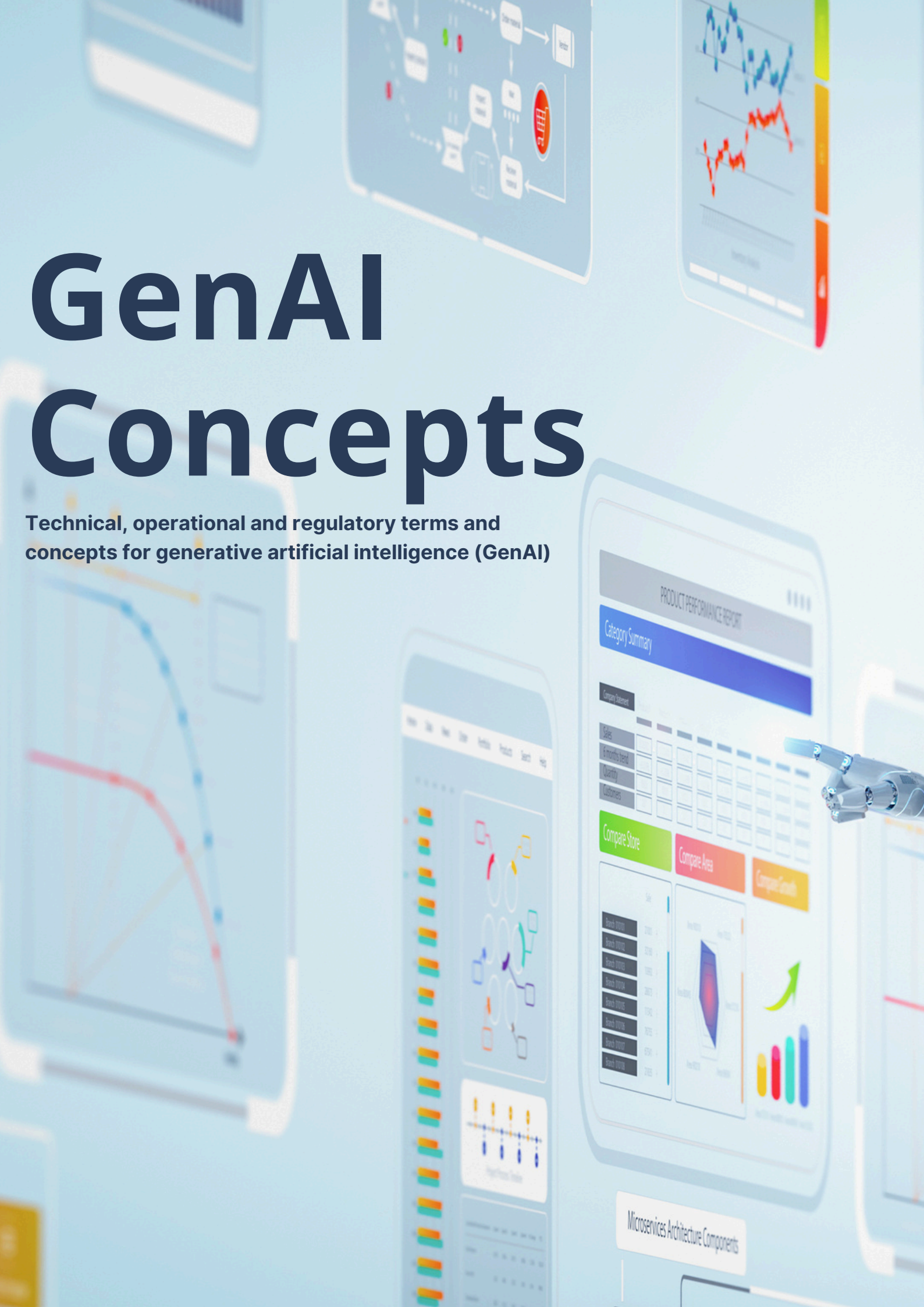
GenAI Concepts

Technical, operational and regulatory terms and concepts for generative artificial intelligence (GenAI)



Contents

Introduction

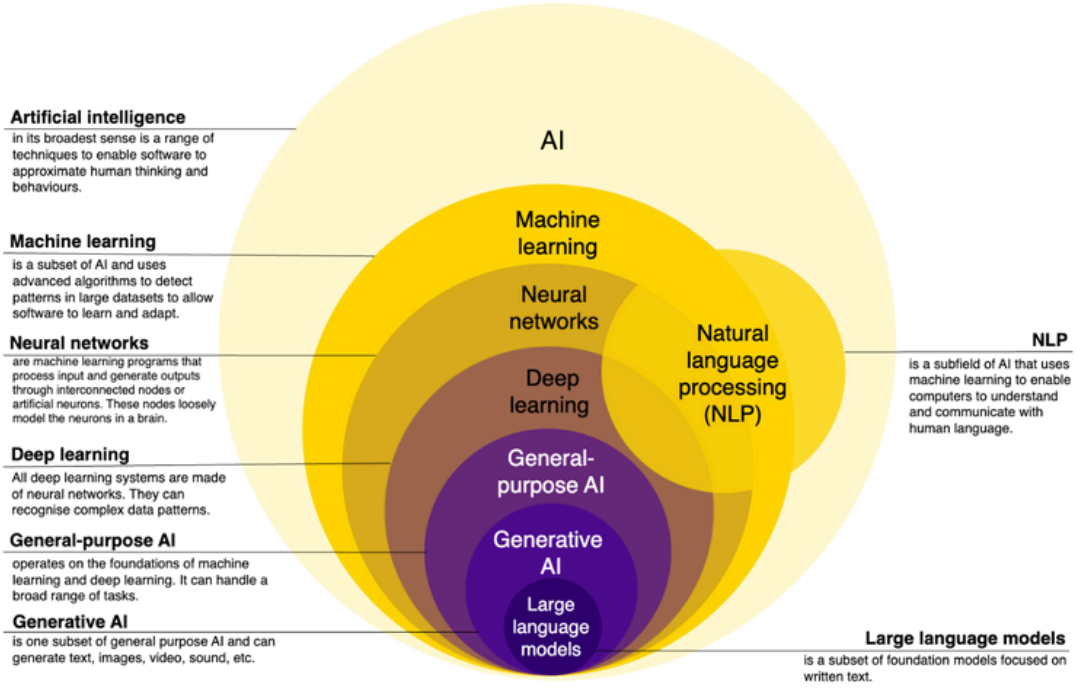
Generative AI products and services, such as OpenAI's ChatGPT, Alphabet's Gemini, and Microsoft's Copilot, have sparked substantial interest in the private and public sectors. Organisations are already experimenting with integrated AI services provided by big tech firms, as well as custom procurements from smaller software companies. In the absence of comprehensive AI regulation in Australia, deploying these systems in a responsible, ethical and legally compliant way demands a deep understanding of how they function and the legal and ethical challenges they raise.

But generative AI (GenAI) is not simple. AI software systems are diverse and rely on complex supply chains and data flows. Coming to grips with the technical, operational and regulatory vocabularies that have emerged around GenAI is a considerable challenge. To help entities interested in GenAI deployments, this publication outlines 42 concepts fundamental to AI software systems. Each concept is illustrated through descriptions, examples and real-world use cases, with accessible language and visual elements to accommodate a diverse range of stakeholders and readerships.

GenAI Concepts can also be accessed as a regularly updated website at decodingdatascience.com

Technical

The following diagram demonstrates the interconnections among fundamental technical terms in the field of artificial intelligence (AI).



Using this diagram as a starting point, this section begins with basic GenAI concepts relevant to daily applications, including prompt and prompt engineering, large language models (LLMs), knowledge cut-off dates and chatbots.

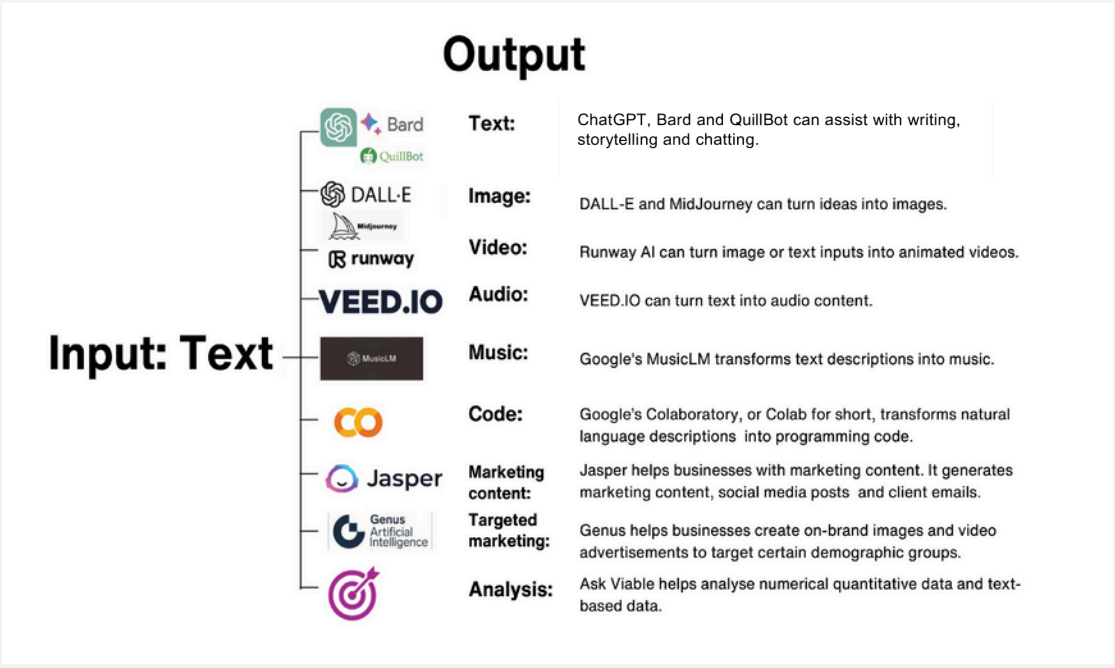
It then moves on to more advanced terms that are often encountered in official documents, such as general-purpose AI, foundation models and frontier AI.

The section further explores specialised technical terms, encompassing open-source and closed-source LLMs, transformer architecture, transfer learning, tokens, reinforcement learning from human feedback (RLHF), diffusion models and inference techniques.

Generative AI or GenAI

Generative AI or GenAI are both short for generative artificial intelligence. These are software systems that create content as text, images, music, audio and videos based on a user's 'prompts'.

Types of GenAI models in the market



Prompt and prompt engineering

A prompt is an instruction, query or command that a user enters into a GenAI interface to request a response from the system.

Because GenAI systems produce text through statistical predictions of the most likely next words in a sentence, the responses that GenAI systems produce may not always be the same. This is why GenAI outputs are sometimes described as non-deterministic.

Prompt engineering is the practice of writing inputs for GenAI tools so that the tool can produce optimal outputs.

> Example
>
> Basic structure of a prompt:
>
> Acting as a [role] perform [task] in [format] in the [style]

Acting as a [role]	Perform a [task]	Show as [format]	In the [style]
Each role provides context and background.	*The task should be clear and specific.*	*How you would like to structure the information generated.*	*This is optional. It shows what tone you'd like to use for the information.*
Marketer Advertiser Copywriter Accountant Lawyer Financial analyst English professor Journalist Project manager Manager Engineer Recruiter ...	Headline Presentation Webinar Blog post Book outline Email sequence Social media campaign Product description Cover letter Summary TikTok, YouTube or Instagram Reel video script Sales page / ad copy ...	Table List With bullet points Summary HTML Code Spreadsheet CSV file Plain text file Rich text PDF Markdown Word cloud ...	Formal Poetic Enthusiastic Shakespearian Accessible Basic English Scientific Objective Neutral Pop culture

As a [life influencer (the role)], create a [blog post (the task)] about the benefits of daily exercise in a [PDF (the format)] in an [accessible and enthusiastic tone (the style)].

> Tip
> If the task is too complicated, break it down into steps to make it easier for the GenAI model to read and interpret.

Machine learning

Machine learning (sometimes seen as ML) is a set of techniques for creating algorithms so that computational systems can learn from data.

A machine learning algorithm is a set of rules or processes that helps an AI system to perform specific tasks, such as finding patterns in data or making predictions based on inputs. In this way, the model's behaviour reflects the data or the learning experience.

Higher quality data helps algorithms improve their accuracy in various tasks, such as recognising faces in photos, predicting the weather or recommending products to buy.

Types of machine learning

- Supervised machine learning:
  An algorithm is provided with labelled data, for example, a collection of pictures of animals with labels for each animal in each picture. The algorithm learns from these examples and tries to predict the correct labels for new, unseen data.
- Unsupervised machine learning:
  An algorithm is provided with unlabelled data and then it tries to find patterns or structures in the data on its own.
- Hybrid machine learning:
  Combines elements of supervised and unsupervised learning approaches and sometimes different types of algorithms to leverage the strengths of different methods based on the developer's needs.

Applications

- Healthcare: ProMed (the Program for Monitoring Emerging Diseases) offers an online real-time data analysis and reporting system showing outbreaks of infectious diseases worldwide.
- Finance: Fraud detection, automated trading activities and financial advisory services for investors.
- Marketing: Product recommendations and news feeds on social media services.

- Transportation: Self-driving vehicles, real-time tracking and last-mile delivery optimisation.

Large language models (LLMs)

Large language models (LLMs) are data transformation systems. They are trained with large numbers of parameters, which are numerical values that developers adjust to shape the inputs and outputs of AI models. When a user inputs a prompt, the model generates text content in response.

LLMs are trained on extremely large datasets sourced from websites, blogs, forums and news articles etc. They contain millions or billions of parameters. In the case of OpenAI's generative pre-trained transformer (GPT) models, the first large language model GPT-1 contains 117 millions parameters, GPT-2 contains 1.5 billion parameters and GPT-3 contains 175 billion parameters. Increased parameters often lead to more complex models that can handle more complicated tasks and generate more nuanced text.

Users interact with GPT models through interfaces like ChatGPT. This feedback loop allows the model to continuously learn and improve over time based on user feedback.

Examples

Some notable LLMs are:

- OpenAI's GPT series of models (GPT-1, GPT-2, GPT-3, GPT-3.5 and GPT-4) used in ChatGPT and Microsoft Copilot
- Google's PaLM (Pathways Language Model) and Gemini
- xAI's Grok
- Meta's Llama (Large Language Model Meta AI) family of open source models
- Antropic's Claude models

Related term: Machine learning

Knowledge cut-off date

The knowledge cut-off date of a GenAI model is the date when the training data for a specific LLM was last updated. It defines the limitations of the model's understanding and knowledge.

Ecological

Training, deploying and using AI systems contribute to the global CO2 emissions. Typically, more powerful AI models require more energy. The servers that power AI models also generate considerable heat and are often water-cooled. The amount of water needed to train an AI model is immense. A team of researchers disclosed that “training GPT-3 in Microsoft’s state-of-the-art U.S. data centers can directly evaporate 700,000 liters of clean freshwater, but such information has been kept a secret.” And running GPT-3 inference for 10-50 queries evaporate 500 mililitres of water depending on when and where the model is hosted.

LLMs are among the biggest machine learning models, spanning up to hundreds of millions of parameters, requiring millions of GPU (graphic processing units) hours to train and emitting carbon in the process.

DECODING DATA SCIENCE

THE 15 MASTERCLASS INCLUDED

Part of the AI Residency Program

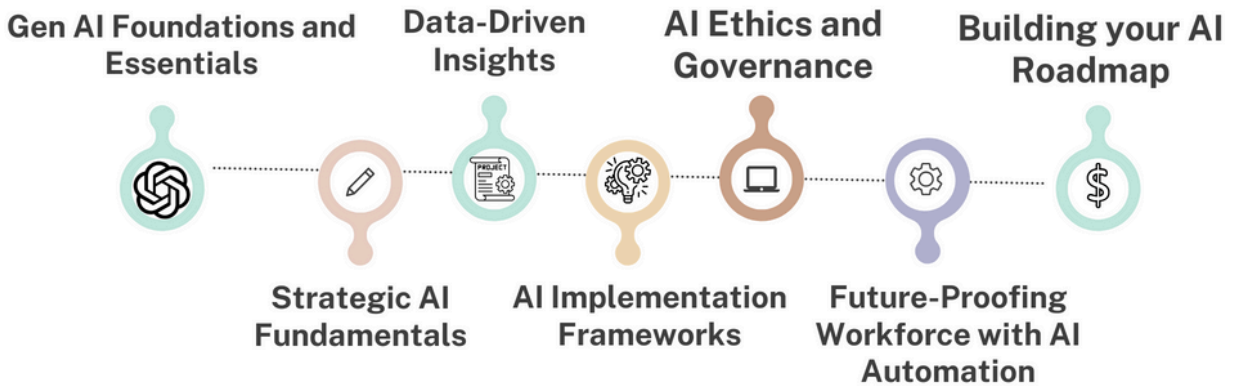
New Cohort Starting this 18 JAN 2025

00	Essential Prerequisition: Basic Python and What is Gen AI?
01	Application Architecture UI Building, Git and Github
02	Intro to API, FAST API Building Apps with APIs (1st Project)
03	Intro to LLM, How it Works Playground
04	Generative AI Configuration & Different LLM (2nd Project)
05	Building LLM Wrappers, Function Calls and Data Integration
06	Hugging Face Building an Open Source Project
07	Llama Index Retrieval Augmented Generation
08	Intro to Vector Embedding, Vector db
09	Advanced RAG vs Fine Tuning
10	Wrapping Fine Tuning and Intro To Agents
11	Langchain , Components and Agents
12	Putting all Together and Capstone Project Orientation
13	One on One Mentoring WIth Industry Experts and Coaches
14	One on One Mentoring WIth Industry Experts and Coaches
15	Weekly Live Workshops and Doubt Clearing Projects
101	Linkedin, Github and Portfolio Optimization
102	Aiding in Internship / Job / Bussiness Opportunities

THE 6 MASTERCLASS INCLUDED

Part of the AI for Bosses Program

New Cohort Starting 7th December 2024



00	Essential Prerequisition: What is Gen AI? and AI @ Work
01	AI Fundamentals, Business Strategy, and Everyday Tasks
02	Data, AI-Driven Decision Making, and Research Automation
03	AI Implementation and Change Management
04	AI Ethics, Risks, and Governance
05	AI and Workforce Strategy
06	Developing Your AI Roadmap

www.decodingdatascience.com/aiforbosses

DECODING DATA SCIENCE

Community Profile

SCAN ME

What Does The Community Provide?

Generative AI

- ✅ Generative AI (chatGPT) for Business
- ✅ Prompt Engineering for Developers
- ✅ Langchain for AI App Development
- ✅ What are Large Language Models
- ✅ Gemini MultiModal LLM

Recordings

- ✅ 50+ Outcome-based Workshops
- ✅ 45+ AI Community Meetup Recordings
- ✅ 10+ Python Projects Videos
- ✅ AI & DS Career & Learning Webinar Series

Data Science Courses

- ✅ Basic Excel For Data Science
- ✅ Basic SQL For AI/Data Science
- ✅ Basic Python for AI/Data Jobs
- ✅ Advanced Python for AI/DS Jobs
- ✅ Basic PowerBI for AI/Data Science
- ✅ Machine Learning
- ✅ Knowledge Shorts
- ✅ Linkedin Growth Mastery

Resources

- ✅ Generative AI Resources
- ✅ Sample Datasets & Projects
- ✅ Sample Reviewed Resume
- ✅ Ready to use Resume Template
- ✅ Linkedin Profile Optimization
- ✅ Essential SQL Documents
- ✅ Essential Python Documents
- ✅ Machine Learning Documents

Every week we have live Zoom calls, Physical Meetups and LinkedIn Audio events and WhatsApp discussions. All calls are recorded and archived.

nas.io/aiguild

What Next? Join the Free AI Community

Artificial Intelligence

2,810 members

SCAN ME

Free

- Three weekly events
- Live workshops
- Knowledge Shorts 50+ Videos
- Basic AI & DS courses
- DS & AI materials
- Webinar recording
- Guidance from experts
- 24 by 7 Whatsapp & Discord
- Latest ai Discussion & More...

decodingdatascience.com/ai