

BeCode AI challenge: Natural Language Processing

Contact us
hello@humain.ai
+32 477 41 25 16

Humain NV
Veldkant 33A, 2550 Kontich, Belgium
BE0807730480



Document outline

1. About us

1. How we got started
2. Humain's business model in a nutshell
3. Humain's ecosystem

2. Our NLP challenge: extract AI usecases out of organisational reports

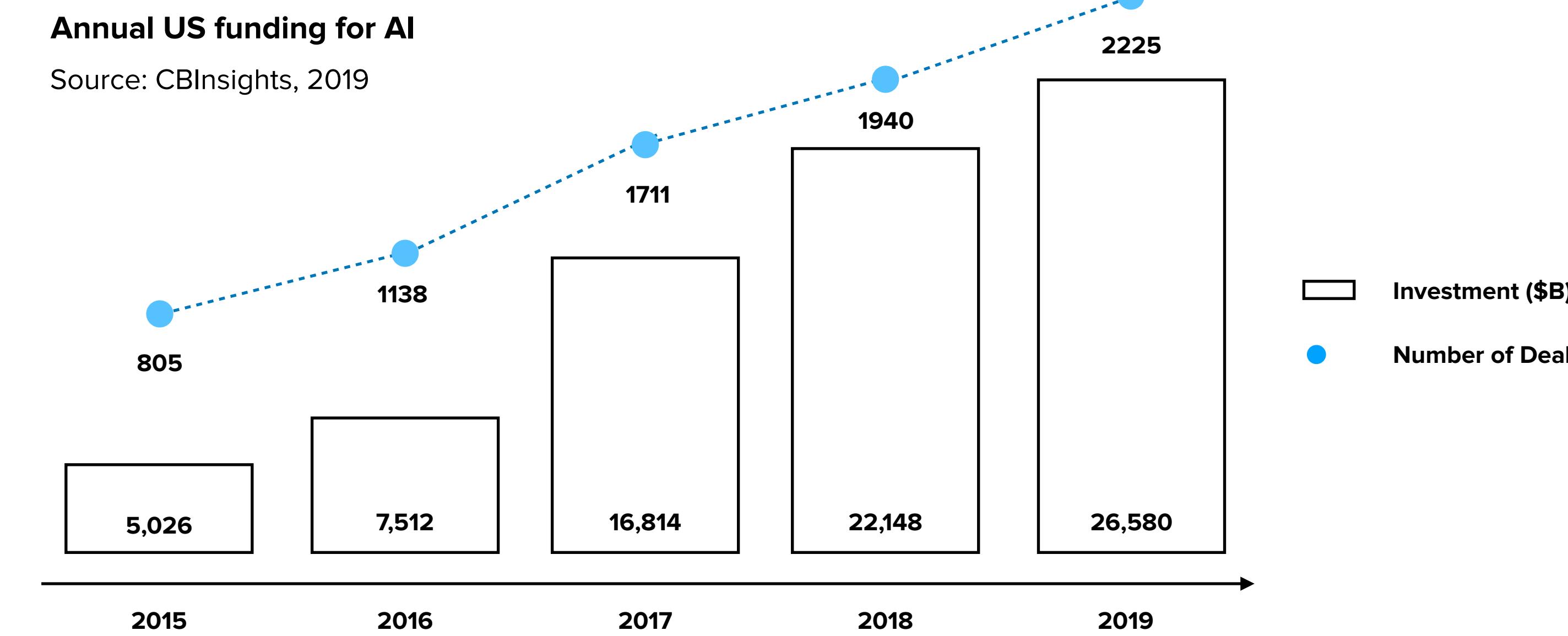
1. Data provided
2. Domains to consider (as a starting point)
3. Focals
 1. Business: focus on product & business case
 2. Data science: focus on data / model selection

3. What we care about (we are hiring)

1. End-to-end
2. Good storytelling
3. Creativity

How we got started: AI projects don't land

Strategic importance of AI is understood, but core adoption lags behind



- AI is a strategic priority for 83% of businesses.

Source: Forbes, 2017

- Machine Learning and AI are the most significant upcoming investment for 61% of businesses.

Source: memSQL, 2018

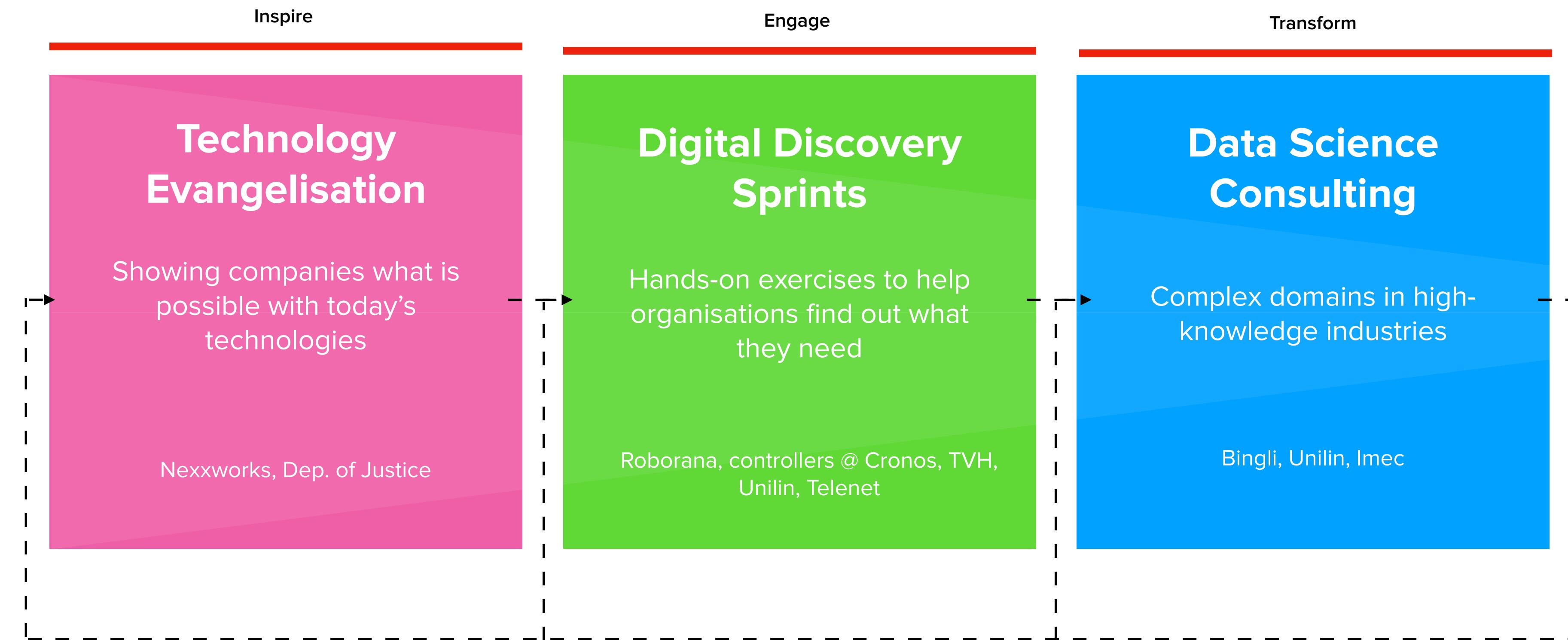
- Only 4% of companies are actively using AI processes to enable advanced tasks.

Source: Microsoft, 2019

Humain's three verticals

H U M A I N

Our business model in a nutshell

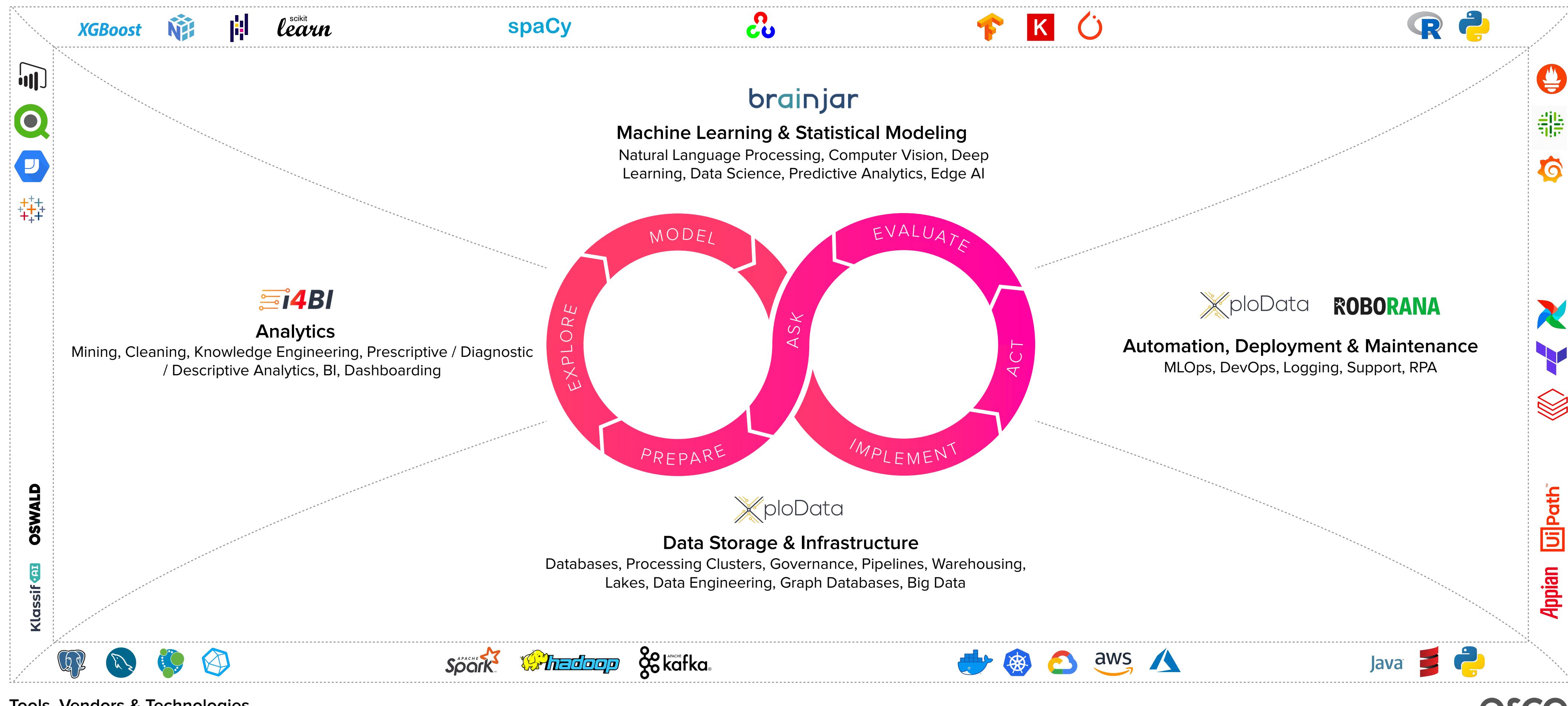


validated Data Science business cases <=> happy clients <=> more projects

The AI Value Chain

H U M A I N

The optimal treatment after the PPP syndrome (PoC, Pilot or Prototype)



H U M A I N

Data Strategy & Data Science Consultancy

Workshops, Business Strategy, Adoption, Training, Product Development, Data Science

Data provided

An opiniated scrape might be necessary



custom_entities.csv

A list of 88 AI usecase entities you can use do
NER / Topic Modeling



ai_usecases.xlsx

A curated list of 122 AI usecases to learn
context. Extracted from a McKinsey report



news.json

Data scraped from [wired.com](#), [AInews](#),
[BBC.com](#) and when searching for AI.



reports directory

A directory of company reports you can use to
test your model on.

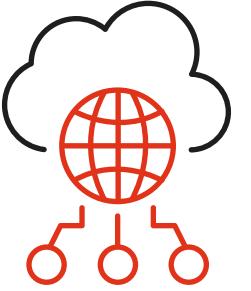


cases.json

Data scraped from an AI usecase website. We
use it on ours too.

Scraping & productivity tips

Having good data is 80% of the effort



Use a pre-trained cloud model

E.g. IBM has a good NER tool + API that you can use to do NER classification. You can also use it as a pre-processing tool for your own pipelines.

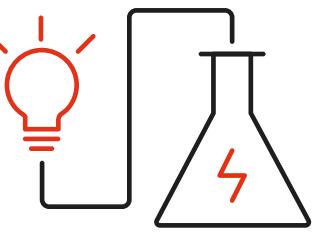


Use search engine results for your context

e.g. Look for your entity / context on Google or other websites and save the search results

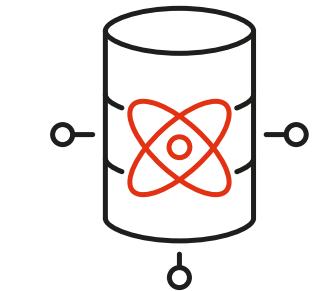
Choosing a focal

You can choose one or the other, or both



Business Focus

Focus on building an NLP end-product & business case with a demonstrator.



Data Science Focus

Evaluate different approaches of data extraction, problem representation, model evaluation.

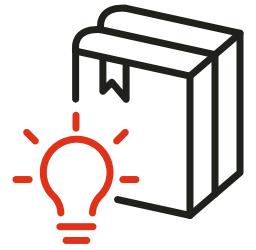
Domains to consider

An opiniated scrape might be necessary



Topic Modeling

Methods to organize, understand and summarize large collections of textual information. E.g. LDA



Name-Entity Recognition

Locate and classify named entities mentioned in unstructured text into pre-defined categories. E.g. SpaCy's fast statistical NER



Multi-class classification

Good old multi-class classification on text-embeddings.

End-to-end

H U M A I N

- Make sure to also put time in your presentation, business case, learnings, evaluation, documentation
- A demo is a plus, but a good story is **more important**
- Demo + good story = **killer**
- Only working code = **no-go**

Storytelling template

Answer these questions to get started

1. Who is your hero / audience? Who uses your NLP tool?
2. What are the challenges they are facing?
3. What do they need to overcome these challenges?
4. What tool / technology are they going to use to overcome this problem?

Your NLP tool

5. How is this going to happen?
6. What / how will we celebrate at the end?

Creativity

H U M A I N

To all the crazy ones, the misfits, the rebels, the troublemakers, the round pegs in the square holes.



Doing something original is strongly encouraged!

Questions? Grab some coffee & let's hack!

Contact us
hello@humain.ai
+32 477 41 25 16

Humain NV
Veldkant 33A, 2550 Kontich, Belgium
BE0807730480

