



Domain-specific NLP pipelines

Aleksandra Vercauteren

<https://www.faktion.com> • info@faktion.com



NLP use cases

When do you need an NLP pipeline?

Resource intensive and/or repetitive tasks on textual data:

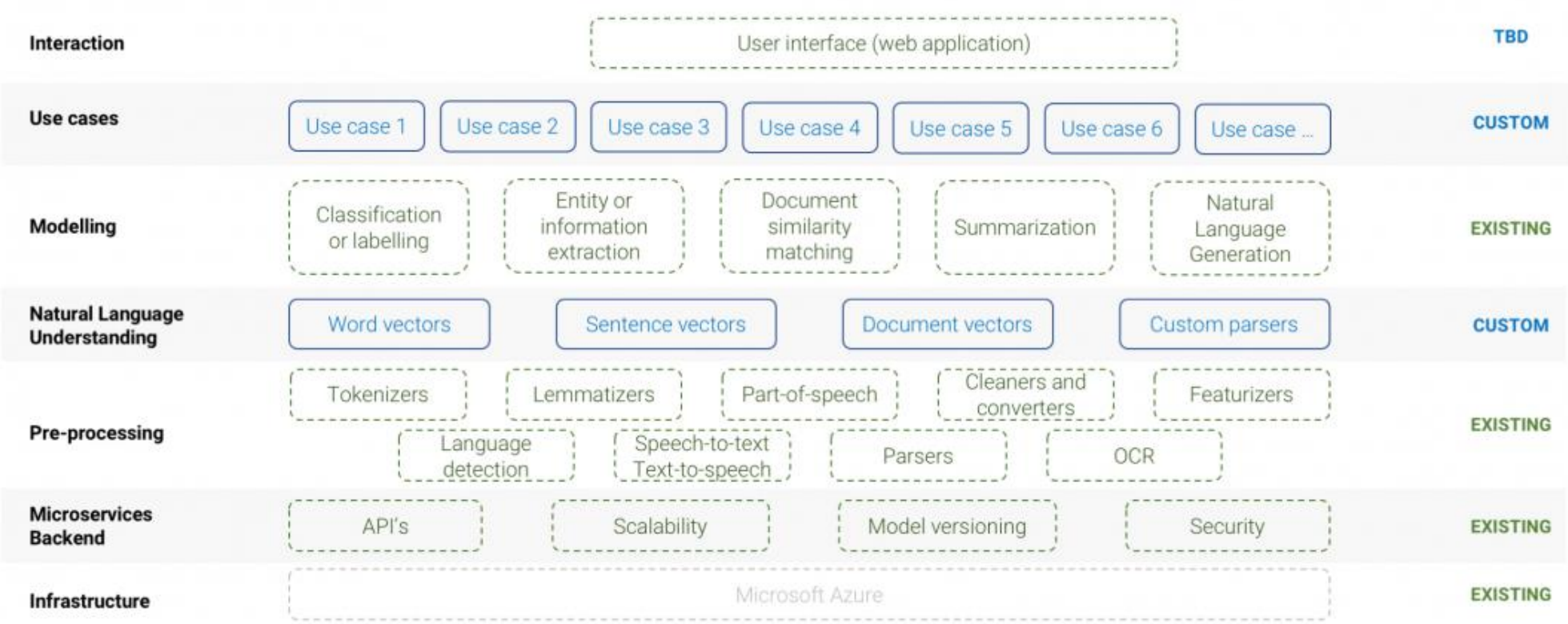
- Call center:
 - 40% are requests for information
 - Employees answer the same questions over and over again
- Order management:
 - orders come in in several formats: phone, email, ...
 - every order has to be manually inserted in Order Management System
- Reviews:
 - What are clients happy about?
 - What could be better?



Building blocks of an NLP pipeline

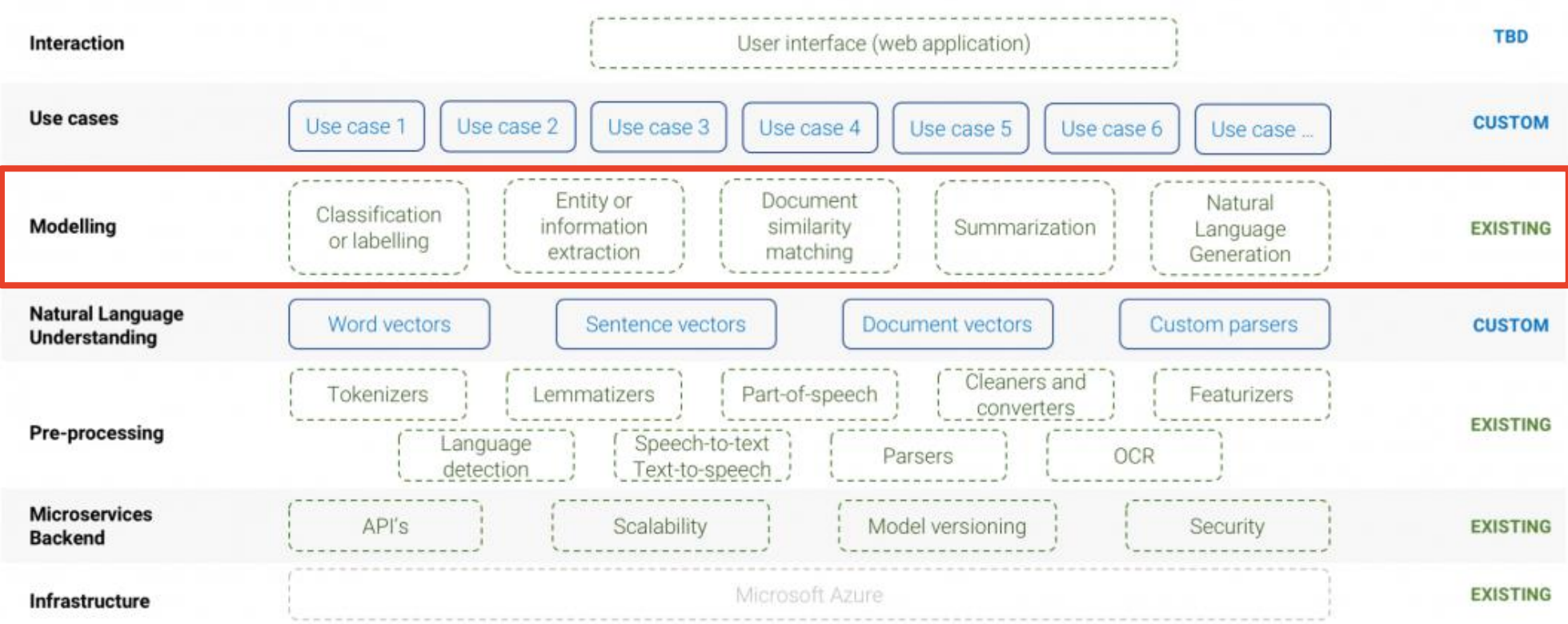
Building blocks

Solution architecture



Building blocks

Solution architecture

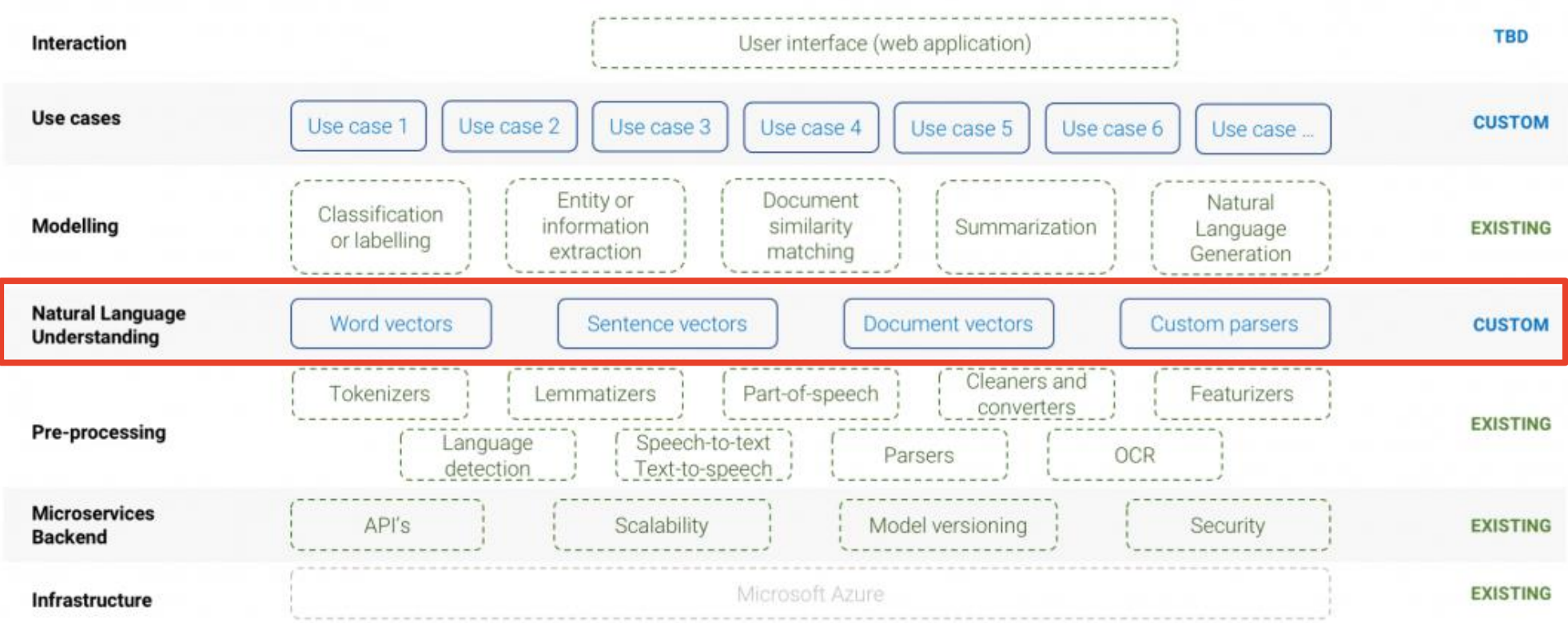


NLP Use cases

Technique	Use cases
Classification or labelling	Chatbots in customer support, FAQ, sales, ... Sentiment analysis in customer support, reviews, ... Automatic tagging, routing and answering support tickets Tagging of customer complaints
Entity extraction	Chatbots Extract to-do lists, tasks, responsibilities, deadlines etc. from emails, meeting notes, ...
Document matching	Resume matching Q&A matching
Summarization	Information overload Newsletters Question answering based on help docs
Natural Language Generation	Question Answering with Knowledge Database Automatic translation

Building blocks

Solution architecture

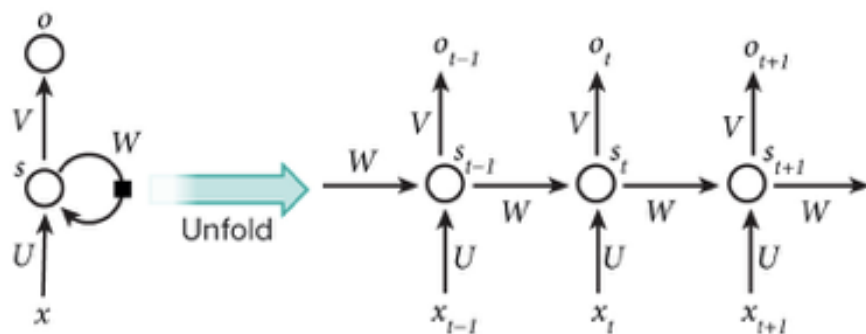


Word, sentence and document vectors

NLP models are trained on set of annotated expressions. They learn the relation between a linguistic pattern and a label.

“Learn” = solve an equation that maps expressions to a label.

 We need numbers!



Word vectors or embeddings

Numerical vectors that describe the meaning of the word. Trained on large text corpora (eg. Wikipedia) and huge amounts of processing power.

300 'meaning' dimensions

word	Female	name	regalness	...
King	0	0.1	1	...
Queen	1	0.1	1	...
Porsche	0.3	0.2	0.3	...
Fiat 500	0.6	0	0	...
Lieselotte	1	1	0.1	...
Elizabeth	1	1	0.4	...
...

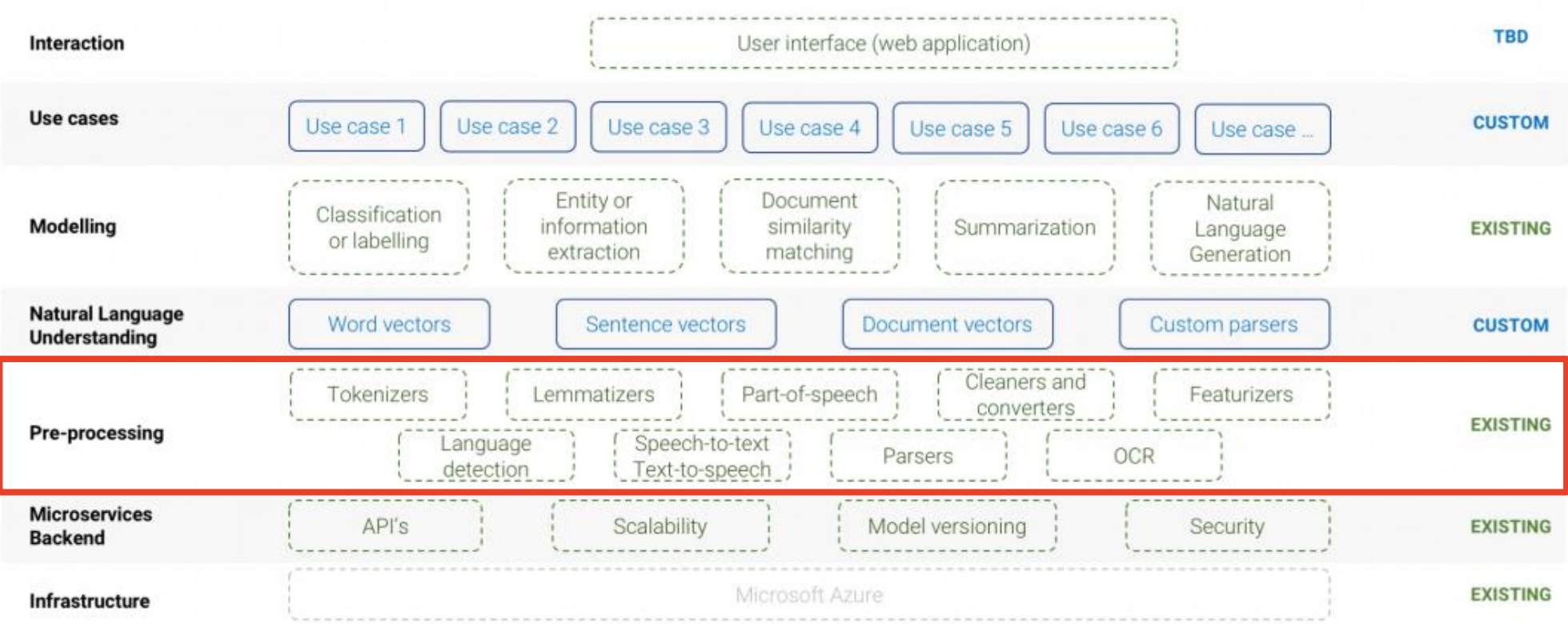
1.6 million words

complaint:

```
[-0.041264 0.026875 0.021691 0.040996 0.066634 0.079733 0.022150 0.021975 -  
0.029170 -0.084697 -0.082365 0.065289 0.085305 -0.082154 -0.064156 0.036492 -  
0.036538 0.047131 0.051098 -0.036164 -0.023157 0.021665 0.082819 0.077477 ...]
```

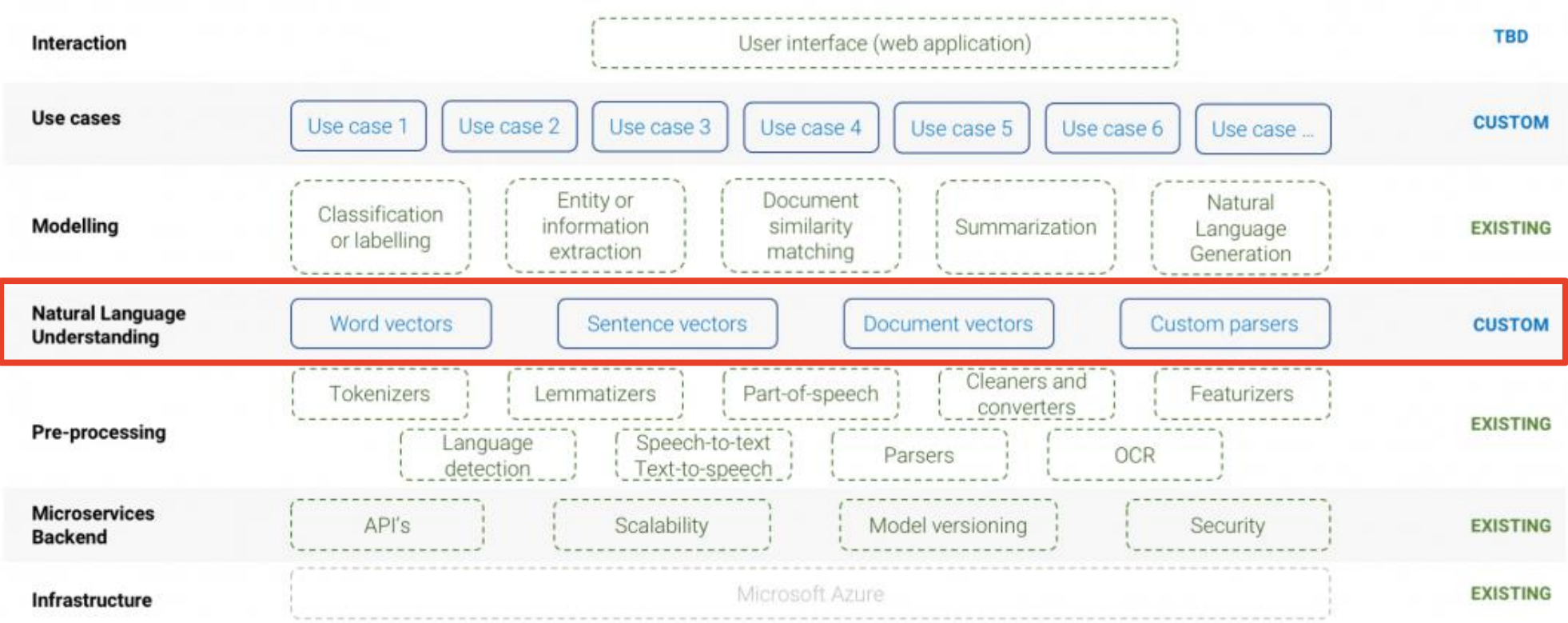
Building blocks

Solution architecture



Building blocks

Solution architecture





Domain-specific pipelines

Domain-specific pipelines

Jargon-filled language:

- Financial documents like prospectuses, annual reports, shareholder letters,...
- Legal documents like contracts, legislation, ...
- Technical manuals
- R&D lab reports

Domain-specific pipelines

Jargon-filled language:

- Financial documents like prospectuses, annual reports, shareholder letters,...
- Legal documents like contracts, legislation, ...
- Technical manuals
- R&D lab reports



Domain-specific pipelines

Non-standard language use:

- Radio communication (police, air traffic control)

Dispatcher: Adam Twelve code five.

Adam Twelve: Twelve, code five, go ahead.

Dispatcher: I'm showing a warrant on your party, Doe, John Q., date of birth three five of sixty, showing physical as white male, six foot, two-eighty, blond and blue, break--

- Text messages



Domain-specific pipelines

Additional preprocessing needed:

- Multi-language documents
- Resume - job matching
- AI project management tool based on email
- ...

Domain-specific pipelines

Singular use cases:

- Early-onset dementia detection
- Writing coach
- Pitching coach
- Spell corrector

Accuracy is paramount



Questions?



FAKTION



Thank you



FAKTION