



Wir helfen Unternehmen bei der Einbettung
moderner Text- und Sprachverarbeitung
innerhalb ihrer Produkte und Prozesse





We are Kern AI.



We are Kern AI. A company founded in 2020, starting at Hasso Plattner Institute.



We are Kern AI. A company founded in 2020, starting at Hasso Plattner Institute.
We focus on natural language processing for companies.



We are Kern AI. A company founded in 2020, starting at Hasso Plattner Institute.
We focus on natural language processing for companies.



Company pitch

We are Kern AI. A company founded in 2020, starting at Hasso Plattner Institute.
We focus on natural language processing for companies.



There is something special about this data.



Textual data is by design not meant for computers to be analyzed. It is unstructured.



Company pitch

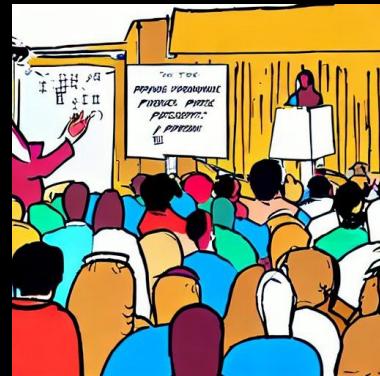
We are Kern AI. A company founded in 2020, starting at Hasso Plattner Institute.
We focus on natural language processing for companies.

VS.

Type	Value
Company name	Kern AI
Founding date	2020
Origin	Hasso Plattner Institute



Why is unstructured data (text) so important?



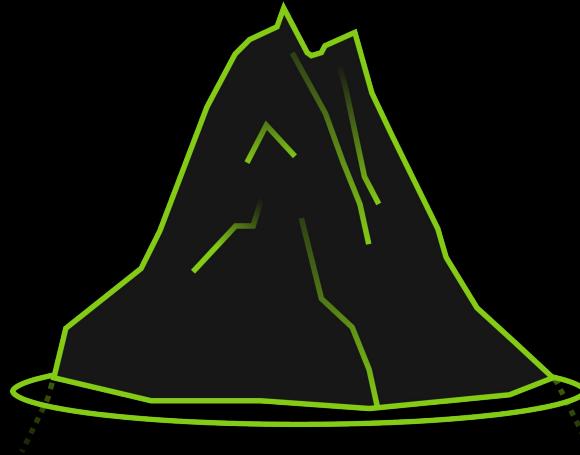
The strength of humankind is to articulate complex scenarios in natural language.

In other words: the strength of humankind is to process unstructured information.

all images were generated using Stable Diffusion by providing a textual description of the scenery.



Human interaction is via natural language. Businesses run on language.



PDFs

Emails

Websites / News

Circa 80% of business data is unstructured.



Why is unstructured data so **difficult** to process?



Because a computer understands 0 and 1.

 $127 + 4$

11111110

+

00100000

=

11000001

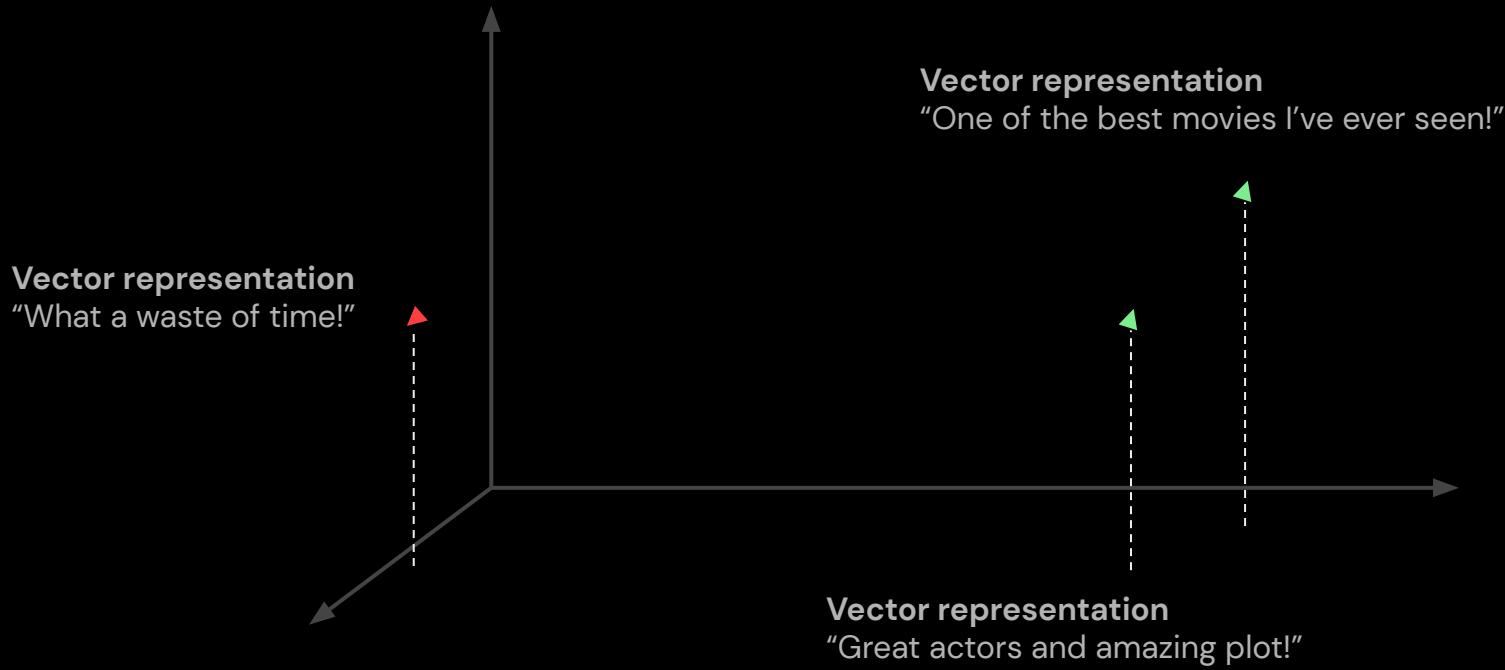

$$127 + 4 = 131$$

But how can I compare
"This summer is so lovely"
with
"My favourite time of the year is from June to August"?

But how can I compare
"Let's sit down at this lovely bank"
with
"I think we need to rob a bank"?



2017, a research team proposes a paper to algorithmically *learn* contextual dependencies of sentences.





In short, this changed everything in natural language processing.



In short, this changed everything in natural language processing.



Classifications



Extract



Search



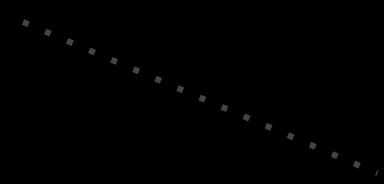
Generate



Company pitch

We are Kern AI. A company founded in 2020, starting at Hasso Plattner Institute.
We focus on natural language processing for companies.

[... Our platform helps developers embed NLP into products and processes.]



Company pitch

We are Kern NLP. A company founded in 2021, starting at TU Munich.
We focus on processing text for businesses.

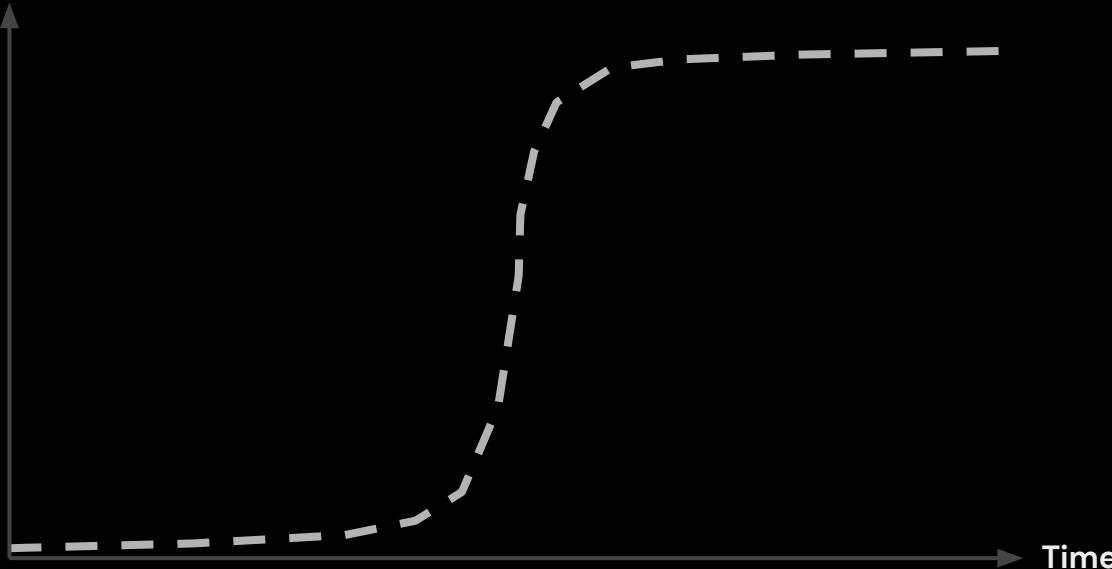


Why am I telling you about this?

What does this mean for you, and generally insurers?



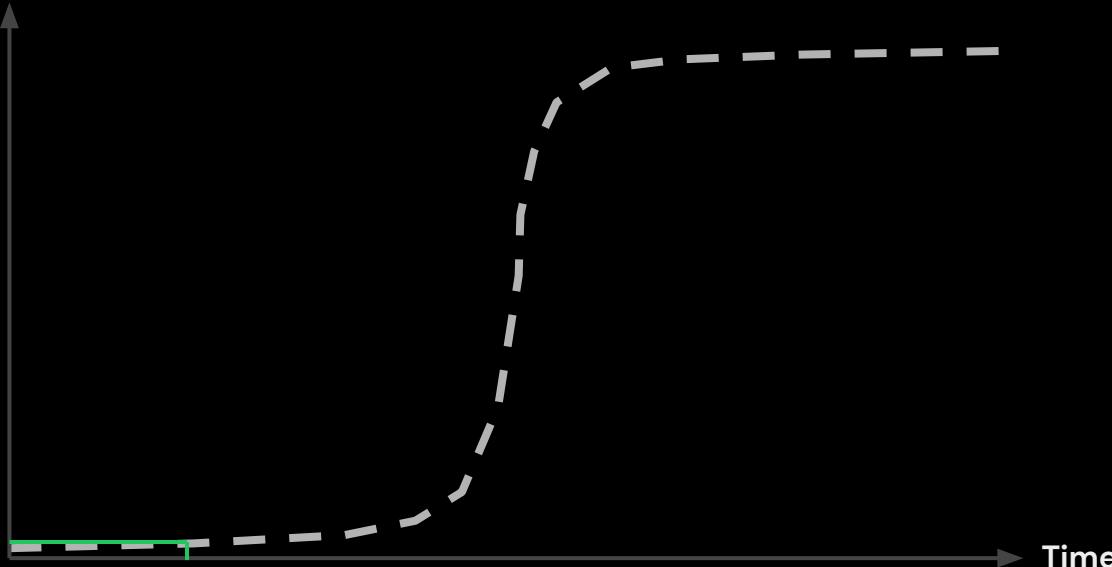
Adoption



Adoption of core technologies are shaped like a Sigmoid function.

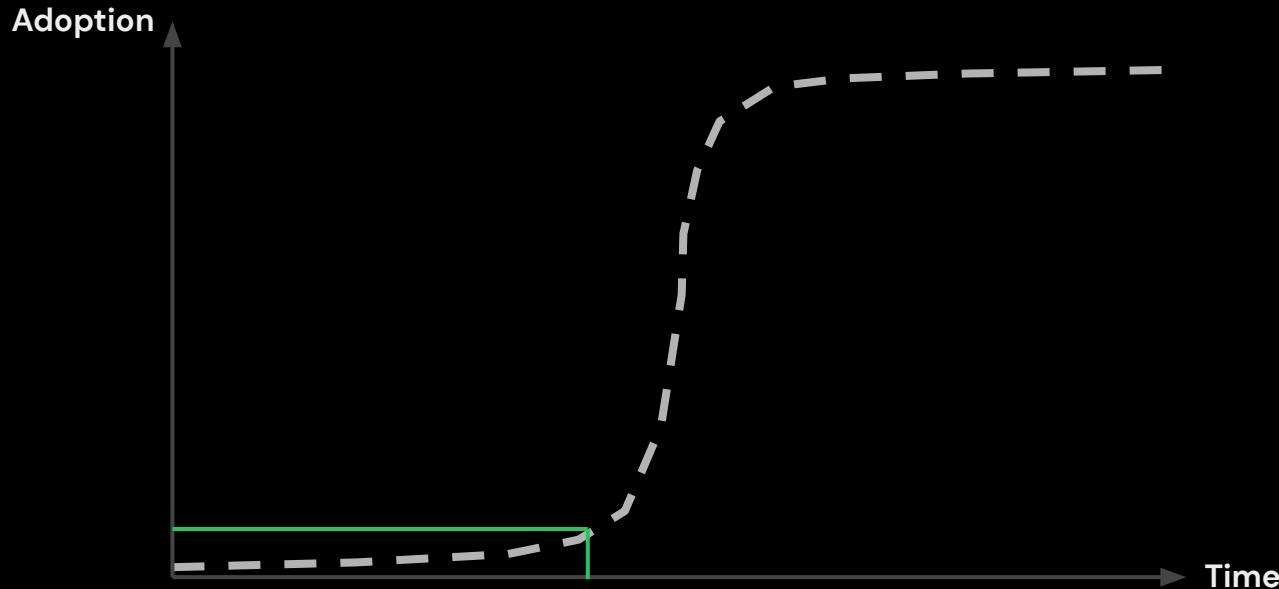


Adoption



Mostly researchers and technicians use the technology.

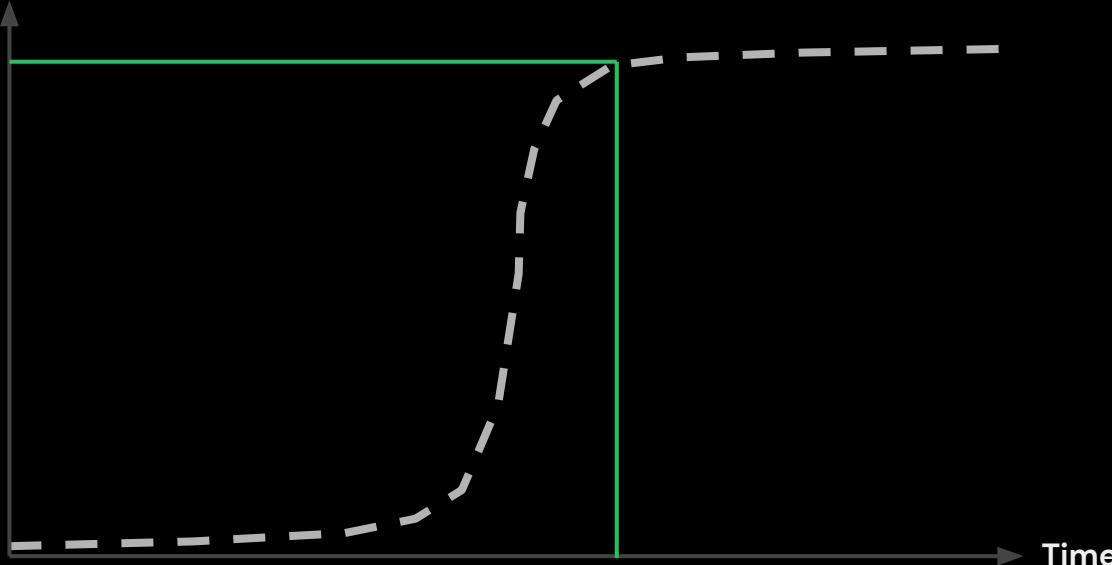
(2017: "Attention is all you need" paper describes new algorithmic architecture to learn context in data)



Mass awareness. People are beginning to realize the potential.
(2022: OpenAI releases ChatGPT. My barber tells me about how he used it.)



Adoption



The technology is widely used across almost any thinkable use case.
(2027: NLP is a core driver in digitizing business processes globally)



Businesses need to re-think which processes,
and which parts of processes, to (semi-)automate.



Three areas in which *not* using NLP harms your business.



Operating at scale

In processes requiring knowledge from experts about your company.

You are limited by a shortage of skilled labor, and need to find ways to improve operations.



Gaining crucial insights

To make informed decisions, you need to derive data to understand actual problems.

You are missing out on tons of insights from both the internet and your internal textual data.



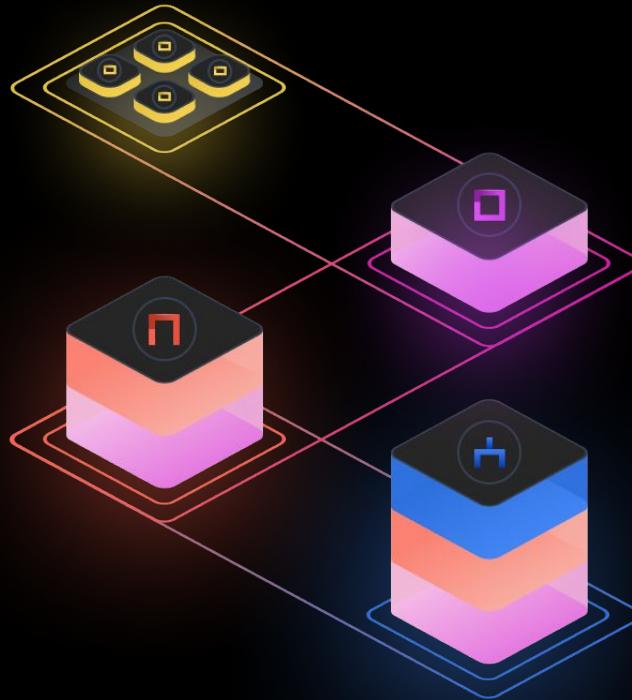
Building modern products

Without a doubt, modern products will use NLP at scale to provide new user interfaces.

Building such products requires training data of highest quality and documentation.



We offer the solution to turn your company into an NLP automation and analytics pioneer!



From idea to prototype within a day.

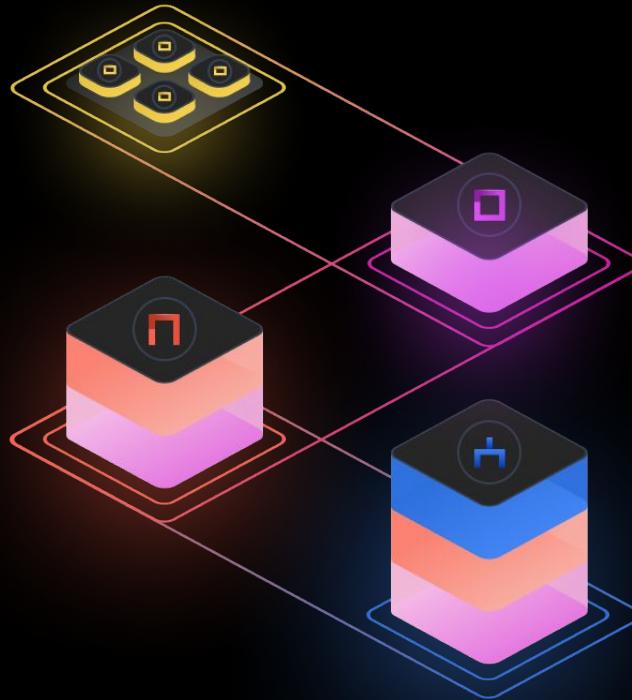
From prototype to scalable solution in weeks.



Security is a top priority at Kern AI. We are ISO 27001 certified.



We offer the solution to turn your company into an NLP automation and analytics pioneer!



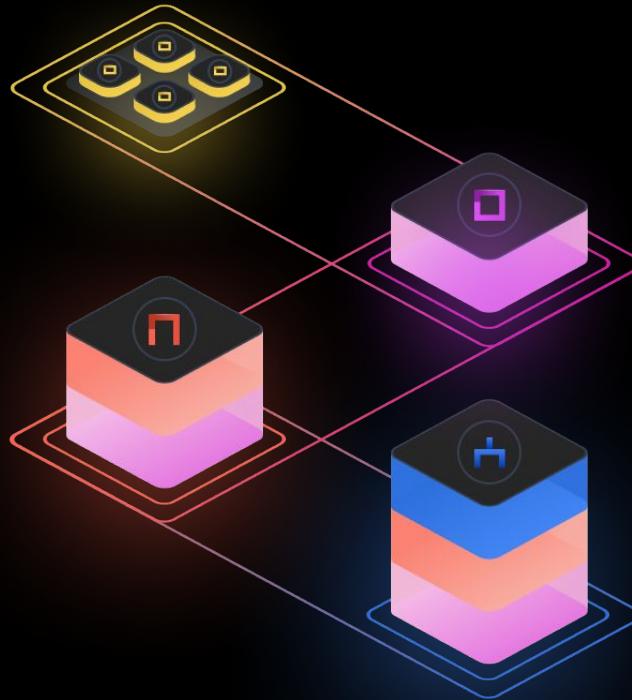
Flexible architecture to fit complex requirements.
Ranging from open-source database to full end-to-end workflows.



Security is a top priority at Kern AI. We are ISO 27001 certified.



We offer the solution to turn your company into an NLP automation and analytics pioneer!



In the cloud and on-premise.
With best practices on data-centric AI embedded.



Security is a top priority at Kern AI. We are ISO 27001 certified.



refinery

refinery turns your raw data into actionable training data.

In enables developers to understand where data quality is causing issues, and gives them the tools to orchestrate workloads.

The screenshot displays the refinery interface for managing conversational AI data. On the left, a sidebar provides navigation and configuration options:

- Existing data slices:** A list containing "random_ord..." with a delete icon.
- Filter:** A section for filtering and ordering data.
- Attributes:** Filter on attributes of your records.
- Users:** Filter manual labels by creation user.
- Labeling task: intent_indicator:** Choose from anything related to intent_indicator.
- Labeling task: intent:** Choose from anything related to intent.
- Result Order:** Order your results.
- Comments:** Filter on comments.
- Connect by:** Radio buttons for OR or AND.

Below the sidebar are buttons for "Save data slice" and "Find outliers". The bottom of the sidebar shows the version "v1.8.0".

The main area shows two data slices:

- 1,800 Records:** This slice contains the following data:

running_id	text
1	what do i need to do to apply for a visa card

TYPE	TASK	LABEL	AMOUNT	AVG.CONFIDENCE
LF: ikp_new_card	intent	new card	1	
AL: DistilbertLR	intent	new card	1	
Weak Supervision	intent	new card	1	100%
Manual	intent	new card	1	
Manual	intent_indicator	new card	1	
Manual	intent_indicator	card provider	1	
- Find similar records:** A button to continue with this record.
- running_id 2:** This slice contains the following data:

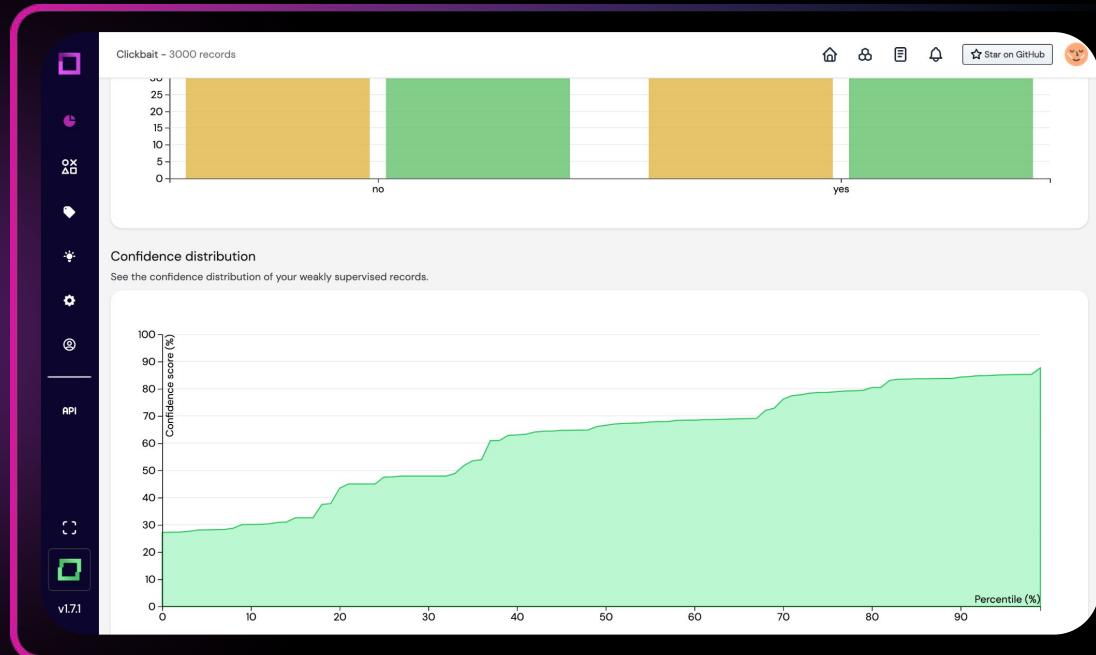
running_id	text
2	my disney visa card is stolen



refinery

refinery turns your raw data into actionable training data.

In enables developers to understand where data quality is causing issues, and gives them the tools to orchestrate workloads.





bricks is a content library of more than 70 ready-made NLP enrichments, e.g. sentiment analysis or text complexity.

Zero setup required; turn ideas into an implementation in seconds.

The screenshot shows the 'bricks' content library interface. At the top, there's a search bar labeled 'Search any module' and a '74 bricks' count. Below the search bar, there are several cards representing different NLP functions:

- Address extraction**: Python function, New. Detects the location/addresses and returns them in a list.
- Aspect extractor**: Python function, New. Matches aspects in a text to positive or negative sentiment.
- Color code extraction**: Python function, New. Extracts CSS colors from a text.
- Credit card extraction**: Python function, New. Extracts the credit/debit card number from a text.
- Crftagger**: Active learner, New. Conditional Random Fields tagger.
- Date extraction**: Python function, New. Detects dates in a text and returns them in a list.
- Digit extraction**: Python function, New. Extracts digits of variable length.
- Email extraction**: Python function, New. Detects emails in a text and returns them in a list.
- Filepath extraction**: Python function, New. Extracts a path from a string.
- Gazetteer extraction**: Python function, New. Detects full entities in a text based on some hints.
- Goodbye extraction**: Python function, New. Extracts all the words associated to farewell.
- Gpt3 information extraction**: Premium, New. Uses OpenAI's GPT-3 model to extract keyword from a text.
- Hashtag extraction**: Python function, New.
- Phone extraction**: Python function, New.



refinery and gates

With **gates**, it takes you 3 clicks to deploy a production-ready NLP model to your cloud or self-hosted premises.

Predictions are partially explained, so that you can see *why* a text has been voted as positive.

The screenshot shows the Refinery and Gates interface. On the left, there's a sidebar with icons for Home, Projects, and Help, and a status bar at the bottom indicating 'V1.0.0'. The main area is titled 'Clickbait - 3000 records' and shows an 'API Status' section with 'Initial' and 'Open config' buttons. Below this is a code editor with Python code for making a POST request to an NLP endpoint. To the right of the code editor are sections for 'Example input' (containing JSON sample data) and 'Run sample' (a button). Further down is an 'Endpoint response' section with the placeholder text 'The endpoint's response will appear here.'

```
1 import requests
2
3 url = "http://localhost:4455/commercial/v1/predict/project/0c2cb3af-f5ee-4a67-9072-644224425d"
4
5 headers = {
6     "accept": "application/json",
7     "content-type": "application/json",
8     # make sure to replace <YOUR API KEY> with your own API key
9     "Authorization": "0c2cb3af-f5ee-4a67-9072-644224425d7c <YOUR_API_KEY>"
10 }
11
12 # replace with your own example data
13 example_data = {
14     "headline": "Mike Tyson set to retire after loss",
15     "running_id": 0
16 }
17
18 response = requests.post(url, headers=headers, json=example_data)
19
20 print(response.text)
```

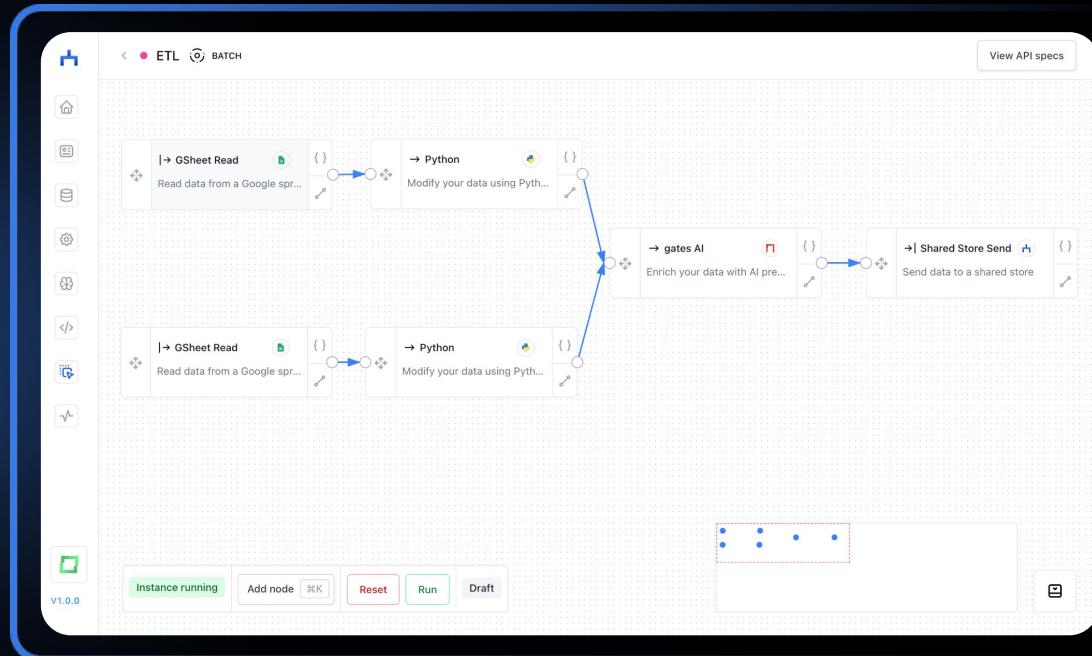


refinery and workflow - for real-time or batch-processes that need to understand natural language.

workflow is the orchestration layer for your natural language-driven tasks.

Define simple or complex conditional flows and directly integrate your textual automations with your IT landscape.

Process automation, but for language-driven cases.





Solve challenges ranging from operational workflows such as scaling customer service tasks ...



At Evolution Time Critical, the platform is used to automatically synchronize hundreds of incoming emails each day for just-in-time requests with the company's internal transport management system.



... to analytical processes such as uncovering insights of market data for decision making ...



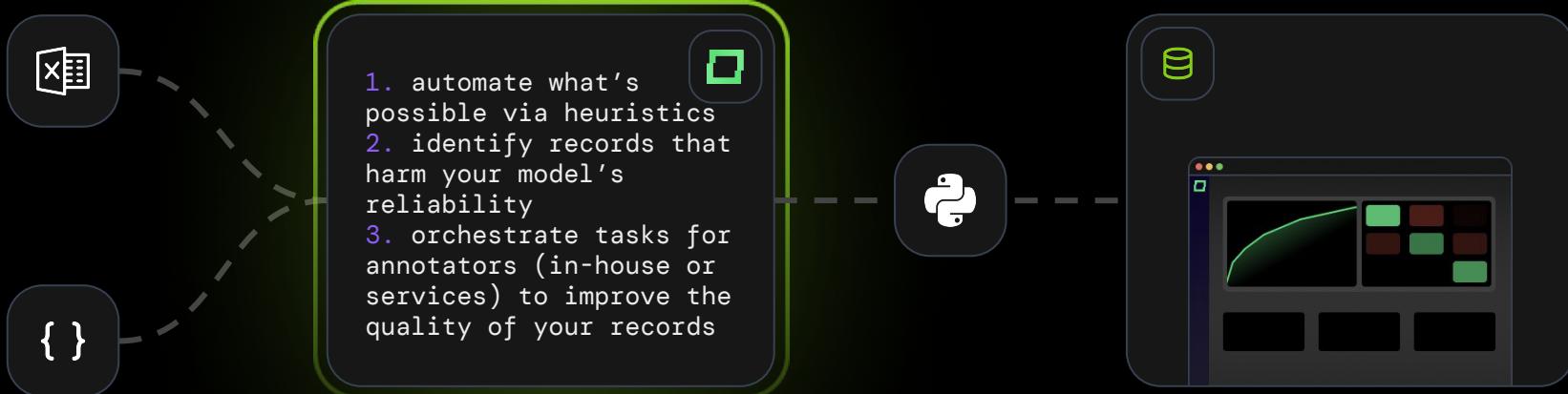
Uncovered insights for better decision making.

E.g. in a large healthcare company, the platform is used to automatically understand market trends and user insights coming in from different social media platforms, aggregated in a dashboard.



..., and even building the fundament for your own sophisticated data-driven NLP products!

</>



Documented and scalable training data to create reliable predictions.

Used by the data science team of Barmenia, a German insurance provider. "With Kern AI, we will be able to further improve the machine selection of conspicuous medical invoices that are steered out into the final expert reviews at Barmenia."



Your net benefits of introducing our platform to your business:



Operating at scale

You can now start to automate even highly complex tasks that require human interaction.

Assist your operators and experts with NLP.

Respond to clients fast and in-detail, scale your operations.



Gaining crucial insights

You can now retrieve insights hidden inside the internet-scale of texts.

Be the first to realize market trends, react to urgent events before it is too late, and make generally better decisions backed by NLP insights.



Building modern products

You can now build products without having to worry about the data infrastructure first.

Go-to-market faster than you can imagine. Iterate on ideas to find meaningful products. Build reliable user experiences from real collaboration.



Not only you will love the platform, but so will its users. Our platform excels in its user experience.



Blazing fast results

Build initial prototypes for complex NLP applications from raw data within hours.



Designed for collaboration

Clear responsibilities within our platform enable a smooth collaboration from the first day.



As extensive as you need

As a database, as an API, or as a workflow orchestration with custom integrations.



Steep learning curve

Our platform is designed to spark creativity, and to introduce best practices to users.



refinery is open-source. You can download and use it immediately. Like thousands of other devs.

Used by data science teams small and large:

DocuSign

Barmenia
Versicherungen

crowddev

co:here

HPI
Hasso
Plattner
Institut
Digital Engineering • Universität Potsdam

SAMSUNG

In a terminal, run these two commands.
That's it. **refinery** is running on your machine.

```
pip install kern-refinery  
refinery start
```



Our pricing is designed to fit your requirements flexibly. In the cloud, and on your premises.



In the cloud



Self-hosted

Growth 149€/mo

- ✓ Full platform access
- ✓ Unlimited users
- ✓ 3,000 incl. credits*/mo
- ✓ Priority support

Enterprise from 999€/mo

- ✓ Full platform access
- ✓ Unlimited users
- ✓ Custom credits*/mo
- ✓ Custom integrations
- ✓ Dedicated support

Enterprise on request

- ✓ Full platform access
- ✓ Unlimited users
- ✓ On your own servers
- ✓ Dedicated support and trainings

*1 credit = 1 real-time operation, or 100 batched operations, or 1MB storage of data per month. More details on our [website](#).



Backed by top-tier investors from Europe, including renowned business leaders

We just raised €2.7 million in seed funding. Read about our announcement on [TechCrunch](#).

Seedcamp **another.vc** **xdeck.**

faber



Nicolas Peters

Co-Founder & Ex-CTO
Signavio

Gerrit de Veer

Senior Vice President
Sales MEE Signavio

Marcus Nagel

Ex-CEO Zurich
Deutschland, Senior Advisor PE

Sebastian Heinz

CEO statworx

Julius Heinz

CFO statworx



I'm happy to discuss with you which challenges your business can solve using NLP technologies.



Johannes Hötter,
CEO and Co-Founder

With Kern AI, you invest in
NLP capabilities for your business.
Let's discuss how this could work for you.



<https://cal.com/jhoetter>



johannes.hoetter@kern.ai



<https://www.linkedin.com/in/johanneshotter/>





5 steps to set up a project in our platform.

Connect a dataset

E.g. Excel spreadsheets, JSON or CSV.
Alternatively, you can pull data directly from one of our integrations.

1

2

3

4

5



The screenshot shows the Clickbait dataset interface. On the left, a sidebar provides navigation and filtering options. The main area displays two tables of data with applied filters.

Filter Options:

- Existing filters (eye icon)
- Attributes (gear icon)
- Users (person icon)
- Labeling task: indicators (lightbulb icon)
- Labeling task: clickbait (lightbulb icon)
- Result Order (gear icon)
- Comments (comment icon)
- API (API icon)
- Connect by (radio buttons for OR or AND)
- Save filter (button)
- Update filter (button)
- Find outliers (button)

Table 1 (Top):

running_id	headline	TYPE	TASK	LABEL	AMOUNT	Avg. Confidence
2534	Brie Larson's Response To Being Called An "It Girl" Will Make Your Day	AL: DistilbertLR	clickbait	yes	1	
		AL: DistilbertTree	clickbait	yes	1	
		LF: lkp_clickbait_expression	clickbait	yes	1	
		Weak Supervision	clickbait	yes	1	85.01%

Table 2 (Bottom):

running_id	headline	TYPE	TASK	LABEL	AMOUNT	Avg. Confidence
2535	British police launch computer hacking investigation	AL: DistilbertLR	clickbait	no	1	
		AL: DistilbertTree	clickbait	no	1	
		Weak Supervision	clickbait	no	1	63.04%



5 steps to set up a project in our platform.

Label reference data

You can either do this in the built-in editor, or via other labeling tools.
Also, you can upload existing labeled data.

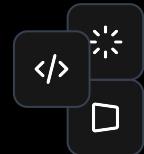
1

2

3

4

5



The screenshot shows the Clickbait dataset with 3000 records. A specific record is selected with the ID 2724. The headline field contains the text "Five dead in **Calgary** murder-suicide". The Full Record panel shows the following details:

TYPE	CREATED BY	LABEL
Weak Supervision	clickbait	Lina Lumburovska
		<input type="checkbox"/> no

The interface includes a sidebar with icons for Data browser, Record IDE, API, and Help. At the bottom, there are buttons for Delete Record, Hide heuristic labels (with a checked checkbox), Automatically get next record (with a checked checkbox), Display of label options (with a value of 5), and a page number indicator (50).

50



5 steps to set up a project in our platform.

Set up automations

1

2

3

4

5

You can program and integrate automations in form of heuristics. These include Python functions, regular expressions, active learning models and more.



Clickbait - 3000 records

Go back DistilbertTree [Finished] 00:00:08

```
7 def __init__(self):
8     params = {
9         "criterion": ["gini", "entropy"],
10        "max_depth": [5, 10, None]
11    }
12    self.model = GridSearchCV(DecisionTreeClassifier(), params, cv=3)
13
14    @params_fit(
15        embedding_name = "headline-classification-distilbert-base-uncased", # pick this from the options above
16        train_test_split = 0.5 # we currently have this fixed, but you'll soon be able to specify this individually!
17    )
18    def fit(self, embeddings, labels):
19        self.model.fit(embeddings, labels)
20
21    @params_inference(
22        min confidence = 0.6,
```

Run Run + weakly supervise

Container Logs 📺 Please send this log to the support if you face problems with your heuristic

```
1: 5.7.2022, 15:10:57 - Preparing data for machine learning.
2: 5.7.2022, 15:10:57 - Running classifier.
3: 5.7.2022, 15:10:57 - Finished execution.
4: 5.7.2022, 15:10:57 - Writing results to the database.
5: 5.7.2022, 15:10:59 - Finished writing.
```

API

Statistics

LABEL	EST. PRECISION	COVERAGE	HITS	CONFLICTS	OVERLAPS
no	80 %	1149	1149	344	690
yes	91.67 %	1851	1851	464	1449

v1.7.1



5 steps to set up a project in our platform.

Monitor and improve

You now have enriched data with metadata like the heuristic indices or confidence scores to analyze the overall quality of your data, and to further drill down into specific filters.

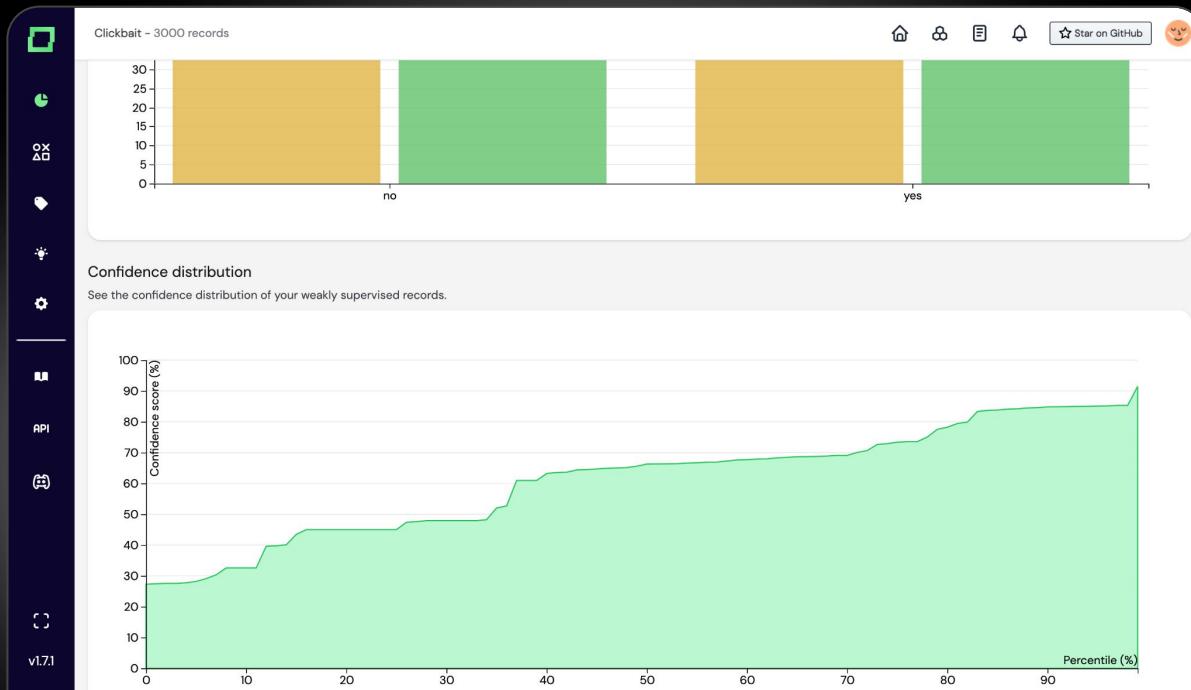
1

2

3

4

5





5 steps to set up a project in our platform.

Integrate or deploy

You can decide whether you want to build your own model on top of the enriched data, or rather use the gates API to use refinery automations as your model's inference engine.

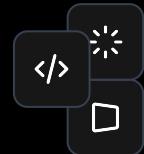
1

2

3

4

5



The screenshot shows the 'gates' API interface. At the top, there are tabs for 'API Status' (Running) and 'Open config'. Below this, a code editor displays Python code for making a POST request to the NLP endpoint:

```
1 import requests
2
3 url = "http://localhost:4455/commercial/predict/project/3005b077-2ff2-401a-9ab1-6b1efe93d1c7"
4
5 headers = {
6     "accept": "application/json",
7     "content-type": "application/json",
8     # make sure to replace <YOUR API KEY> with your own API key
9     "Authorization": "3005b077-2ff2-401a-9ab1-6b1efe93d1c7 <YOUR_API_KEY>"
10 }
11
12 # replace with your own example data
13 example_data = {
14     "headline": "Mike Tyson set to retire after loss",
15     "running_id": 0
16 }
17
18 response = requests.post(url, headers=headers, json=example_data)
19
20 print(response.text)
```

To the right, there are sections for 'Example input' and 'Endpoint response'. The 'Example input' section contains JSON data:

```
1 {
2     "headline": "Mike Tyson set to retire after loss",
3     "running_id": 0
4 }
```

The 'Endpoint response' section shows the JSON output from the API call:

```
1 {
2     "clickbait": {
3         "prediction": "no",
4         "confidence": 0.9999212818567548,
5         "heuristics": [
6             {
7                 "name": "DistilbertLR",
8                 "prediction": "no",
9                 "confidence": 0.9126641218106772
10             },
11             {
12                 "name": "DistilbertTree",
13                 "prediction": "no",
14                 "confidence": 1
15             }
16         ]
17     }
18 }
```

At the bottom left, it says 'V1.0.0'.