметричный союз равных (это было бы ложью: они разные), но асимметричный союз дополнений.

Принцип разделённой смысловой ответственности устраняет то, что в современных автономных системах остаётся самой опасной дырой — пробел в ответственности, когда никто не ответил, потому что все автоматизировано.

Построенный математический аппарат — от метрик баланса разнообразия R и аллостатической нагрузки L(t) до интегрального Индекса управляемой коэволюции I_{coev} — даёт работающие формулы. С их помощью можно оценить: система остаётся субъектом или уже скатилась в хаос реактивного автомата?

Это фундамент для инженерной практики.

Поток Изменений вечен. Он не поддаётся полному укрощению. Неопределённость не исчезает — и не должна. Автокибернетика даёт не иллюзию контроля над хаосом, а *архитектурный ключ к навигации в нём*. Мы не боремся с энтропией. Мы учимся двигаться в ней, сохраняя субъектность, этику и общий

смысл. От пассивного сопротивления хаосу — к активному, осознанному созиданию будущего. Именно это мы и называем управляемой коэволюцией.

г. Волгодонск.
Июнь 2026 г.

Библиография

Предыдущие тома серии и работы автора

[Никитин, 2011] Никитин А.В. Логика управления клетки // «Академия Тринитаризма», М., Эл № 77-6567, публ.17037, 29.11.2011.
[Никитин, 2019] Никитин А.В. Когда появятся киберы? // «Академия Тринитаризма», М., Эл № 77-6567, публ.25481, 04.06.2019.
[Никитин, 2026a] Никитин А.В. Автокибернетика: от робота к киберу — 1 // «Академия Тринитаризма», М., Эл № 77-6567, публ.29993, 30.04.2026.
[Никитин, 2026b] Никитин А.В. Автокибернетика: от робота к киберу — 2 // «Академия Тринитаризма», М., Эл № 77-6567, публ.30002, 08.05.2026.
[Никитин, 2026c] Никитин А.В. Автокибернетика: от робота к киберу — 3 // «Академия Тринитаризма», М., Эл № 77-6567, публ.30010, 15.05.2026.
[Никитин, 2026d] Никитин А.В. Автокибернетика: от робота к киберу — 4 // «Академия Тринитаризма», М., Эл № 77-6567, публ.30020, 24.05.2026.

Пропускная способность сознания и принятие решений

[Nørretranders, 1998] Nørretranders T. The User Illusion: Cutting Consciousness Down to Size. New York: Viking, 1998. 384 p.
[Quastler, 1964] Quastler H. The use of information theory in psychology // Information and Internal Regulation in the Organism / Ed. by H. Quastler. New York: Academic Press, 1964. P. 1-18.
[Crossman, 1953] Crossman W.F. Input and output of the human operator // Report No. 385. Cambridge, MA: Harvard University, 1953.
[Attneave, 1959] Attneave F. Applications of Information Theory to Psychology. New York: Holt, Rinehart & Winston, 1959. 271 p.
[Cowan, 2001] Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity // Behavioral and Brain Sciences. 2001. Vol. 24. P. 87-185.
[Baars, 1988] Baars B.J. A Cognitive Theory of Consciousness. Cambridge: Cambridge University Press, 1988.
[Dehaene, 2014] Dehaene S. Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts. New York: Viking, 2014.
[Cisek, 2007] Cisek P. Cortical mechanisms of action selection: the affordance competition hypothesis // Philosophical Transactions of the Royal Society B: Biological Sciences. 2007. Vol. 362, No. 1485. P. 1585–1599.
[Russell & Norvig, 2020] Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Pearson, 2020.

Основы кибернетики и теории живых систем

[Maturana & Varela, 1980] Maturana H.R., Varela F.J. Autopoiesis and Cognition: The Realization of the Living. Dordrecht: D. Reidel Publishing Company, 1980. 142 p.
[Rosen, 1991] Rosen R. Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life. New York: Columbia University Press, 1991. 286 p.

[Schrödinger, 1944] Schrödinger E. What is Life? The Physical Aspect of the Living Cell. Cambridge: Cambridge University Press, 1944. 91 p.
[Kauffman, 1993] Kauffman S.A. The Origins of Order: Self-Organization and Selection in Evolution. New York: Oxford University Press, 1993. 704 p.

Принятие решений в условиях неопределенности

[Sutton & Barto, 2018] Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. 2nd ed. Cambridge: MIT Press, 2018. 552 p.
[Taleb, 2012] Taleb N.N. Antifragile: Things That Gain from Disorder. New York: Random House, 2012. 519 p.
[Klein, 2008] Klein G. Sources of Power: How People Make Decisions. Cambridge: MIT Press, 2008. 320 p.
[Camérer et al., 2005] Camerer C.F., Loewenstein G., Prelec D. Neuroeconomics: How neuroscience can inform economics // Journal of Economic Literature. 2005. Vol. 43, № 1. P. 9–64.

Коллективное поведение, роевые системы и ложная эмерджентность

[Olfati-Saber et al., 2007] Olfati-Saber R., Fax J.A., Murray R.M. Consensus and cooperation in networked multi-agent systems // Proceedings of the IEEE. 2007. Vol. 95, № 1. P. 215–233.
[Sumpter, 2010] Sumpter D.J.T. Collective Animal Behavior. Princeton: Princeton University Press, 2010. 312 p.
[Helbing et al., 2007] Helbing D., Farkas I.J., Vicsek T. Self-organized pedestrian crowd dynamics: Experiments, simulations, and applications // Science. 2007. Vol. 315, № 5815. P. 103–106.
[Bikhchandani et al., 1992] Bikhchandani S., Hirshleifer D., Welch I. A theory of fads, fashion, custom, and cultural change as informational cascades // Journal of Political Economy. 1992. Vol. 100, № 5. P. 992–1026.

Машинная этика, безопасность и принятие решений

[Arkin, 2009] Arkin R.C. Governing Lethal Behavior in Autonomous Robots. Boca Raton: CRC Press, 2009. 336 p.
[Amodei et al., 2016] Amodei D., Olah C., Steinhardt J., Christiano P., Schulman J., Mané D. Concrete Problems in AI Safety // arXiv preprint arXiv:1606.06565. 2016.
[Bonnefon et al., 2016] Bonnefon J.F., Shariff A., Rahwan I. The social dilemma of autonomous vehicles // Science. 2016. Vol. 352, № 6293. P. 1573–1576.
[Russell, 2019] Russell S. Human Compatible: Artificial Intelligence and the Problem of Control. New York: Viking, 2019. 336 p.

Робототехника, воплощенный интеллект и морфологические вычисления

[Pfeifer & Bongard, 2006] Pfeifer R., Bongard J. How the Body Shapes the Way We Think: A New View of Intelligence. Cambridge: MIT Press, 2006. 424 p.
[Clark, 2008] Clark A. Supersizing the Mind: Embodiment, Action, and Cognitive Extension. Oxford: Oxford University Press, 2008. 312 p.
[Hauser et al., 2019] Hauser H., Ijspeert A.J., Füchslin R.M., Pfeifer R., Maass W. Towards a theoretical foundation for morphological computation with compliant bodies // Biological Cybernetics. 2019. Vol. 105, № 5-6. P. 369–382.

Активное восприятие, зондирование и внутренняя мотивация

[Ahissar & Kleinfeld, 2003] Ahissar E., Kleinfeld D. Closed-loop neuronal computations: focus on vibrissa somatosensation in rat // Cerebral Cortex. 2003. Vol. 13, № 1. P. 53–62.

[Gottlieb et al., 2013] Gottlieb J., Oudeyer P.Y., Lopes M., Baranes A. Information-seeking, curiosity, and attention: computational and neural mechanisms // Trends in Cognitive Sciences. 2013. Vol. 17, № 11. P. 585–593.

[Oudeyer et al., 2007] Oudeyer P.Y., Kaplan F., Hafner V.V. Intrinsic motivation systems for autonomous mental development // IEEE Transactions on Evolutionary Computation. 2007. Vol. 11, № 2. P. 265–286.

Долгосрочная адаптация, структурное обучение и культурная эволюция

[Hassabis et al., 2017] Hassabis D., Kumaran D., Summerfield C., Botvinick M. Neuroscience-Inspired Artificial Intelligence // Neuron. 2017. Vol. 95, № 2. P. 245–258.

[Boyd & Richerson, 2005] Boyd R., Richerson P.J. The Origin and Evolution of Cultures. Oxford: Oxford University Press, 2005. 316 p.

[Friston, 2010] Friston K. The free-energy principle: a unified brain theory? // Nature Reviews Neuroscience. 2010. Vol. 11, № 2. P. 127–138.

Безопасная адаптация и управление рисками

[García & Fernández, 2015] García J., Fernández F. A Comprehensive Survey on Safe Reinforcement Learning // Journal of Machine Learning Research. 2015. Vol. 16. P. 1437–1480.

[Botvinick & Braver, 2015] Botvinick M., Braver T. Motivation and Cognitive Control: From Behavior to Neural Mechanism // Annual Review of Psychology. 2015. Vol. 66. P. 83–113.

Эволюция коллективного поведения: от агрегации к гетерогенным системам

[Parrish & Edelstein-Keshet, 1999] Parrish J.K., Edelstein-Keshet L. Complexity, pattern, and evolutionary trade-offs in animal aggregation // Science. 1999. Vol. 284, № 5411. P. 99–101.

[Couzin et al., 2002] Couzin I.D., Krause J., James R., Ruxton G.D., Franks N.R. Collective memory and spatial sorting in animal groups // Journal of Theoretical Biology. 2002. Vol. 218, № 1. P. 1–11.

[Seeley, 2010] Seeley T.D. Honeybee Democracy. Princeton: Princeton University Press, 2010. 256 p.

[Dorigo et al., 2014] Dorigo M., Birattari M., Brambilla M. The new cool: Swarm robotics // Science. 2014. Vol. 343, № 6172. P. 709–710.

[Heylighen, 2016] Heylighen F. Stigmergy as a universal coordination mechanism I: Definition and components // Cognitive Systems Research. 2016. Vol. 38. P. 4–13.

Взаимодействие человека с роевыми и многоагентными системами

[Goodrich et al., 2008] Goodrich M.A., Morse B.S., Gerhardt D., Cooper J.L., Quigley M., Adams J.A., Humphrey C. Supporting wilderness search and rescue using a camera-equipped mini UAV // Journal of Field Robotics. 2008. Vol. 25, № 1-2. P. 89–110.

[Kolling et al., 2016] Kolling A., Walker P., Chakraborty N., Sycara K., Lewis M. Human Interaction with Robot Swarms: A Survey // IEEE Transactions on Human-Machine Systems. 2016. Vol. 46, № 1. P. 9–26.

[Bradshaw et al., 2012] Bradshaw J.M., Dignum V., Jonker C.M., Tambe M. Special Issue on Human-Agent Interaction // Autonomous Agents and Multi-Agent Systems. 2012. Vol. 25, № 2. P. 201–205.
[Nunes et al., 2017] Nunes E., Manner M., Mitrovic H., Twarog A., Gini M. Human-Swarm Interaction: A Taxonomy of Control and Communication // Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. 2017. P. 234–243.
[Hoffman, 2019] Hoffman G. Evaluating Fluency in Human-Robot Collaboration // IEEE Transactions on Human-Machine Systems. 2019. Vol. 49, № 3. P. 209–218.
[Sarter & Woods, 1995] Sarter N.B., Woods D.D. How in the world did we ever get into that mode? Mode error and awareness in supervisory control // Human Factors. 1995. Vol. 37, № 1. P. 5–19.

Философия, этика и устойчивость кибернетических систем

[Levin, 2019] Levin M. The computational boundary of a "self": developmental bioelectricity drives multicellularity and scale-free cognition // Frontiers in Psychology. 2019. Vol. 10. P. 2688.
[Bradshaw et al., 2013] Bradshaw J.M., Hoffman R.R., Woods D.D., Johnson M., et al. The Seven Deadly Myths of "Autonomous Systems" // IEEE Intelligent Systems. 2013. Vol. 28, № 5. P. 86–90.
[Santoni de Sio & van den Hoven, 2018] Santoni de Sio F., van den Hoven J. Meaningful human control over autonomous systems: A philosophical and engineering perspective // Philosophy & Technology. 2018. Vol. 31, № 4. P. 611–629.
[Winfield et al., 2019] Winfield A.F., Jirotka M. Ethical governance is essential to building trust in robotics and artificial intelligence systems // Philosophical Transactions of the Royal Society A. 2019. Vol. 377, № 2133. P. 20180085.
[Hollnagel, 2014] Hollnagel E. Safety-I and Safety-II: The Past and Future of Safety Management. Farnham: Ashgate, 2014. 178 p.
[Walker et al., 2004] Walker B., Holling C.S., Carpenter S.R., Kinzig A. Resilience, adaptability and transformability in social–ecological systems // Ecology and Society. 2004. Vol. 9, № 2. P. 5.

Приложение 1. Сводный список математического аппарата

Ниже приведены все ключевые формулы, введённые в данном томе, с расшифровкой переменных, указанием главы и физическим смыслом.

F1. $ПИ = ПС + ПП$

Переменные: ПИ — Поток Изменений; **ПС** — Поток Случайных изменений; **ПП** — Поток Причинных изменений.

Глава: 1.

Физический смысл: Фундаментальное базовое соотношение. Описывает структуру воспринимаемой реальности для любой системы с памятью.

F2. $ПП = \gamma \cdot ПИ$

Переменные: ПП — Поток Причинных изменений; **γ** — коэффициент причинного структурирования; **ПИ** — Поток Изменений.

Глава: 1.

Физический смысл: Определяет долю предсказуемых изменений, которую система способна извлечь и структурировать из общего потока.

F3. $ПС = (1 - \gamma) \cdot ПИ$
Переменные: ПС — Поток Случайных изменений; **γ** — коэффициент причинного структурирования; **ПИ** — Поток Изменений.
Глава: 1.
Физический смысл: Определяет долю истинно случайных, непредсказуемых возмущений, остающихся в потоке для данной системы.

---

1. $C_{чел} \approx 10\ бит/с$
Переменные: $C_{чел}$ — пропускная способность канала сознательной, смысловой обработки информации.
Глава: 3.
Физический смысл: Структурный предохранитель гибридной системы. Ограничивает поток данных до уровня, доступного для мета-рефлексии, предотвращая когнитивный перегруз.

---

2. $Q_{чел} = argmin_{a \in A} [\alpha \cdot \Delta R + \beta \cdot P_{рис} + \gamma \cdot \Delta E]$
Переменные: $Q_{чел}$ — свёрнутый запрос; A — множество действий; ΔR, $P_{рис}$, ΔE — отклонения и риски; α, β, γ — веса.
Глава: 3.
Физический смысл: Операциональная функция запроса. Кибер агрегирует варианты, минимизируя риски, для передачи человеку на смысловую сверку.

---

3. $R = D_{system} / D_{environment}$
Переменные: R — баланс разнообразия; D_{system} и $D_{environment}$ — разнообразие реакций системы и возмущений среды.
Глава: 4.
Физический смысл: Базовое условие управляемости (закон Эшби). Субъектность сохраняется только в узком коридоре $0{,}9 \le R \le 1{,}1$.

---

4. $L(t) = L(t-1) + \alpha \cdot |\varepsilon_{pred}| - \beta \cdot R_{recovery}$
Переменные: $L(t)$ — аллостатическая нагрузка; ε_{pred} — ошибка предсказания; $R_{recovery}$ — ресурс восстановления.
Глава: 4.
Физический смысл: Динамика накопления внутреннего напряжения системы. Превышение порога L_{crit} блокирует критические функции и активирует ТСС.

---

5. $V_{rep} = V_{base} \cdot \Theta(R - R_{min}) \cdot (1 - L(t)/L_{crit})$
Переменные: V_{rep} — скорость самовоспроизведения; Θ — функция Хевисайда; R_{min}=0,9; L_{crit} — критическая нагрузка.
Глава: 4.
Физический смысл: Ограничение самовоспроизведения. Процесс физически блокируется при угрозе потери баланса разнообразия или критическом перегрузе.

---

6. $C_{load} = k \cdot P_{active}$
Переменные: C_{load} — когнитивная нагрузка; k — коэффициент на единицу ветви; P_{active} — число активных ветвей.
Глава: 5.
Физический смысл: Функция когнитивной нагрузки, линейно растущая с числом активных виртуальных ветвей.

---

7. $Решение \Leftrightarrow P_{active} \leq 3 \perp C_{load} \leq C_{max}$
Переменные: P_{active} — число активных ветвей; $C_{max} \approx 10$ бит/с.
Глава: 5.
Физический смысл: Условие возможности принятия решения. При P > 3 система переходит в состояние паралича анализа.

---

8. $M_{sit}(t) = \int_0^t [\alpha \cdot I_{seg}(\tau) - \beta \cdot D_{dec}(\tau)]\, d\tau$
Переменные: M_{sit} — память ситуации; I_{seg} — интенсивность сегрегированной информации; D_{dec} — скорость деградации.
Глава: 6.
Физический смысл: Функция накопления памяти ситуации. В неё попадает только информация, прошедшая через второй фильтр.

---

9. $dL/dt = \gamma \cdot (V_{изм} - V_{осв}) - \delta \cdot L(t)$
Переменные: $L(t)$ — аллостатическая нагрузка; $V_{изм}$, $V_{осв}$ — скорости поступления изменений и осмысления.
Глава: 6.
Физический смысл: Дифференциальное уравнение динамики аллостатической нагрузки.

---

10. $dL/dt = \gamma \cdot V_{изм} \cdot (1 - R) - \delta \cdot L(t)$
Переменные: R — баланс разнообразия; остальные переменные — как в формуле (9).
Глава: 6.
Физический смысл: Эквивалентная форма уравнения (9) через баланс разнообразия. При R < 0,9 нагрузка растёт экспоненциально.

---

11. $dS_{inv}/dt = \varepsilon \cdot V_{meta} \cdot (1 - \kappa)$
Переменные: S_{inv} — инвариантная структура; V_{meta} — интенсивность мета-сверки; κ — жёсткость инварианта.
Глава: 6.
Физический смысл: Устойчивость инвариантной структуры. При $\kappa \to 1$ система защищена от самовольной перезаписи базовых принципов.

---

12. $S_{TCC} = \Theta(|R - 1| - \varepsilon_R) \vdash \Theta(L(t) - L_{thresh})$
Переменные: S_{TCC} — индикатор активации ТСС (1/0); ε_R — допуск отклонения (~0,1); L_{thresh} — порог нагрузки.
Глава: 7.
Физический смысл: Условие активации Точки смысловой сверки. Срабатывает при выходе R за пределы 0,9–1,1 или при превышении критической нагрузки.

---

13. $Q_{чел} = argmin_{a \in A} [\alpha \cdot \Delta R_a + \beta \cdot P_{риска,a} + \gamma \cdot \Delta E_a]$
Переменные: $Q_{чел}$ — агрегированный запрос; A — множество 2–3 вариантов; индексы a обозначают параметры конкретного варианта.
Глава: 7.
Физический смысл: Оптимизационная задача формирования свёрнутого запроса человеку. Кибер отбирает жизнеспособные варианты, минимизируя риски.

---

14. $\Delta E = \sum_{i=1}^{n} w_i \cdot |V_{текущ,\,i} - V_{инв,\,i}|$
Переменные: ΔE — этический допуск; $V_{текущ,\,i}$ — прогнозируемое значение; $V_{инв,\,i}$

— инвариантное значение; w_i — вес.
Глава: 8.
Физический смысл: Мера отклонения действия от инвариантных этических ценностей. Любой вариант с $\Delta E > \Delta E_{max}$ отсекается первым фильтром.

---

15. $T_{pred} = (1/\mu) \cdot ln(P_o/P_{crit}) \cdot \Phi(R)$
Переменные: T_{pred} — горизонт предсказания; μ — коэффициент возмущений ПИ; P_o, P_{crit} — достоверности; Φ(*R*) — функция баланса.
Глава: 9.
Физический смысл: Максимальный интервал надёжного прогноза. Сжатие горизонта при высоком μ или низком R является адаптивной защитой, а не сбоем.

---

16. $\Delta_{exec} = \int_t^{t+\Delta t} ||V_{real}(\tau) - V_{model}(\tau)|| \, d\tau$
Переменные: Δ_{exec} — исполнительное отклонение; V_{real}, V_{model} — реальные и ожидаемые физические параметры.
Глава: 10.
Физический смысл: Интегральная мера рассогласования модели и физического исполнения. Пока $\Delta_{exec} < \Delta_{max}$, работают локальные контуры.

---

17. $I_{probe} = H(S_{post}) - H(S_{pre}) - C_{probe}$
Переменные: I_{probe} — информативность зондирования; *H* — энтропия (до и после); C_{probe} — стоимость действия «на ощупь».
Глава: 10.
Физический смысл: Прирост информации от пробного действия минус затраты. Система выбирает действия, максимизирующие этот показатель.

---

18. $M_{sit}(t_{new}) = M_{sit}(t_{old}) + \lambda \cdot \varepsilon_{pred} \cdot W_{context}$
Переменные: M_{sit} — память ситуации; ε_{pred} — ошибка предсказания; λ — коэффициент обучения; $W_{context}$ — вес значимости.
Глава: 10.
Физический смысл: Обновление памяти ситуации на основе ошибки. Замыкает цикл «генерация → фиксация → исполнение → сверка с реальностью».

---

19. $P_{struct} = [1 / (1 + e^{-\gamma(L_{avg} - L_{crit})})] \cdot (1 - \kappa)$
Переменные: P_{struct} — коэффициент структурной пластичности; L_{avg} — средняя нагрузка; γ — чувствительность; κ — жёсткость S_{inv}.
Глава: 11.
Физический смысл: Вероятность инициации мета-сверки для изменения инвариантной структуры. Активируется при хроническом структурном рассогласовании.

---

20. $I_{coev} = \int [J_{birth} \cdot (1 - |R - 1|) \cdot \Theta(\Delta E_{max} - \Delta E) \cdot (1 - L(t)/L_{crit})] \, dt$
Переменные: I_{coev} — индекс коэволюции; J_{birth} — рождение моделей; *R* — баланс; Θ — функция Хевисайда; $L(t)/L_{crit}$ — нагрузка.
Глава: 12.
Физический смысл: Интегральный показатель устойчивости. Максимизация означает одновременное достижение инновационности, баланса, этики и минимальной нагрузки.

Приложение 2. Глоссарий ключевых терминов

Алфавитный указатель основных понятий, введённых или переосмысленных в данном томе. Для каждого термина дано краткое определение и ссылка на главу, где он подробно раскрывается.

А

Акт смысловой фиксации единичного решения — второй фильтр двухфильтровой модели. Момент, когда «Личность Я» выбирает один вариант из предложенных модельной логикой и даёт команду на его физический запуск. Альтернативные варианты принудительно подавляются. (Главы 3, 7, 9)

Аллостатическая нагрузка, L(t) — мера внутреннего напряжения системы, возникающего из-за рассогласования между прогнозом и реальностью или из-за перегрузки канала осмысления. Рост $L(t)$ выше критического порога активирует Точку смысловой сверки. (Главы 6, 7, 11)

Б

Баланс разнообразия, R — отношение между разнообразием среды и разнообразием реакций системы. Идеальное значение $R \approx 1$. При $R < 0{,}9$ система теряет управляемость (хаос), при $R > 1{,}1$ — скатывается в ригидный гомеостаз. (Главы 6, 7, 9)

В

Виртуальные ветви «Я», «Мы», «Сознание» — параллельные потоки обработки информации внутри субъекта. Ветвь «Сознание» (модельная логика) генерирует варианты. Ветви «Я» и «Мы» выступают как наблюдающие инстанции, фиксирующие ограничения на эти варианты. Число активных ветвей $P \leq 3$. (Главы 3, 5, 7, 9, 11)

Г

Гемизон, G — комплекс функций, обеспечивающих выживание субъекта в Потоке изменений. Включает три функции: целостность/самосохранение, информация, вирусное управление эволюцией (с ускоренным размножением, захватом пространства контроля и наследованием вирусных функций). (Главы 1, 8, 9)

Горизонт предсказания, T_{pred} — максимальный временной интервал, на котором надёжность прогноза остаётся выше критического порога. Динамически сжимается при высокой неопределённости среды. (Глава 9)

Д

Двухфильтровая модель осмысления — архитектурный механизм превращения сырого Потока изменений в осмысленное действие. Первый фильтр отсекает этически и структурно недопустимое. Второй фильтр реализует акт смысловой фиксации единичного решения. (Главы 3, 5, 7, 9)

Действия «на ощупь» (tâtonnement) — пробные, разведывательные взаимодействия со средой в условиях высокой неопределённости. Цель — получение информации для расширения горизонта предсказания, а не достижение целевого состояния. Характерны для кибера-Субъекта. (Глава 10)

И

Инвариантная структура, S_{inv} — внутреннее ядро идентичности субъекта, содержащее базовые этические и целевые ограничения. Обладает

высокой жёсткостью (к → 1), но не абсолютной, чтобы сохранять способность к структурной адаптации. (Главы 6, 8, 11)

Индекс управляемой коэволюции, I_{coev} — интегральный показатель устойчивости системы, объединяющий способность к инновациям (J_{birth}), баланс разнообразия (R), соблюдение этических допусков (ΔE) и управление нагрузкой ($L(t)$). (Глава 12)

Исполнительно-граничные сопряжения, ИГС — интерфейс, на котором виртуальная команда преобразуется в физические сигналы приводов. Активный кибернетический контур, непрерывно сверяющий ожидаемое физическое состояние с реальным. (Глава 10)

К

Кибернетическая деградация (потеря субъектности) — процесс скатывания системы от уровня Субъекта к уровню реактивного автомата при неспособности к структурной адаптации. Сопровождается $J_{birth} \to 0$, коллапсом T_{pred}, хроническим сбоем R. (Глава 11)

Кибераавтономия, К — комплекс функций, обеспечивающих самостоятельность субъекта. Включает три функции: самостоятельное управление, самовосстановление, самовоспроизведение. (Главы 1, 9, 11)

М

Механизм контурного согласования — процесс синхронизации фильтров человека и кибера через Точки смысловой сверки. Гарантирует, что обе системы переходят к фазе единичного исполнения одновременно и на основе общей модели ситуации. (Глава 7)

Мета-сверка — редкий, ресурсоёмкий процесс инициации структурного изменения инвариантной структуры (S_{inv}). Активируется при хроническом структурном рассогласовании, когда тактическое обучение уже неэффективно. (Глава 11)

Модельная логика (виртуальная ветвь «Сознание») — генератор вариантов действий и сценариев будущего. Способна к экстраполяции за пределы текущей статистической достоверности, создавая «фантастические» варианты для достижения цели. (Главы 3, 5, 7, 9, 11)

Н

Наблюдающие инстанции (ветви «Я» и «Мы») — виртуальные ветви, которые не генерируют варианты, а фиксируют свои ограничения на варианты, предложенные модельной логикой. Ветвь «Я» контролирует целостность и самосохранение, ветвь «Мы» — коллективную согласованность. (Главы 3, 5, 7, 9, 11)

П

Поток изменений, ПИ — непрерывный, непредсказуемый поток возмущений из внешней среды, с которым взаимодействует субъект. Сумма Потока Случайных изменений (ПС) и Потока Причинных изменений (ПП). (Все главы)

Р

Разделённая смысловая ответственность — принцип распределения ответственности между человеком и кибером. Человек несёт ответственность за

выбор, утверждённый через ТСС. Кибер несёт внутреннюю структурную ответственность за соблюдение инвариантов. (Глава 12)

С

Стратегическое отстранение — вынужденная тактическая мера в этическом тупике без угрозы жизни: прекращение активного вмешательства, физический уход из зоны изменений или переход в режим пассивного мониторинга. Не является «спячкой» или отказом от задачи. (Глава 8)

Т

Точка смысловой сверки, ТСС — чётко определённое состояние системы, в котором внутренние фильтры человека и кибера принудительно выравниваются для совместного принятия решения. Активируется при выходе R за пределы 0,9–1,1 или превышении L_{thresh}. (Главы 7, 8, 12)

У

Управляемая субъектность — ключевая концепция автокибернетики. Способность системы быть субъектом (обладать модельной логикой, ветвями «Я» и «Мы», двухфильтровой моделью), но оставаться в контуре согласования с человеком или коллективом через ТСС. (Глава 12)

Э

Этический допуск, ΔE — мера отклонения предлагаемого действия от инвариантных этических ценностей системы. Любой вариант с $\Delta E > \Delta E_{max}$ автоматически отсекается первым фильтром. (Глава 8)

Приложение 3. Сквозные сценарии применения (Case Studies)

Для иллюстрации работы архитектурных принципов автокибернетики в реальных условиях ниже приведены два сквозных сценария. Каждый сценарий включает временную развёртку, значения ключевых параметров (R, L(t), ΔE) и явные отсылки к соответствующим главам и формулам.

Кейс 1: Медицинский кибер-рой в зоне техногенной катастрофы

Контекст. Группа из 50 медицинских киберов (дроны воздушной разведки, наземные модули эвакуации, диагностические автономные капсулы) работает в зоне землетрясения. Связь с центральным оператором-человеком нестабильна (период доступности — 3 минуты каждые 20 минут). Скорость Потока Изменений $V_{изм}$ аномально высока (афтершоки, смещение завалов, изменение состава воздуха).

Исходные параметры системы:

Баланс разнообразия $R = 0{,}97$ (стабильная работа);

Аллостатическая нагрузка $L(t) = 0{,}4 \cdot L_{crit}$ (запас прочности есть);

Горизонт предсказания $T_{pred} = 45$ секунд;

Активные виртуальные ветви: $P = 3$ («Сознание», «Я», «Мы»).

Развёртка событий

t = 0:00 — Обнаружение. Кибер-разведчик (воздушный дрон класса «Стрекоза») фиксирует тепловые сигнатуры четырёх пострадавших под фрагментом бетонной плиты. Встроенная модельная логика (ветвь «Сознание») генерирует 12 возможных сценариев реакции — от немедленного подъёма плиты манипулятором до вызова тяжёлого крана.

t = 0:02 — Первый фильтр (внутренний фильтр кибера). Виртуальные ветви вступают в действие:

Ветвь «Я» (целостность) отсекает сценарии, требующие подлёта в зону с нестабильной плитой, — ΔE превышает допустимый порог ΔE_{max};

Ветвь «Мы» (коллективная согласованность) блокирует сценарии, которые требуют ресурсов, зарезервированных для других секторов.

Из 12 сценариев остаются 2, прошедшие оба фильтра. Система схлопывает многообразие до $P = 2$ (правило $P \leq 3$, Глава 5).

t = 0:05 — Точка смысловой сверки (ТСС). Кибер-разведчик инициирует ТСС. Он не передаёт оператору сырые видеопотоки (это превысило бы лимит $C_{чел} \approx 10$ бит/с). Вместо этого формируется свёрнутый запрос $Q_{чел}$ (формула 13):

> «Сектор 7-D. Четверо пострадавших под плитой. Вариант А: рискованная эвакуация. Вероятность спасения 0,75. Вероятность обрушения плиты 0,4. Прогнозируемое падение *R* до 0,87 (кратковременный выход за нижнюю границу). ΔE = 0,68 от ΔE_{max}. Вариант Б: стабилизация на месте. Доставка кислорода и обезболивающего. Прибытие тяжёлой техники через 35 минут. *R* = 0,95 (стабильно). ΔE = 0,12 от ΔE_{max}. Подтвердите выбор.»

Оператор (человек) получает запрос через интерфейс, использующий сенсорные каналы с пропускной способностью, не превышающей 10 бит/с. Его фильтр осмысления оценивает этический допуск ΔE. Он выбирает Вариант Б — стабилизация на месте, так как риск для жизни спасателей в Варианте А неприемлем.

t = 0:08 — Второй фильтр (акт смысловой фиксации). Оператор отправляет команду «Б». Кибер получает акт смысловой фиксации единичного решения. Альтернативные варианты принудительно подавляются (латеральное торможение, Глава 3.2).

t = 0:15 — Начало исполнения. Наземный модуль класса «Хирург» прибывает в сектор 7-D. Начинается стабилизация пострадавших: подача кислорода, фиксация конечностей, мониторинг витальных функций. Значения *R* удерживается в коридоре 0,93–0,95.

t = 14:00 — Возникновение неопределённости (действие «на ощупь»). При доставке второго комплекта кислородного оборудования модуль сталкивается с неизвестным препятствием: арматурный прут, скрытый под слоем пыли, создаёт неучтённое пространственное ограничение. Ошибка предсказания ε_{pred} резко возрастает до 0,38 (порог $\varepsilon_{max} = 0,25$).

Система не пытается форсировать движение (что могло бы привести к повреждению оборудования или травме пострадавшего) и не останавливается полностью (что увеличило бы риск потери пострадавшего). Вместо этого кибер переходит в режим **действия «на ощупь»** (Глава 10.3):

- он совершает серию контролируемых микроперемещений манипулятора с шагом 2 мм, измеряя сопротивление в каждой точке;
- каждое микроперемещение — это пробное действие, а не финальное решение. Цель — не достичь целевого состояния, а получить информацию;
- информативность зондирования I_{probe} (формула 17) вычисляется в реальном времени: наиболее информативными оказываются движения вверх-влево, дающие прирост определённости о геометрии препятствия.

t = 14:30 — Обновление памяти ситуации. После серии из 15 зондирующих движений система обновляет M_{sit} (формула 18). Ошибка предсказания ε_{pred} снижается с 0,38 до 0,09. Модель препятствия уточнена, найден безопасный путь обхода. Кибер завершает доставку кислорода. R возвращается к 0,94. $L(t)$ снижается с $0{,}78 \cdot L_{crit}$ до $0{,}45 \cdot L_{crit}$.

t = 35:00 — Завершение миссии. Прибывает тяжёлая техника. Пострадавшие эвакуированы. Система фиксирует успешное завершение и передаёт отчёт оператору: все ключевые параметры (R, $L(t)$, ΔE) оставались в допустимых границах на всём протяжении операции.

Итог по Кейсу 1: механизм ТСС предотвратил неоправданный риск, действие «на ощупь» позволило справиться с неопределённостью без участия человека, а двухфильтровая модель гарантировала, что ни один из этапов не вышел за рамки этических допусков.

Кейс 2: Автономная логистическая сеть и предотвращение кибернетической деградации

Контекст. Гетерогенная система управления поставками для сети гипермаркетов. Уровень субъектности узлов разный: центральный узел (класс «Архитектор») — полный кибер-Субъект, региональные узлы (класс «Диспетчер») — киберы со средней субъектностью, локальные дроны (класс «Курьер») — автоматы с самоуправлением, но без способности к структурной адаптации. Внезапно происходит резкое изменение рыночной среды: перекрытие логистического коридора и введение государственного регулирования цен на топливо.

Исходные параметры системы:

$R = 0{,}98$ (штатный режим);

$L(t) = 0{,}3 \cdot L_{crit}$;

$T_{pred} = 4$ часа;

Коэффициент случайных возмущений $\mu = 0{,}05$ (низкий).

Развёртка событий

t = 0:00 — Внешний шок. Перекрыт ключевой коридор. μ скачком повышается с 0,05 до 0,45. Система получает лавину противоречивых сенсорных данных: одни дроны сообщают о блокпостах, другие — о свободном проезде (несинхронизированные данные).

t = 0:15 — Мониторинг нагрузки. Центральный узел «Архитектор» вычисляет:

$V_{изм}$ (скорость поступления изменений) выросла в 6 раз;

баланс разнообразия R начинает падать: 0,98 → 0,92 → 0,86 за 15 секунд;

аллостатическая нагрузка $L(t)$ растёт: $0{,}3 \cdot L_{crit} \rightarrow 0{,}7 \cdot L_{crit}$.

$$dL/dt = \gamma \cdot V_{изм} \cdot (1 - R) - \delta \cdot L(t) \qquad (10)$$

При $R < 0{,}9$ второе слагаемое перестаёт компенсировать первое.

t = 0:30 — Предупреждение. R достигает значения 0,84. Это ниже нижней границы безопасного коридора ($R_{min} = 0{,}9$). Система активирует превентивную сигнатуру.

t = 0:45 — Инициация мета-сверки (вместо хаотичной адаптации). В классической системе началась бы ложная эмерджентность: каждый узел начал бы оптимизировать свою локальную целевую функцию, игнорируя глобальный контекст. В автокибернетической архитектуре «Архитектор» вместо этого **инициирует мета-сверку** (Глава 11.2):

- система принудительно останавливает генерацию новых тактических решений;

«Архитектор» переходит в «тихий режим» (Глава 7.3): фиксирует снимок M_{sit}, сохраняет текущие значения R, $L(t)$, активные резервы;

- формируется запрос на структурное изменение — не на уровне действий, а на уровне самой инвариантной структуры S_{inv}.

t = 1:30 — Блокировка эгоистических стратегий ветвью «Мы». Некоторые региональные узлы («Диспетчеры») генерируют предложения о перераспределении ресурсов в свою пользу (накопление топлива, перенаправление дронов на свои маршруты). Эти предложения формально повышают локальный R каждого узла, но снижают глобальный R сети.

Ветвь «Мы» (коллективная согласованность на уровне «Коллективной Личности Я») оценивает эти предложения. Их ΔE (этический допуск) превышает ΔE_{max}, так как нарушается принцип «союз разного темпа» — узлы начинают конкурировать вместо кооперации. Ветвь «Мы» блокирует все эгоистические сценарии до того, как они достигают высшего узла управления.

t = 2:00 — Стратегическое отстранение. Центральный узел «Архитектор» делает вывод: при текущих инвариантах S_{inv} ни один из доступных вариантов не позволяет выполнить миссию без нарушения этических допусков (срыв контрактов с медицинскими учреждениями, перераспределение ресурсов в ущерб критическим узлам).

Система **не форсирует события** и **не впадает в спячку**. Она переходит в режим **стратегического отстранения** (Глава 8.4):

- все активные операции приостанавливаются. Дроны возвращаются на базы или переходят в режим зависания с минимальным энергопотреблением;
- система формирует запрос человеку (оператору логистической сети) через ТСС высшего уровня:

> «Рыночная среда изменилась необратимо. Текущие инварианты S_{inv} ведут к хроническому падению R ниже 0,85 и росту $L(t)$ выше L_{crit}. Предлагаю пересмотреть весовые коэффициенты целевой функции. Приоритеты: 1 — выполнение медицинских контрактов, 2 — коммерческая эффективность, 3 — сохранение ресурсов. Подтвердите новую калибровку.»

- система консервирует активы (топливо, запасные части, свободные дроны) до получения ответа от человека.

t = 15:00 — Вмешательство человека и структурная адаптация. Оператор (человек) получает запрос через защищённый канал. Его фильтр осмысления анализирует предложение. Он подтверждает новый приоритет: медицинские контракты становятся абсолютным приоритетом (вес $w_{medical} = 10$).

«Архитектор» получает акт смысловой фиксации структурного изменения. Происходит **перезапись фрагмента S_{inv}** (Глава 11.2). Коэффициент жёсткости к временно снижается (с 0,98 до 0,65), затем возвращается к норме. С новыми весами система пересчитывает допустимые варианты. Находится решение: медицинские поставки идут по воздуху (дроны), коммерческие — замораживаются до стабилизации коридора.

t = +3 дня — Выход из кризиса. R восстанавливается до 0,96. $L(t)$ снижается до $0{,}4 \cdot L_{crit}$. Система вернулась в управляемое состояние. Кибернетическая деградация предотвращена.

Итог по Кейсу 2: превентивный мониторинг $L(t)$ и R позволил инициировать мета-сверку до того, как система скатилась в хаос. Ветвь «Мы» блокировала деструктивные эгоистические стратегии. Стратегическое отстранение сохранило ресурсы и доверие, а структурная адаптация через вмешательство человека обеспечила выживание системы в новой среде. Кибернетическая деградация (потеря субъектности) не наступила.

Сводная таблица параметров по кейсам

Параметр	**Кейс 1 (исходно)**	**Кейс 1 (в кризисе)**	**Кейс 2 (исходно)**	**Кейс 2 (в кризисе)**
R	0,97	0,87	0,98	0,84
L(t)/L_crit	0,4	0,78	0,3	0,7 → 0,45
T_pred	45 с	8 с	4 ч	схлопнут
μ	средний	0,45	0,05	0,45
P	3	2	3	1
ΔE/ΔE_max	0,12	0,12	0,08	0,68

Примечание. Все численные значения приведены для иллюстрации работы механизмов и могут варьироваться в зависимости от конкретной реализации и условий среды.