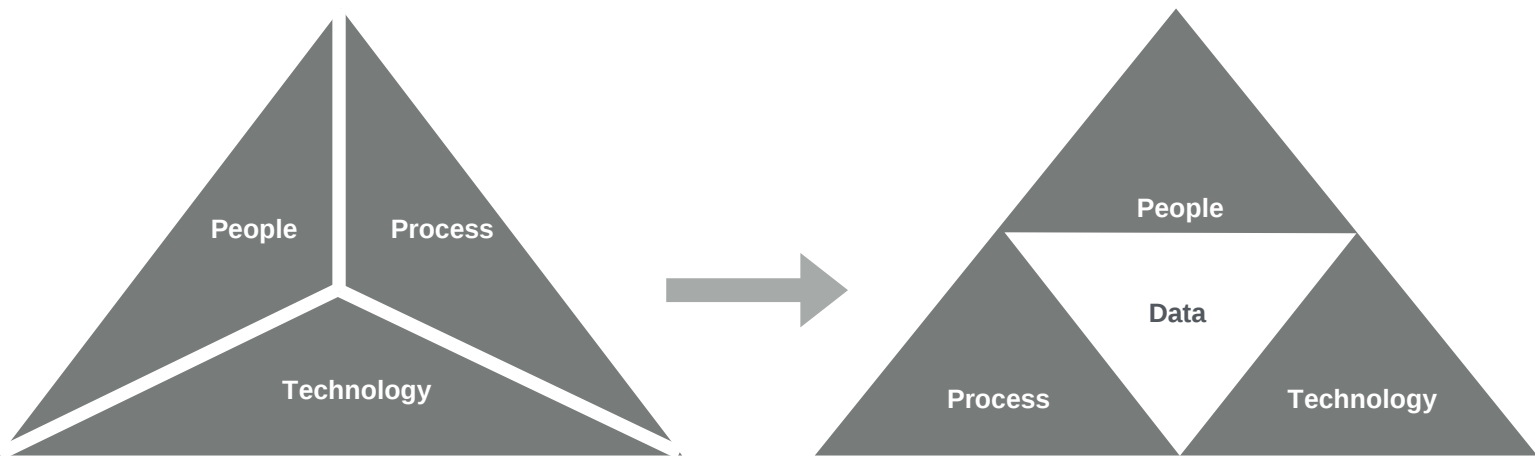


Machine Translation in Clinical Trial Development

Rida Liu

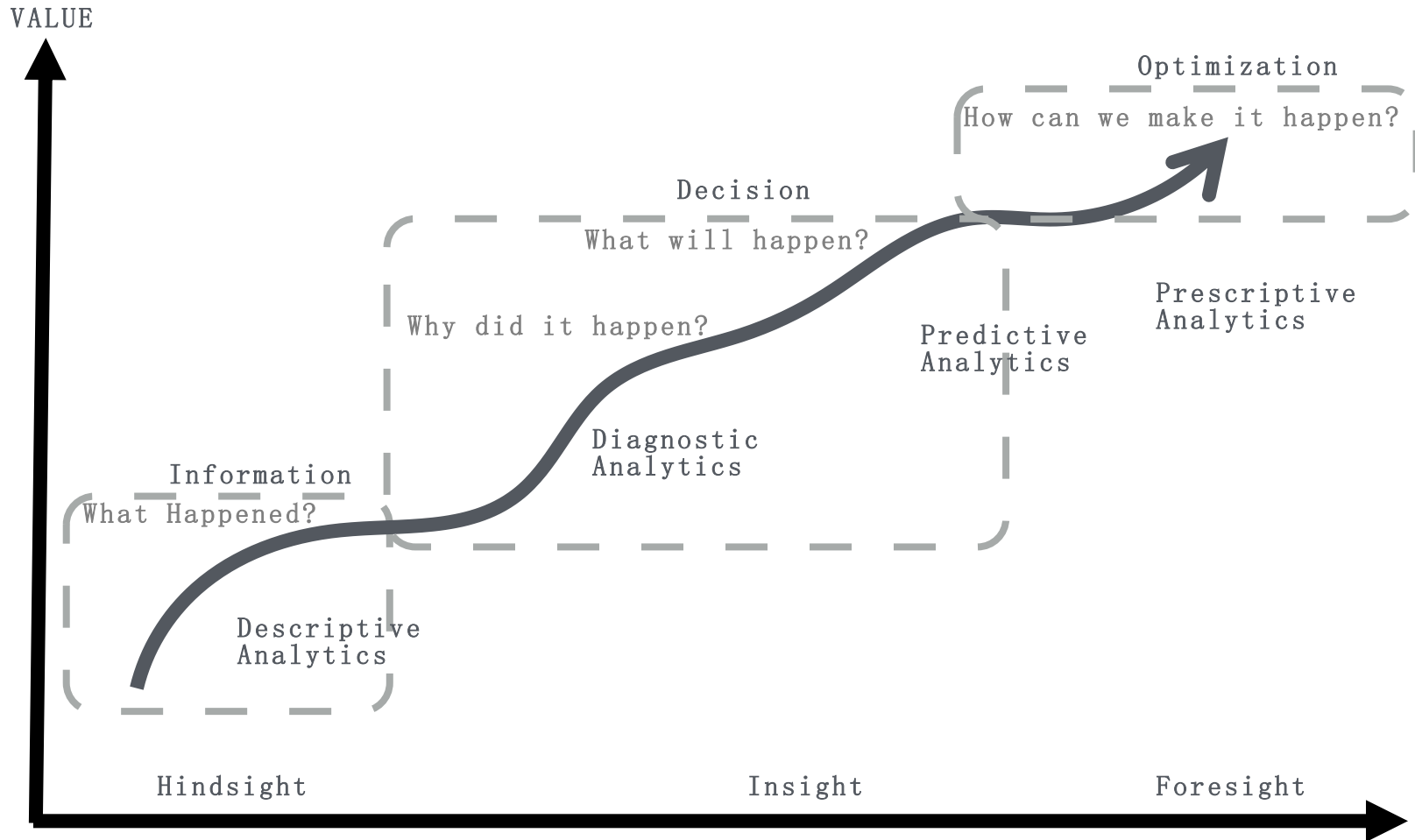
Sanofi-CSO-BICC

Data as the New Core Capability of Digital Business



Gartner defines data literacy as: The ability to **read**, **write** and **communicate** data in context, including an **understanding** of data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case application and resulting value.

4 Stages of Data Analytics Maturity



Use machine learning to solve a variety of business challenges

Sales & Marketing

- Churn Reduction
- Customer Acquisition
- Lead Scoring
- Product Recommendation
- Campaign Optimization
- Customer Segmentation
- Next Best offer/Action

Operations

- Predictive Maintenance
- Load Forecasting
- Inventory/Demand Optimization
- Product recommendation
- Price Optimization
- Manufacturing Process Optimization
- Quality Management
- Yield Management

Fraud & Risk

- Fraud And Abuse Detection
- Claims Analysis
- Collection and Delinquency
- Credit Scoring
- Operation Risk Modeling
- Crime Threat
- Revenue and Loss Analysis

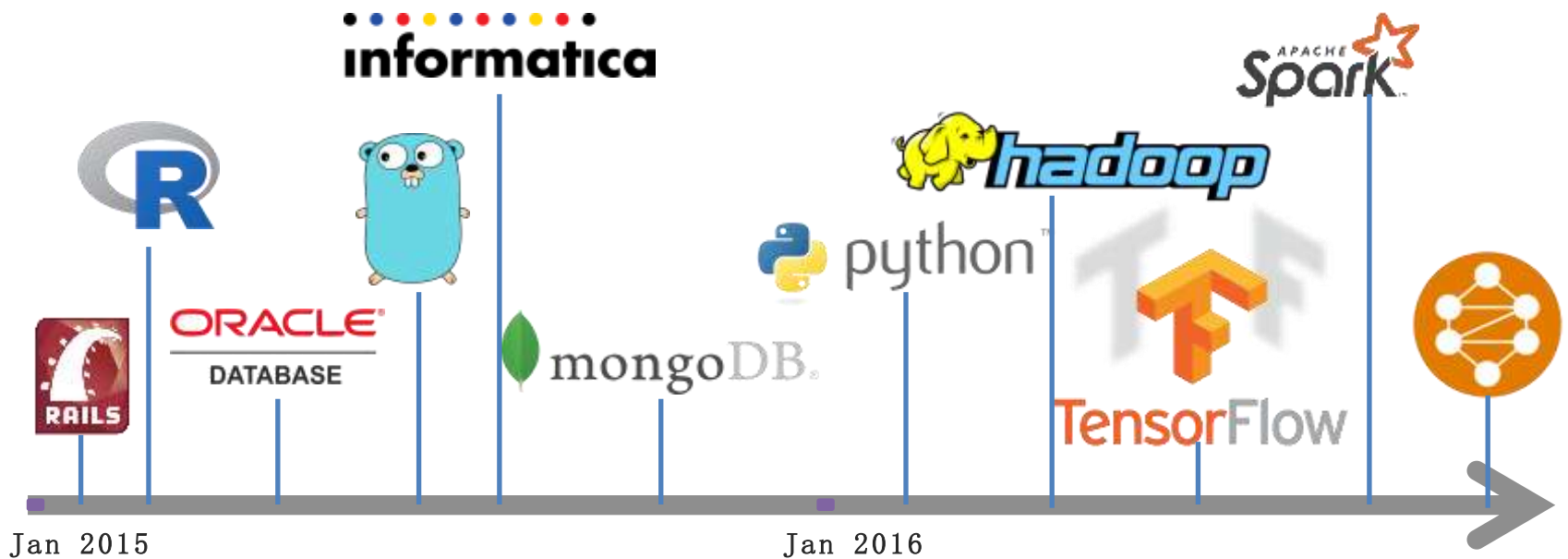
Finance & HR

- Cash Flow and Forecasting
- Budgeting Simulation
- Profitability and Margin Analysis
- Financial Risk Modeling
- Employee Retention Modeling
- Succession Planning

Other Sectors

- Life Sciences
- Healthcare
- Media
- Higher Education
- Public Sector/Social Sciences
- Construction and Mining
- Travel and Hospitality
- Big Data and IoT

Team's Technology Roadmap



Translation Service



Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi
`yonghui,schuster,zhifengc,qvl,mnorouzi@google.com`

Machine Translation at Booking.com: Journey and Lessons Learned

Pavel Levin
Booking.com
Amsterdam
`pavel.levin@booking.com`

Nishikant Dhanuka
Booking.com
Amsterdam
`nishikant.dhanuka@booking.com`

Maxim Khalilov
Booking.com
Amsterdam
`maxim.khalilov@booking.com`

Deep Visual-Semantic Alignments for Generating Image Descriptions

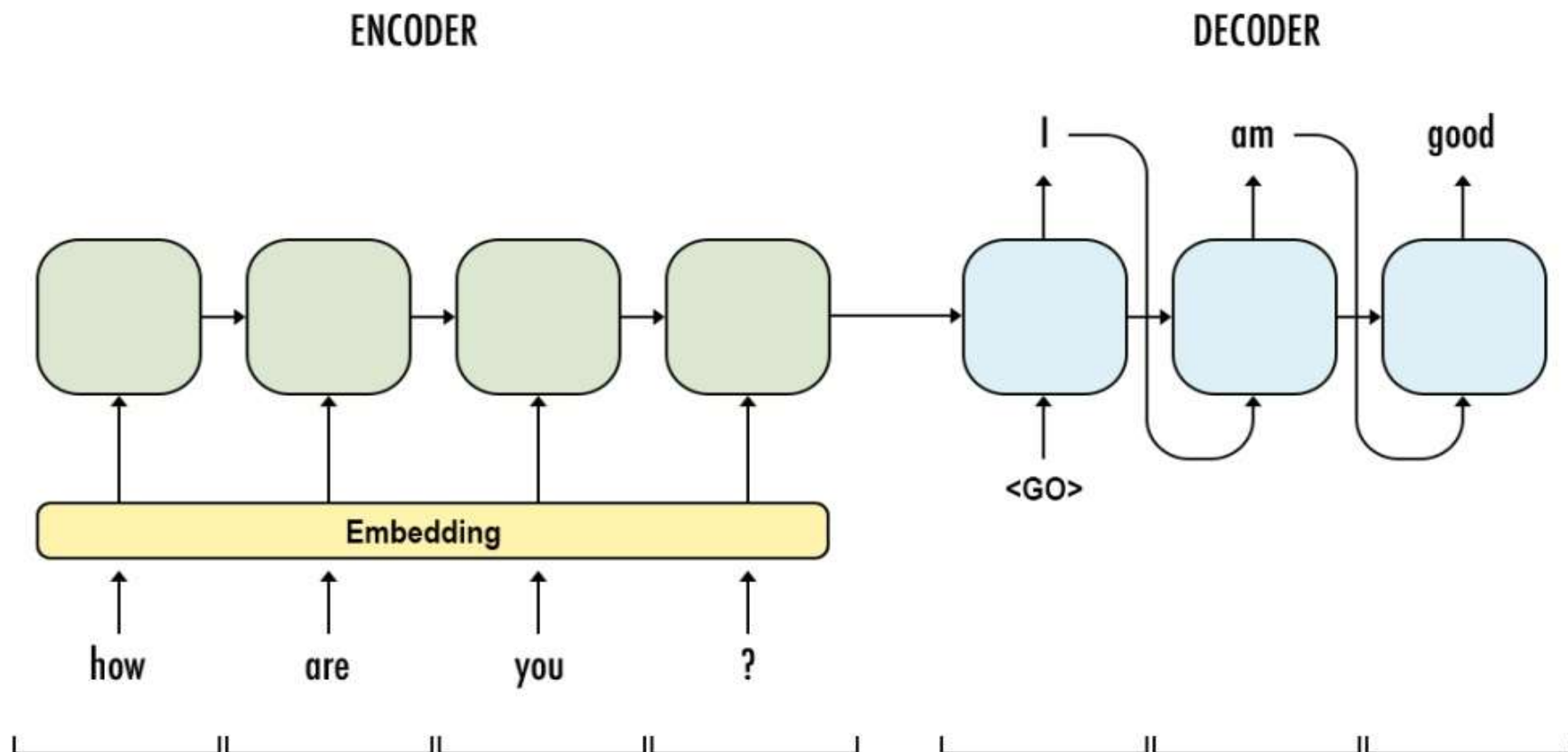
Abstract

We present a model that generates natural language descriptions of images and their regions. Our approach leverages datasets of images and their sentence descriptions to learn about the inter-modal correspondences between language and visual data. Our alignment model is based on a novel combination of Convolutional Neural Networks over image regions, bidirectional Recurrent Neural Networks over sentences, and a structured objective that aligns the two modalities through a multimodal embedding. We then describe a Multimodal Recurrent Neural Network architecture that uses the inferred alignments to learn to generate novel descriptions of image regions. We demonstrate that our alignment model produces state of the art results in retrieval experiments on Flickr8K, Flickr30K and MSCOCO datasets. We then show that the generated descriptions significantly outperform retrieval baselines on both full images and on a new dataset of region-level annotations.

Data Source



Seq2Seq (Recurrent Neural Network)



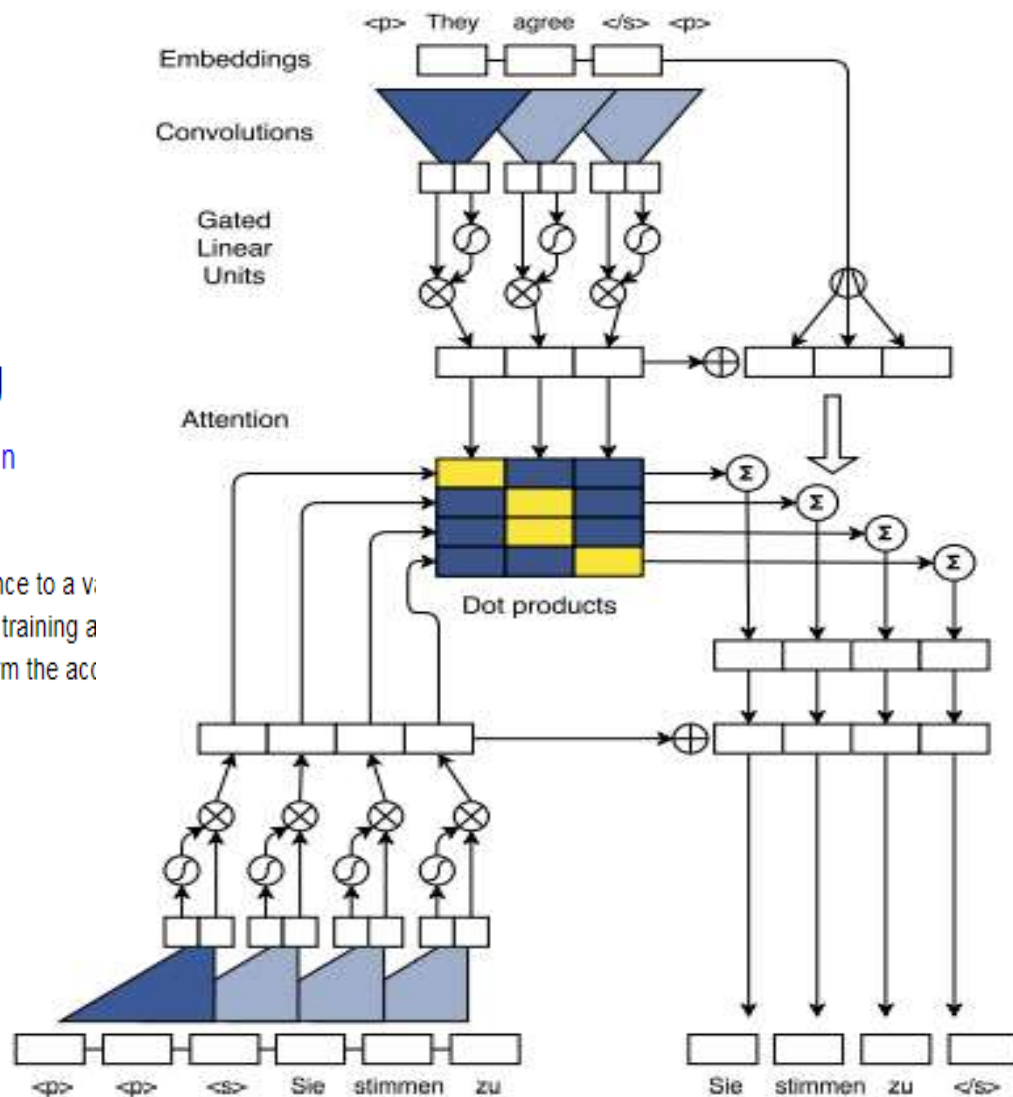
FairSeq (Convolutional Neural Network)

Convolutional Sequence to Sequence Learning

Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann N. Dauphin

(Submitted on 8 May 2017 (v1), last revised 25 Jul 2017 (this version, v3))

The prevalent approach to sequence to sequence learning maps an input sequence to a v. recurrent models, computations over all elements can be fully parallelized during training and we equip each decoder layer with a separate attention module. We outperform the acc. speed, both on GPU and CPU.



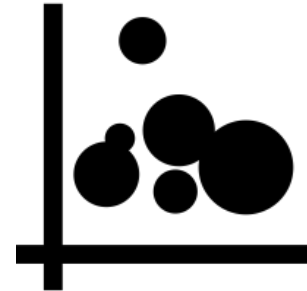
Performance



Community



Monitoring Visualization



Mercury

Business Scenario

- Clinical Study Report (hybrid)
- Audit report (For documentation, non-official)
- Drug Product
-

Best Scenario

- Repeat, Urgent, Multilingual

Technology Exploration

▶ Tokenize

小龙女说我也想过过过儿过过的生活

小龙女 说 我 也 想 过 过 过 儿 过 过 的 生 活

用毒毒毒蛇 毒蛇会不会被毒死

用 毒 毒 毒蛇 毒蛇 会 不 会 被 毒死

▶ NER(Named-entity recognition)



Figure 1: An example of NER application on an example text

Technology Exploration

▶ Memory



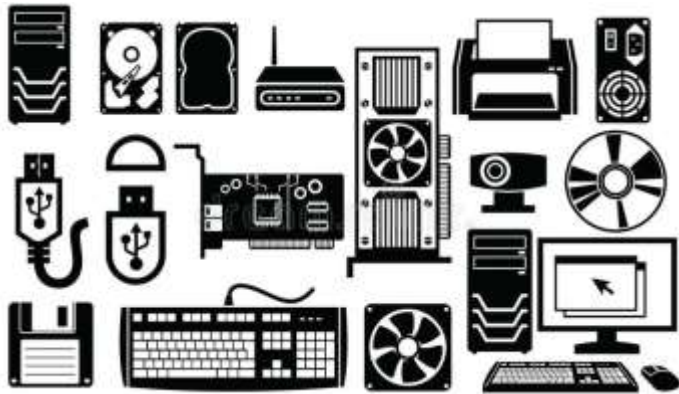
elasticsearch

▶ Synonym mapping

"united states, u s a, united states of america => usa"

Challenge

► Hardware utilization rate



► Data Quality

10011000101001101

11100011100110001

1000011001111

100110001001111

111011110011000

10100110010011111



Run the model with could

Launch bicc-image 4. Launch

Exec environment

Docker image: ✓

* Type: ✓ CPU: ✓ RAM: ✓ GPU: ✓

AWS Region: ✓

* Disk (Gb): ?

Price type: ✓

Timeout (min): ✓

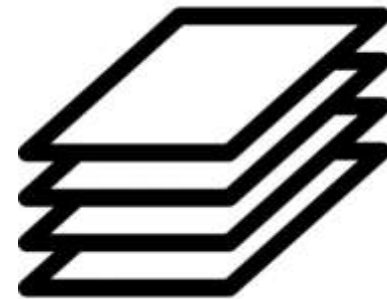
Cmd template: ☐ Start idle ?

Challenge

► Hyperparameter tuning

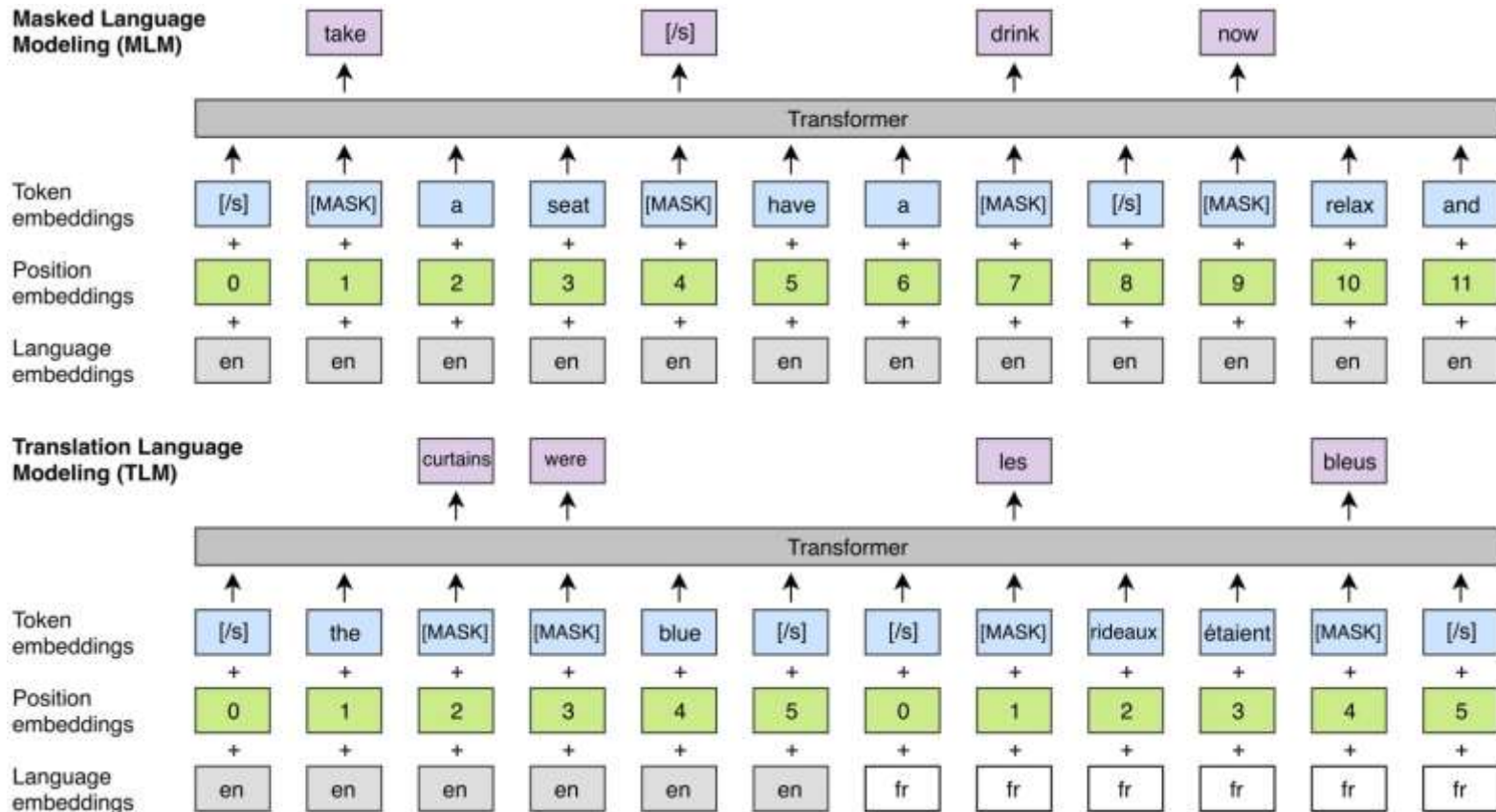


► Continuous model Training

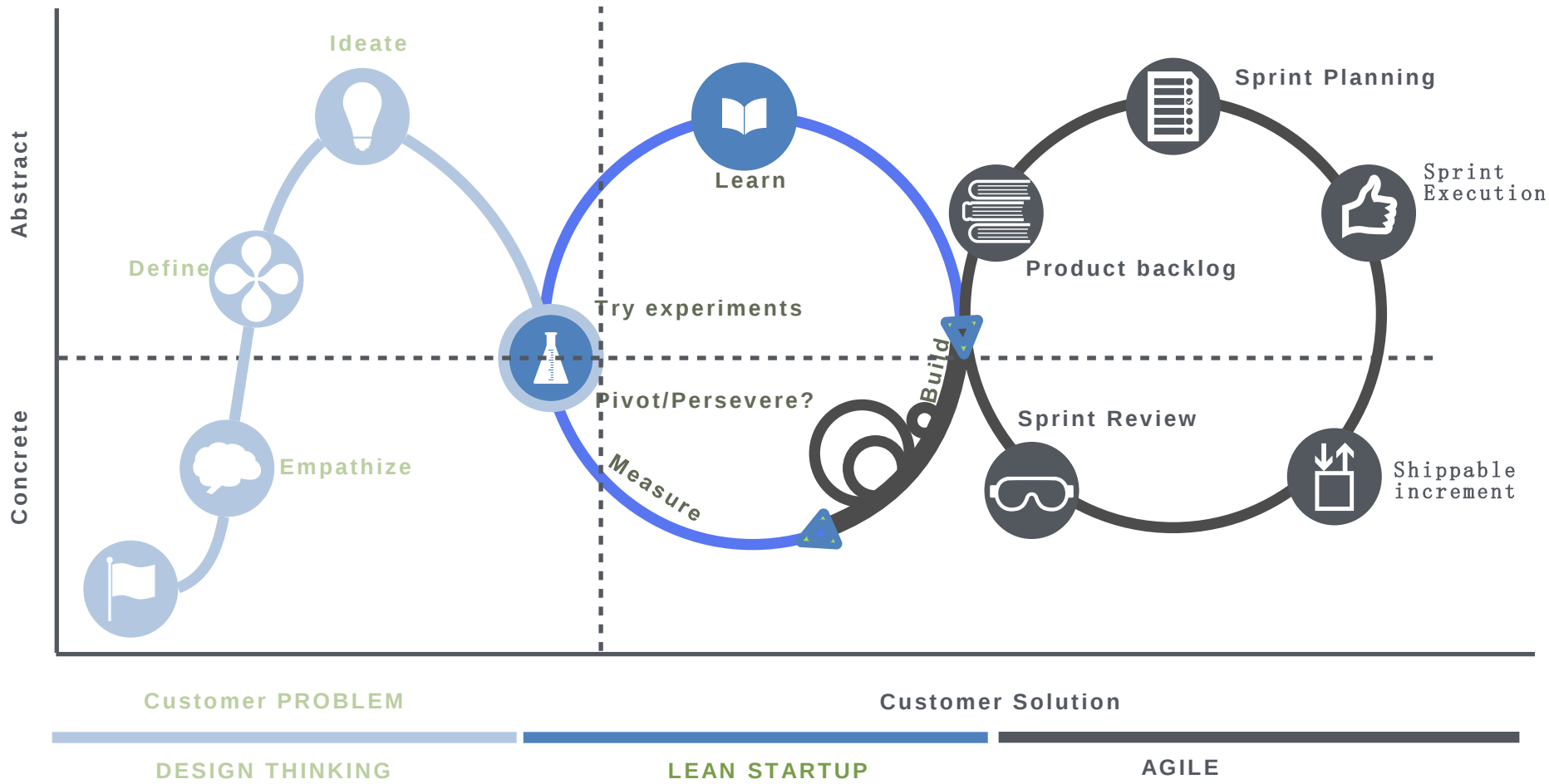


Challenge

▶ Leverage Pre-training word vectors



Combine Design Thinking, Lean Startup and Agile to Raise our Game in Digital Business



Q & A

 Thanks