Applied statistics and Machine learning in Python with subsurface applications

Enrico Riccardi, University of Stavanger

Mar 2, 2024

University of Stavanger

Big Data

Big data has been one of the main keywords to represent the latest digital innovation (before AI with gen AI)

Big data implies Machine learning applications based on:

1. Large volume of data
2. Fast process speed
3. Variety of data types and resolution

Examples:

- IoT
- Social Media
- Internet Marketplaces (Amazon)
- Healthcare
- Transport

Challenges

Big data has to face three main problems:

1. Storage: strategies to dump useless data and/or store continuously generated large data volumes
2. Analysis: data has to be indexed and analysed on-the-fly
3. Privacy: proprietary data access has to be limited

As one can note, these challenges are still present in gen AI.

Goal

Big data was considered a revolution, where machine could take over human intelligence thanks to the immense quantity of data available and on the data processing speed.

Now (2024) big data is considered a simple base for gen AI and it is pretty much forgotten by most.

Conclusion

Big data is a good example of data science marketing.

Generative AI

A generative AI model is a type of artificial intelligence that is designed to generate new content, based on the data it has been trained on.

It started in 1932, with the **mechanical brain** by Georges Artsrouni that was supposed to translate automatically between languages,

Here a nice recaps of Generative AI and its storyline

Generative AI

Key characteristics of generative AI models include:

1. Learning from Data: They are trained on large datasets, enabling them to learn patterns, styles, or features inherent in the data.
2. Generating New Content: Generative models can create new data instances. For example, a model trained on a dataset of paintings can generate new images in the style of those paintings.

Trained generative models are thus able to input information at a low resolution/dimension and give output with a much greater dimensionality.

Applications

Here a list of possible applications:

- Images/video: Image generation, Super-resolution, Deep fakes.
- Music: noise filter, voice and music generation, voice deep fake.
- Text(LLM): chatGPT, bard, Gemini, etc.
- Chemistry: DeepMind (Alpha Fold).
- Coding (co-pilot)
- Speech
- Attacks and Hacking (Security testing)
- Generating training sets
- And many more

Science fiction?

This is scary:

1. Virtual best friends
2. Medical images to show diseases consequences
3. Synthetic data for digital twins
4. Preemptive suggestions (e.g. driving)
5. Matrix

Problems (currently)

New possibilities do not come with side effects.

1. Lack of transparency: how the output is generated, and why?
2. Accuracy: a lot of hallucinations
3. Bias: human biases are kept, supported and eventually increased
4. Intellectual properties (IP): who owns what is produced?
5. Cybersecutiry and frauds: mass cyber attacks can be created
6. Sustainability: massive quantity of electricity is used
7. Responsibility (who to blame?): Will AI get citizenship everywhere?

Where generative AI is ?

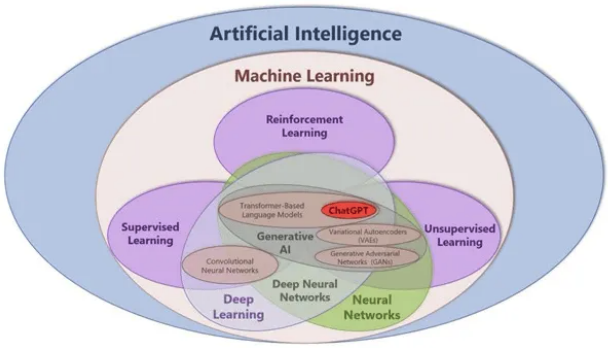


Image: https://iot-analytics.com

Structure of generative AI

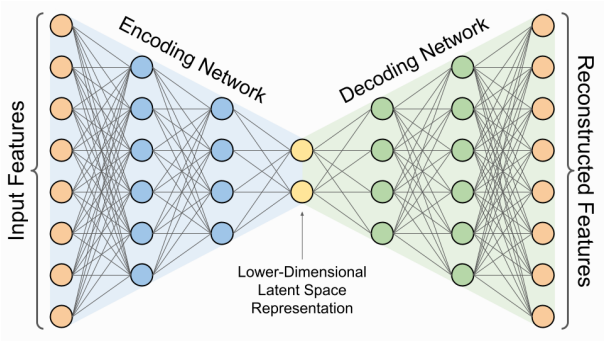


Image: https://www.rapidops.com

A new field?

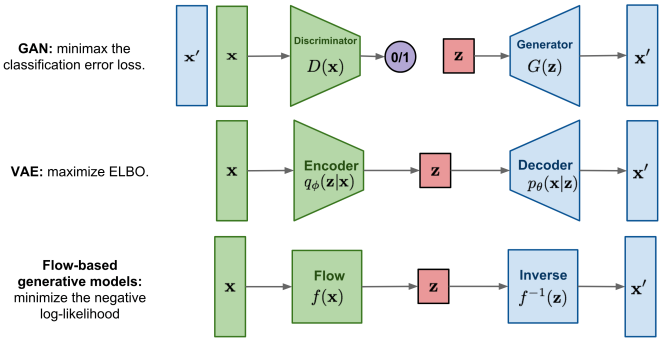
Generative AI is actually a new evolution.

It is based on Neural Network, and in comprises a set of advanced tools (numerical recipes):

1. Generative Adversarial Networks
2. Generative Pre-trained Transformers
3. Variational Autoencoders
4. Conditional Variational Autoencoders
5. Autoencoders

Types of generative AI

It is quite an advanced technique



Source: Lilian Weng

Advanced AI

Multimodal AI

An AI that processes and interprets multiple types of data inputs — such as text, images, audio, and video

M-AI can make decisions or provide insights (it thus combines unsupervised, supervised and reinforcement learning).

It uses an integrated set of different neural networks with different architectures each of which is designed to handle specific types of input data.

Applications:

- Autonomous Vehicles
- Healthcare Diagnostics
- Humanoids
- Interactive machine
- Augmented data

Generative Adversarial Networks (GANs)

Definition

A class of machine learning frameworks where two neural networks contest with each other to generate new, synthetic instances of data that can pass for real data.

In other words, a machine is able to teach from another.

Something to worry

Citation

With great power comes great responsibility (spiderman)

Several advanced AI application can be really worrisome:

- AI war-drone
- votation frauds
- propaganda campaigns
- job market catastrophe
- energy and material demanding

AI trends (2024)

- Generative AI
- Augmented reality
- Ethical and explainable AI
- Democratic AI
- Sustainable AI

AI potential (2024)