KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Kairntech Use Case: A Contract Analysis NLP Pipeline

Kairntech webinar, Feb 28, 2023

info@kairntech.com

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Introduction

- *NLP approaches have made huge progress on many use cases in the recent years*
- *Many previously hard to implement scenarios can now be addressed with impressive quality*
- *We focus today on how to combine various approaches to implement a contract analysis process*

Kairntech: “No code / Low code”

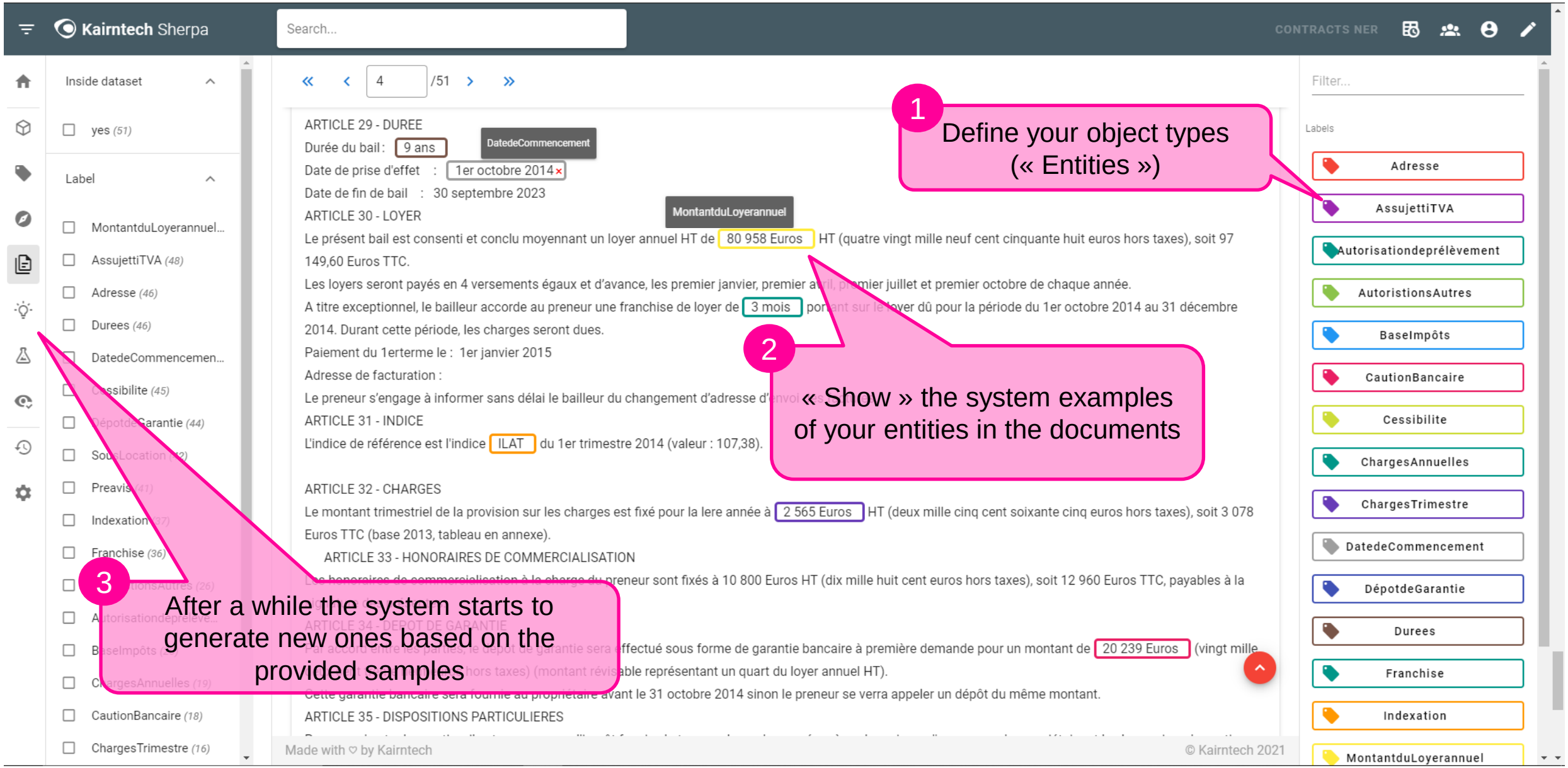
KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

No code / Low code:

Enable users to make use of NLP/AI without requiring them to be programmers

Adding your expertise

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Scenario today: Contract Analysis

- Due Diligence, Audit
- Tedious if done manually: Reviewing numerous complex/lengthy documents
- Time and cost vs Completeness!
- So: The scenario is:
  - We need to analyse Leasing Contracts
  - But these contracts are buried in a larger collection of other contracts
  - So we first need to find the Leasing Contracts in the larger set
  - Then extract key information from these

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Our task requires the combination of different approaches:

1. Document Categorization: Determine the type of a document
2. Entity Extraction: Which specific entities are mentioned in the text?
3. Combine individual analysis steps to a larger conditional processing pipeline

Contract Analysis Pipeline: Categorize, then extract if category is right

A contract analysis pipeline

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Import documents

Different unsorted contracts

Manually label documents: LeaseContract or not?

Train ML models

Train categorization model

Apply model on corpus → select only LeaseContracts

Manually label documents: Landlord, tenant, date, object?

Train entity extraction model

Combine models into Pipeline: Categorize → right category? → extract entities

Final pipeline, implementing the full process

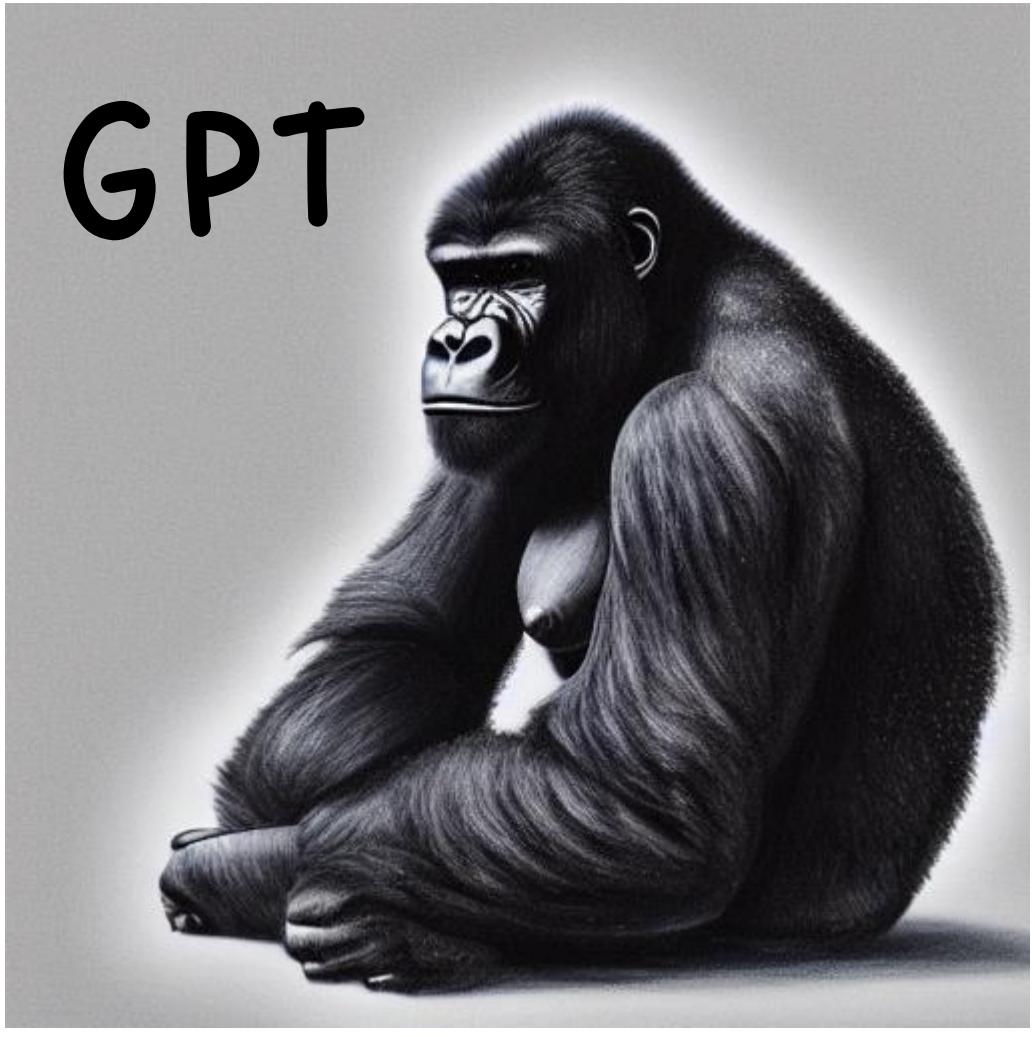
All inside the system, not requiring any coding

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Demo

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Whenever you talk about NLP these days:



KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

GPT – needs perhaps no introduction

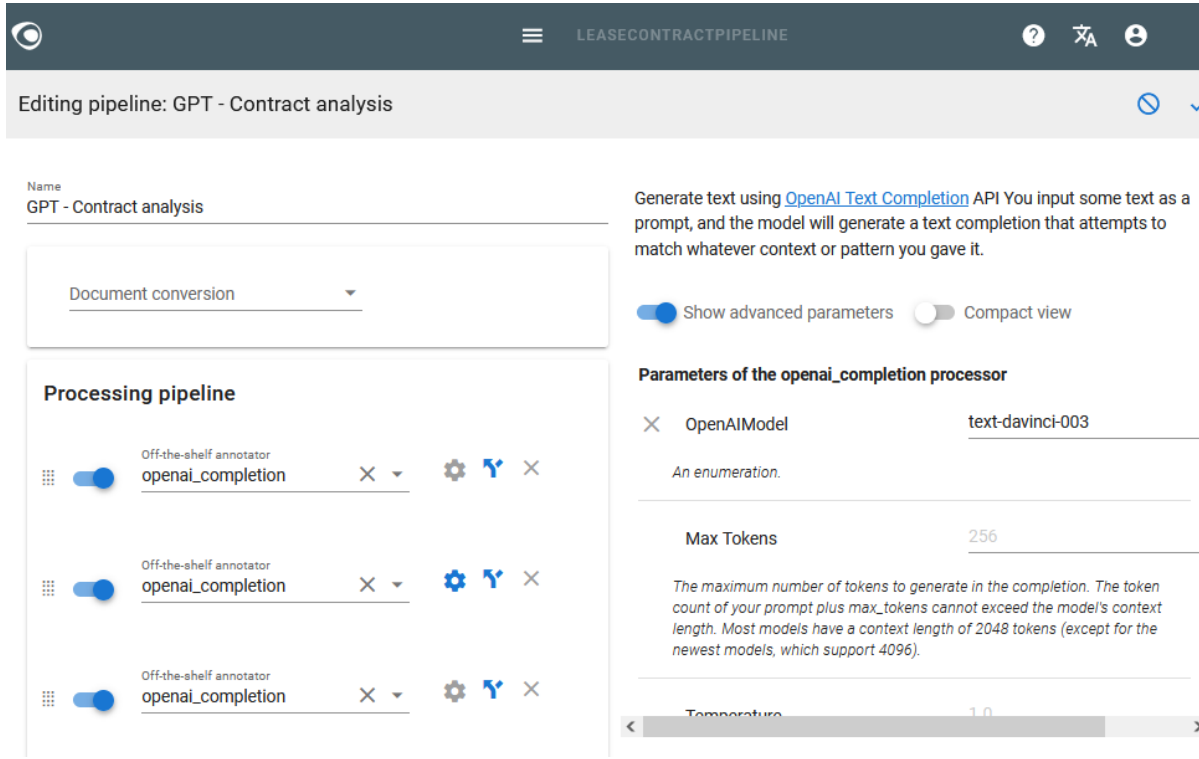
- Most popular example of a new class of NLP models ("LLMs – Large Language Models").
- Fastest growing piece of software ever: >100mio users after only eight weeks
- Impressive results on a wide range of use cases: Text generation, question&answer, conversations
- Proprietory (openAI – Microsoft), others in preparation (Google, Facebook, …)
- Apparent potential to disrupt many business models

- … potential for our use case here too (Contract Analysis)?

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Kairntech as a multi component platform

- Kairntech integrates wide variety of components and models
  - DeepL, Spacy, Flair, BERT …
- … allowing to combine them into custom NLP pipelines
- And of course we have done the same with GPT

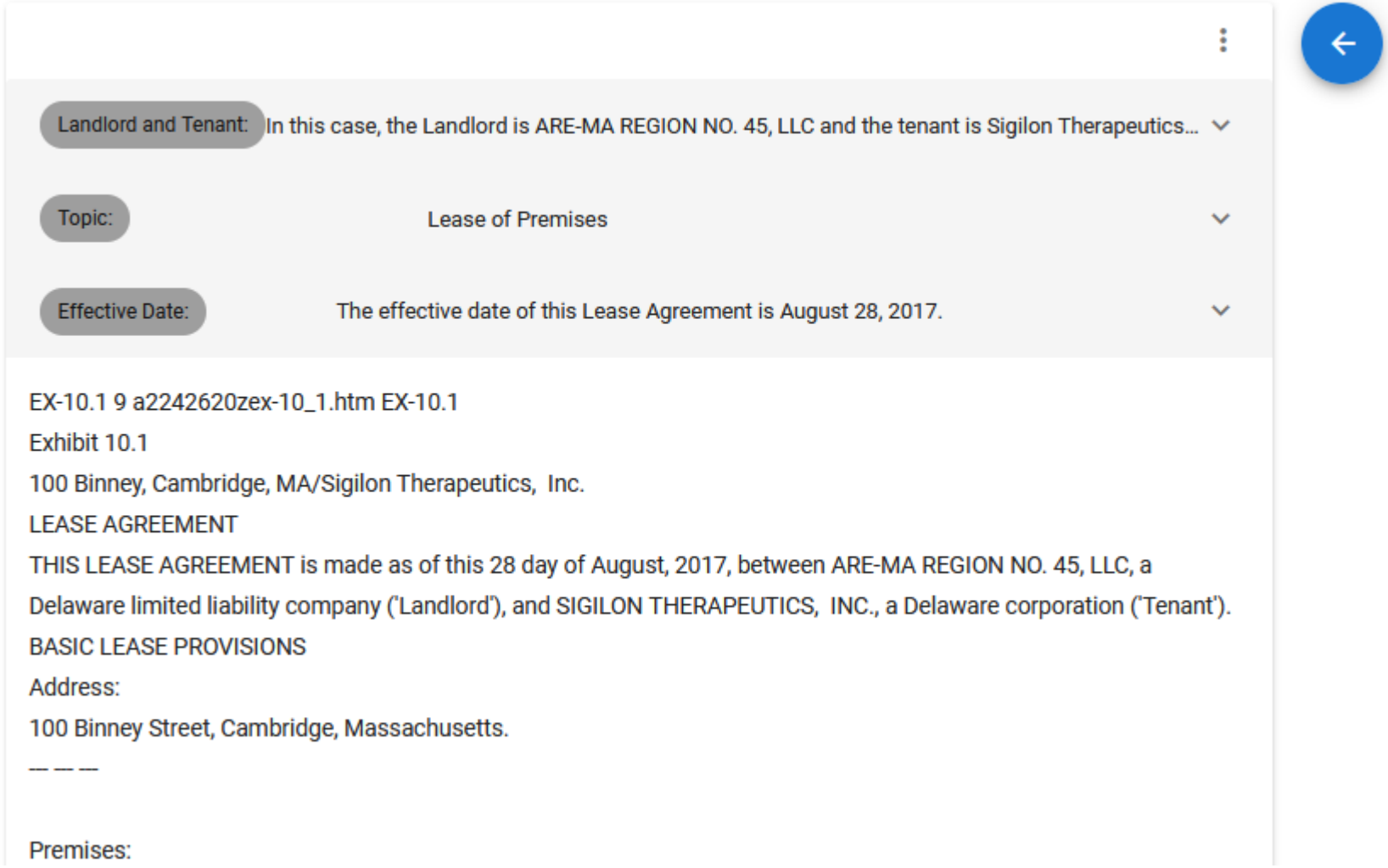


KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

"Prompt":

- "Who is the landlord an
- "What is the topic of the
- "What is the effective da

Sample result:



KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

“Nice results! So: Was that it? Can we all go home now?”

Perhaps not quite yet!

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

GPT without doubt has the potential to disrupt many use cases. To consider when thinking about deploying GPT-powered approaches:

- Cost?
  - Model “Davinci”: 0,02$/1000 tokens text.
  - Sample contract (50000 tokens) → 1$ per GPT analysis
  - Adapting (retraining) GPT-like models on your content / your vocabularies most often prohibitively complex → combine with other approaches.

- Confidentiality?
  - Currently GPT models can normally be run only on openAI infrastructure
  - Comparison with newer, smaller (open source?) models?

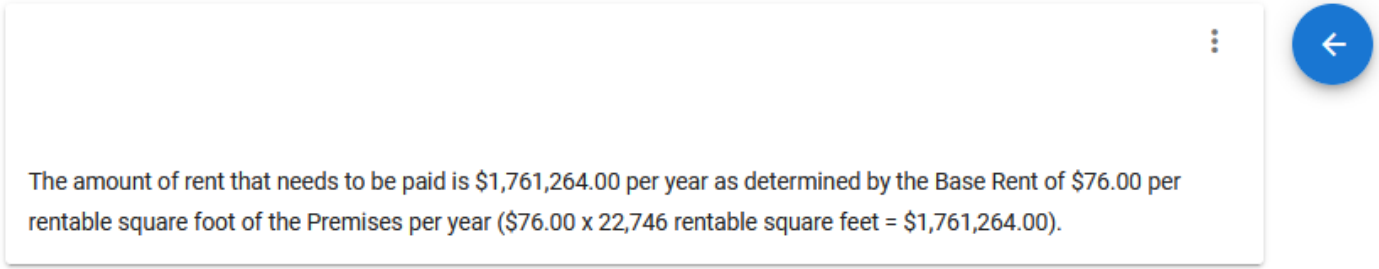
- Speed?
- Accuracy?
  - GPT-like models often seem to “hallucinate” (return unjustified answers)
  - Parameter tuning
  - “Prompt design” vs finetuning: When to favor what?

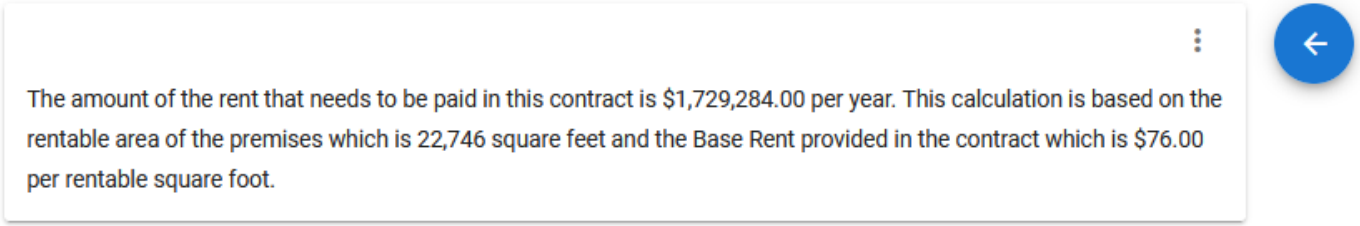
- Very dynamic landscape today: new models being published almost on a weekly basis.

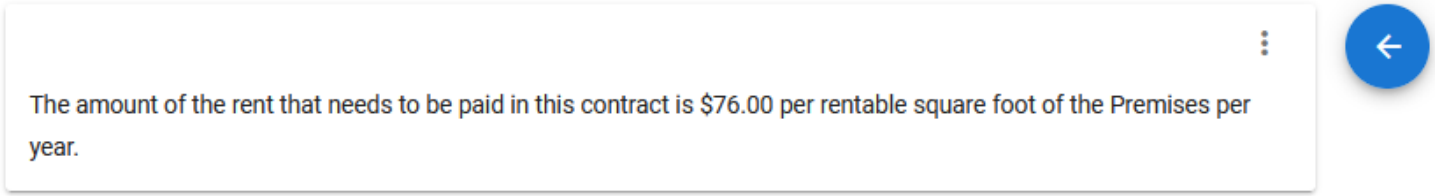
KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

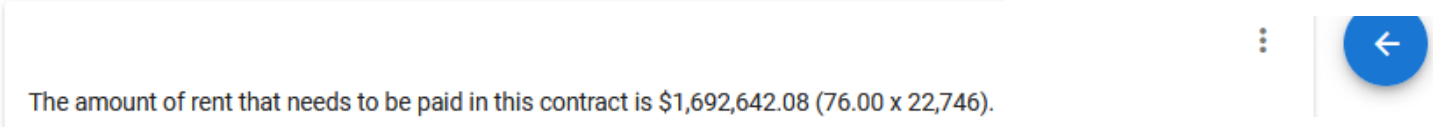
To consider(continued)

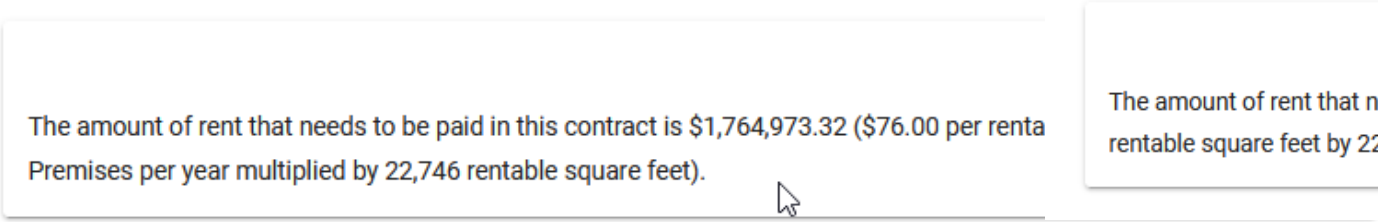
- Randomness
  - Default GPT returns different results on subsequent identical requests – even where no variation makes sense
  - (There are ways to reduce that: turn down “temperature”)
  - Example here: Asking GPT for the total yearly rent:

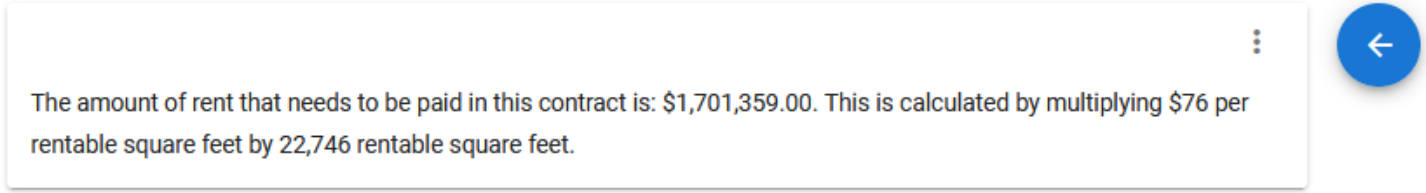












KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

To consider (continued):

- The entity types in our example (Landlord, tenant, date, address) are very common, GPT picked them up out of the box
- Often, Contract analysis is about more detailed and sophisticated concepts
- Adding these in a GPT-powered scenario often requires finetuning GPT models (with at least 200-300 examples)
- Use Kairntech for generating this finetuning data set (and using it to finetune GPT, where this makes sense)?

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

GPT (and comparable models) are major piece of technology, affecting many use cases. A fast succession of leaps can be expected in the months and years to come, further extending capabilities and impact.

It should be seen as a tool, to be assessed with its strengths and limitations in order to determine whether, where and how to best be employed on your use case.

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

And now?

- Learn more about Kairntech and what we do
  - Website with use cases, blog, product description: www.kairntech.com
  - Product video: https://kairntech.com/doc/video/
  - Documentation, FAQ: www.kairntech.com/doc
- Interested to try it out yourself? (no costs, no commitment)
  - Ask for a free trial www.kairntech.com
- Any suggestion, question, comment?
  - info@kairntech.com

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Thank you for your attention!

info@kairntech.com