**Этап 1: Сбор агента и создание базы знаний**

**Цель:**

Собрать все необходимые документы по информационной безопасности, преобразовать их в текстовый формат (TXT), и загрузить в GPT-агент для обеспечения работы с нормативной базой.

**Шаги по реализации**

**1. Сбор документов**

1. **Источники для сбора:**
   * Официальные порталы:
     + [pravo.gov.ru](http://pravo.gov.ru) (законодательство РФ).
     + [duma.gov.ru](http://duma.gov.ru) (законопроекты и нормативные акты).
     + Сайты ФСТЭК, ФСБ, Роскомнадзора.
   * Профильные ресурсы:
     + Хабр, CISO Club (рекомендации, практика).
   * Внутренние регламенты организаций (при наличии).
2. **Перечень обязательных документов:**
   * Федеральные законы (ФЗ № 152, № 149, № 187 и т.д.).
   * Указы Президента РФ (например, Указ № 400 о Стратегии нацбезопасности).
   * Постановления Правительства.
   * ГОСТы и отраслевые стандарты.
   * Методические рекомендации регуляторов (ФСТЭК, ФСБ, Роскомнадзор).

**2. Преобразование в текстовый формат**

1. **Инструменты для конвертации PDF/Word в TXT:**
   * **Программы:**
     + Adobe Acrobat Pro: Экспорт PDF в TXT.
     + LibreOffice: Экспорт DOC/DOCX в текстовый формат.
     + ABBYY FineReader: Для сканированных документов.
   * **Python-скрипт для автоматизации:**

python

Копировать код

from PyPDF2 import PdfReader

def pdf\_to\_txt(input\_path, output\_path):

reader = PdfReader(input\_path)

with open(output\_path, "w", encoding="utf-8") as file:

for page in reader.pages:

file.write(page.extract\_text())

# Пример использования

pdf\_to\_txt("document.pdf", "document.txt")

1. **Очистка текста:**
   * Удаление ненужных элементов (водяные знаки, номера страниц).
   * Структурирование текста по разделам и статьям.

**3. Организация документов**

1. **Создание структуры папок:**

javascript

Копировать код

/knowledge\_base/

/laws/

/standards/

/internal\_policies/

/methodologies/

1. **Именование файлов:**
   * Используй понятные и уникальные названия, например:
     + FZ-152-personal-data.txt
     + GOST-R-57580-2022.txt
2. **Создание индекса документов:**
   * Составь файл (например, index.txt) с описанием каждого документа:

kotlin

Копировать код

FZ-152-personal-data.txt: Федеральный закон "О персональных данных"

GOST-R-57580-2022.txt: ГОСТ Р 57580-2022 "Требования к ИБ"

**4. Загрузка в GPT-агент**

1. **Проверка лимитов:**
   * Убедись, что объем текста в каждом документе не превышает ограничений GPT (например, 3 000 символов на фрагмент).
2. **Разбиение документов на части:**
   * Если документ превышает лимит, раздели его на логические блоки.
3. **Загрузка через интерфейс или API:**
   * **Через интерфейс:** Если GPT-агент поддерживает загрузку через панель управления, используй это.
   * **Через API:**

python

Копировать код

import openai

openai.api\_key = "your\_api\_key"

def upload\_document(text, document\_name):

response = openai.File.create(

file=text,

purpose="fine-tune"

)

print(f"Загружен документ: {document\_name}")

# Пример использования

with open("document.txt", "r", encoding="utf-8") as file:

upload\_document(file.read(), "FZ-152-personal-data")

**5. Тестирование и настройка**

1. **Проверка работы агента:**
   * Задавай тестовые вопросы, например:
     + "Какие требования устанавливает ФЗ № 152 к обработке персональных данных?"
     + "Какова ответственность за нарушение требований ГОСТ Р 57580-2022?"
2. **Обучение агента:**
   * Добавь примеры вопросов и ответов, чтобы улучшить понимание контекста.
3. **Обратная связь:**
   * Протестируй с реальными пользователями и собери обратную связь.

**6. Поддержка актуальности**

1. **Мониторинг изменений:**
   * Настрой регулярное сканирование правовых порталов на обновления.
2. **Автоматизация обновлений:**
   * Разработай скрипты для загрузки и замены устаревших документов.
3. **Рассылка уведомлений:**
   * Включи функцию уведомлений о новых или изменённых документах.

**Ожидаемый результат**

* Полностью собранная и структурированная база нормативных документов.
* GPT-агент, способный отвечать на вопросы по ИБ и предоставлять информацию из базы.
* Автоматизированный процесс обновления данных для обеспечения актуальности.

Готов продолжить с детализацией следующих этапов или помочь с конкретным кодом для автоматизации!