Natural Language Processing for Information Retrieval and Information Management in Engineering

Dr. Thomas Kinkeldei
ROKIN

ROKIN

ABOUT ROKIN

MUNICH STARTUP
established 2020

Thomas

Roberto

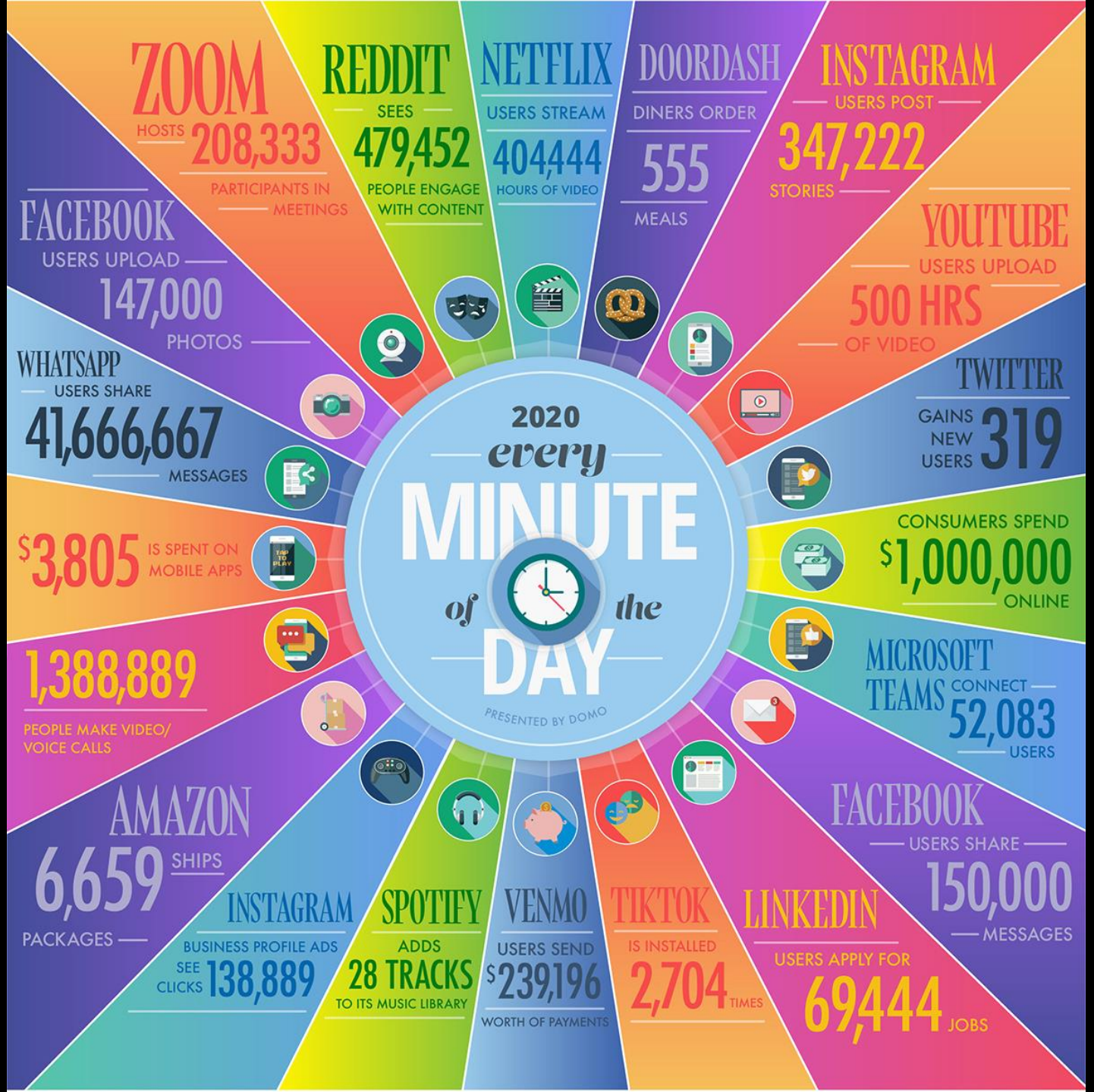
Engineering meets **NLP**

thomas.kinkeldei@rokin.tech

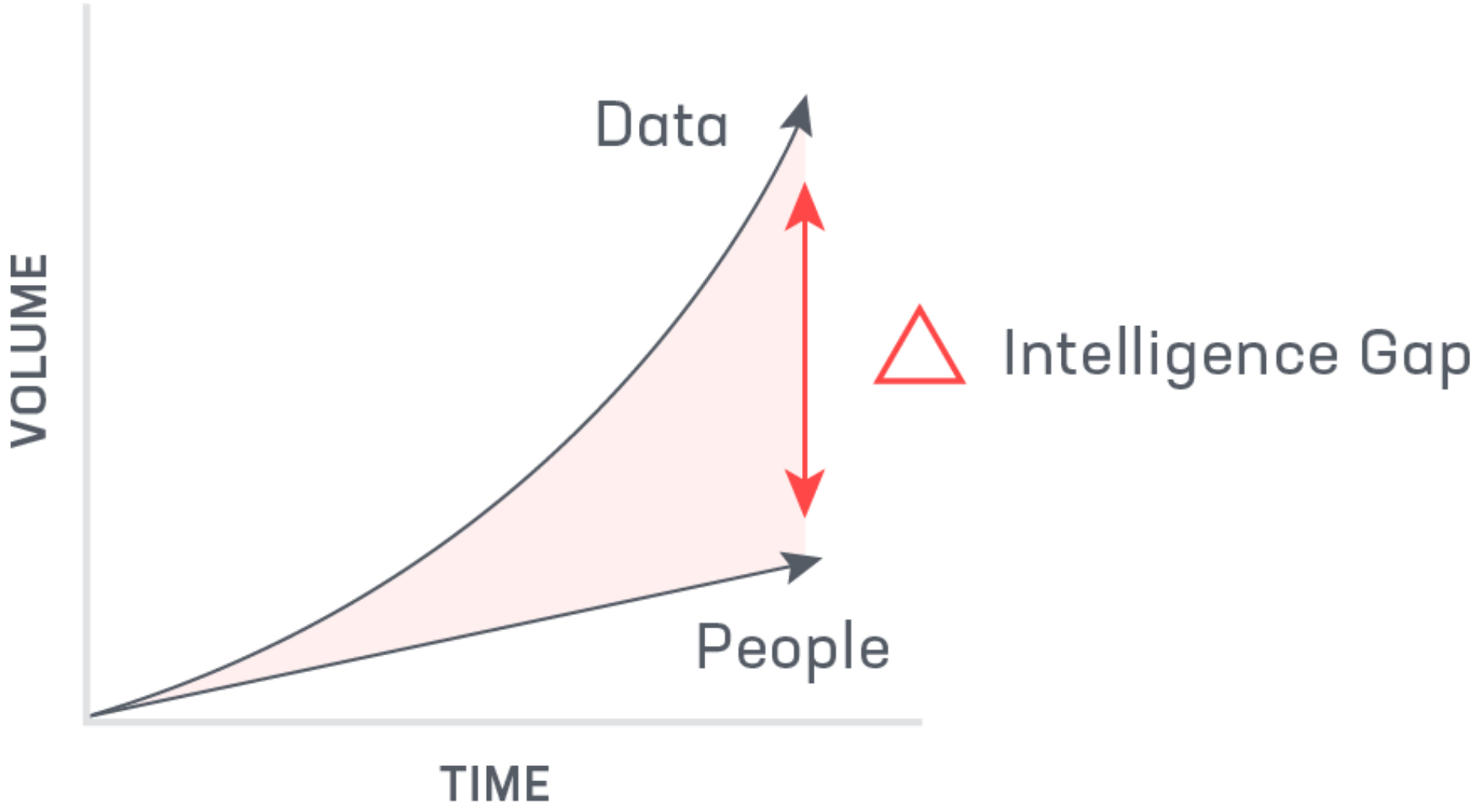
"Information is the oil of the 21st century, and analytics is the combustion engine"

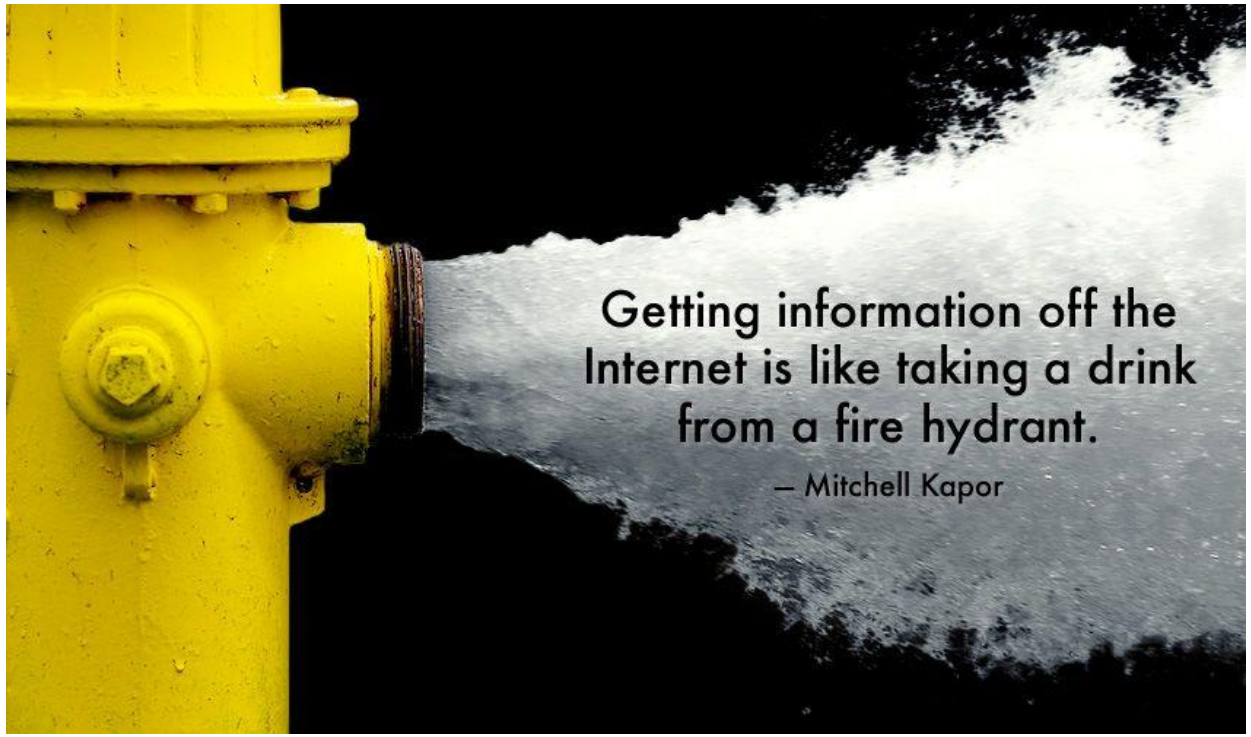
Peter Sondergaard

www.steps.tn


2020
every
MINUTE
of the
DAY
PRESENTED BY DOMO
ZOOM
HOSTS 208,333
PARTICIPANTS IN MEETINGS
REDDIT
SEES 479,452
PEOPLE ENGAGE WITH CONTENT
NETFLIX
USERS STREAM 404,444
HOURS OF VIDEO
DOORDASH
DINERS ORDER 555
MEALS
INSTAGRAM
USERS POST 347,222
STORIES
FACEBOOK
USERS UPLOAD 147,000
PHOTOS
YOUTUBE
USERS UPLOAD 500 HRS
OF VIDEO
WHATSAPP
USERS SHARE 41,666,667
MESSAGES
TWITTER
GAINS NEW USERS 319
$3,805 IS SPENT ON MOBILE APPS
CONSUMERS SPEND $1,000,000 ONLINE
1,388,889
PEOPLE MAKE VIDEO/ VOICE CALLS
MICROSOFT TEAMS CONNECT 52,083 USERS
AMAZON
6,659 SHIPS
PACKAGES
FACEBOOK
USERS SHARE 150,000
MESSAGES
INSTAGRAM
BUSINESS PROFILE ADS SEE CLICKS 138,889
SPOTIFY
ADDS 28 TRACKS
TO ITS MUSIC LIBRARY
VENMO
USERS SEND $239,196
WORTH OF PAYMENTS
TIKTOK
IS INSTALLED 2,704 TIMES
LINKEDIN
USERS APPLY FOR 69,444 JOBS

INTELLIGENCE GAP




Getting information off the Internet is like taking a drink from a fire hydrant.
— Mitchell Kapor

Google

sensor messung schichtdicke

Alle | Shopping | Bilder | News | Videos | Mehr

Ungefähr 93.300 Ergebnisse (0,31 Sekunden)

www.micro-epsilon.de › ... › Suche nach Messgrößen › Dicke

Schichtdicke berührungslos messen | Micro-Epsilon

Durch verrechnen beider Signale ergibt sich die **Schichtdicke**. Dieses Verfa
mit zwei **Sensoren** unterschiedlicher Prinzipien wird von ...

frtmetrology.com › schichtdicke

Schichtdickenmessung mit FRT - FRT GmbH

Einfacher Einstieg in die optische, berührungslose **Messung** der **Schichtdic**
kommen optische **Schichtdickensensoren** mit interferometrischen ...

de.wikipedia.org › wiki › Schichtdickenmessung

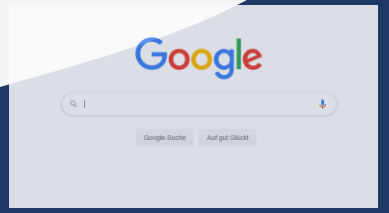
18 billion Euros
yearly expenses for German companies
searching for relevant information

INFORMATIONRETRIEVAL OF ENGINEERS

Tradefairs

Trade publications

Search engines

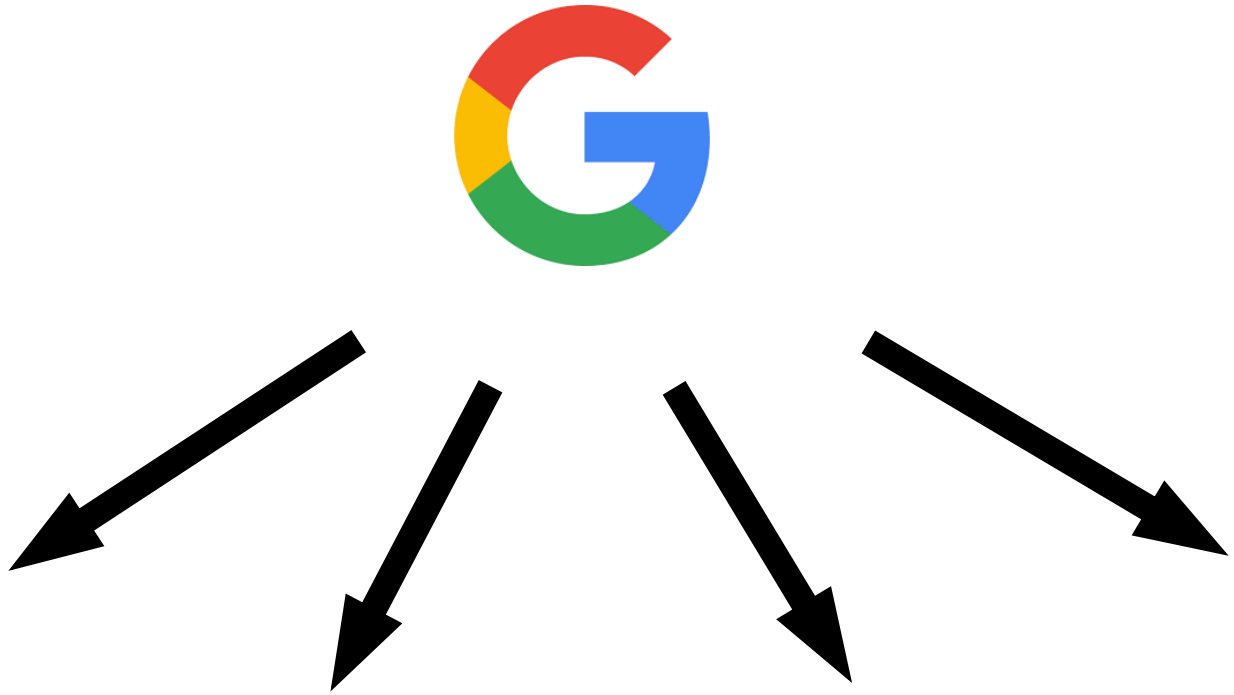


Datavolume

1990

2020

PLATFORM SPECIALIZATION



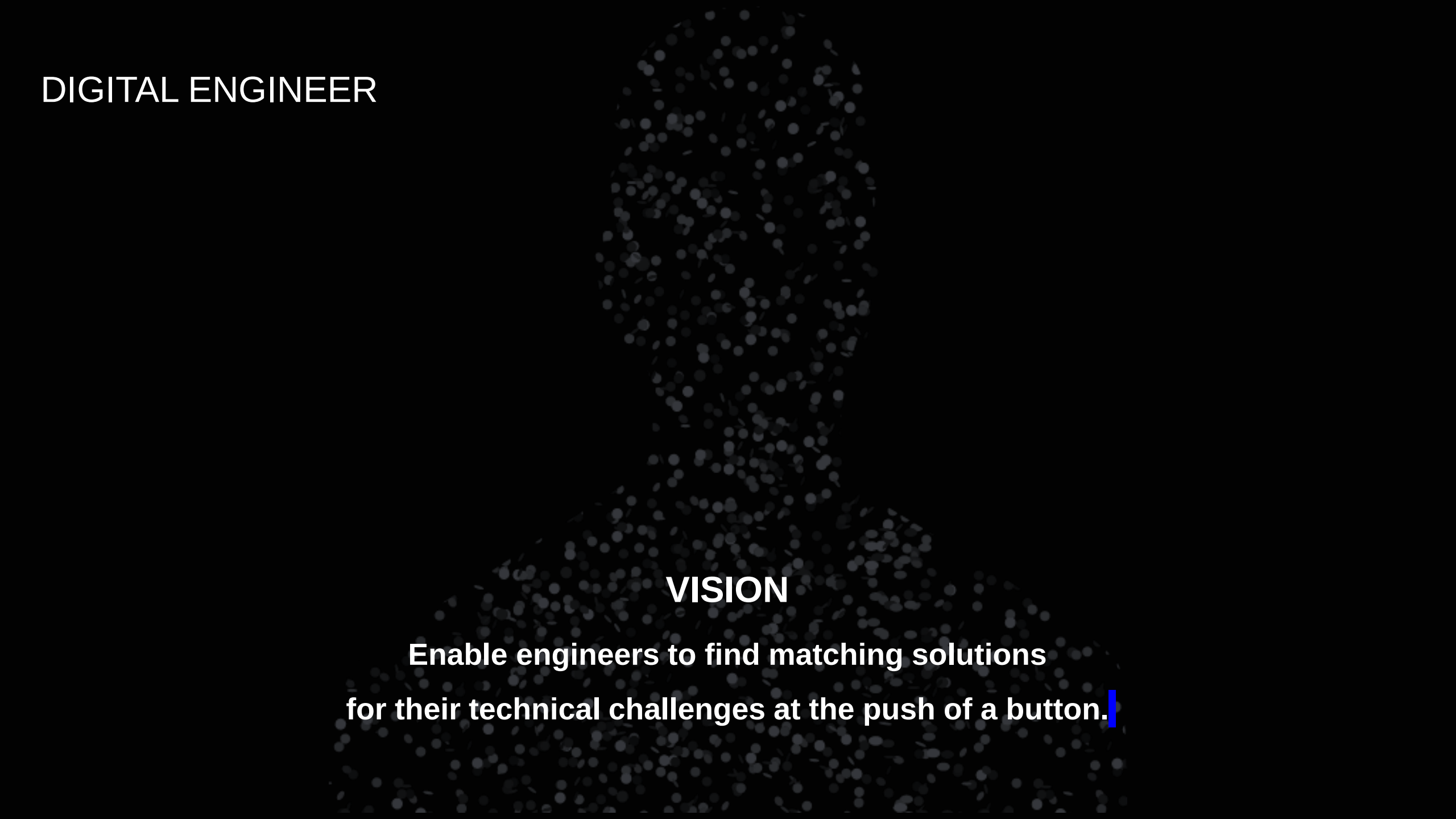
Booking.com

IMMOBILIEN SCOUT 24

komoot

tinder

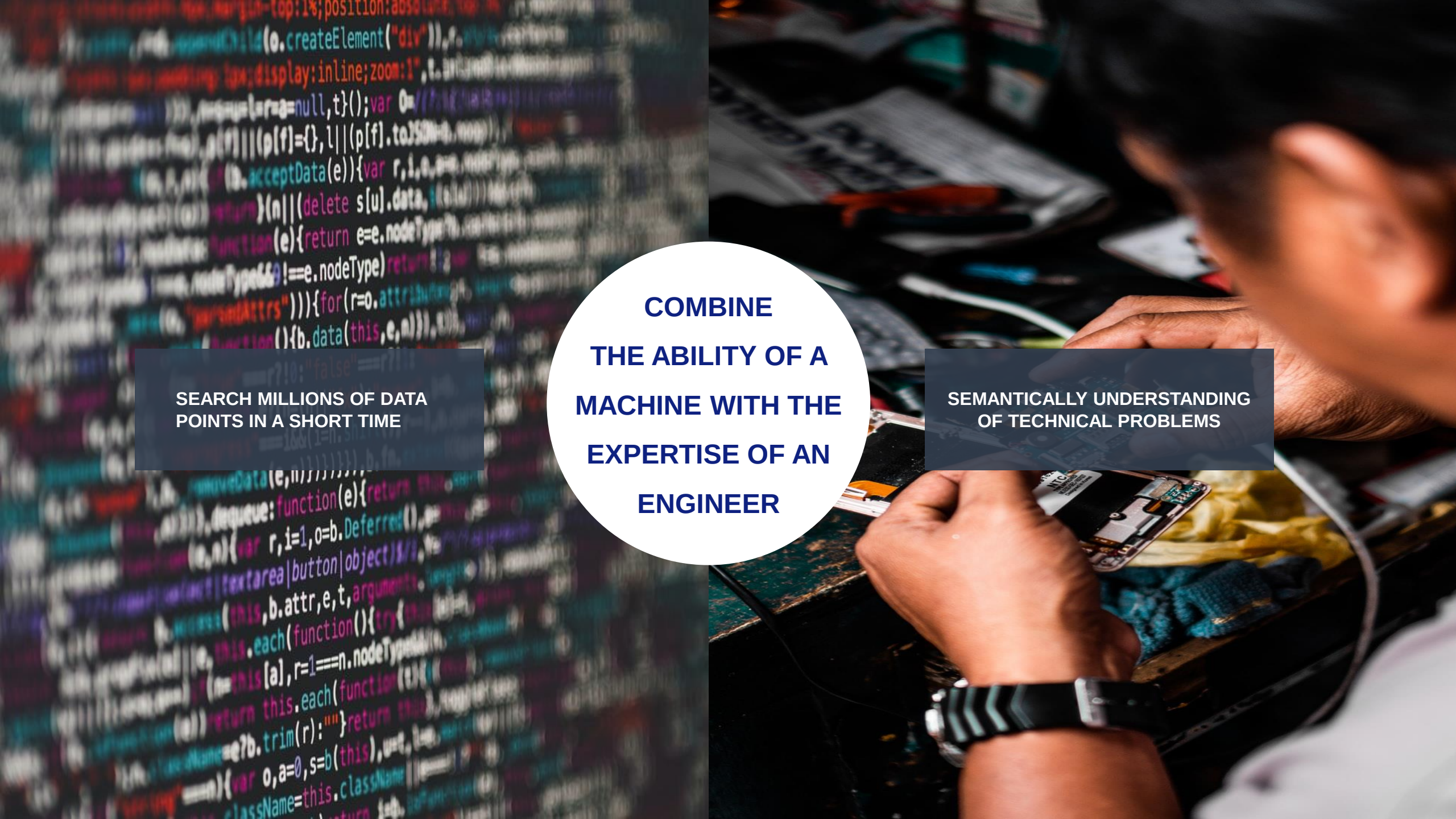
DIGITAL ENGINEER



VISION

Enable engineers to find matching solutions

for their technical challenges at the push of a button.

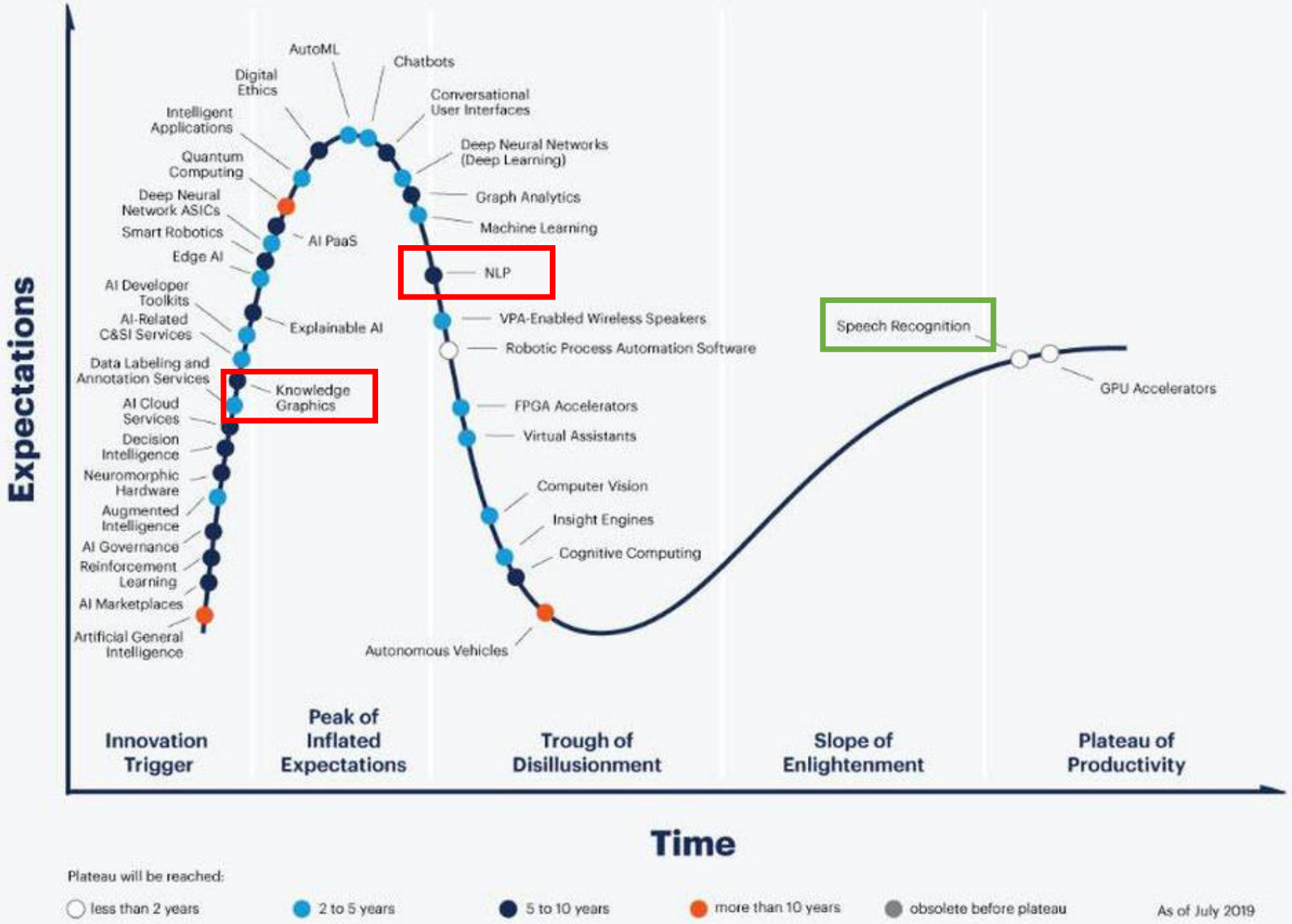


COMBINE

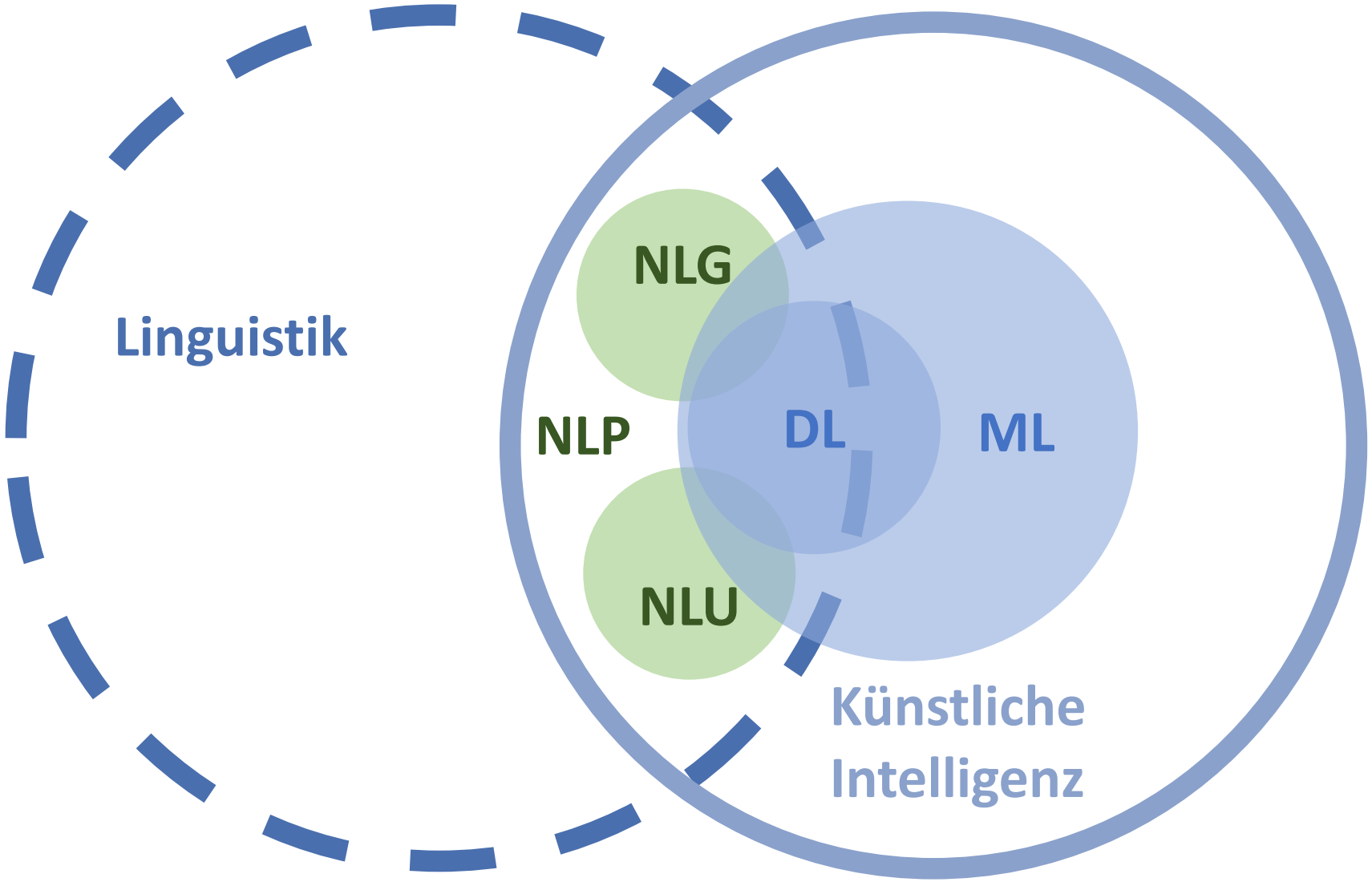
THE ABILITY OF A MACHINE WITH THE EXPERTISE OF AN ENGINEER

SEARCH MILLIONS OF DATA POINTS IN A SHORT TIME

SEMANTICALLY UNDERSTANDING OF TECHNICAL PROBLEMS

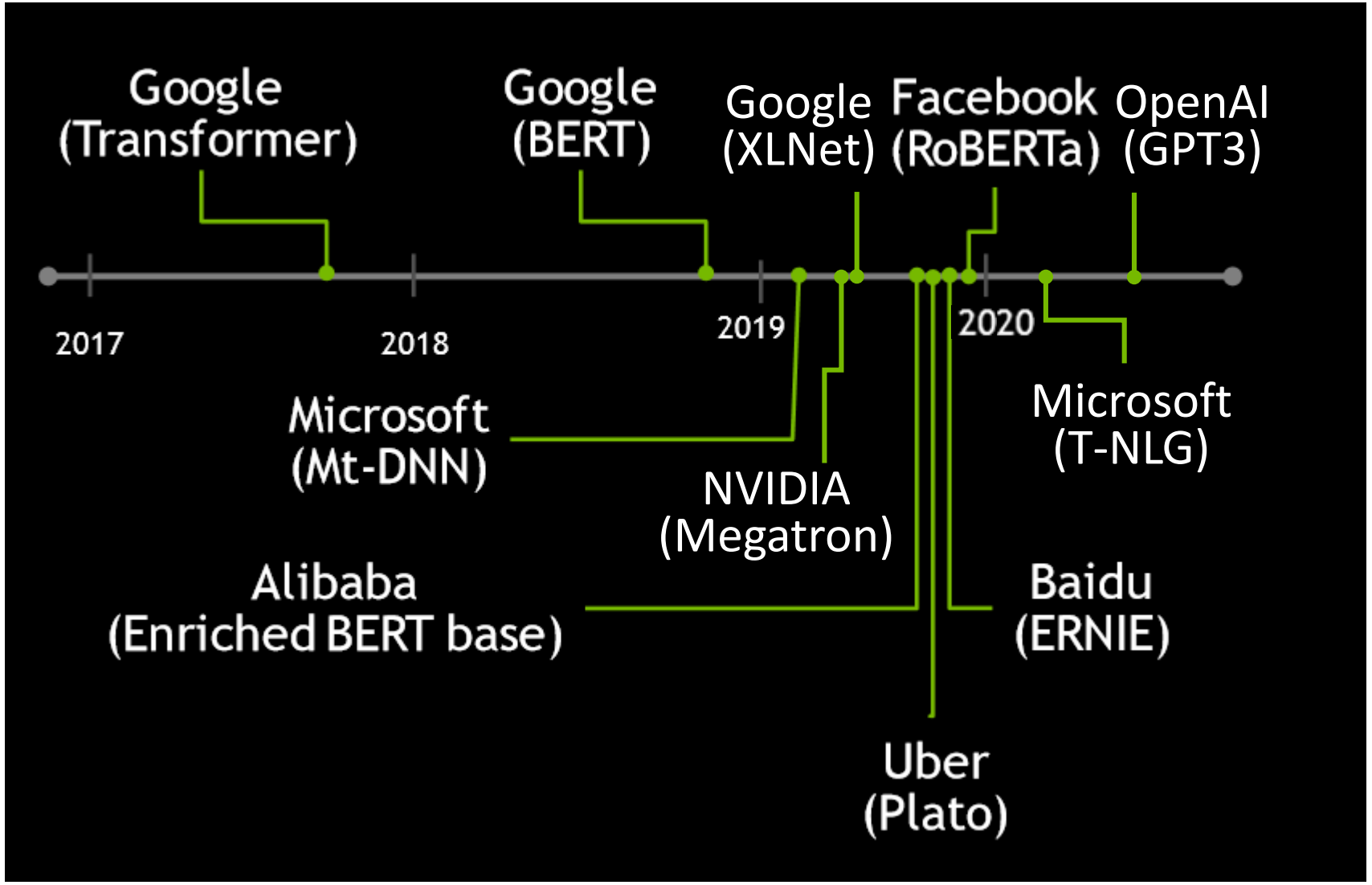

Expectations
AutoML
Chatbots
Digital Ethics
Conversational User Interfaces
Intelligent Applications
Deep Neural Networks (Deep Learning)
Quantum Computing
Graph Analytics
Deep Neural Network ASICs
Machine Learning
Smart Robotics
AI PaaS
Edge AI
NLP
AI Developer Toolkits
VPA-Enabled Wireless Speakers
AI-Related C&SI Services
Explainable AI
Robotic Process Automation Software
Speech Recognition
Data Labeling and Annotation Services
Knowledge Graphics
GPU Accelerators
AI Cloud Services
FPGA Accelerators
Decision Intelligence
Virtual Assistants
Neuromorphic Hardware
Computer Vision
Augmented Intelligence
Insight Engines
AI Governance
Cognitive Computing
Reinforcement Learning
AI Marketplaces
Artificial General Intelligence
Autonomous Vehicles
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
Time
Plateau will be reached:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau
As of July 2019

ENABLER NR. 1: NATURAL LANGUAGE PROCESSING



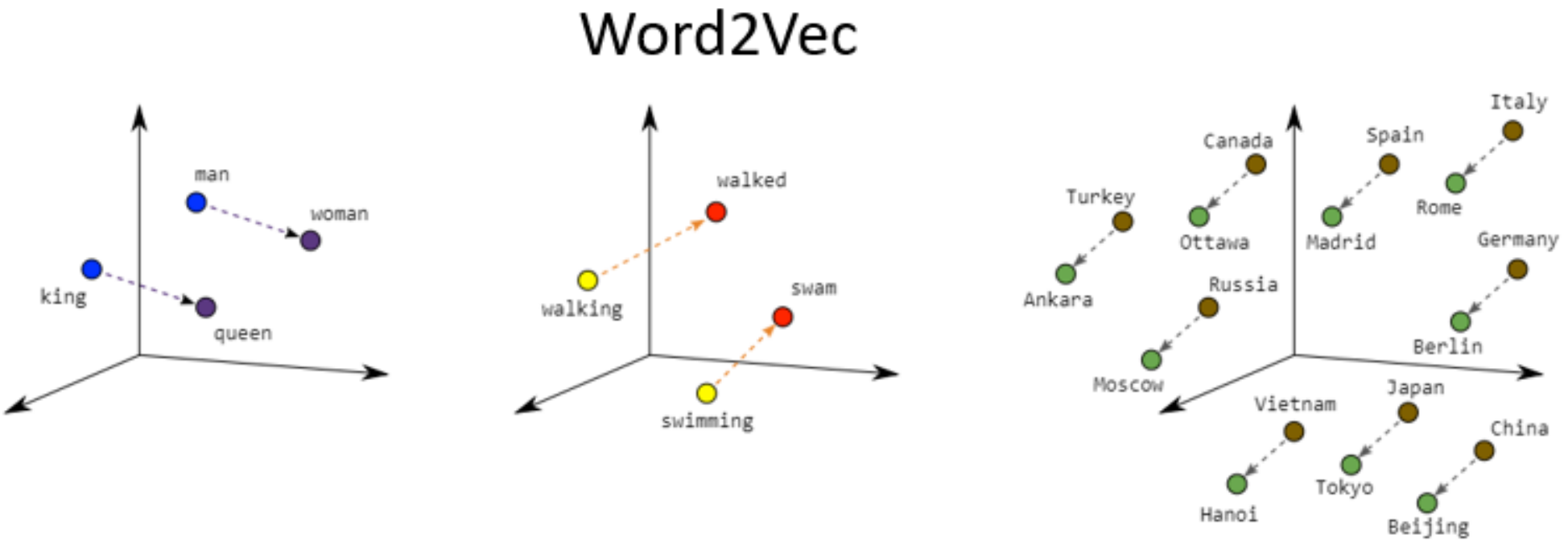
ROKIN

NLP - MILESTONES



ROKIN

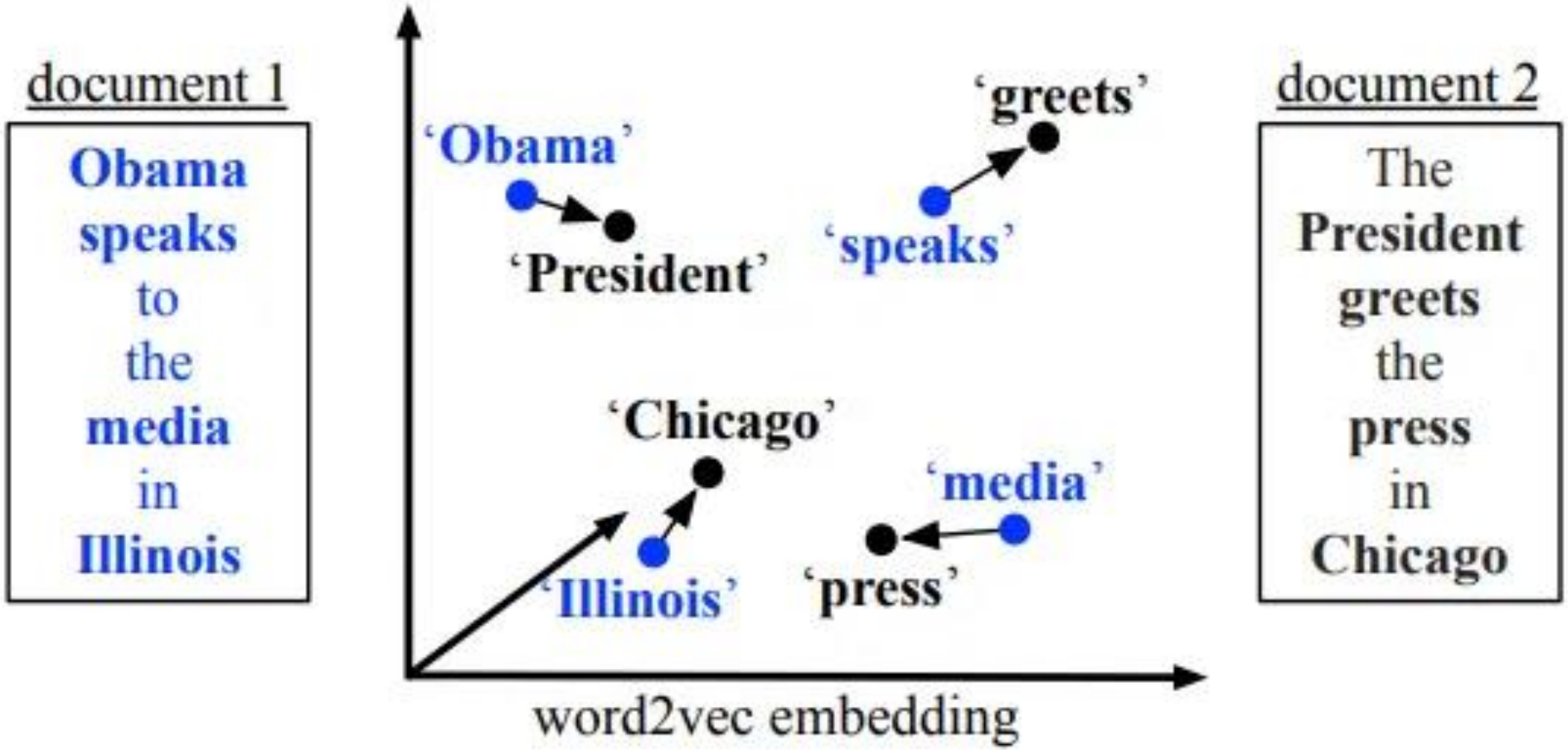
PAPER WORD2VEC FINDING SIMILAR STUFF



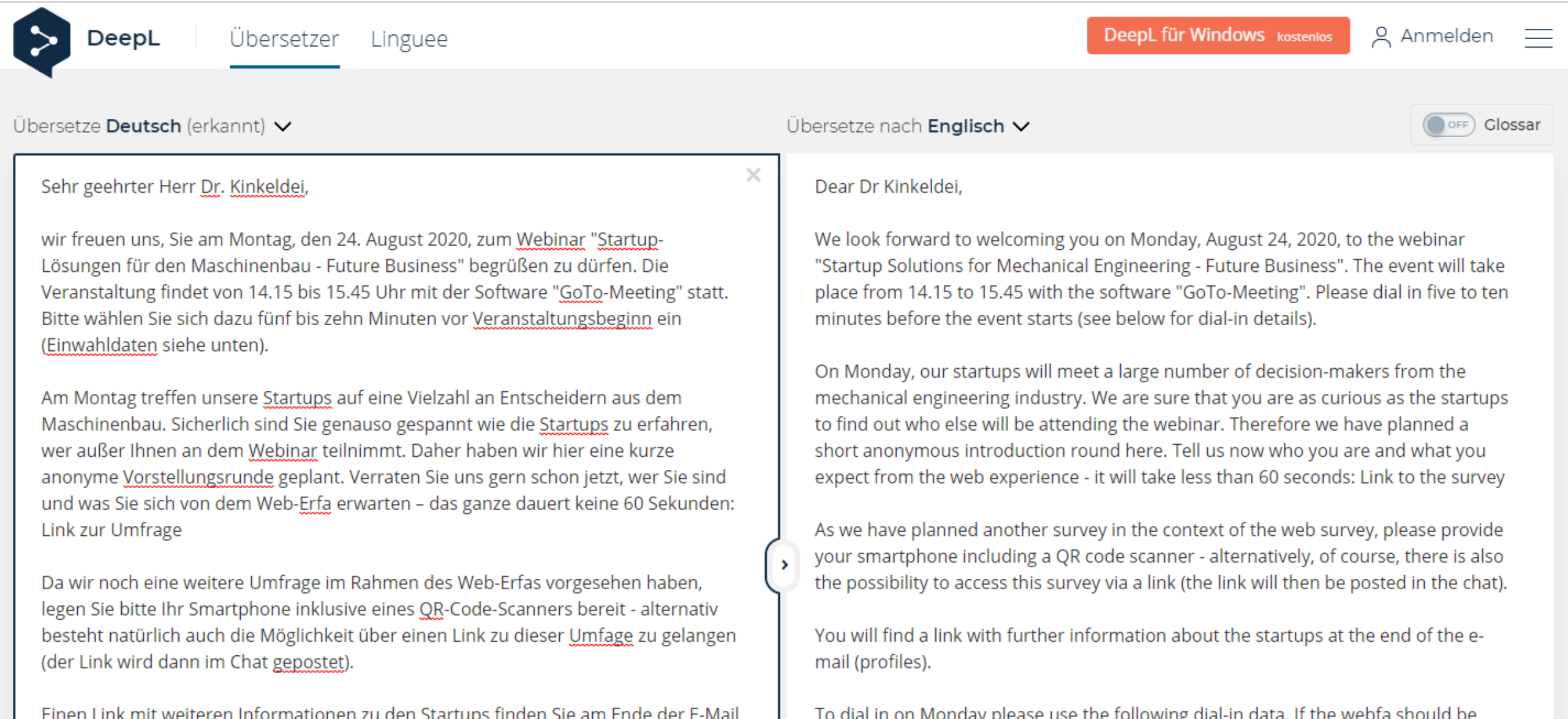
Male-Female

Verb Tense

Country-Capital


document 1
Obama speaks to the media in Illinois
'Obama'
'President'
'greets'
'speaks'
'Chicago'
'Illinois'
'media'
'press'
word2vec embedding
document 2
The President greets the press in Chicago

NLP IN OUR DAILY LIFE

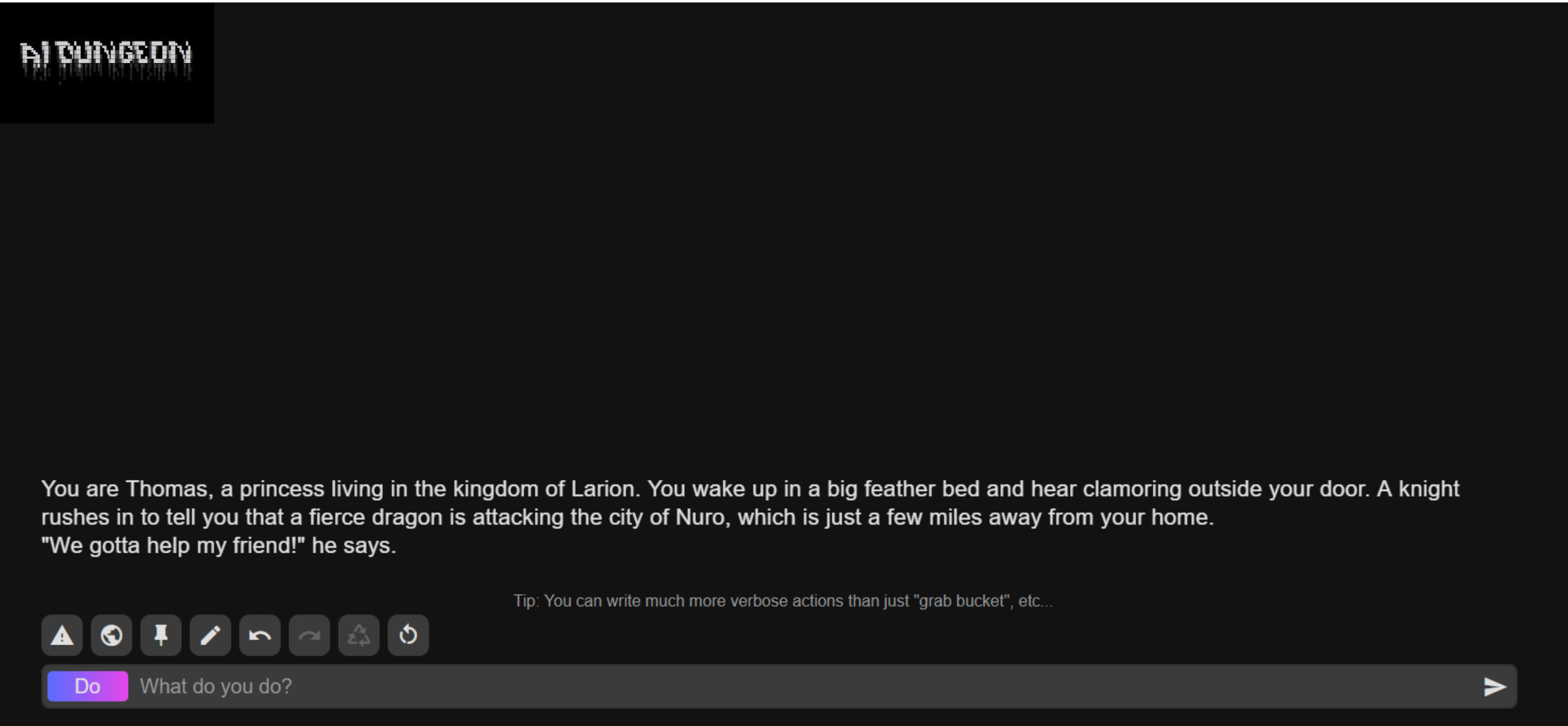


TRANSFORMER IN ZUKUNFT

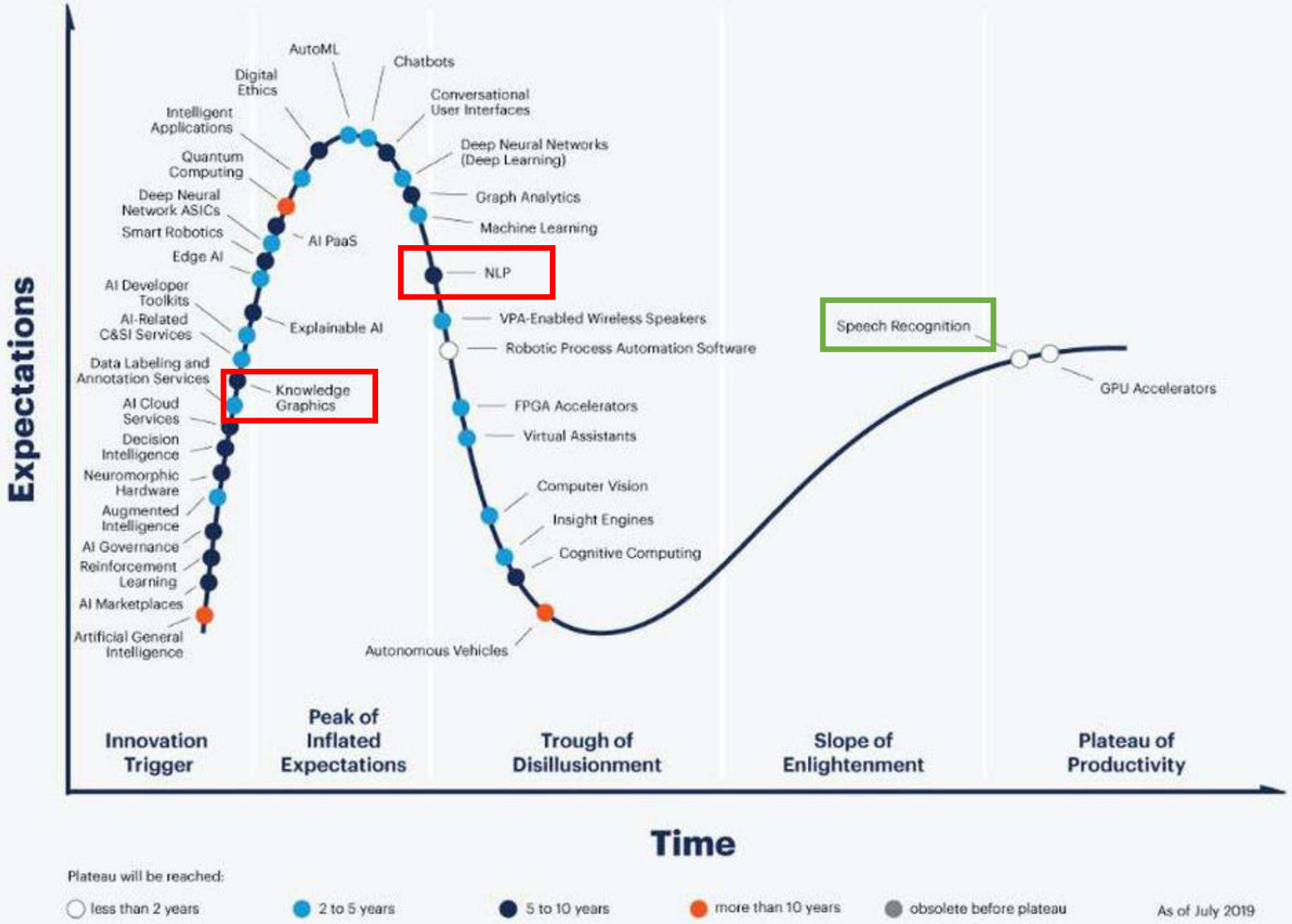
SYSTEM PROMPT (HUMAN-WRITTEN)	*A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.*
MODEL COMPLETION (MACHINE-WRITTEN, FIRST TRY)	The incident occurred on the downtown train line, which runs from Covington and Ashland stations. In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief. “The theft of this nuclear material will have significant negative consequences on public and environmental health, our workforce and the economy of our nation,” said Tom Hicks, the U.S. Energy Secretary, in a statement. “Our top priority is to secure the theft and ensure it doesn’t happen again.” The stolen material was taken from the University of Cincinnati’s Research Triangle Park nuclear research site, according to a news release from Department officials. The Nuclear Regulatory Commission did not immediately release any information. According to the release, the U.S. Department of Energy’s Office of Nuclear Material Safety and Security is leading that team’s investigation.

ROKIN

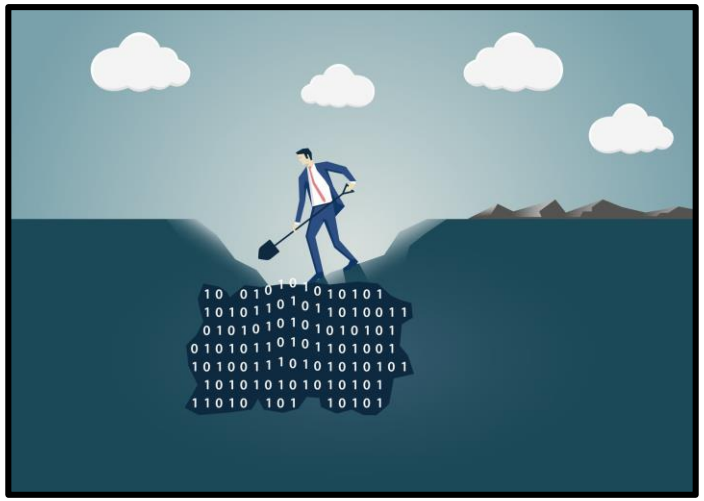
NLP IN OUR DAILY LIFE



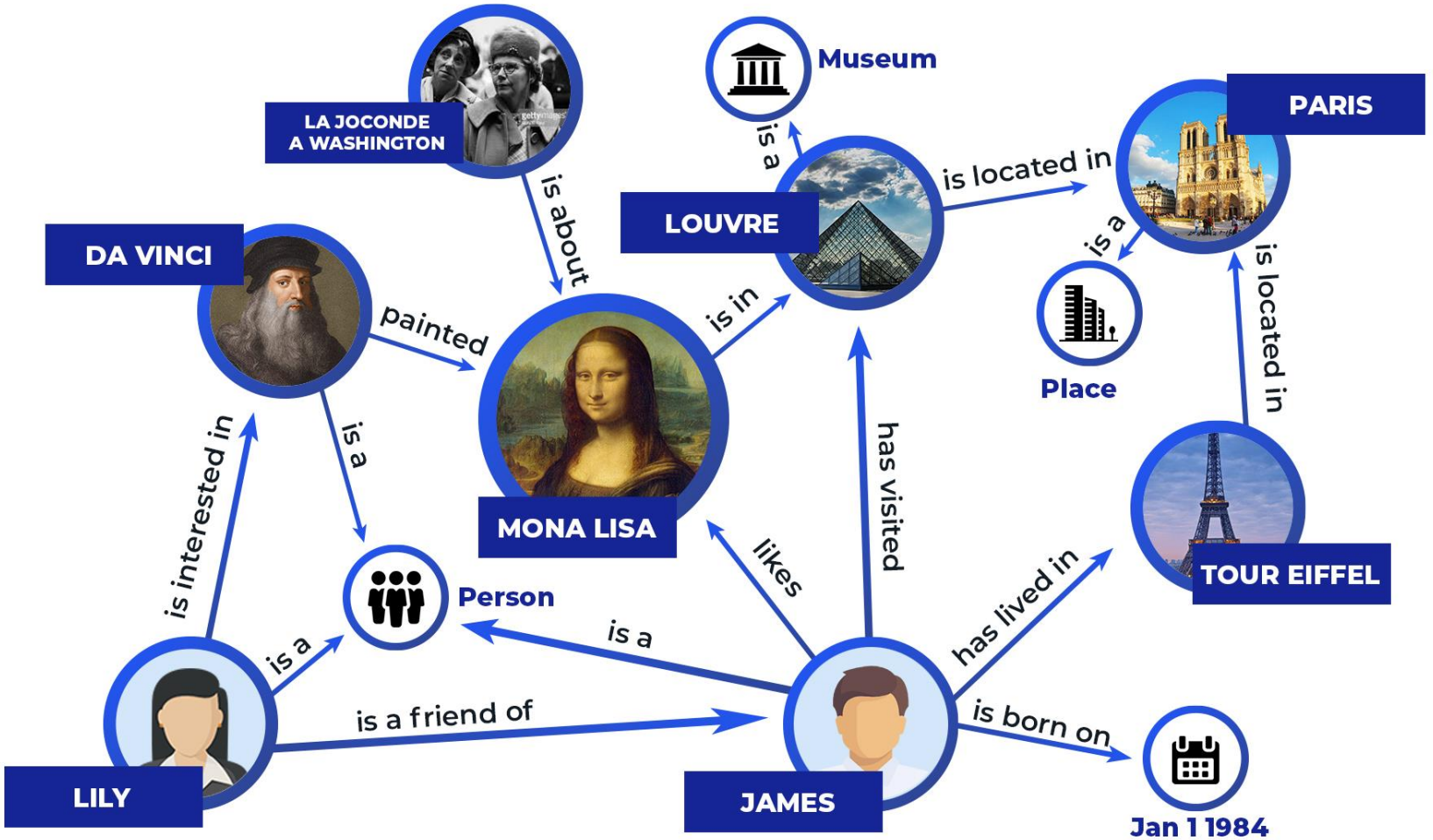
NLP IN OUR DAILY LIFE


Expectations
AutoML
Chatbots
Digital Ethics
Conversational User Interfaces
Intelligent Applications
Deep Neural Networks (Deep Learning)
Quantum Computing
Graph Analytics
Deep Neural Network ASICs
Machine Learning
Smart Robotics
AI PaaS
Edge AI
NLP
AI Developer Toolkits
VPA-Enabled Wireless Speakers
AI-Related C&SI Services
Explainable AI
Robotic Process Automation Software
Speech Recognition
Data Labeling and Annotation Services
Knowledge Graphics
GPU Accelerators
AI Cloud Services
FPGA Accelerators
Decision Intelligence
Virtual Assistants
Neuromorphic Hardware
Computer Vision
Augmented Intelligence
Insight Engines
AI Governance
Cognitive Computing
Reinforcement Learning
AI Marketplaces
Artificial General Intelligence
Autonomous Vehicles
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
Time
Plateau will be reached:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau
As of July 2019

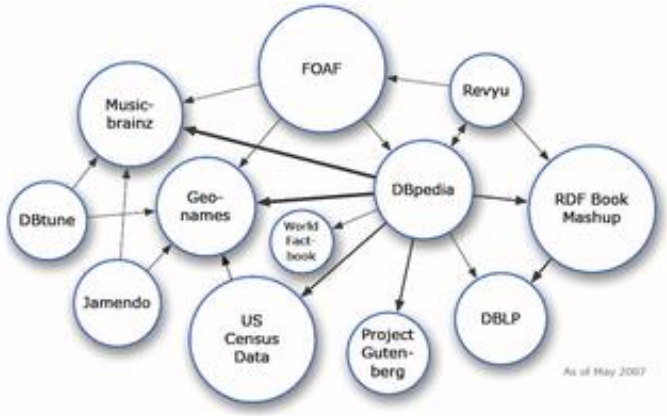
DATA PARADOXON



ENABLER NR. 2: KNOWLEDGE GRAPHS



GROWTH OF KNOWLEDGE GRAPHS

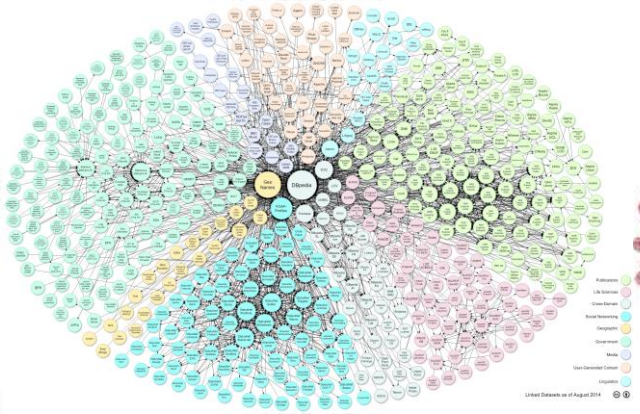


2007

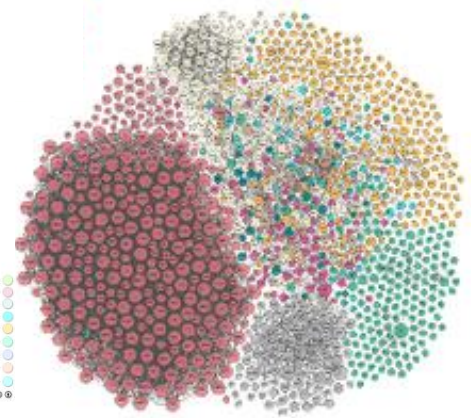
2008

2009

2011



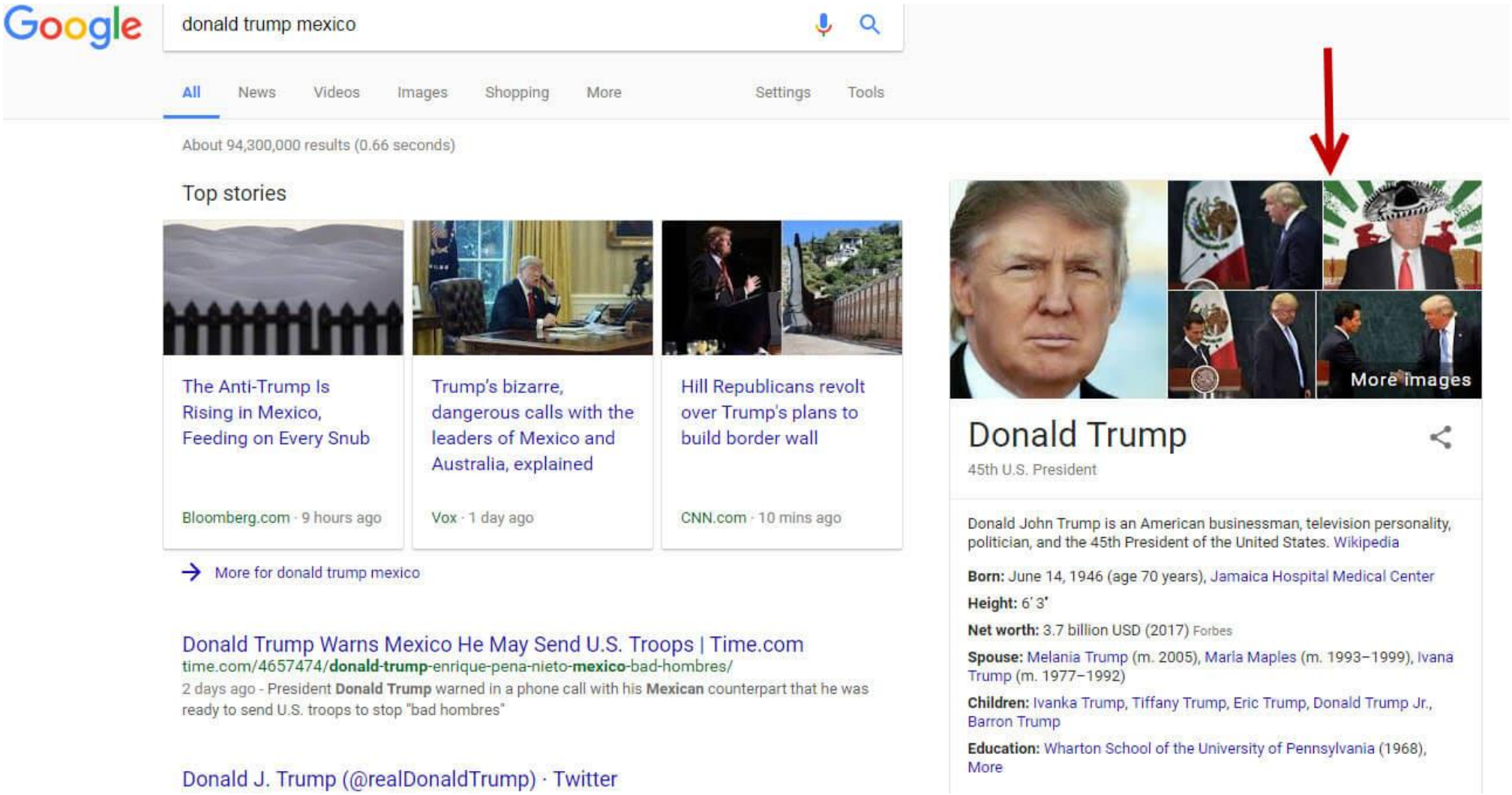
2014



2017

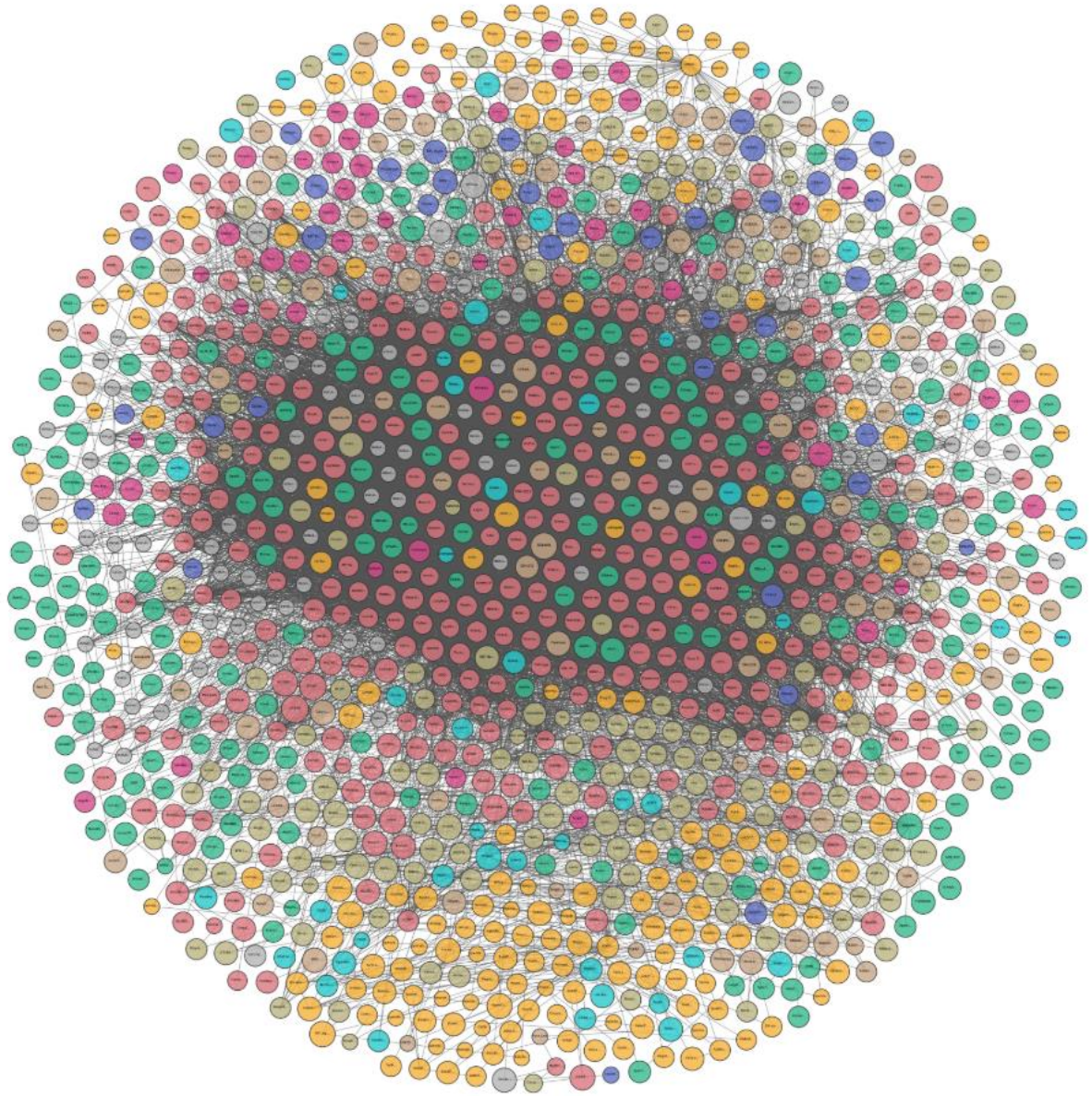
Beispiele von Knowledge Graphen

Google Knowledge Graph in unserem täglichen gebrauch

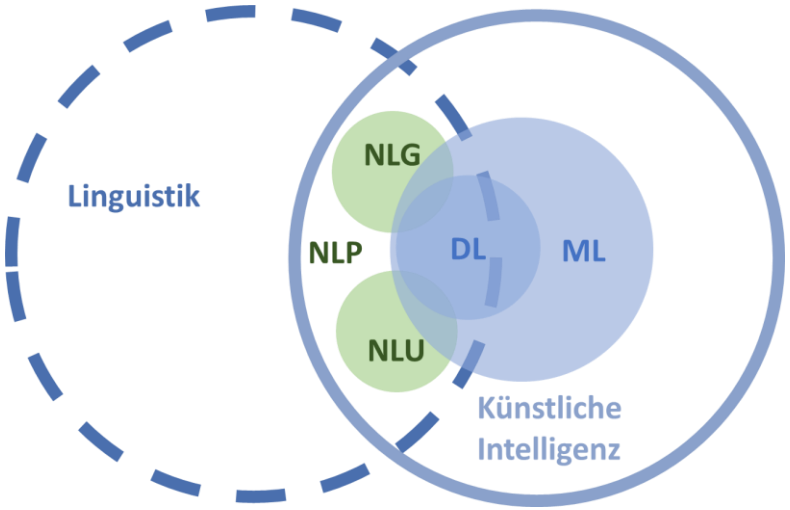


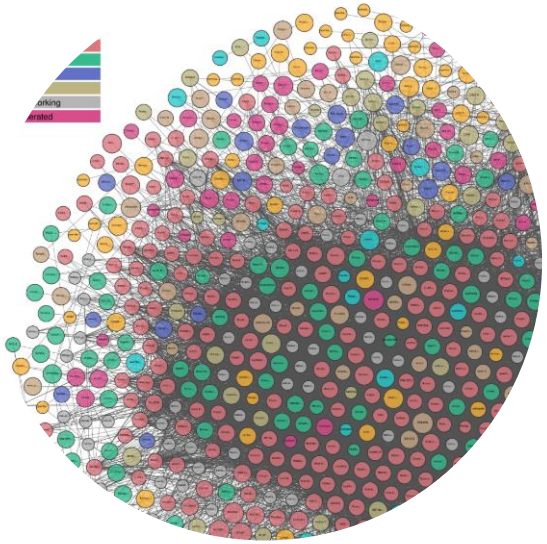
Legend

- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated



NLP FOR INFORMATION RETRIEVAL & MANAGEMENT





ROKIN

PAPER WORD2VEC FOR ENGINEERING

Table 2: Semantically most similar words for "ai" in the Word2Vec model and our Engineering model

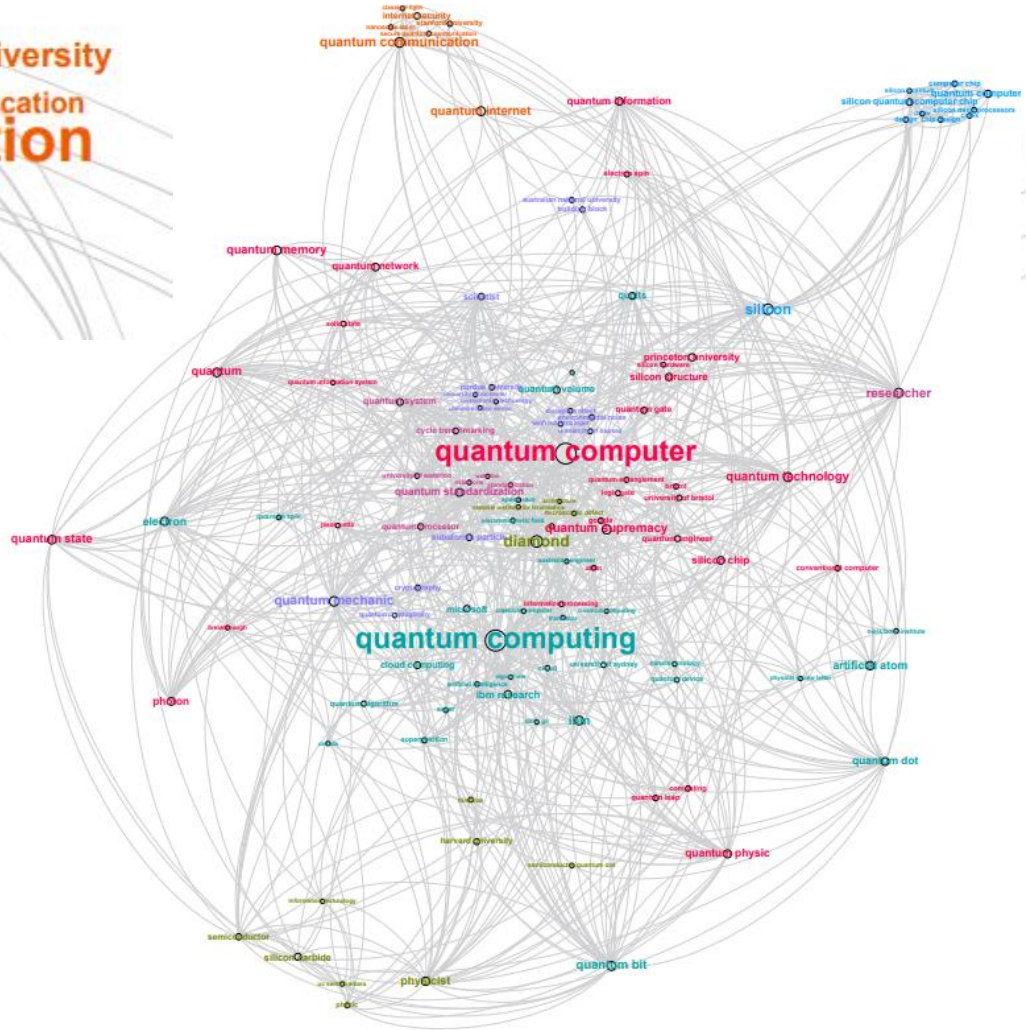
#	Word2Vec	Similarity	Engineering	Similarity
1	che	0.63	artifical_intelligence	0.93
2	te	0.62	machine_learning	0.89
3	essere	0.59	deep_learning	0.85
4	é	0.59	computer_vision	0.77
5	mai	0.59	conversational_ai	0.71
6	voi	0.59	analytics	0.71
7	tutto	0.58	predictive_analtics	0.70
8	ti	0.58	ai_ml	0.70
9	tutti i	0.58	cognitive_computing	0.70
10	questo	0.58	image_recognition	0.67

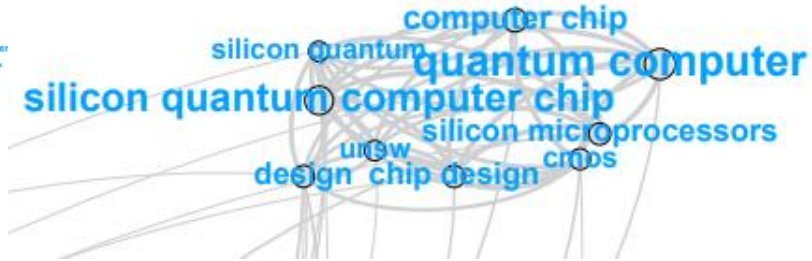
RESEARCH ARTICLE CLUSTERING WITH TRANSFORMER

clideo.com

ROKIN

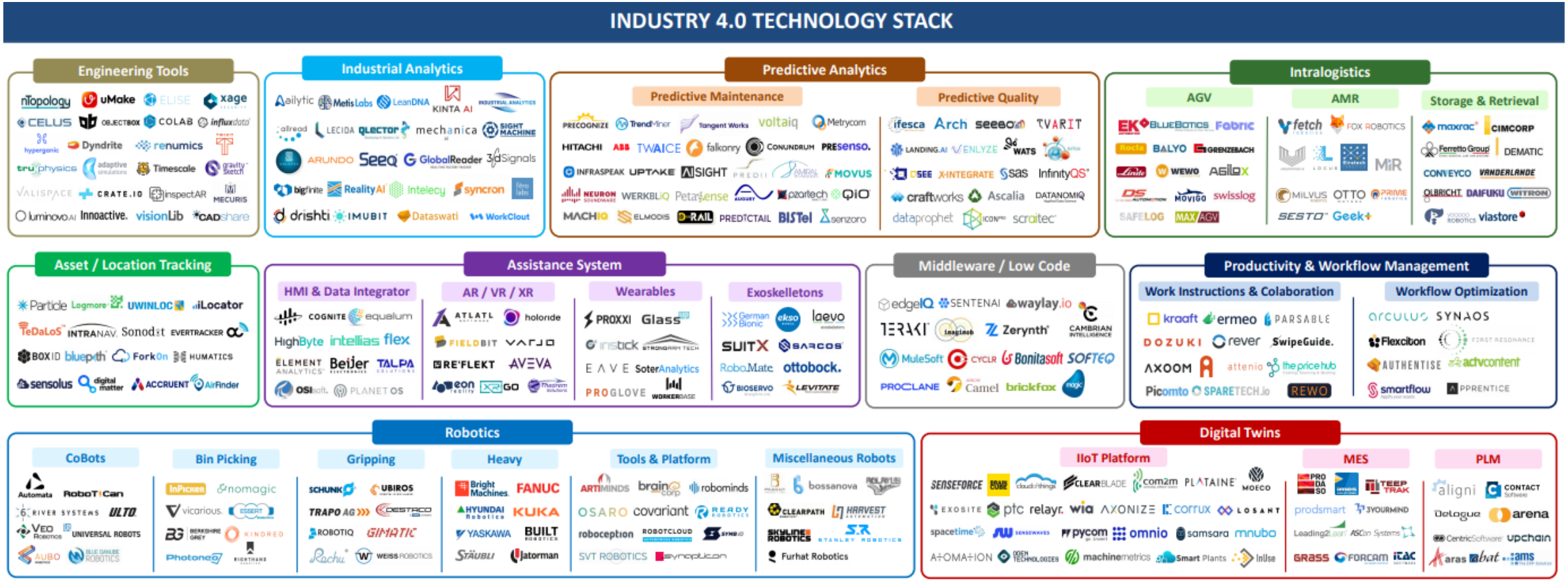
INFORMATION EXTRACTION





ROKIN

TECHNOLOGIE MAPPING



https://www.rokin.tech/industry40technologystack

ROKIN

LIMITATIONS OF NLP AND KG