Бауманский учебный центр
Специалист.ru
Мировые стандарты обучения

Человек в контуре ИИ – как выявить реальную потребность и внедрить ИИ помощников

Данил Динцис

Динцис Данил Юрьевич



- Общий опыт в ИТ отрасли 30 лет
- Управление проектами, программами и портфелями ~20 лет
  Значимые клиенты: Сургутнефтегаз, ХоумКредит, АШАН, Альфабанк, Сименс-Нокиа телеком, Пепсико СНГ, ОКТЕЛЛ, Мясницкий ряд
- Опыт вывода на рынок ИТ продуктов (международный и российский) в компании, которая начинала как ИТ стартап;
- Ментор стартапов
- Доктор техн. наук по системному анализу информационных систем
  Профессиональные сертификации:
  PfMP, PgMP, PMP, DevOps Product Owner,
  ITIL Managing Professional, Strategic Leader, Expert
- DASA DevOps и ScrumStudy Product Owner, Accredited PO trainer
  ScrumStudy Agile Master, Trainer
- Св. 30 публикаций по тематике курса. Академический редактор книги «CRM и машинное обучение»

ИИ усиливает то, что есть

В 3 раза

быстрее возникают сбои

Автоматизация на основе ИИ, внедренная в хаотичные процессы, выходит из строя в 3 раза быстрее, чем ручные операции

67 %

неудачных проектов в сфере ИИ

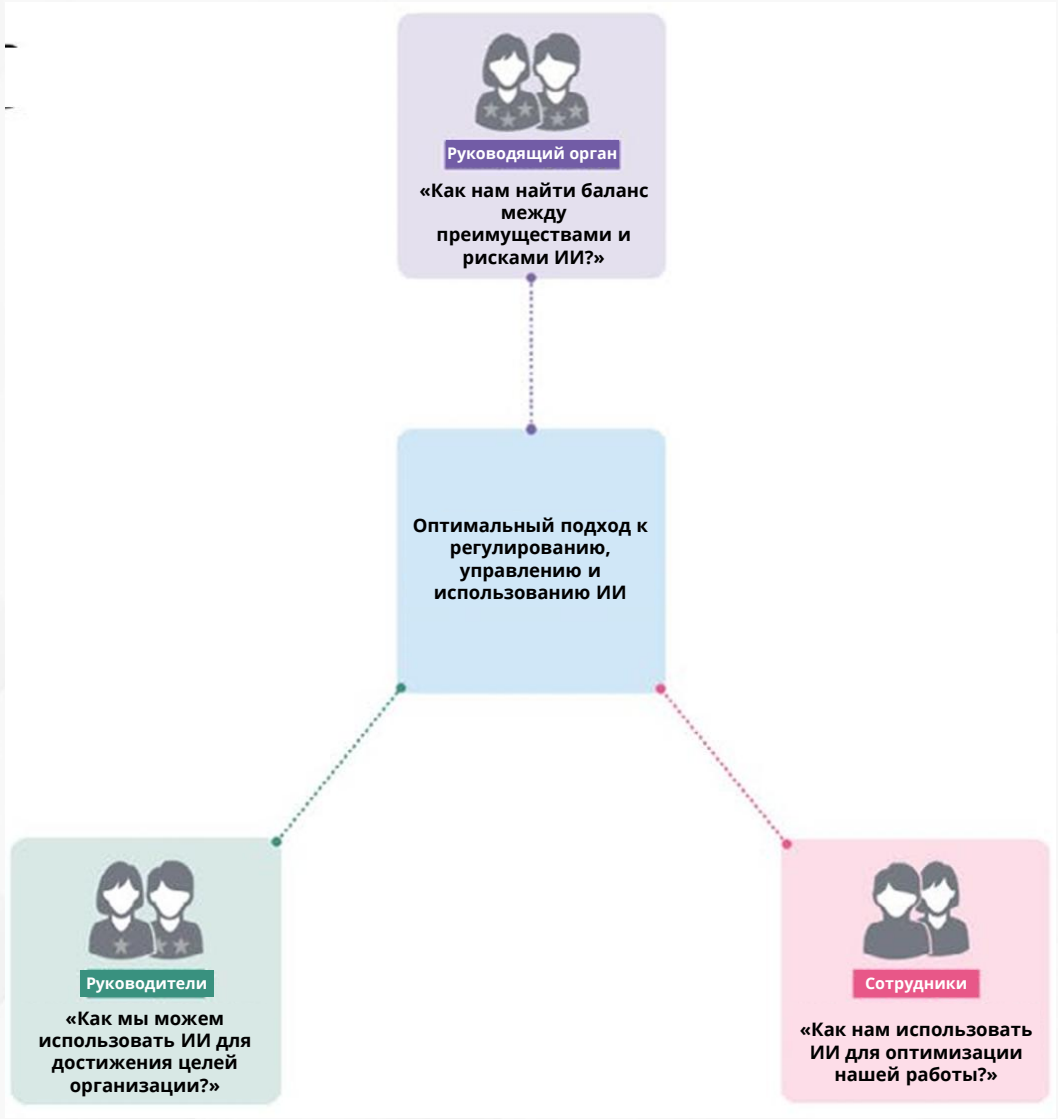
инициатив в области ИИ/автоматизации заканчиваются неудачей в первый год — главная причина — низкое качество данных

40 %

отклонение запросов

возможный показатель отклонения запросов с помощью ИИ — но только при наличии надлежащей базы данных

Кто вовлечен



Понимание типового спроса

Аспект «Объем»

КОЛИЧЕСТВЕННЫЙ

Какие типы запросов/инцидентов/тикетов встречаются чаще всего? Используйте данные за последние 90 дней. Отсортируйте по количеству тикетов. Топ-10 — это кандидаты для автоматизации.

Аспект «Проблемы»

КАЧЕСТВЕННЫЙ

Какие типы вызывают наибольшие задержки, эскалации или разочарование операторов? Проведите опрос среди операторов и проанализируйте данные о нарушениях SLA по категориям.

Аспект «Повторяемость»

ФИЛЬТР

Какие типы достаточно предсказуемы, чтобы их можно было стандартизировать? Если способы решения сильно различаются, автоматизируйте маршрутизацию, а не само решение.

Типы ИИ как сотрудника

Модель	Что это означает	Примеры использования
ИИ как инструмент	Люди непосредственно управляют ИИ; ИИ не проявляет инициативы.	Инструмент проверки грамматики или переводчик, используемый по запросу.
ИИ как помощник	ИИ поддерживает людей, которые сохраняют полную власть при принятии решений.	«Второй пилот» составляет черновики писем, подводит итоги встреч или предлагает дальнейшие шаги.
ИИ как коллега	ИИ работает в рамках согласованных границ вместе с людьми и может выполнять отдельные этапы рабочего процесса.	ИИ с агентными способностями управляет заявками в службу поддержки, собирает информацию, составляет черновики ответов и эскалирует исключительные ситуации.
Персонал, использующий ИИ	Несколько ИИ-помощников, агентов, автоматизированных процессов и систем с поддержкой ИИ работают в различных функциях и рабочих процессах.	ИИ поддерживает отбор персонала, обслуживание клиентов, финансовое прогнозирование, сортировку угроз кибербезопасности и оперативный мониторинг.

Что **меняется,** когда **ИИ** становится **коллегой**:

- Ему нужна чётко определённая **роль**; за что он отвечает?
- Ему нужны **границы**; какие решения он может принимать самостоятельно?
- Ему нужен **контроль**; кто и как проверяет его работу?
- Он генерирует **результаты**, которые затрагивают реальных людей; кто несёт ответственность?
- Он может давать **сбои или вести себя непредсказуемо**; каков порядок эскалации?

ИИ может выглядеть как коллега в рабочем процессе, но ответственность по-прежнему лежит на организации и ответственных лицах.

Пример AIOps (ИИОпс)

Модель	Что это означает	Пример AIOps
ИИ, позиционируемый как инструмент	Люди непосредственно управляют ИИ; ИИ не проявляет инициативы.	Аналитик по эксплуатации открывает панель управления AIOps, чтобы просмотреть взаимосвязанные оповещения, паттерны аномалий или показатели работоспособности сервисов.
ИИ, позиционируемый как помощник	ИИ поддерживает людей, которые сохраняют полную власть при принятии решений.	AIOps рекомендует вероятную причину инцидента, предлагает меры по устранению или составляет черновик отчета об инциденте, но решение о том, что делать, принимает человек.
ИИ, позиционируемый как коллега	ИИ работает в рамках согласованных границ вместе с людьми и может выполнять отдельные этапы рабочего процесса.	Агентный AIOps обнаруживает повторяющийся паттерн инцидента, открывает заявку, назначает приоритет, запускает утвержденный скрипт устранения, уведомляет владельца сервиса и эскалирует проблему, если восстановление не удается.
AIOps становится частью персонала, использующего ИИ, когда подобные агенты работают в рамках нескольких функций, рабочих процессов, поставщиков и областей принятия решений.		

Вопрос управления заключается не в том, полезен ли AIOps, **а в том, какой автономией и полномочиями он обладает, какие ограничения применяются, когда необходимо вмешательство человека, какие доказательства сохраняются и кто несет ответственность за операционные результаты.**

Пример проектного управления

Модель	Что это означает	Пример управления проектами
ИИ, позиционируемый как инструмент	Люди непосредственно управляют ИИ; ИИ не проявляет инициативы.	Руководитель проекта использует ИИ по мере необходимости для составления отчёта с основными выводами, подготовки сводки материалов для совета директоров или переформатирования плана.
ИИ, позиционируемый как помощник	ИИ поддерживает людей, которые сохраняют полную власть при принятии решений.	ИИ оценивает экономическое обоснование, моделирует варианты ресурсов и выгод по связанным проектам, а также выявляет проблемы с графиком и риски, но решение принимает руководитель.
ИИ, позиционируемый как коллега	ИИ работает в рамках согласованных границ вместе с людьми и может выполнять отдельные этапы рабочего процесса.	ИИ отслеживает ход работ с учётом допустимых отклонений, обновляет реестр рисков, составляет отчёты об исключениях и передаёт вопросы на рассмотрение уполномоченному лицу для принятия решений на этапе или в рамках программы.
Проектный ИИ становится частью персонала, использующего ИИ, когда аналогичные агенты координируют работу в рамках проектов, программ и портфелей, согласовывая выгоды, зависимости и ресурсы сразу по многим инициативам.		

Вопрос управления заключается не в том, помогает ли ИИ в реализации проектов, **а в том, какой автономией он обладает в пределах согласованных допустимых отклонений, когда должны вмешиваться ответственные лица, какие доказательства сохраняются для принятия решений на этапах и по выгодам, и кто несет ответственность за бизнес-обоснование и результаты.**

Управленческий разрыв

Управленческий разрыв

У большинства организаций есть **политики**. У немногих налажено **управление**.

- Наличие политики в области ИИ — это не то же самое, что наличие системы управления ИИ

Что мы понимаем под управлением?

- Управление подразумевает руководство, контроль, подотчетность и гарантии; это не просто набор правил

Где находится ваша организация на этом спектре?

Повышение зрелости системы управления →



Примеры ошибочного поведения ИИ

Цена неконтролируемого ИИ может быть **огромной** (как для организаций, так и для специалистов). Три примера показывают, как отсутствие надлежащего управления может превратить преимущества ИИ в угрозы для организации.

Управление персоналом

ИИ-система для подбора персонала, обученная на исторических данных о найме, систематически ставила определённых соискателей в невыгодное положение.

Что пошло не так?

- предвзятые обучающие данные
- отсутствие проверки предвзятости
- ограниченный контроль человека

Последствия для бизнеса

- иски о дискриминации
- расследование со стороны регулирующих органов
- ущерб репутации

Финансы

ИИ-помощник сгенерировал неверный финансовый анализ, который повлиял на инвестиционные решения.

Что пошло не так?

- низкое качество данных
- модель не прошла валидацию
- неадекватный мониторинг

Последствия для бизнеса

- финансовые убытки
- неудачные инвестиционные решения
- потеря доверия стейкхолдеров

Правоохранительные органы

Модель прогнозируемого полицейского контроля усилила исторические модели полицейской деятельности вместо того, чтобы снизить уровень преступности.

Что пошло не так?

- неясная система подотчётности
- ограниченная прозрачность
- отсутствие оценки воздействия

Последствия для бизнеса

- недоверие со стороны общественности
- юридические проблемы
- отмена политики

10 факторов управления ИИ

1. Подход к изменениям
2. Показатели успеха
3. Требования к интеграции
4. Толерантность к рискам
5. Стабильность внешней среды
6. Скорость принятия решений
7. Требования к соответствию
8. Разнообразие стейкхолдеров
9. Сфера применения системы управления
10. Возможности системы управления



6 компонентов системы управления ИИ

Компонент	Описание
Направление	Определяет и регулярно корректирует стратегическое направление использования ИИ для обеспечения его соответствия целям и задачам организации, созданию ценности и меняющимся условиям ведения бизнеса.
Политика	Определяет решения, ожидания, ограничения и принципы работы, которые определяют порядок использования, управления и контроля ИИ во всей организации.
Ответственность	Обеспечивает, чтобы руководящий орган оставался подотчетным за использование ИИ, демонстрируя соблюдение требований, прозрачность, надежность и соответствие ожиданиям заинтересованных сторон.
Делегирование	Распределяет полномочия и ответственность за использование ИИ через структуры управления, которые обеспечивают надлежащий надзор и подотчетность во всей организации и ее экосистеме.
Эффективность	Устанавливает и отслеживает ожидания и показатели эффективности, чтобы гарантировать, что ИИ способствует созданию ценности, сохраняет устойчивость и поддерживает совершенствование и адаптацию организации.
Возможности	Определяет, регулирует и управляет возможностями ИИ, необходимыми для поддержки и реализации продуктов, услуг, процессов и стратегических целей организации.

Типы ИИ с точки зрения управления

Существует несколько различных типов ИИ, каждый из которых предназначен для решения определенных задач. На практике эти типы часто пересекаются и взаимодействуют в рамках одной системы.



Узкоспециализированный ИИ / ИИ для конкретных задач
ИИ, предназначенный для выполнения конкретной задачи или набора связанных задач в определённом контексте.

Генеративный ИИ (GenAI)
Системы ИИ, способные создавать новый контент, такой как текст, изображения, код или другие результаты, на основе шаблонов, выученных из существующих данных.

Агентный ИИ
Системы ИИ, предназначенные для действий от имени пользователей путем автономного выполнения заранее определённых задач в согласованных пределах, при этом оставаясь под контролем и управлением человека.

Различные типы ИИ поддерживают разные задачи, модели поведения и потребности в управлении.

Типы ИИ с точки зрения управления

Правильный выбор ИИ поможет вам понять:

- среду **рисков**
- **ограничения** каждого подхода к ИИ
- необходимый **уровень** контроля со стороны человека

Узкоспециализированный ИИ

предназначен для решения конкретной задачи в определённом контексте.
Пример: направление запроса в нужную команду.

Генеративный ИИ

создаёт новый контент на основе закономерностей, выявленных из существующих данных.
Пример: составление черновика письма клиенту.

Агентный ИИ

действует от имени пользователей, автономно выполняя заранее определенные задачи в согласованных пределах, оставаясь при этом под контролем человека.
Пример: полное оформление записи на обслуживание.

Чем выше степень автономности или влияния, тем более тщательный контроль требуется.

Существует несколько различных типов ИИ, каждый из которых подходит для выполнения определенных задач. В таблице 2.1 представлены распространенные типы ИИ.

Тепловая карта ИИ рисков

Аспекты управления	Риски, связанные с ИИ	Создание	Курирование	Разъяснение	Познание	Коммуникация	Координация
Автономность ИИ и принятие решений	Автономность в принятии решений	Низкий	Низкий	Низкий	Средний	Средний	Высокий
	Прозрачность и объясняемость	Средний	Средний	Средний	Высокий	Средний	Высокий
Этичный и ответственный ИИ	Риски предвзятости и справедливости	Высокий	Средний	Высокий	Высокий	Высокий	Средний
	Галлюцинации и ошибки в генерируемых результатах	Высокий	Низкий	Средний	Средний	Высокий	Средний
Качество и безопасность данных ИИ	Несоответствие контроля доступа между исходными данными и результатами ИИ	Высокий	Средний	Средний	Средний	Высокий	Низкий
	Доступность и качество данных для обучения ИИ	Высокий	Высокий	Средний	Высокий	Средний	Высокий
Соблюдение нормативных требований и операционные практики	Правовые, нормативные риски и риски несоблюдения требований	Высокий	Средний	Средний	Высокий	Высокий	Высокий
	Операционные риски и риски жизненного цикла	Средний	Средний	Средний	Высокий	Средний	Высокий
	Риски управления ИИ в расширенной цепочке поставок и со стороны третьих лиц	Средний	Средний	Средний	Средний	Средний	Средний
	Организационная готовность и навыки	Средний	Низкий	Низкий	Средний	Средний	Средний

Пример типового риска: «Теневой ИИ»

- «Теневой» AI-AI используется без надзора, контроля и одобрения.
- Это приводит к пробелам в контроле и повышает организационные риски.
- Это искажает процесс принятия решений из-за неуправляемых и недостоверных результатов.
- Контроль и соразмерное управление позволяют вернуть ситуацию в нужное русло.



Выработка мер реагирования. Человек в цикле



Типовые меры реагирования

Риски, связанные с ИИ	Пример возможных мер контроля
Автономность в принятии решений	• Разработать политики, предусматривающие контроль с участием человека (human-in-the-loop) или постоянное участие человека (human-on-the-loop).
Прозрачность и объясняемость	• Определить ожидания в отношении объясняемости на основе профиля рисков системы и требовать наличие описаний моделей, подробно излагающих их назначение, ограничения и подход к интерпретации.
Риски, связанные с предвзятостью и справедливостью	• Проводить аудит предвзятости в отношении обучающих данных и моделей ИИ.
Галлюцинации и ошибки в сгенерированных результатах.	• Установить процессы проверки и валидации критически важных результатов ИИ силами человека.
Несоответствие в контроле доступа между исходными данными и результатами работы ИИ	• Избегать применения моделей ИИ к особо конфиденциальным данным, если не внедрены эффективные механизмы обеспечения безопасности.
Доступность и качество данных для обучения ИИ	• Мониторить производительность моделей на наличие признаков проблем, связанных с данными.
Правовые, нормативные риски и риски несоблюдения требований	• Интегрировать нормативные требования в процесс проектирования, закупки и развертывания систем ИИ.
Операционные риски и риски, связанные с жизненным циклом	• Разработать надежные процедуры обновления, переобучения и вывода моделей из эксплуатации.
Риски, связанные с управлением ИИ в расширенной цепочке поставок и со стороны третьих лиц	• Проводить комплексную проверку практик управления ИИ у поставщиков.
Организационная готовность и навыки	• Обеспечить обучение по вопросам грамотности в области ИИ и этики.

Факторы, повышающие доверие

Надежные знания + релевантный контекст + утвержденные действия = эффективное самообслуживание с помощью ИИ

Почему знания имеют значение

Обоснованные ответы
ИИ отвечает, используя утвержденные статьи, известные ошибки, сервисную информацию и руководства по эксплуатации.

Релевантные рекомендации
Намерение пользователя сопоставляется с подходящей статьей, этапом диагностики, запросом или действием.

Безопасное самообслуживание
Пользователи получают контролируемые инструкции, а не сгенерированные или неутвержденные способы устранения неполадок.

Доверие и принятие
Актуальный, понятный и эффективный контент побуждает пользователей возвращаться к самообслуживанию.

Как ИИ укрепляет базу знаний

Создание
Создание черновых статей и кратких резюме простым языком на основе проверенных результатов устранения неполадок.

Кураторство
Выявление дубликатов, устаревшего, противоречивого или пересекающегося контента.

Оценка качества
Выявление низких оценок, неудачных поисковых запросов, повторных эскалаций и статей с низким уровнем решения проблем.

Улучшение и поддержка
Предложения по более чёткой формулировке, улучшению тегов, добавлению недостающего контента, датам пересмотра и статьям, подлежащим удалению.

Человеческий контроль, поддержка ИИ: ИИ составляет черновики, извлекает информацию и выявляет проблемы; владельцы знаний проверяют, утверждают, оценивают и удаляют контент.

Оценка готовности к внедрению ИИ

Качество данных	Ясность рабочего процесса	База знаний	Ответственность	Контроль со стороны человека
❑ Категории ≤ 25 и структурированные	❑ Документированы 10 наиболее распространённых типов запросов	❑ Статьи охватывают 20 наиболее распространённых проблем	❑ Назначен ответственный за автоматизацию	❑ Определены условия отмены
❑ Приоритет: единообразная матрица 2×2	❑ Процесс соблюдается последовательно	❑ У каждой статьи указан ответственный	❑ Назначен ответственный за базу знаний	❑ Ведётся журнал аудита
❑ Коды закрытия: охват > 80 %	❑ Исключения документируются	❑ Установлен и соблюдается график проверок	❑ Назначен ответственный за эскалацию	❑ План отката протестирован

0–5 выполненных критериев: НЕ ГОТОВО. 6–10 выполненных критериев: ЧАСТИЧНО ГОТОВО. 11–15 выполненных критериев: ГОТОВО К АВТОМАТИЗАЦИИ

Выбор процессов – кандидатов на внедрение ИИ

Пошаговый метод отбора
Используйте его перед запуском любого пилотного проекта с ИИ — это позволит избежать выбора задач исключительно по объему.

1. **Исключите задачи с высоким риском**
   Серьезные инциденты, доступ к конфиденциальной информации, сбои в работе магазинов
2. **Оцените каждый тип**
   Объем, повторяемость, качество данных, ясность процесса, низкий риск
3. **Распределите задачи по категориям**
   Автоматизировать сейчас / С помощью ИИ / Исправить в первую очередь / Под контролем человека
4. **Определите границы**
   Что ИИ может делать, чего не должен делать и когда следует остановиться
5. **Выберите 2–3 пилотных проекта**
   Небольшой масштаб, измеримый исходный уровень, четкая возможность отката

Пример результата: наиболее востребованные типы запросов, преобразованные в решения ИИ
Пример для службы поддержки в сфере розничной торговли. Первый пилотный проект должен быть безопасным, повторяемым, контролируемым и легко откатываемым.

Тип запроса	Решение	Почему
Сброс пароля	Автоматизировать сейчас	Большой объём, простые шаги, низкий операционный риск
Разблокировка учётной записи	С механизмами контроля	Повторяемо, но требует проверки личности и безопасности
Запрос ПО	Автоматический рабочий процесс	Стандартный путь утверждения и чёткие шаги выполнения
Проблема с входом в POS	Сначала сортировка по приоритету	Большой объём, но первопричины варьируются в зависимости от магазина/системы
Доступ для нового сотрудника	Исправить сначала	Путь утверждения и ответственность варьируются в зависимости от отдела
Проблема с сетью магазина	Под руководством человека	Серьёзные последствия для бизнеса в случае неправильной классификации или задержки

Простые критерии оценки по шкале от 1 до 5: Объём | Повторяемость | Качество данных | Ясность процесса | Низкий риск

Первые кандидаты должны набирать высокие баллы по всем пяти критериям, а не только по объёму заявок.

Автоматизируйте предсказуемую работу с низким риском. Усиливайте поддержку при выполнении сложных задач. Оставляйте за людьми право принимать решения, требующие глубокого анализа.

Человек в цикле ИИ.
Когда решение обязательно за человеком

Автоматические триггеры для человека

БЕЗ ИСКЛЮЧЕНИЙ

- Инцидент P1/P2 → обязательное уведомление руководителя
- Новый тип проблемы без соответствующей категории → очередь аналитика
- Уровень уверенности ИИ < 85% → сортировка человеком
- Любое действие, затрагивающее безопасность данных → проверка службой безопасности

Протокол отмены решения

ПОЛНОМОЧИЯ АГЕНТА

- Любой агент может отменить любое решение ИИ в любое время
- Отмена решения регистрируется в журнале — штрафные санкции не применяются
- Шаблоны отмены решений еженедельно анализируются руководителем отдела автоматизации
- Частые отмены решения инициируют пересмотр правил в течение 48 часов

Аудит и подотчетность

УПРАВЛЕНИЕ

- Каждое действие ИИ сопровождается аудиторским следом
- Кто/что принял решение, с какой степенью достоверности, в какое время
- Ежемесячный отчет по аудиту руководителю службы поддержки
- ITIL: требования к аудиту управления услугами выполнены

Непрерывная калибровка

УЛУЧШЕНИЕ

- Еженедельно: частота отмены решений, % ошибочных маршрутизаций, точность ИИ
- Ежемесячно: производительность модели по сравнению с базовым показателем
- Ежеквартально: обязательное переобучение или перекалибровка
- Ежегодно: полный обзор системы управления с участием руководства

Разберем детали на курсах

- Организация работы и управления ИИ и большими данными (Big Data) в современной компании
- ИИ для руководителей: от инструментов к стратегии