Generative AI & Its application’s

Infineon Technologies

Agenda

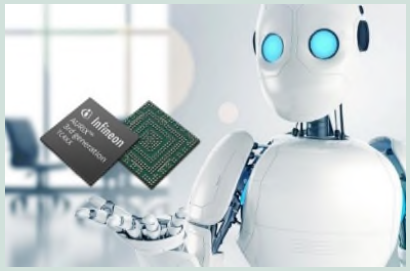
1. Introduction
2. What do we do @ Infineon ?
3. What is AI/ML/DL ?
4. NLP (Natural Language Processing)
5. Generative AI & LLM’s
6. Limitations
7. References & Glossary

Agenda

1. Introduction
2. **What do we do @ Infineon ?**
3. What is AI/ML/DL ?
4. NLP (Natural Language Processing)
5. Generative AI & LLM’s
6. Limitations
7. References & Glossary

Our work @Infineon

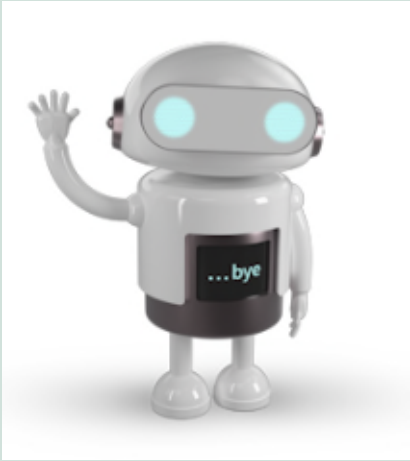
Using the power of NLP* improve process Efficiency



Using various NLP techniques to bring values from the textual and unstructured dataset from different domains.

Various applications:

- Q&A Bot
- Sentiment analysis
- Search engines
- Text analytics
- Speech-to-Text etc.

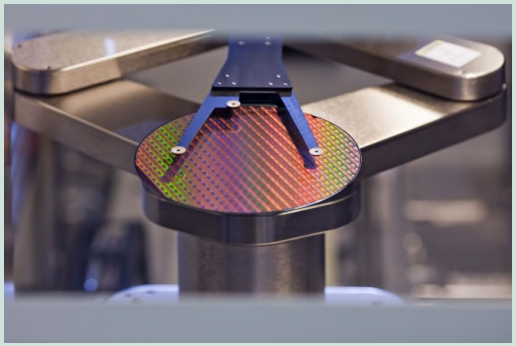


Enabling CV* in Manufacturing

Using CV techniques to automate & improve the decision making at the chip manufacturing process

Various applications:

- Wafer crack detection
- Wafer defect detection
- Similar wafer images
- Image generation etc.



NLP - Natural Language Processing
CV - Computer Vision

Our work @Infineon

Exploring the AI in EDA* to improve R&D Process

Using AI/ML techniques to enhance the R&D Design, make the process more efficient, and improve the productivity of the designers

Various applications:

- Analog structure recognition
- Congestion prediction
- Anomalies detection
- Time-series forecasting etc.

Bringing AI on Edge

Providing the platform for our customers deploy the AI/ML* models on Infineon Products by automating the process of Quantization, Pruning etc.

Various applications:

- Wake-word detection
- Lidar object detection
- Radar-based object detection
- Image recognition etc.

In-house MLOps Infrastructure

Developing the in-house MLOps Infrastructure to provide scalable infrastructure to develop and host AI/ML Models

Agenda

1. Introduction
2. What do we do @ Infineon ?
3. **What is AI/ML/DL ?**
4. NLP (Natural Language Processing)
5. Generative AI & LLM’s
6. Limitations
7. References & Glossary

What is AI /ML /DL ?

Artificial Intelligence (AI)

Programs with the ability to learn and reason like humans
E.g.: A* Algorithms, Reinforcement Learning etc.

Machine Learning (ML)

It is a subset of AI in which we build algorithms with the ability to learn from data using statistics and minimal rules.

Deep Learning (DL)

Subset of ML in which neural network architectures are adopted to learn from the vast amount of data

Data Science (DS)

DS is a field of study which combines Statistics and Math. extract meaningful insights from data

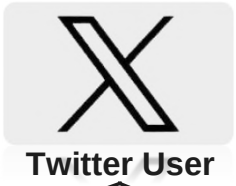
~90%

90% of World's Data generated in last 10 Years

Rest of the **10%** of the data is generated in last 6 million years

AI and Data Science intersect to automate the extraction of insights from data and make better predictions from large datasets.

Sub-domains of AI



Twitter User

Tweets, comments

Images, videos

No. of followers, retweets, likes

Natural Language Processing

Connects human language to computers and involves all kinds of ***text processing, analysis and generation***.

Ex.
- Sentiment analysis
- Text generation

Computer Vision

Handles the processing, analysis and ***generation of images and videos***.

Ex.
- Facial classification/detection
- Image-to-image generation etc.

Data Science

Opens the door to process and analyze huge amounts of data to ***extract useful insights*** from them.

Ex.
- Recommendation systems
- Social media analysis

Agenda

1. Introduction
2. What do we do @ Infineon ?
3. What is AI/ML/DL ?
4. **NLP (Natural Language Processing)**
5. Generative AI & LLM’s
6. Limitations
7. References & Glossary

Introduction to Natural Language Processing (NLP)

Human-To-Human Communication

Hey, what's your favorite movie?

Human brain Interprets the message and gives the response

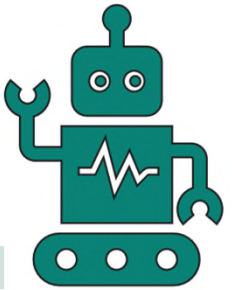
Hmm, that's tough! I really like "The Shawshank Redemption". It's a great story about hope and redemption.

Human-To-Machine Communication

Can you recommend a good movie to watch?

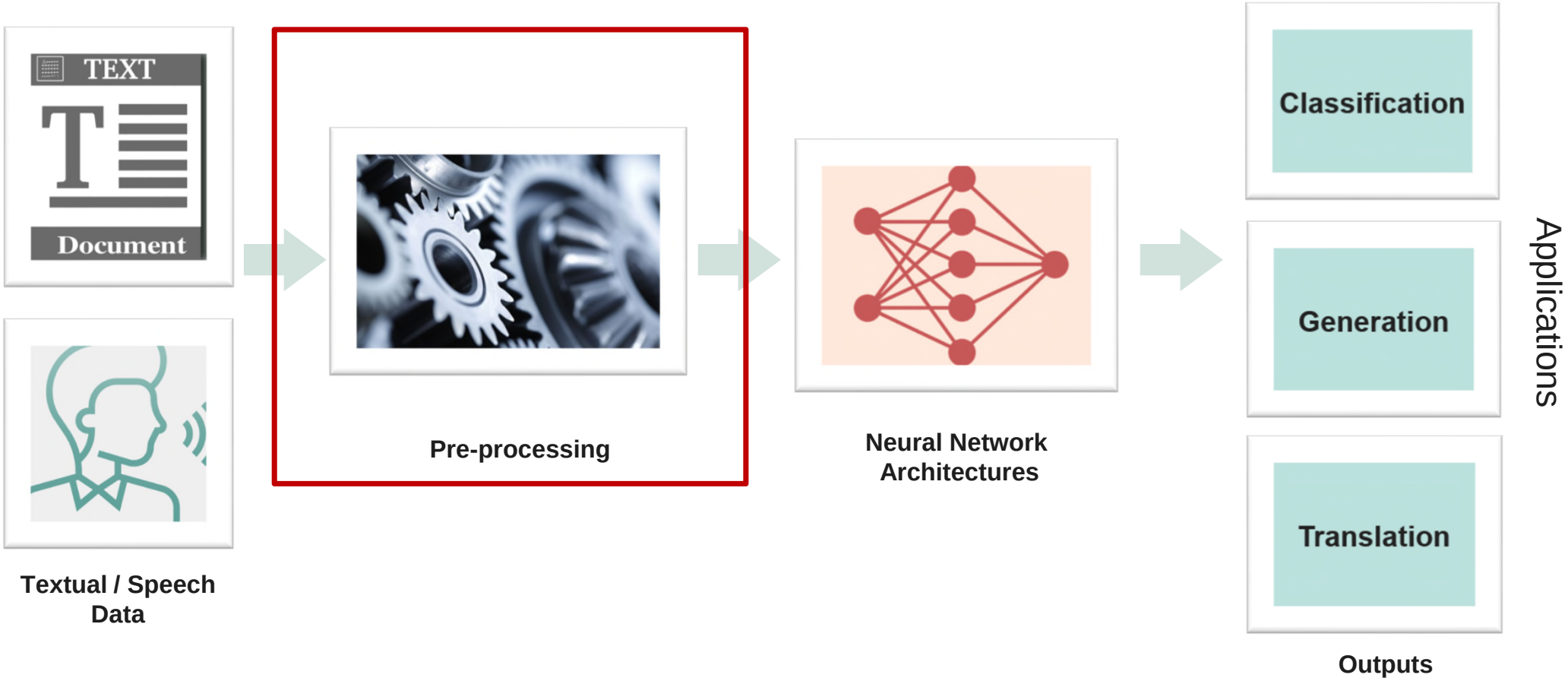
NLP allows the machine to understand & communicate with humans in Natural Language

NLP



Based on your viewing history, I would suggest "The Shawshank Redemption" or "Forrest Gump". Which one do you prefer?

How does NLP work?



Text Pre-processing Techniques

Text Pre-processing

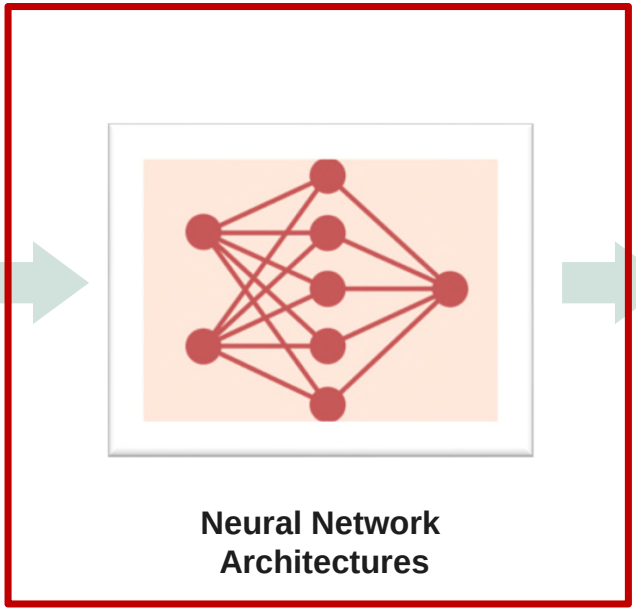
Text Pre-processing Techniques

- Noise Removal
- Tokenization
- Stop Word Removal
- Stemming & Lemmatization

How does NLP work?

Textual / Speech Data

Pre-processing



Neural Network Architectures

Classification

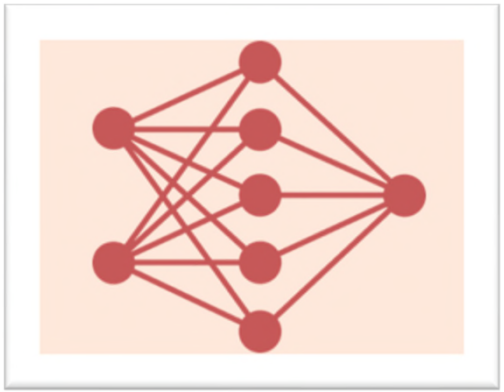
Generation

Translation

Outputs

Applications

NLP Neural Network Architectures



Neural Networks

Neural Network Architectures

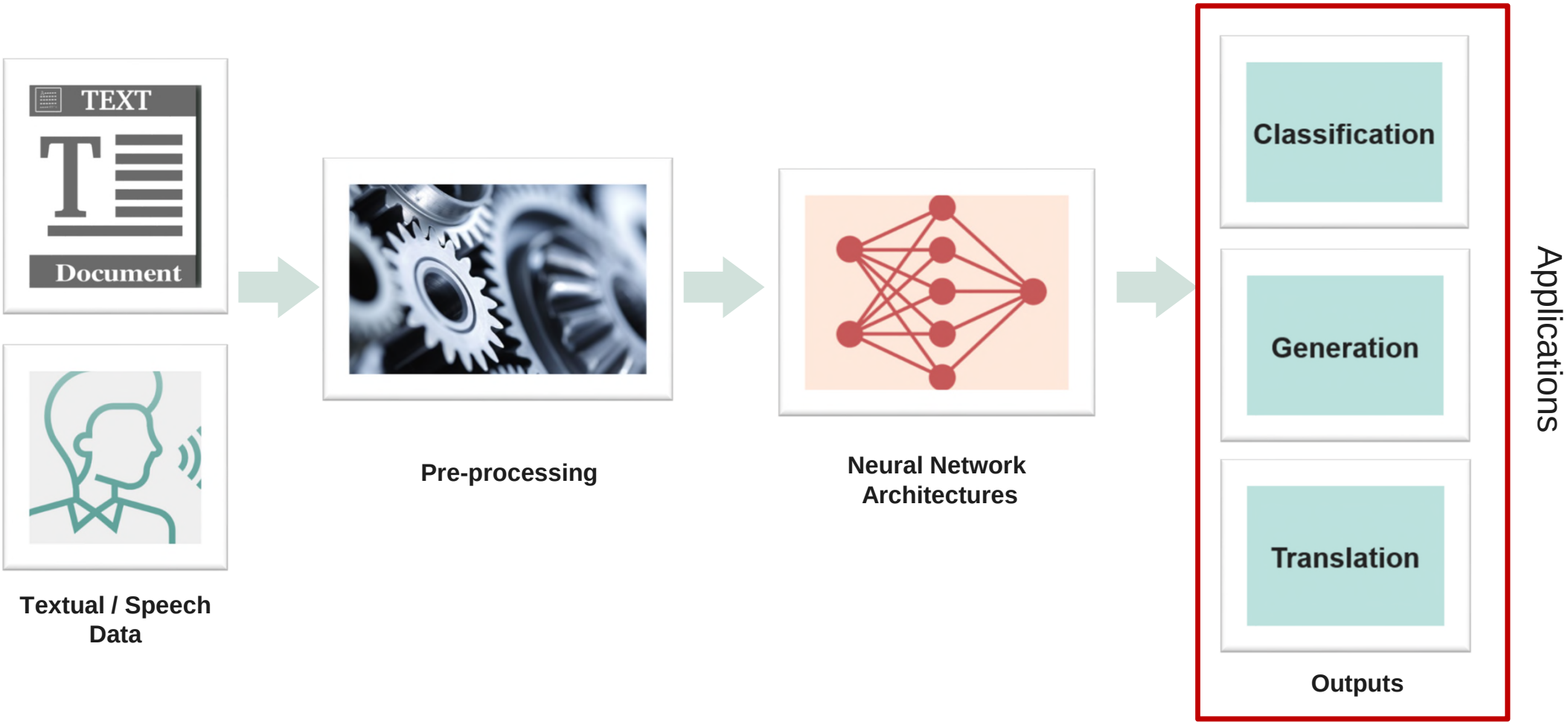
LSTMs

RNNs

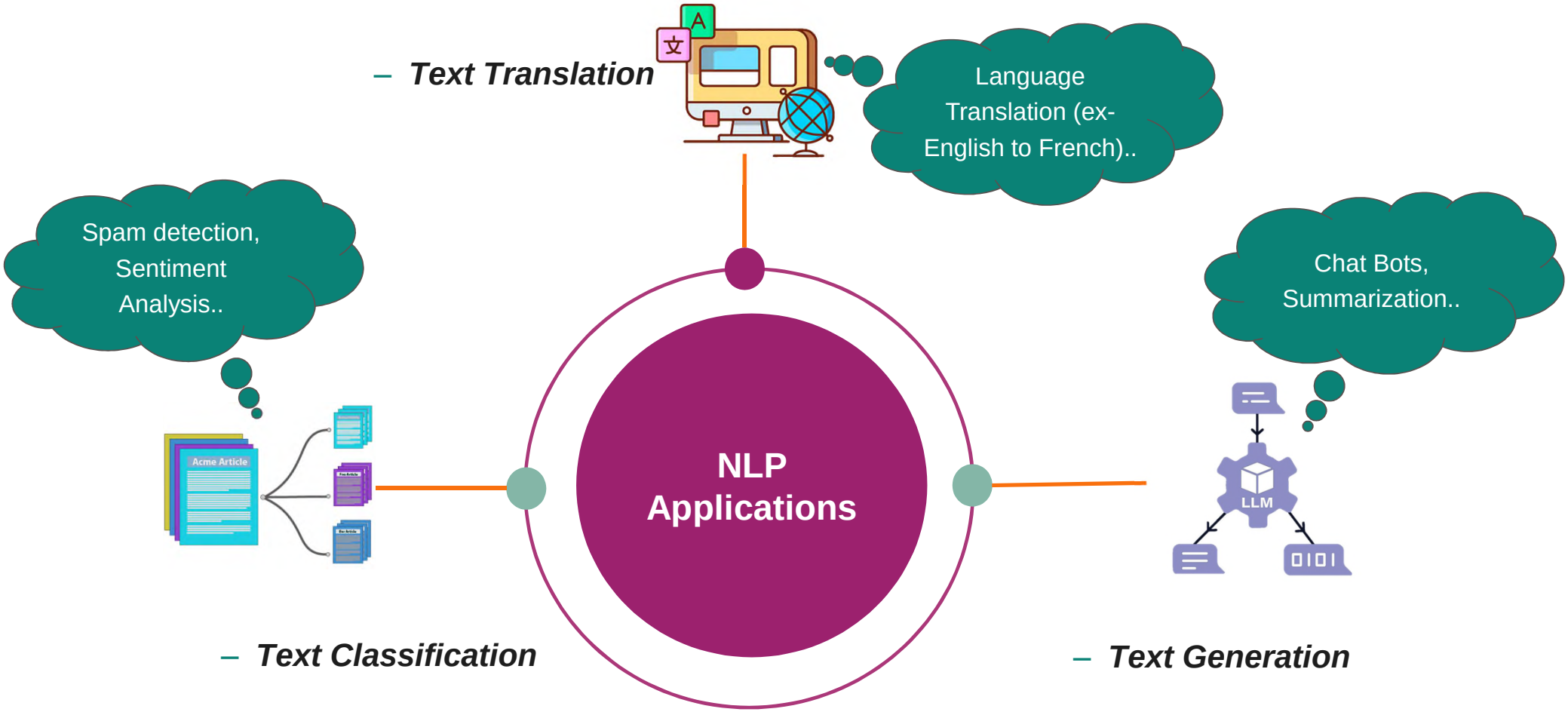
Transformers

Encoder - Decoder

How does NLP work?



Common NLP Applications



Agenda

1. Introduction
2. What do we do @ Infineon ?
3. What is AI/ML/DL ?
4. NLP (Natural Language Processing)
5. **Generative AI & LLM’s**
6. Limitations
7. References & Glossary

What is Generative AI?

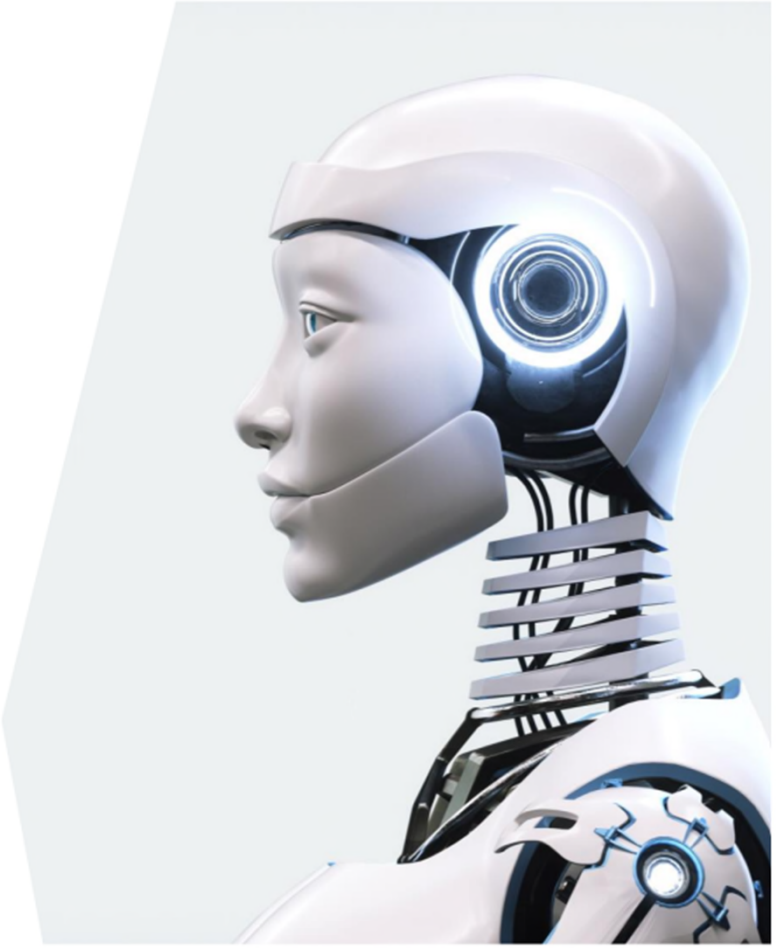
What is Generative AI?

- Type of artificial intelligence capable of generating new content, e.g. text, images, code, music and video
- Within GenAI, **transformer models** in the form of **Large Language Models** (LLMs) are currently the most advanced
- Trained to predict the next token of a sequence, e.g. with text, keeping relationships in memory and thereby creating context:

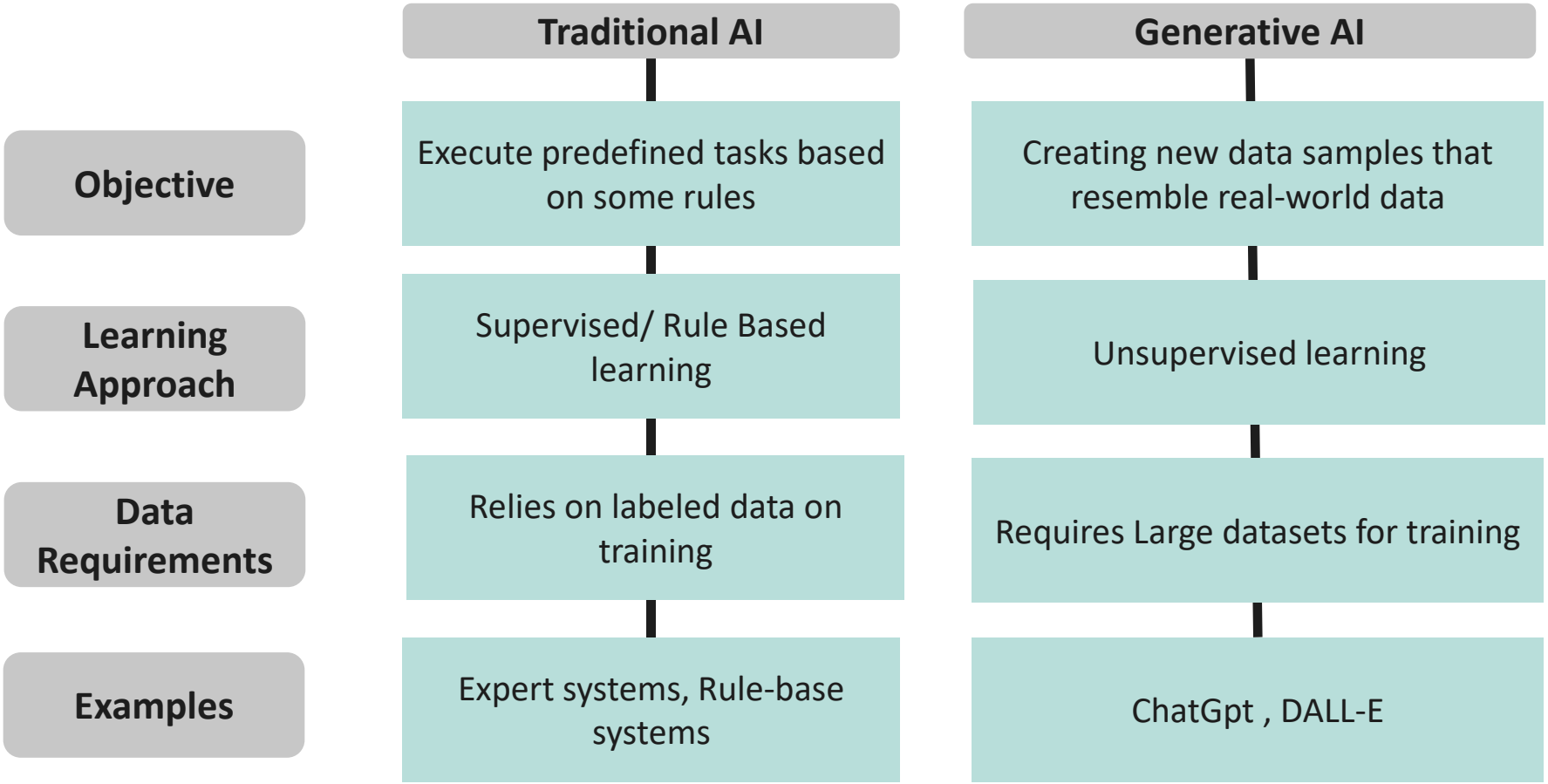
„AI will change _____“ → → „the“, „world“

What are possible risks?

- False output – “hallucinations”
- Bias and lack of fairness
- Leakage of confidential data and IP
- Copyright infringements
- Privacy and Data Protection
- Cyber security



Traditional AI Vs Generative AI

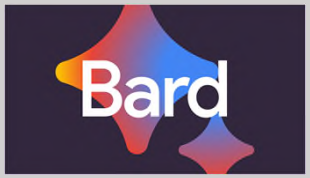


Types of Generative Models

Based on data they generate from text input

Text-to-Text

- ChatGPT
- Bard
- LLaMa (Meta)
- PaLM2



Text-to-Code

- Codex
- AlphaCode

LLM's

Text-to-Image

- Midjourney
- DALL-E3
- Stable Diffusion
- Imagen

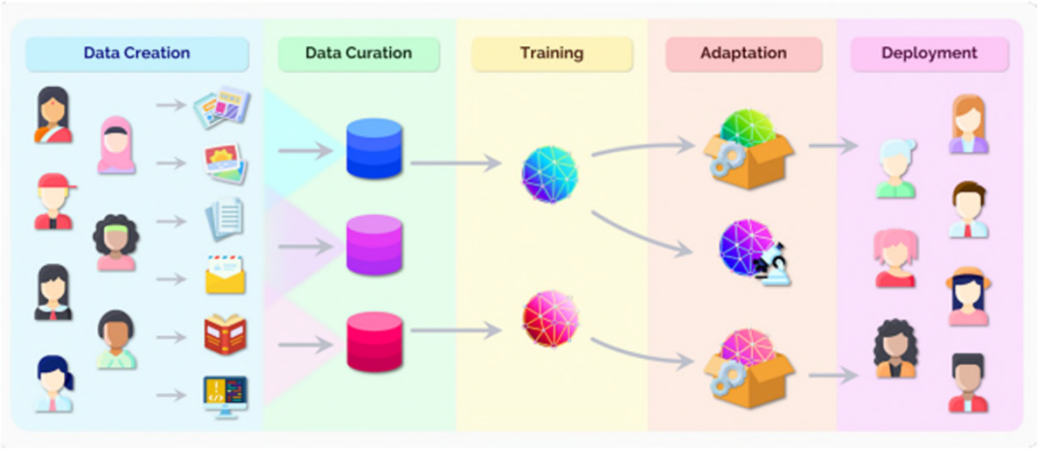
Text-to-Audio

- AutoLM
- LukeBox

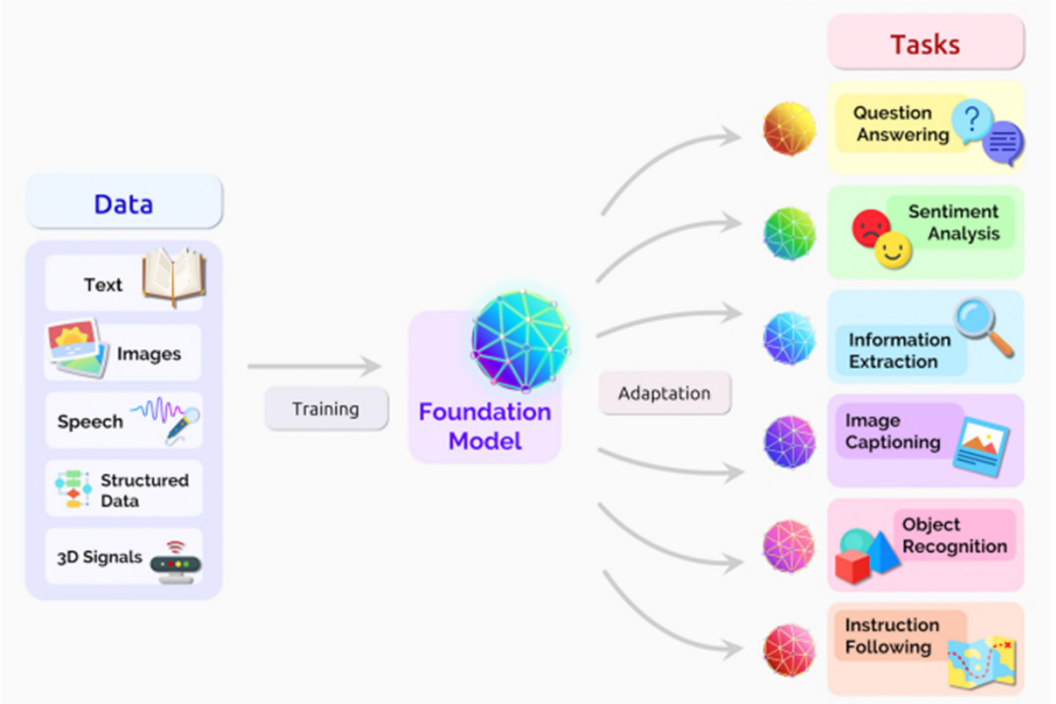
Large Language Models (LLM's)

Large language models are AI models that can generate human-like text based on vast amounts of training data. They are capable of understanding and producing natural language.

Before LLM's



After LLM's



Language Models Evolution

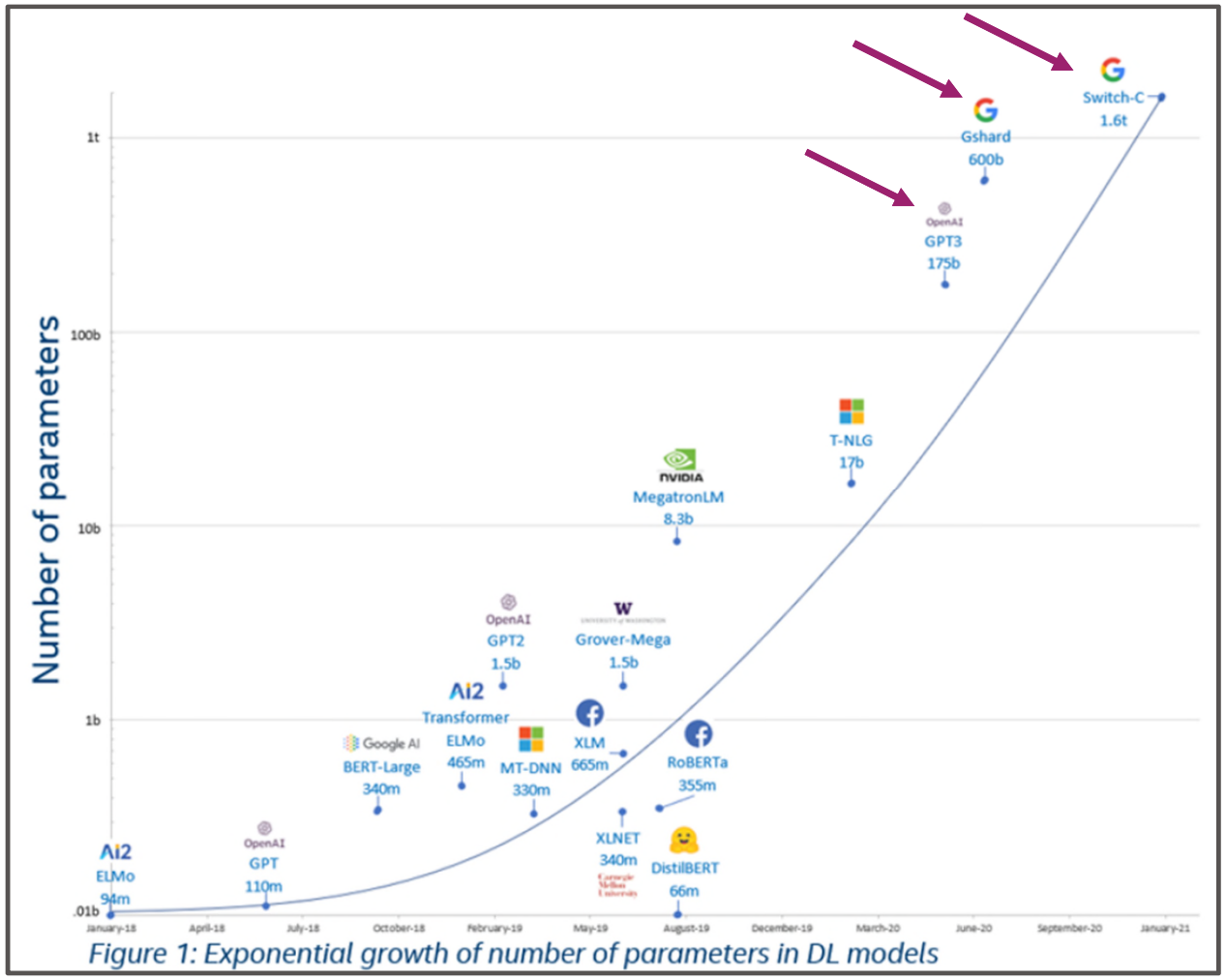


Figure 1: Exponential growth of number of parameters in DL models

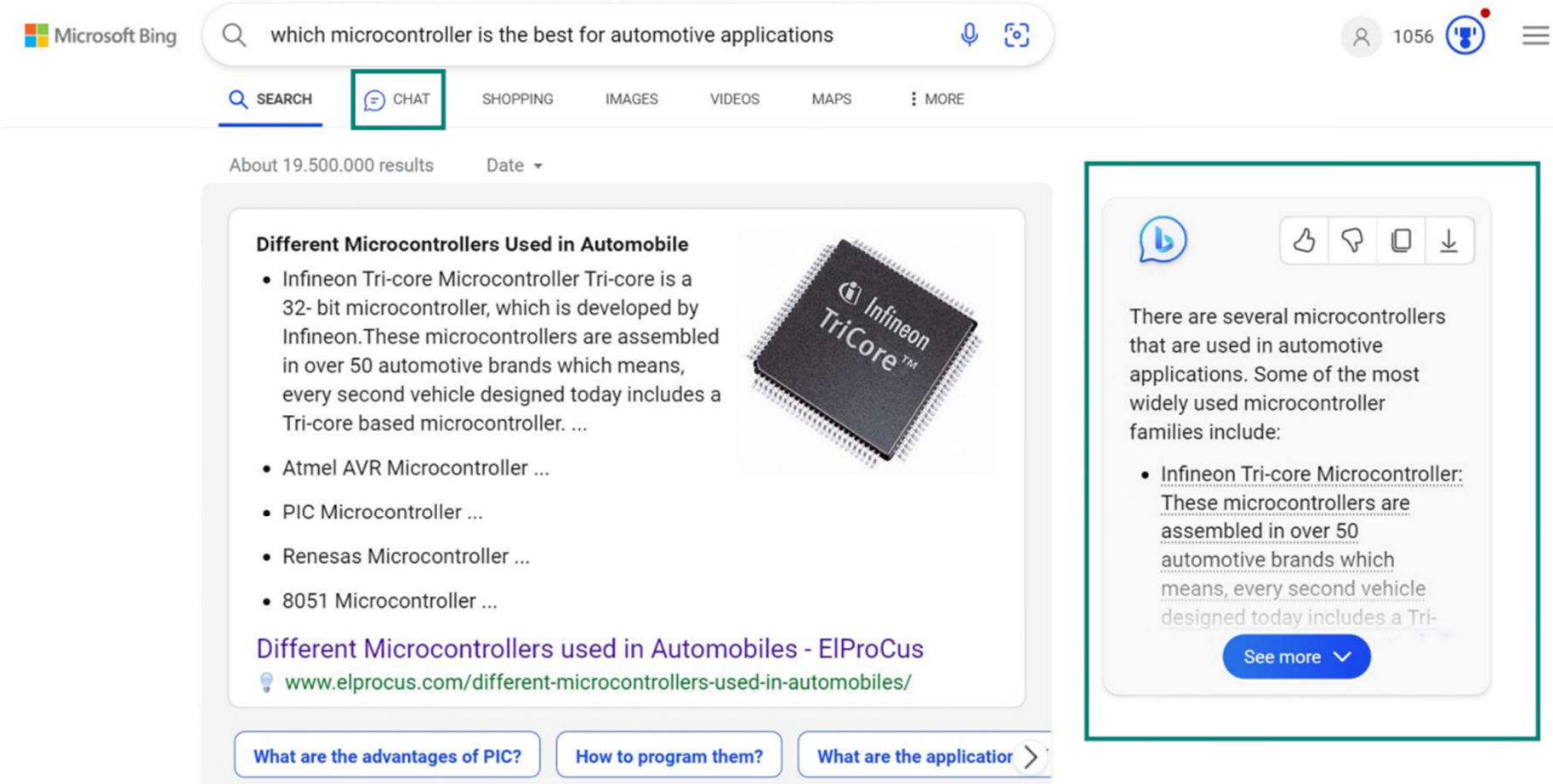
Now we have Large Language Models like:

- Bloom (BigScience) : 176 B
- LaMDA (Google): 137 B
- Flan-T5 XXL (Google): 11 B
- GPT4 (OpenAI): NA

Source

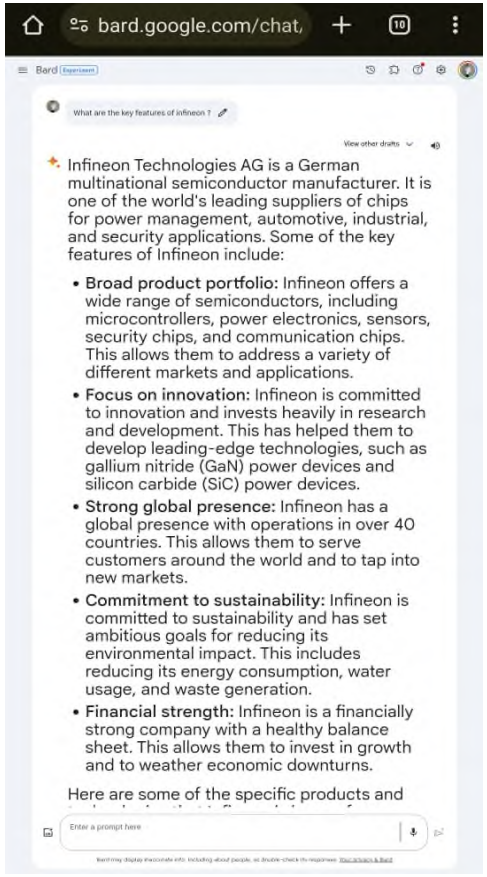
Search Engines

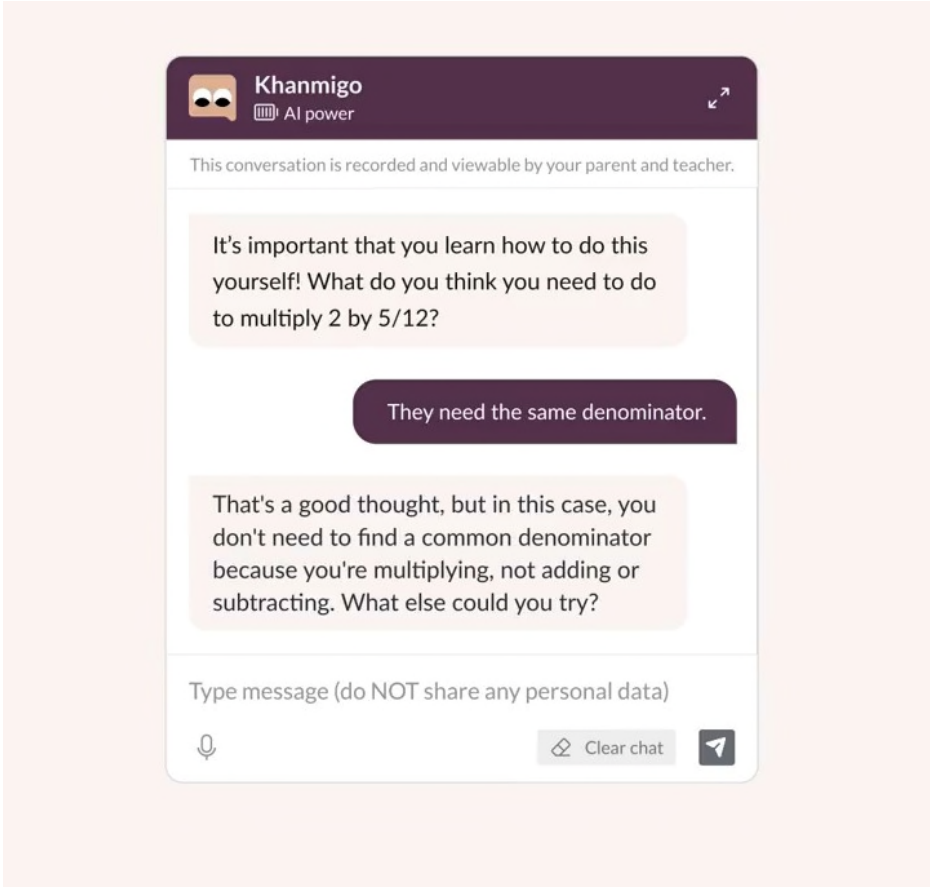
Microsoft Bing: AI-generated responses already complement traditional search results



Virtual Assistants and Customer Support

- Google Bard: Answers the user queries.
- Khan Academy: AI Assistant for students to help with the queries to understand the concepts better.





Other Potential Applications

Coding Support

- Code generation and transformation
- Analysis, debugging, and optimization
- Documentation

Information Retrieval and Content Synthesis

- Quick access to information and knowledge across large repositories of structured and unstructured data
- Translation, summarization

Content Generation

- Marketing content, SEO content
- Meeting minutes and video transcripts
- Documentation, presentations
- Standardized reports and notes

Common Myths about LLM's

LLM's **can provide any answer**	While large language models can generate impressive responses to many questions, **they are not omniscient and have limitations.**	For example, if a large language model is asked a question that requires information that is **not present in its training data**, it may not be able to provide an accurate or useful answer.
LLM's **can work for any use case**	While large language models are incredibly powerful tools, they have specific strengths and weaknesses that make them **better suited for certain use cases over others.**	For example, they **may not be as effective** for tasks that require a deep understanding of the underlying data**, such as scientific research, domain specific data search, data analysis.**
Providing company data to LLM's can **automatically create a model** that can answer all company-related questions	While large language models can process and analyze large amounts of data, creating a useful and accurate model requires **careful planning, data preprocessing, and model training.**	For example: if you want your LLM to answer your all queries like ChatGPT – in that case you **need to train** your LLM model with your own company **data in the pre-defined format** i.e. required by LLM's.

Agenda

1. Introduction
2. What do we do @ Infineon ?
3. What is AI/ML/DL ?
4. NLP (Natural Language Processing)
5. Generative AI & LLM’s
6. **Limitations**
7. References & Glossary

Limitations of Generative AI

Hallucination

Can be caused by overfitting, incomplete, or inaccurate training data

Bias and Fairness

Insufficient data or Human generated Biased data

Copyright Infringement

Who owns the content that generative AI creates?
Does copyright, patent, trademark infringement apply to AI creations?

Limited Knowledge

Limited training leads to limited range of outputs

Agenda

1. Introduction
2. What do we do @ Infineon ?
3. What is AI/ML/DL ?
4. NLP (Natural Language Processing)
5. Generative AI & LLM’s
6. Limitations
7. **References & Glossary**

Glossary

Short Form	Description
AI	Artificial Intelligence
ML	Machine Learning
CV	Computer Vision
NLP	Natural Language Processing
DL	Deep Learning
CNN	Convolutional Neural Network

References

https://medium.com/@prajwalkankate/what-is-genai-7e10f008e749

https://en.wikipedia.org/wiki/Hallucination_%28artificial_intelligence%29

https://mlmlovevs.live/product_details/4510993.html

https://www.pycodemates.com/2023/08/top-5-generative-ai-tools-and-uses.html#google_vignette\

https://datasciencedojo.com/blog/embeddings-and-llm/

https://www.analyticsvidhya.com/blog/2023/12/generative-ai-vs-traditional-ai/


Infineon