

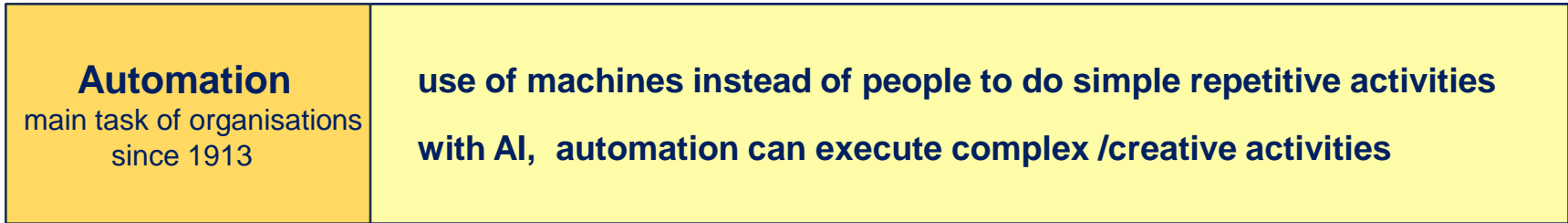


Status of AI and Gen AI applications in Procurement and Supply Chain Management

Giovanni Atti

Former President of ADACI – Member of the Board of IFPSM





Developent of automation from 1965 to 2050

Technology	First movers	Increase of annual productivity
<ul style="list-style-type: none"> ● Basic automation¹ <small>manufacturing + main company processes</small> 	1965 - 2010 <small>1st movers- late movers</small>	0,3% ÷ 1.2% <small>≠ levels of adoption</small>
<ul style="list-style-type: none"> ● Digitalization Industry 4.0 <small>enabling technologies</small> 	2016 - 2035	0.6% ÷ 1,6%
<ul style="list-style-type: none"> ● Smart automation <small>Digitalization strenghtened by AI</small> 	2022 - 2040	0.2% ÷ 1.2%
<ul style="list-style-type: none"> ● Hyperautomation <small>Automation of all company processes extended to key customers /suppliers and strenghtened by AI</small> 	2025 - 2050	0.3% ÷ 1.2%



With reference to the application moment Gartner distinguishes between:

- first movers,
- fast followers,
- majority followers,
- laggard followers (late movers)
wait until technology is adopted by majority

some technologies never adopted by a small percentage of companies

Ref: McKinsey Global Institute 2023 A future that works
 integrated with Information from other sources



**Artificial
Intelligence
1956¹**

System performing some human functions with different levels of autonomy



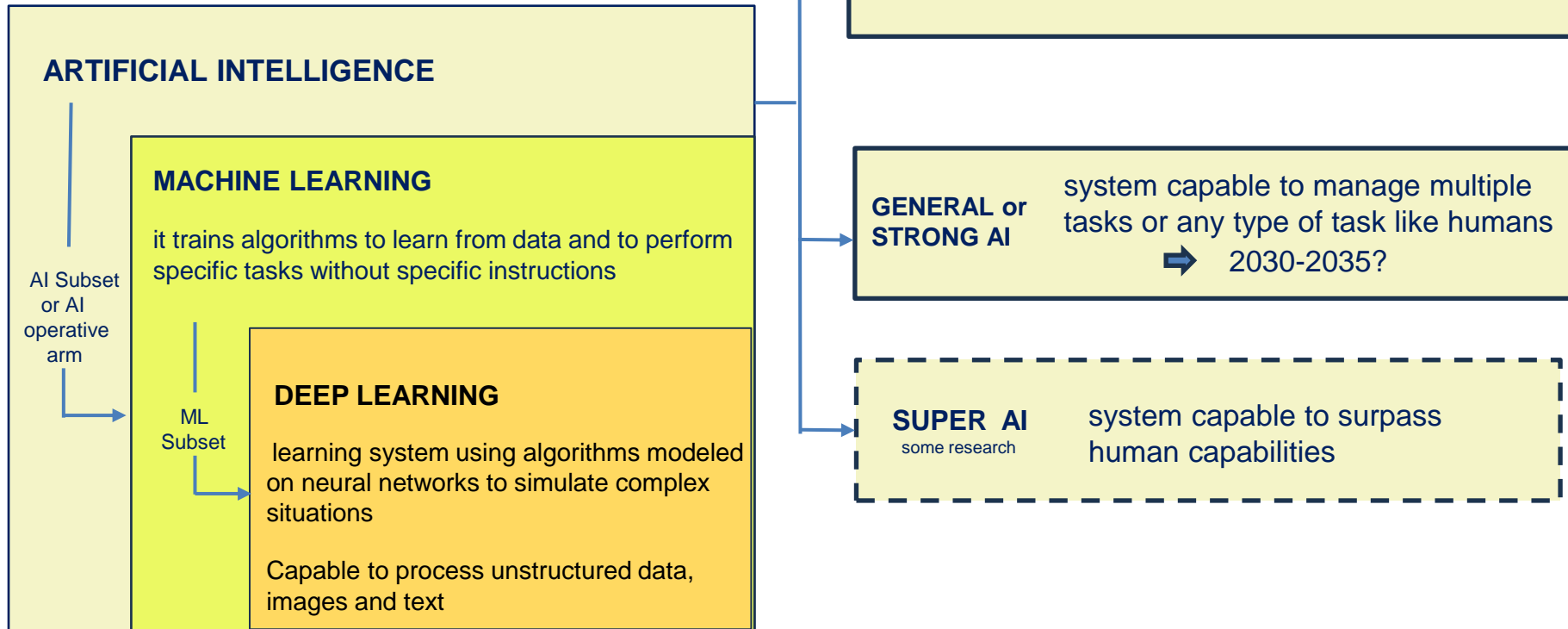
it generates predictions, content, patterns recommendations or decisions that can improve the outcome of our activities

AI tools: • **big data**

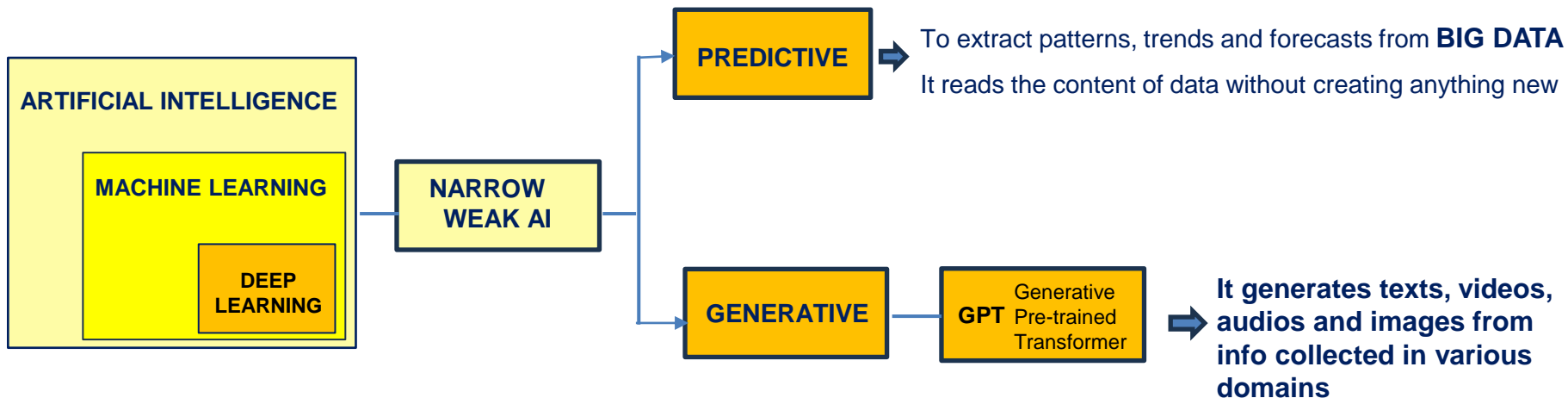
- **models, algorithms, LLM** designed, trained, tested, improved and validated - some fail the test
- **powerful computing systems** (GPUs² - especially for Deep Learning applications)

- **new factor of value production**
- **it learns from data**, extracting logics, patterns, forecasts, anomalies and similarities
- **it ensures excellent results if supervised by humans**
 - **if not carefully trained, its output may diverge from expectations** (not always smart)

1. Birth of research project on AI, Dartmouth USA
2. Graphics Processor Unit allowing parallel processing



Most common types of AI used in P&SCM



Its Large Language Models are trained on large volumes of information and predict words and their sequence in a specific context (statistical approach)

when we write a WhatsApp message, the smartphone suggests the next word

Big Data huge amount of data that organisations prepare and process to catch their insights using high performing computers



large volumes from different domains,
wide variety frequently stored in different data systems
frequently updated

The adoption of AI in a company begins with understanding company data,

Usually **only producers of large volume of goods, incorporating sensors - have big data**

They rarely exist in Procurement where

patterns and trends are detected by buyers from 'little data' available

Evolution of procurement task

AI allows buyers, supported by virtual assistants, to focus on strategies and high added-value activities





Sam Altman CEO di OpenAI
Creator of ChatGPT Nov. 2022

Artificial Generative Intelligence (AGI)



Generative AI

deep-learning applications
generating texts, images,
audios, videos

from a huge amount of
informations

Instead of interpreting company data, it creates new content relying on pre-existing information

Built on large language models (LLMs) and trained on a vast corpus of data

GPT-3 175 bl of parameters (50 ml of books) GPT-4 1767 bl of parameters,
\$100 ml spent for training, using 10.000 GPUs

It generates content in response to prompts

adequate prompts generate better answer /text

It reads unstructured data and can be used by everyone

It operates in **probabilistic terms** and predicts which words and in which order to respond to prompt

Within 2040, 2/3 of occupations will be impacted by Gen AI automation
impact similar to that of desktop computer, Internet, smartphone

Generative AI



It requires a unified data infrastructure, a comprehensive project and the workforce reskilling
Gen AI applications usually are managed in cloud

LLMs produce widely different answers when same question is repeated

Lack of consistency /Allucination (incorrect responses)

LLMs reproduce the biases found in Internet (they are trained with Internet informations)

Interpretability: it is difficult to understand why a particular response is generated

Accuracy: from 75% to 90% (% of correct answers)

Its performance depends on quantity & quality of data

Maturity in 3-5 years

McKinsey 2023 survey: inaccurate results identified by 913 respondents
56%

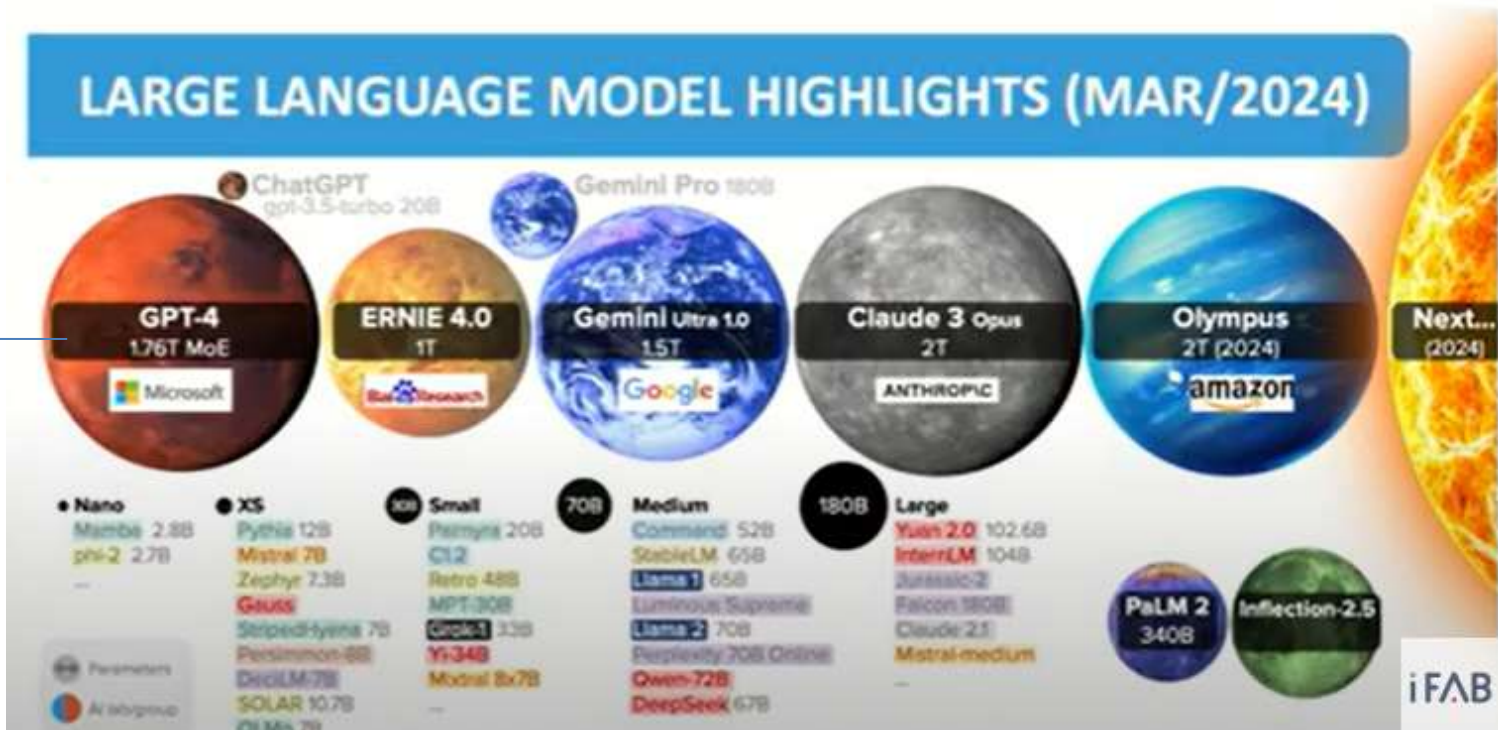
In future, Gen AI used as assistant supporting and improving human work capabilities

↳ job augmentation integration of AI technologies to empower workforce

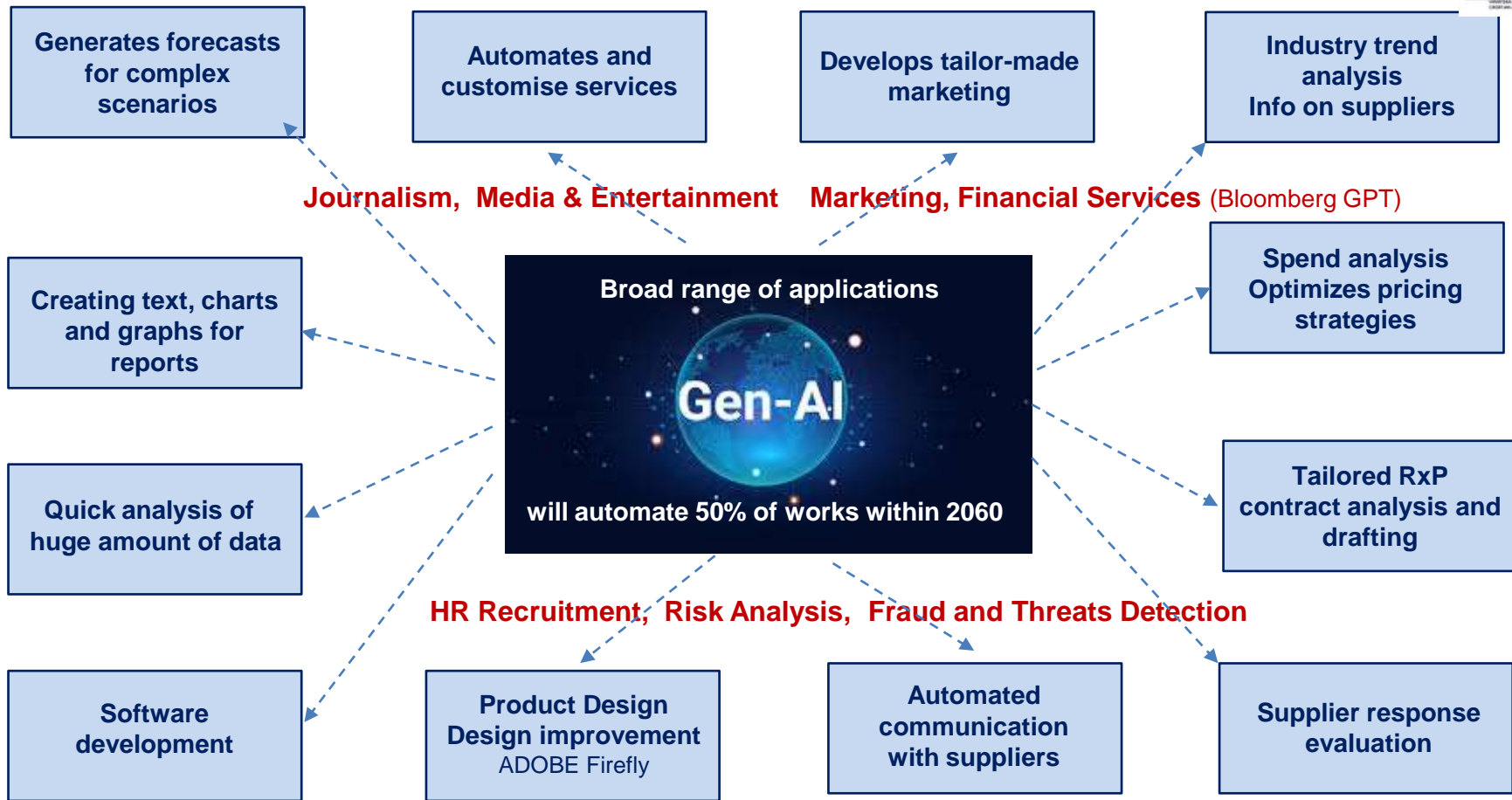
Each service provider has its own LLM and use cases

McKinsey analysed 63 use cases, if implemented they could add \$ 2.6÷ \$ 4.4 trillion of GDP

Major GPTs existing on the market (Generative Pre-trained Transformers)



OpenAI spends \$40 ml a month to process the queries entered by users
Its data centers consume up to 5 million liters of water per week (cooling systems)



Open source or proprietary LLMs?

Organisations could use open-source technology to build their own LLMs:

- protecting their own data and IP
- avoiding to use of platforms they don't control (lack of visibility)

Feature	Open source – alternatives to GPT 4	Proprietary
Models	Publicly available	Kept private
Performance	Variable: depends on internal expertise	High
Transparency	High: full source code access	Low: restricted by vendor
Cost effectiveness	High: no license fee	Low: high licensing and usage fees
Easy to use	Less support and lack of infrastructure	Infrastructure and support services provided
Data governance	High: full control over data	Low: vendor control data governance
Community support	High: broad community contribution	Low: reliance on vendor support
Bias identification	Transparency in training sources	Difficulty to identify bias for lack of transparency
Support and updates	Require expertise	Vendor driven updates (not users)

All AI applications:

- **imply a unified data infrastructure: efficient, scalable, well-governed and future-proof**
allowing collection, selection, storage, and analysis of data
minimizing the need to move data and ensuring privacy
- **require high-performance computers (GPUs)**

LLMs are learning engines that **don't differentiate true from false. "If there's an inaccurate or out-of-date information, they memorize it**

companies should pay attention on what goes into the model, it must be
free from errors and complete

Good-quality data sets for training, validation and testing **require the implementation of appropriate data governance systems** (orchestration)

Companies have to customize models with their own data integrated into their applications

Gen AI risks/criticalities

Impact on work

Monitoring and
evaluating
performances

+



**Intellectual
property
violations**

Civil liability
link to IP
violations

**Violation of
Privacy**
Exposure of info on
people, contracts,
products design that
could be hacked

**Inappropriate,
false content**

**Security breaches
Malicious SW**

Bias: algorithms reproduce human errors, prejudices, or societal inequalities present in the data
involuntary distortions (hallucinations)

2024 TELUS Digital survey: 61% of respondents concerned about the spread of inaccurate information

Common questions: Is AI applicable in my company?

Are the benefits greater than costs?

It depends on 5 factors

- **need to improve the internal efficiency**
- **logic and model required** and its integration with data source and SW applications
- **quantity and type of data and information available**
- **culture of management and AI users** (readiness to invest and risk)
- **acceptable level of risk**

To reduce risk of failure

**Pilot test followed by an extended application
with adequate governance system**

1. Need to improve Local Efficiency

Analysis of current processes to identify inefficiencies that AI can address process optimization

Establishment of adequate governance system to ensure compliant application after pilot test
changes in procedures, job roles, and responsibilities

Definition of expected (ROI) comparison of cost vs. benefits ROI not immediate

2. Logic and models required

Model Selection: to extract trends from data or to create content?

the choice defines the type of data required

Can we adopt standard use cases? Which vendor should we work with?

Do we rely on third-party capabilities or do we develop in house models?

Do we want the fine-tuning to customise the LLM to the organisation need?

How do we integrate LLM with data sources and other applications ensuring security?

Skillsets and Expertise required: data scientists, machine learning engineers, and system analysts

Available in our company?

Definition of type of hardware required

3. Data and Information

Can we obtain value from our data?

- **Assess the Quality and Quantity of existing data**

Poor quality prevents reliable results

limited quantity prevents Gen AI applications

large quantities allow fine tuning of LLM

- **Capability to integrate different data sources** (best solution: unified data infrastructure)

preparation of raw data, extraction of those related to the activity considered, chunking¹ and embedding them in a specific database, testing and fine tuning

1. Dividing larger pieces of text or information into smaller units, or "chunks"

4. Culture of management and AI users

- **Readiness to invest in new culture**

need to build skills & capabilities to support the application after pilot test

people are naturally resistant to changes, many are concerned that Gen AI will eliminate their job

- **CIOs must prepare the business case with all functions involved**

- **Project has to be known and accepted by middle management**

5. Acceptable level of risk

- **level of maturity of technology considered**

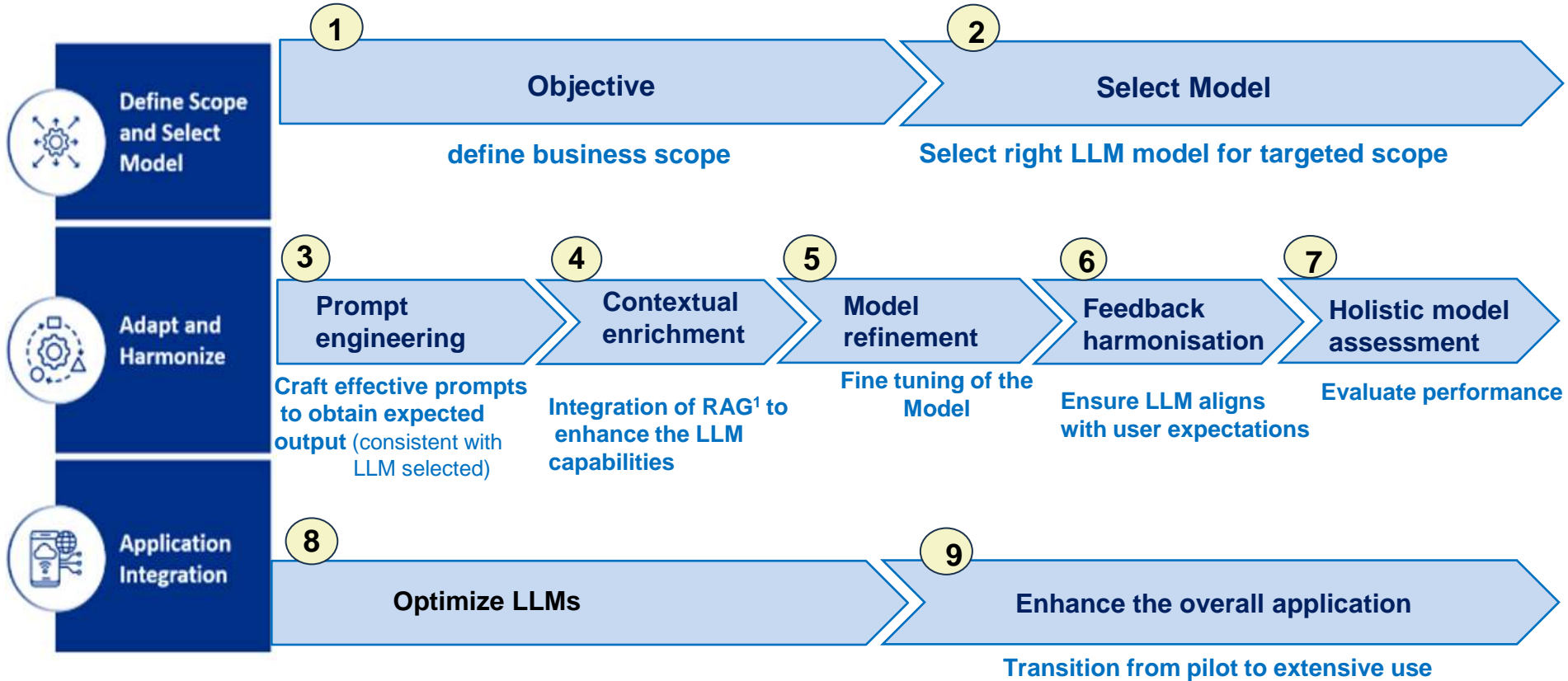
- **allocation of sufficient resources and continuous commitment of entire organization**

- **ability to overcome technical difficulties and resistance to change** especially after pilot test

Gen AI Roadmap for a complex application



it is not like a walk in the park
it is not just a matter of appropriate prompts



1. RAG: Retrieval Augmented Generation

AI Applications by industry /organisations

AI Applications implemented by 16% of product leaders / large corporations (projects launched or completed)
mainly focused on marketing and sales



Gartner 16.11.2023

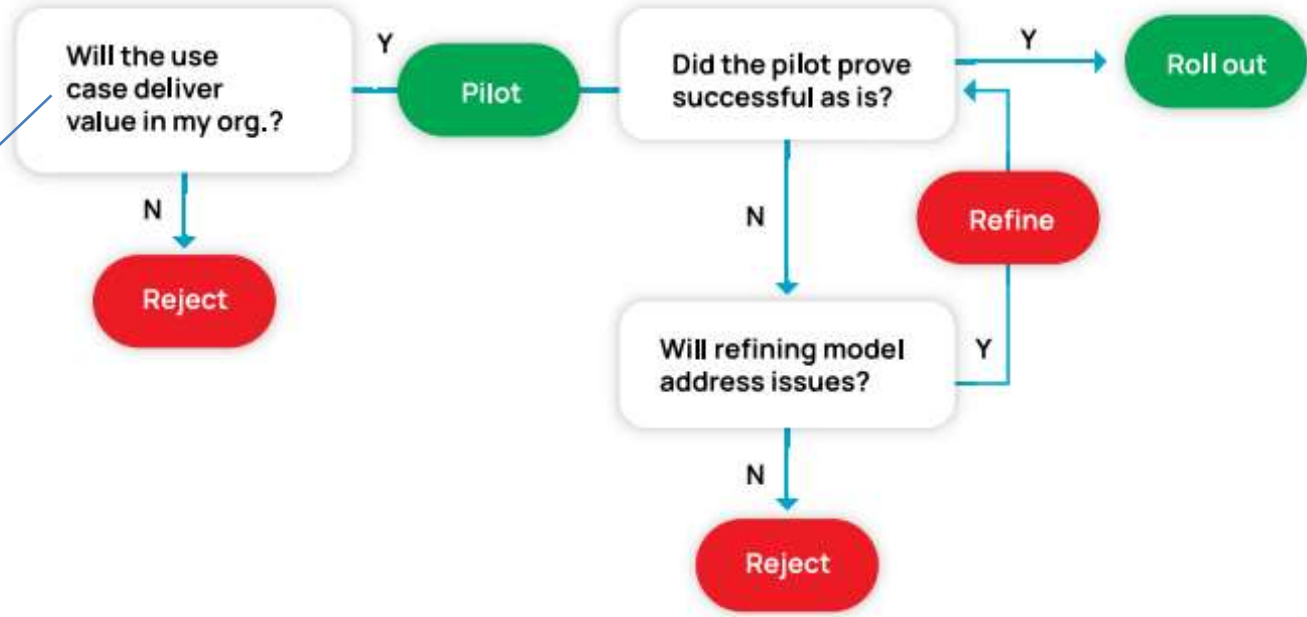
high turnover justifies the investment
availability of skilled resources
positive impact on company image

Minor AI applications by 10% of Manufacturers with less 700-900 employees

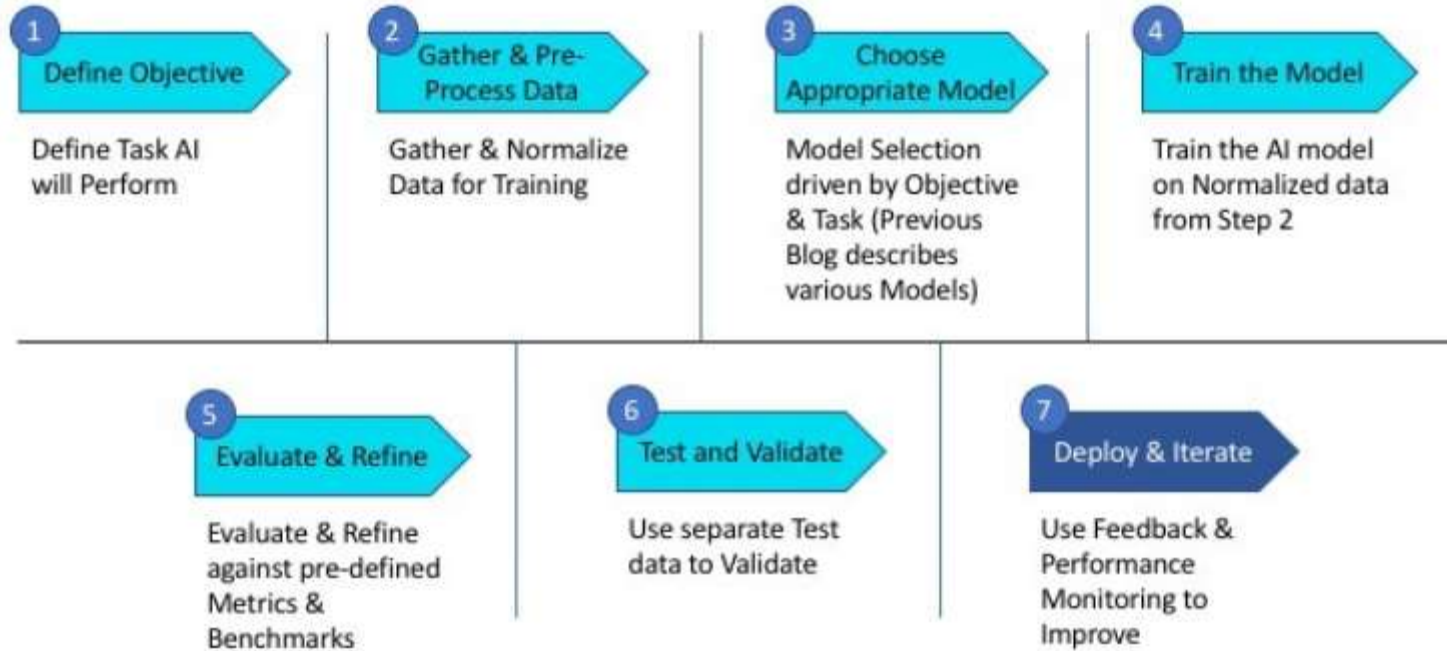
Optional slides that could be used during Q&A session

USE CASE ASSESSMENT APPROACH

Off the shelf LLM model



Generative AI: End-To-End Process



Chat GPT advertising slogan is similar to that of Watson which promised to solve challenging problems and which we no longer hear about

June 2023



Ask me what you want

I am able to find the optimal answer to questions of various kind including medical ones, much more effectively than human mind.

The application fields are endless: law, finance, customer service, weather forecasting, fashion design, tax assistance, etc.

Chat GPT practical guide
for business online



Giovanni Atti 2024

Dejà vu

Watson is able to find the optimal answers to questions of various kind, including medical ones.

The applications fields are endless: law, finance, customer service, weather forecasting, fashion design, tax assistance, etc.

aprile 2011, www.ibm.com
Watson and Healthcare Developer website

How generative AI is different from traditional AI?

Traditional AI	Generative AI
Predictions based on existing data	Creates new content
Best at numerical output	Best at non-numerical output (text, images, sound...)
Specialized use cases	Broad use cases

Gen AI produce outputs that are based on patterns learned from training data

[June 2024 study by Ardent Partners](#) of nearly 400 procurement leaders found 62% believing the impact of AI on procurement in the next 2-3 years will be Transformational or Significant

Users: employees that have access to Gen AI capabilities
suppliers, if authorised

Applications: SW accessed by users where Gen Ai capabilities are made available (source-to-pay, contract mgmt e procurement)

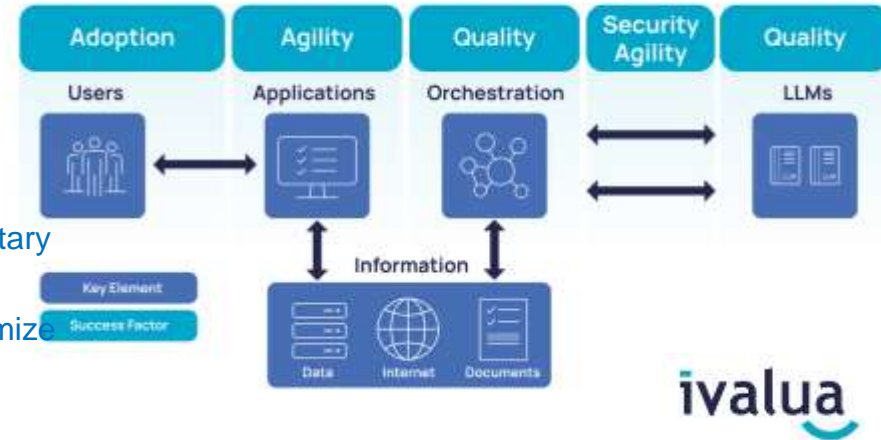
Orchestration: integration of LLM with data sources and complementary applications (API, source-to-pay technology S2P), to allow the expected output, the check the response to minimize the risk of allucinations or erroneous responses

Information: for several use cases, the model itself cannot provide accurate response without having access to various type of information: internal documents, data, contracts, info from Internet (e.g. news of disruptions, supplier information)

LLMs:

- open source** (companies deploy them to their own infrastructure and fine-tuning them to fit their own needs)
- closed source** (proprietary LLMs such as OpenAI's ChatGPT models, Microsoft Azure's Cognitive Services, Google's Gemini, Antropic's Claude – users unable to sudit their behaviour). They come with limited scope for modification to suite specific needs
- private proprietary** in future, large organisations will build their own LLMs based on open source LLMs tailored to their needs

Generative AI Architecture Overview



Success criteria

Communication: people are naturally resistant to changes, many are concerned that Gen AI will eliminate their job

User Access: must be simple, natural, conversational

Trust: errors or hallucinations are real risks

Recommendations

- plan a structured user training programme inclusive of:
 - Prompt engineer training
 - Expectation setting (perfection is impossible, automation will never reach 100%, variability of output quality)
- ensure the technology you select allows fine-tuning or creation of new use cases without vendor dependency

Mitigating security and compliance risks





- data privacy and cybersecurity risk- Use of 3° party LLM expose sensitive information about individuals (employees, contractors, contracts, products design. If users provide such information in prompts or if information pulled from internal documents is stored by the LLM or used to train model, it is possible that the information is hacked or exposed to external users

Complex Use Cases Requires Significant Development

	 GPT-4	 BERT	 Clip	 GPT-4
Domain	Fortune 500 pharma	Top US bank	Global ecommerce	Legal data case study
Use case	Information extraction	Chat intent classification	Image classification	Document classification
Foundation model performance	60%	60%	43%	59%

katonic.ai

Complex Use Cases Requires Significant Development

	 GPT-4	 BERT	 Clip	 GPT-4
Domain	Fortune 500 pharma	Top US bank	Global ecommerce	Legal data case study
Use case	Information extraction	Chat intent classification	Image classification	Document classification
Foundation model performance	60%	60%	43%	59%
Fine Tuned model performance	86%	85%	71%	83%*

katonic.ai

Foundation models are trained on a large quantity of data, working under the maxim "the more data, the better" Performance evaluation does show that more data generally leads to better performance, but other issues arise as data quantity grows.

12 GPT-4 Open-Source Alternatives



1.8.2024

GPT-4 open-source alternatives that can offer similar performance and require fewer computational resources to run. These projects come with instructions, code sources, model weights, datasets, and chatbot UI.

1. ColossalChat



2. Alpaca-LoRA

3. Vicuna

4. GPT4ALL

5. Raven RWKV

6. OpenChatKit

7. OPT

8. Flan-T5-XXL

9. Baize

10. Koala

11. Dolly

12. Open Assistant

Generative Models

Closed Source

GPT-3.5 DALL.E 2

Codex LaMDA

Open Source

CLIP DALL.E 2

Stable Diffusion BLOOM

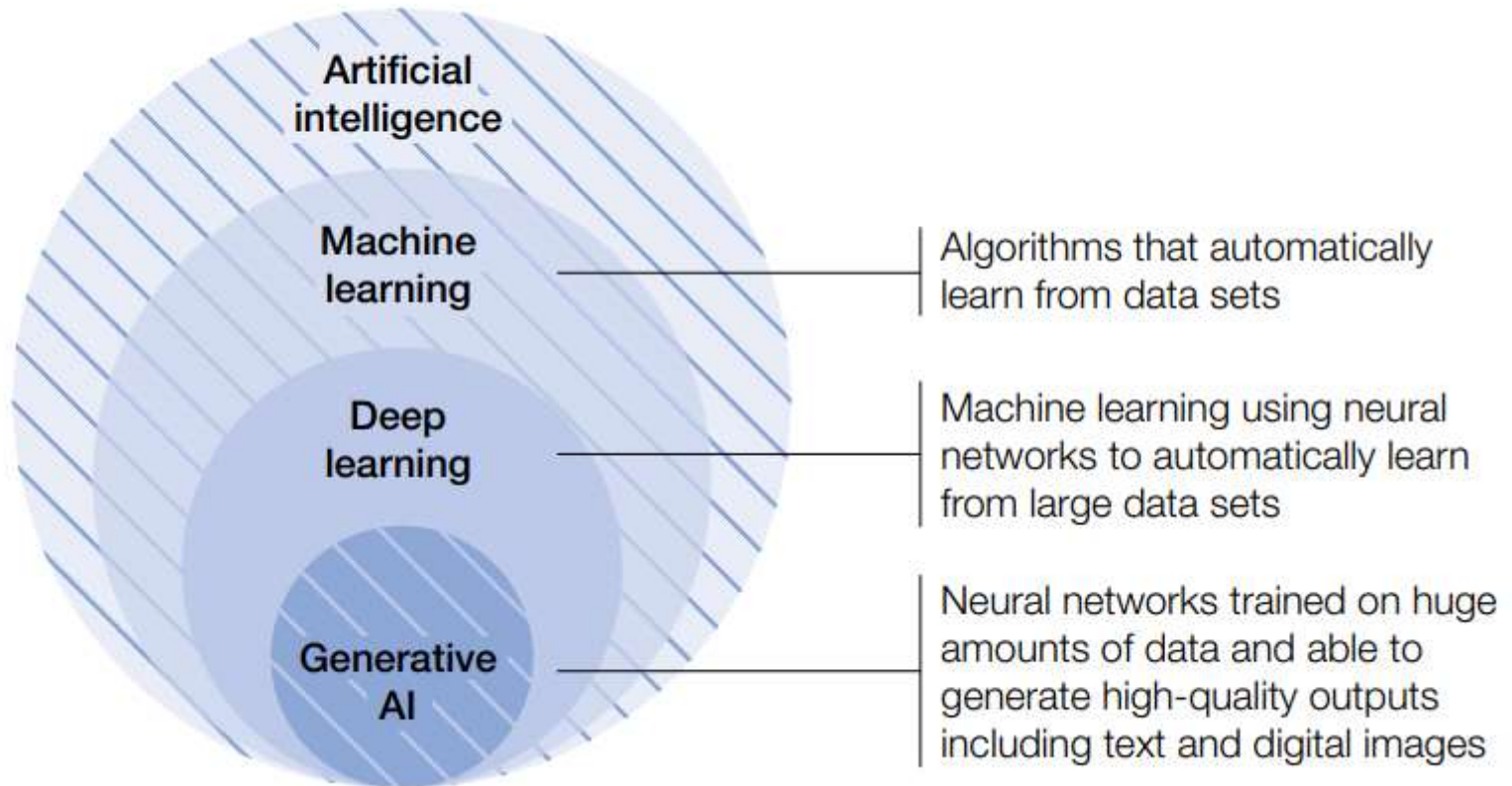
Build Your Own



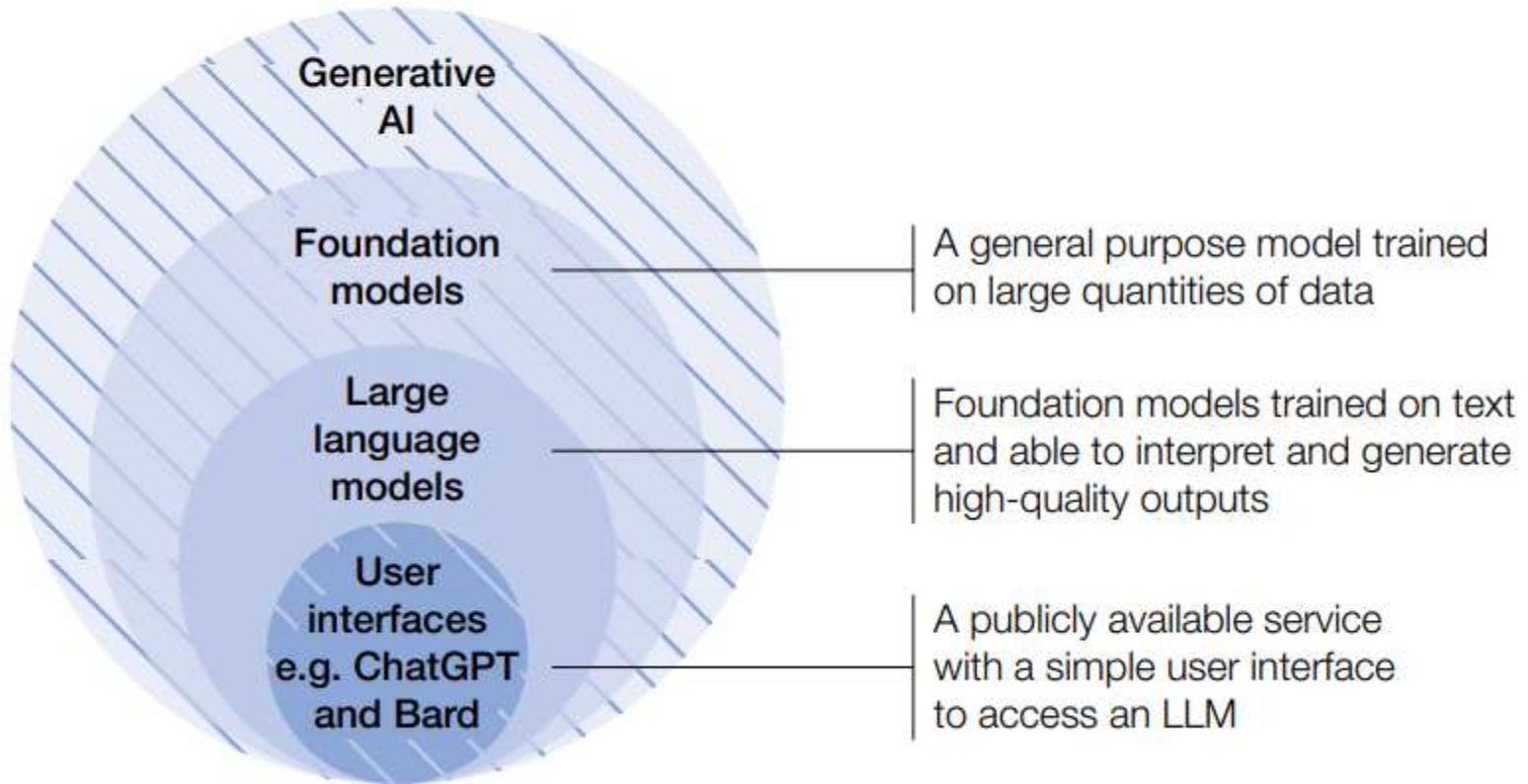
What are the options?

	Option	Explanation	Cost
	Off the Shelf	<i>Subscribe and Use</i> Embracing off-the-shelf tools leveraging LLMs that are already available.	Minimal
	Customise	<i>Consume with Guardrails</i> Build Guard Rails by adding pre and post-processing restrictions to off the shelf LLM's	\$
		<i>Augment</i> Use database lookups to tailor LLMs to an organisation's needs.	\$\$
		<i>Fine Tune</i> Using fine-tuning to tailor LLMs to an organisation's needs	\$\$\$
	Train your own	<i>Build your own</i> Build and Train your model from scratch with your data	\$\$\$\$

Generative AI fits within the broader field of AI as shown below:



Public LLM interfaces fit within the field of generative AI as shown below:



Responsibilities of a Gen AI Developer:

As a Generative AI developer, your responsibilities revolve around creating and fine-tuning AI models that can generate new content, such as images, text, music, or even code. Here's what you might do:

1. **Building Models:** You'll design and build generative models, like GANs or VAEs, that can create new data from existing data. This involves programming and experimenting with different neural network architectures.
2. **Training Models:** You'll gather and prepare data to train your models, ensuring they learn to produce high-quality, realistic outputs. This means working with large datasets and running training processes that can take hours or even days.
3. **Optimizing Performance:** You'll tweak and fine-tune your models to improve their performance, making them faster and more accurate. This involves adjusting parameters, testing different algorithms, and solving any issues that arise.
4. **Testing and Evaluation:** You'll rigorously test your models to ensure they generate the desired output without errors or biases. You'll also compare their performance against benchmarks to see how well they're doing.

Off the Shelf - Benefits and Limitations

Using paid subscriptions or corporate user plans of Generative AI tools like ChatGPT, Jasper, Notion etc. for trial and training of employees **without exposing confidential company data**. Use-cases limited to the generation of low-quality and low-risk content.



BENEFITS

- ▶ Requires the least LLM training technical skills.
- ▶ Cost limited to subscription fees
- ▶ Fastest turnaround time
- ▶ Cost limited to subscription fees
- ▶ Can leverage the best-performing LLMs in the market



LIMITATIONS

- ▶ Limited to publicly available info
- ▶ Cybersecurity Concerns
- ▶ Fabricated Information.
- ▶ Copyright Issues
- ▶ Data Privacy
- ▶ Deepfakes



RECOMMENDATION

- ▶ Acceptable only for trial and training of employees.
- ▶ Strongly recommend avoiding sharing of any confidential information.
- ▶ Good for prototyping apps and exploring what is possible with LLMs.

Customise - Benefits and Limitations

Organisations can boost the capabilities of their applications by integrating them with LLMs by consuming Generative AI and LLM applications through APIs and tailor them, to a small degree, for your own use cases through prompt engineering techniques such as prompt tuning and prefix learning.



BENEFITS

- ▶ Model trained on organisations data which is publicly not available .
- ▶ More affordable than organisations further training ("fine-tuning") an LLM
- ▶ Data security as data resides in your own environment.



LIMITATIONS

- ▶ Not appropriate where the model needs to have a wide-ranging understanding of the content in the knowledge base, as only a limited amount of data can be passed to the LLM.
- ▶ The LLM will only use the data passed to it, along with the user's original query, to construct a response.



RECOMMENDATION

- ▶ An affordable and powerful way to quickly leverage the power of generative Ai for your business
- ▶ An intermediate step for most businesses.

Build your Own - Benefits and Limitations

Organisations training their own LLM gives them a deep moat: superior LLM performance either across horizontal use cases or tailored to your vertical, allowing you to build a sustainable advantage, especially if you create a positive data/feedback loop with LLM deployments.



BENEFITS

- ▶ Specialised models are smaller and can be deployed on significantly cheaper hardware
- ▶ Specialised models are significantly more accurate for the same resource budget
- ▶ Gain full control of training datasets used for the pre-training



LIMITATIONS

- ▶ Very expensive endeavor with high risks. Need cross-domain knowledge spanning from NLP/ML, subject matter expertise, software and hardware expertise.
- ▶ Less efficient than Customise option as it leverages existing LLMs, learning from an entire internet's worth of data and can provide a solid starting point



RECOMMENDATION

- ▶ Best if you need to change model architecture or training dataset from existing pre-trained LLMs.
- ▶ Typically, you have or will have lots of proprietary data associated with your LLM to create a continuous model improvement loop for sustainable competitive advantage

EU AI Act: different risk levels of AI systems

