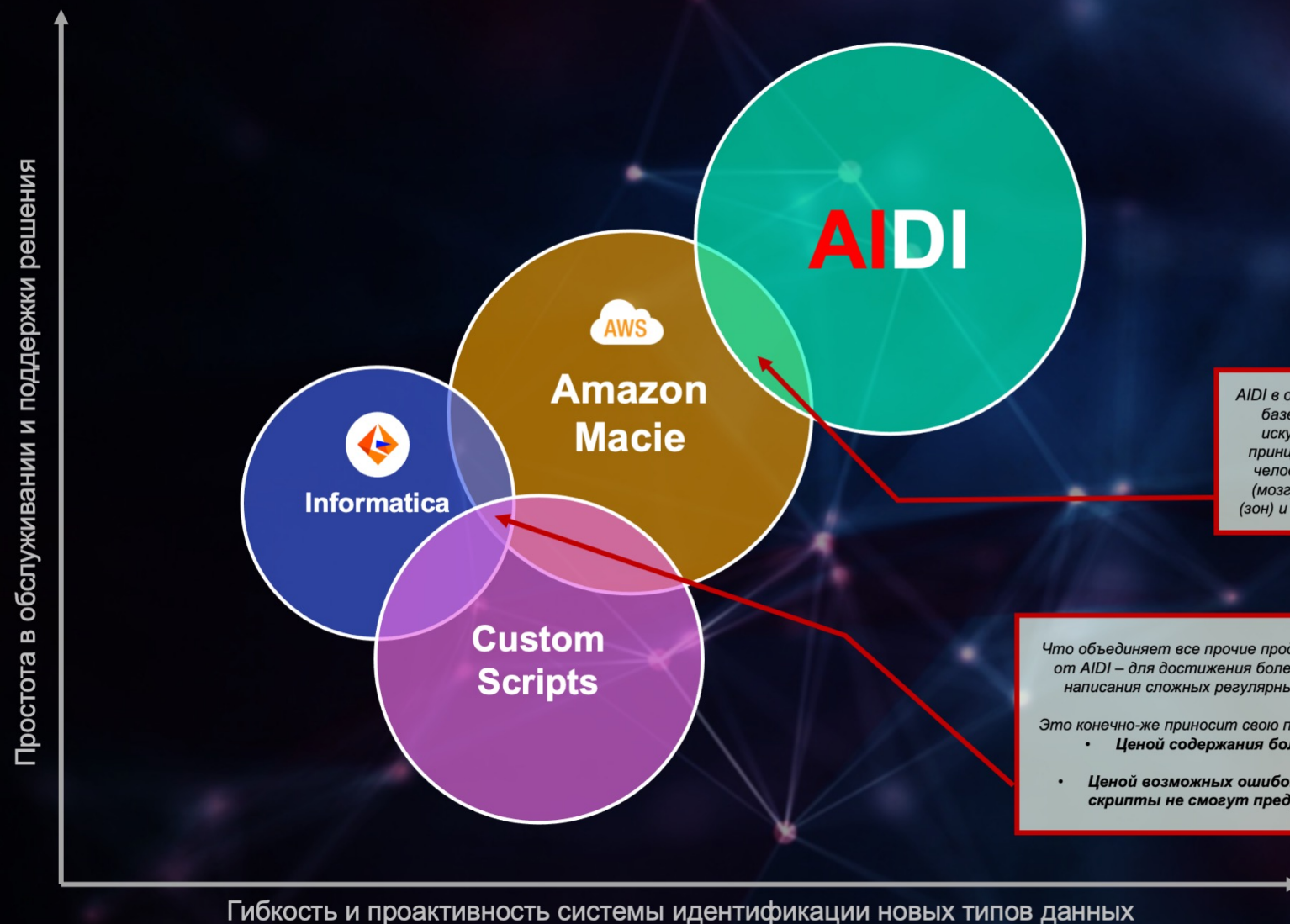




Think Like a human,
see Like a human ...

Artificial Intelligence Data Identification

Анализ рынка продуктов в сфере идентификации и классификации данных



AIDI в отличие от Macie полностью функционирует на базе ряда AI ядер, связанных между собой общей искусственной нейронной сетью. Таким образом принцип работы AIDI очень сильно похож на работу человеческого мозга – глобальная нейронная сеть (мозг) объединяет множество более мелких сетей (зон) и формирует целостные и взвешенные решения.

Что объединяет все прочие продукты по идентификации данных, и в то же самое время отличает их от AIDI – для достижения более-менее качественного результата они требуют от пользователя написания сложных регулярных выражений, либо элементов программного кода для определения нестандартных вариаций данных.

Это конечно-же приносит свою пользу – продукты так или иначе решают свои задачи, но какой ценой?

- Ценой содержания больших команд высококвалифицированных и дорогостоящих специалистов
- Ценой возможных ошибок в случае, если созданные вручную регулярные выражения или скрипты не смогут предугадать внезапно обнаруженные сложные конструкции данных



Artificial Intelligence Data Identification

AIDI - прогрессивная система идентификации типов данных в основе которой лежит машинное обучение.

По умолчанию AIDI обучена идентификации базовых типов данных (города, номера телефонов, адреса электронной почты и так далее), однако по мере использования ее вы сами того не замечая будете обучать ее узнавать все новые и новые типы данных, и при этом вам не потребуется быть программистом.

Ниже представлены некоторые возможности применения AIDI:

- Data Discovery и Data Identification баз данных и текстовых файлов
- Потокковая идентификация данных (REST Full)
- Выявление данных сложных структур и конструкций (блоки программного кода, структуры JSON, XML и так далее)

Какие типы данные могут быть идентифицированы в AIDI?

ЛЮБЫЕ!

Для вашего удобства, а также возможности протестировать работу AIDI «сразу из коробки» ее сознание будет обучено распознавать следующие типы данных:

Тип данных	Примеры данных
Имя	<i>Kevin; Mike; Alena; Ivan ...</i>
Фамилия	<i>Stivens; Sidorov; Watson ...</i>
Web адрес (URL)	<i>https://google.com/ ...</i>
Email	<i>example@domain.net ...</i>
Географический адрес	<i>221B Baker Street ...</i>
Телефонный номер	<i>+375-29-111-12-34; +1(22) 446 45 22 11 ...</i>
Страна	<i>Belarus; USA; China; PLN ...</i>
Штат	<i>Texas; Arkansas; California ...</i>
Город	<i>London; Warsaw; Oslo; NY ...</i>
Название компании (организации)	<i>Google Inc; Novartis Pharma ...</i>
Дата / Время	<i>2020.12.18 15:44, 2012.03.15</i>
Профессия	<i>Doctor; System Administrator ...</i>
Статусы процессов	<i>Completed; Planned; Waiting for support ...</i>
Валюты	<i>BYN; Rubl; USD; Dollar USA ...</i>
Временные интервалы и периодичности	<i>Days; weeks; years; monthly ...</i>

С каждым новым обновлением AIDI стандартный набор распознаваемых типов данных, как и качество идентификации данных будет расти.

Таким образом вы с лёгкостью сможете обучить AIDI распознаванию почти любых типов данных. К примеру она сможет справляться с идентификацией следующего содержимого в ячейках баз данных:

- Json и xml структуры
- Коды на языках программирования
- Языки веб разметки и стилей (HTML/CSS)
- Файлы изображений и прочих документов хранимых в базах данных



Идентификация данных и обучение AIDI

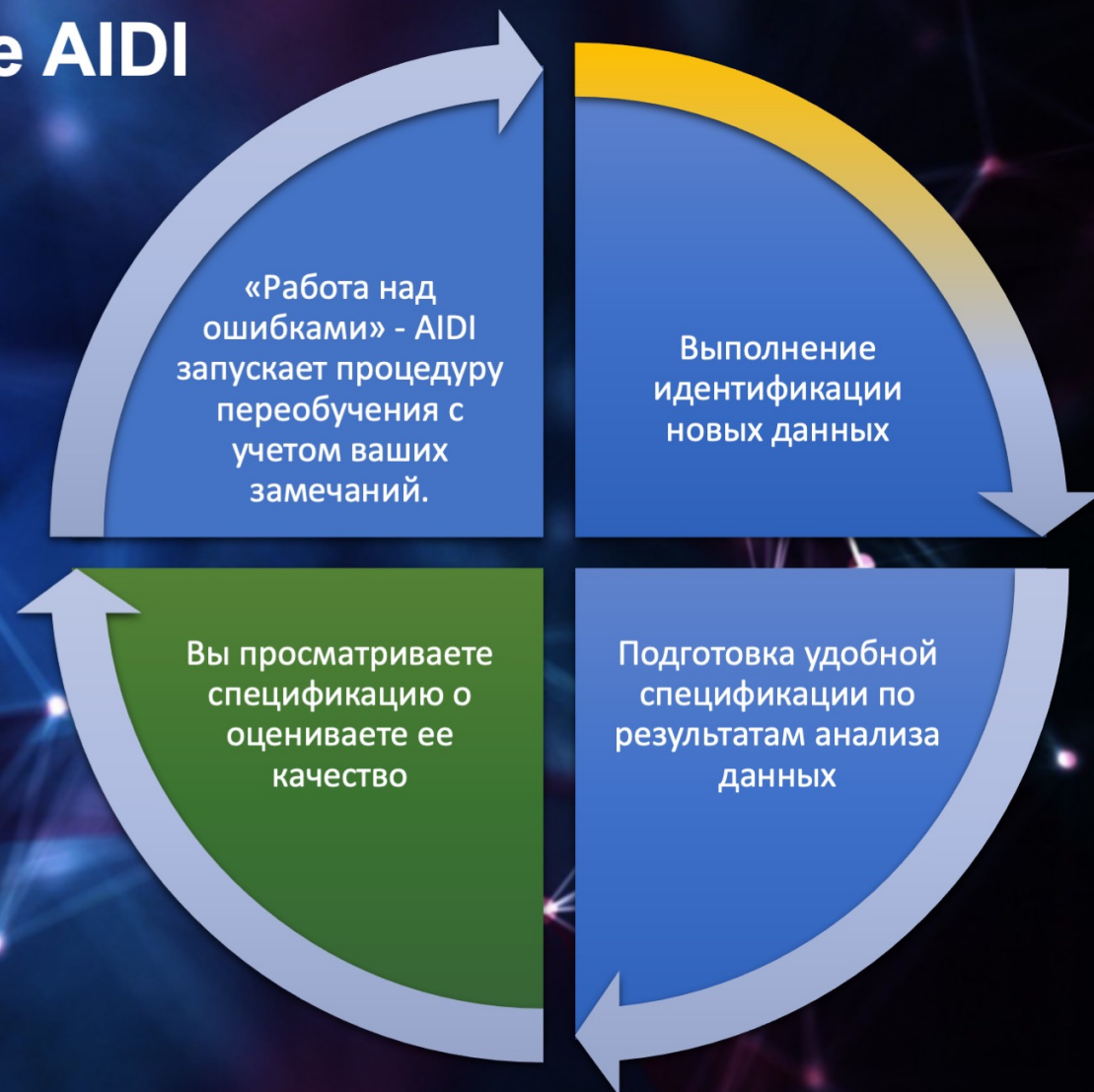
Процессы идентификации данных, а также дополнительного обучения вашей AIDI тесно связаны между собой. В процессе решения ваших рабочих задач AIDI будет корректировать свою работу, выполнять работы над ошибками и повышать свою квалификацию основываясь на ваших оценках ее работы.

Таким образом процесс повышения качества работы AIDI происходит постоянно.

Она самостоятельно отслеживает свои ошибки, стремится разобраться в них и повысить качество своей работы.

При этом ваше участие сводится лишь к удобной верификации качества ее работы в несколько кликов мышкой.

На следующих слайдах более подробно описан каждый из этих этапов.

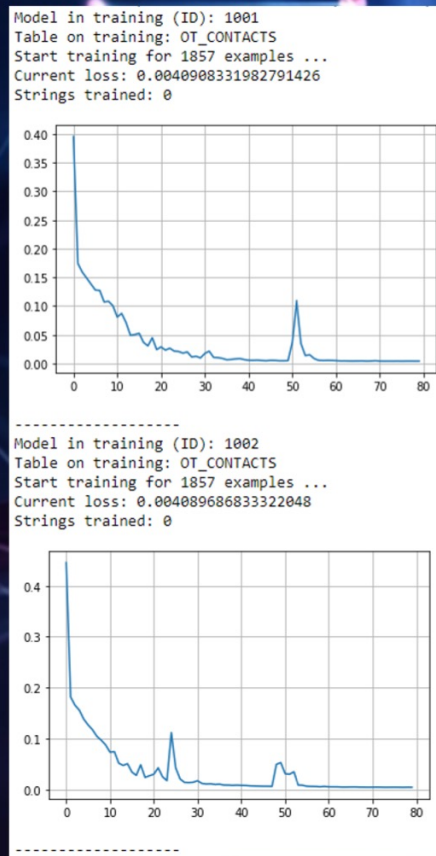


Идентификация данных и обучение AIDI

К процессу обучения, а также его результатам предъявляется ряд обязательных требований качества. Соблюдение этих требований отслеживается системой в автоматическом режиме. К примеру, к ним относятся:

- **Требование о разнообразии данных для обучения** – данные для обучения должны быть максимально уникальными, а также максимально отражать вариативность их представления
- **Требование об снижении вероятностных ошибок** – обучение продолжается до тех пор, пока ошибка идентификации новых данных системой не будет менее 0.001%

На рисунке представлен процесс усвоение нейронными связями новых интерпретаций уже известных типов данных. Как видно из графиков система активно усваивает новые данные на протяжении первых 30 повторений их изучения, а уже к 60 повторению вероятность возникновения ошибок в идентификации данных достигает приемлемых значений.



Накопленный опыт

- Этот опыт не позволил решить последнюю поставленную задачу на 100%
- Количество ошибок идентификации составило более 0,01%



Новые инструкции

- Новые типы данных
- Новые представления уже известных типов данных

Обновленный опыт задействуется в дальнейших итерациях идентификации данных и переобучения

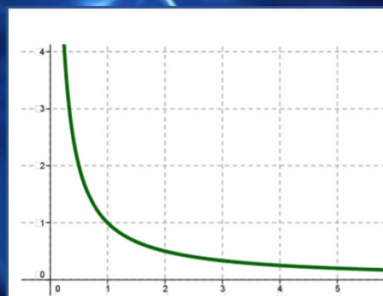
Обновленный опыт

- Успешно решает задачи все предыдущие задачи по идентификации данных
- Уменьшает вероятность возможных ошибок для будущих задач идентификаций

Идентификация данных и обучение AIDI

Column name	Data Examples	Identification Results
PO_CHANGED_ON	07/21/2020 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
DATE_ID	07/30/2013 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
REQUISITIONER	DERIDMA1	Наименование юридического лица 1 ☐ Login 521 (Novartis Pharma) 1 ☒
INVOICE_ENDORSER	DRUTIER1	Наименование юридического лица 1 ☐ Login 521 (Novartis Pharma) 1 ☒
PO_CREATED_BY	STROESO3	Наименование юридического лица 1 ☐ Login 521 (Novartis Pharma) 1 ☒
PO_VALID_FROM	06/28/2019 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
PO_VALID_TO	02/27/2021 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
WBS_ELEMENT_STATUS	OPEN	Наименование юридического лица 1 ☐ Activity Status 1 ☒ First Name 0.51 ☐
VENDOR_NAME	EUROFINS LANCASTER LABORATORIES INC	Наименование юридического лица 1 ☒ Profession 0.82 ☐
COUNTRY	FR	State 1 ☐ Country 1 ☒ Наименование юридического лица 1 ☐
COUNTRY_NAME	France	Country 1 ☒ Наименование юридического лица 1 ☐ Activity Status 0.69 ☐ City 0.59 ☐
FINAL_INVOICE	Closed	First Name 1 ☐ Наименование юридического лица 1 ☐ Activity Status 1 ☒ City 0.99 ☐
PO_DOCUMENT_DATE	11/27/2018 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
FIRST_INVOICE_DATE	07/09/2015 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
FIRST_GR_DATE	01/29/2020 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
LAST_INVOICE_DATE	08/28/2014 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
LAST_GR_DATE	01/26/2021 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
DOCUMENT_CURRENCY	GBP	Наименование юридического лица 1 ☐ Currency 1 ☒ Activity Status 0.57 ☐
LOCAL_CURRENCY	CHF	Наименование юридического лица 1 ☐ Currency 1 ☒ Activity Status 0.67 ☐
INVOICE_DATE	11/30/2009 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
BASE_DATE	01/11/2010 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
CLEARING_DATE	09/28/2018 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
NET_DUE_DATE	07/25/2020 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒

По результатам проведенной идентификации система формирует сводный отчет, в котором информирует пользователя (оператора) о том, к каким классификациям она предлагает отнести те или иные данные. Задача пользователя – выполнить валидацию этих данных и отметить галочкой верно идентифицированные данные.



При этом, процент вовлеченности оператора в контроль качества идентификации будет гиперболически сокращаться с каждой новой процедурой идентификации данных.

При этом уже на первых итерациях идентификации (как продемонстрировано на примере ниже) системе удается идентифицировать столь сложный смысловой объект данных, как название компании.

PO_VALID_FROM	06/28/2019 00:00:00	Telephone, Fax 1 ☒ Date / Time 1 ☒
PO_VALID_TO	02/27/2021 00:00:00	Telephone, Fax 1 ☐ Date / Time 1 ☒
WBS_ELEMENT_STATUS	OPEN	Наименование юридического лица 1 ☐ Activity Status 1 ☒ First Name 0.51 ☐
VENDOR_NAME	EUROFINS LANCASTER LABORATORIES INC	Наименование юридического лица 1 ☒ Profession 0.82 ☐
COUNTRY	FR	State 1 ☐ Country 1 ☒ Наименование юридического лица 1 ☐ Country 1 ☒

Итак, почему AIDI – лучший выбор

1. Неограниченные возможности в идентификации данных
2. Понятный каждому и удобный процесс работы с системой
3. Процесс обучения и повышения качества работы интегрирован в процесс идентификации данных, что существенно ускоряет процесс обучения AIDI
4. Интеллектуальное обучение AIDI позволяет ей самостоятельно адаптироваться под ваши потребности в идентификации данных
5. Нет необходимости в найме высококвалифицированных сотрудников для поддержки инфраструктуры
6. Стоимость продукта на порядок ниже чем у аналогов
7. Успешно проведенные тестирования POC на реальных задачах показали, что *неподготовленное ядро AIDI способна классифицировать данные с точностью более чем в 85%. В свою очередь обученное ядро классифицирует данные с предельной ошибкой менее 0.001% (то есть точность классификации данных составляет более 99,999%).* Это значительно превосходит показатели качества многих других конкурентов.
8. Обширный потребительский рынок – в данном решении уже сейчас заинтересован ряд крупных компаний, уделяющих особое внимание сохранности собственных данных. Все что необходимо – воплотить концептуальное решение в продаваемый продукт.

CERTIFICATE

This certificate confirms that the software based on Artificial Intelligence and Machine Learning “**AIDI**” (Artificial Intelligence Data Identification) has been functionally tested in conditions as close as possible to Production.

The ability of this software to identify various types of data (including complex data types of non-standard structure) was demonstrated and confirmed during this functionality testing process.

Date of PoC: **29 April 2021**

Date of presentation for Novartis
ISRM Data Masking Team: **3 May 2021**

The author, developer and copyright
owner of this software :

Yauhen A. Panimatchanka

Panimatchanka Yauhen
Digitally signed by Panimatchanka Yauhen
DN: dc=com, dc=novartis, ou=external-gold, serialNumber=2416165, cn=Panimatchanka Yauhen
Reason: I am approving this document
Date: 2021.05.04 16:14:57 +03'00'

SME Expert, NBS IT SRM

The validity of this document is confirmed by Novartis with digital signature certificates of signers.

Bradzetskaya Iryna
Digitally signed by Bradzetskaya Iryna
DN: dc=com, dc=novartis, ou=external-gold, serialNumber=2443309, cn=Bradzetskaya Iryna
Date: 2021.05.04 16:09:07 +03'00'

Solution Developer, NBS IT SRM

Software development was carried out in accordance with the current legislation of the Republic of Belarus, as well as international norms for the recognition of copyright:
- solely by the author of the software
- using the author's personal computing resources and in his spare time.

 **NOVARTIS**
PHARMACEUTICALS

 **ІНОВАЦЫІ**
PHARMACEUTICALS

За что вы платите используя другие продукты



По статистике, крупные компании (со штатом сотрудников более 2000 человек) тратят на решении данных задач от 1,5 до 2 миллионов долларов США в год.

Итого: от \$ 178.000 в год

Минимальные требования к инфраструктуре для осуществления анализа 100 гигабайт данных в месяц	
Количество специалистов поддержки инфраструктуры	1
Количество разработчиков техник идентификации данных	2
Количество операторов идентификации данных	2

Стоимость AIDI

Расходы и обязательства

В настоящее время мы активно ищем инвесторов и покупателей прав на дальнейшее распространение AIDI. Если вы заинтересованы в эффективном вложении денежных средств в этот продукт с возможностью получения в дальнейшем прибыли, пожалуйста ознакомьтесь с нашими требованиями, а также примером стратегий получения прибыли от распространения AIDI.

От \$ 25.000

(без НДС)

Приобретение прав на разработку и распространения продукта AIDI

Компенсация расходов понесенных на исследование требований, проектирование и разработку концепции, а также проведение РОС.

Ваши преимущества:

- Приобретая права на разработку и распространение конечного продукта вы получаете уникальный шанс к вам внимания со стороны крупных клиентов, а также возможность дальнейшего сотрудничества и развития данного проекта в соответствии с вашими бизнес потребностями.



На срок от 1 года

Заключения соглашения о дальнейшей разработке и развитии продукта с iTerra

В течении как минимум одного года процесс дальнейшей разработки продукта осуществляется сотрудниками iTerra. В зависимости от поставленных вами целей нами формируется необходимый штат сотрудников и ведется разработка решения. С вашей стороны, соответственно, осуществляется финансирование процесса разработки.

В чем выгода:

- Над уже вашим продуктом продолжают трудиться его основатели – те люди которые понимают процессы его функционирования в мелочах, а соответственно могут с максимальной продуктивностью и качеством вести его дальнейшее развитие.
- Мы как основатели продукта, и соответственно также являясь в дальнейшем совместно с вами выгода приобретателем максимально заинтересованы в том, чтобы продукт был качественным и крайне привлекательным для заказчиков. В связи с этим мы готовы приложить максимальные условия для нашей наиболее совместной выгоды от этой сделки.



10% от всех продаж

Заключение соглашения о распределении прибыли

Данное условие предусматривает уплату денежного вознаграждения в размере 10% от всех продаж лицензий продукта а также оказания услуг по сопровождению продукта третьим лицам.

Обоснование:

- Выплата процентов от суммы перечисленных заказчиками обуславливает низкую стоимость приобретения уникальных прав на разработку и распространения программного комплекса.
- Данный размер процентной ставки, учитывая потенциальный доход не создает существенных экономических нагрузок на приобретателя прав на разработку и распространения программного комплекса AIDI и с лёгкостью покрывается всеми доходами, полученными от продажи лицензий на данное программное обеспечение и оказания услуг по его сопровождению.
- Существование данной программы выплаты является огромным стимулом для iTerra и первоначальных авторов AIDI всячески способствовать ее развитию и повышению ее покупательской привлекательности.



**HOT
DEALS**

Стоимость AIDI

Рекомендуемые стратегии капитализации

На приведенном слайде продемонстрированы рекомендуемые цены на распространение AIDI конечным заказчикам. Данные цены на порядок (в разы) ниже чем у конкурирующих продуктов, дабы сформировать высокую привлекательность AIDI перед потенциальными заказчиками.



Лицензии на использование

Основной доход приносит продажа годовых лицензий на использование вычислительных единиц AIDI.

Пример: Одна лицензия активирует AIDI на одном физическом/виртуальном сервере и максимально задействует 4 ядра CPU и 16 GB RAM.



Рекомендуемая стоимость для новых клиентов:

\$ 25.000 за одну лицензию в год

Basic Offering



Техническая поддержка

Продажа услуг сопровождения AIDI на стороне заказчика, а также консультативной и практической помощи в решении ряда сопутствующих задач.

Пример: В рамках приобретенной лицензии заказчик получает бесплатных 8 часов технической поддержки. Эти часы он может потратить, к примеру, на то чтобы ему была установлена AIDI, либо были выполнены какие-либо обновления. При необходимости привлечения наших сотрудников к выполнению работ в инфраструктуре заказчика, оплата за время оказания этих услуг взимается дополнительно.



Рекомендуемая стоимость для новых клиентов:

От \$ 25 в час



Пакеты расширенного функционала

Подписки на расширенный функционал, приобретаая который заказчик получает дополнительные функциональные возможности. Помимо того, что это приносит дополнительную прибыль, также данный подход открывает гибкие возможности для дальнейшего развития продукта без каких-либо существенных воздействий на его архитектуру и бизнес-модель.

Пример: Разработка и последующая продажа решений по полному циклу маскирования SPI и PII данных в среде заказчика.



Рекомендуемая стоимость для новых клиентов:

\$ 10.000x в год за одну лицензию

Additional Offering

Ожидаемые поступления (в год) от одного клиента:

от \$ 50.000 до \$ 500.000

Остались вопросы?

Свяжитесь с нами и мы ответим на них!



info@ilterra.com



ilterra.com

Artificial Intelligence Data Identification