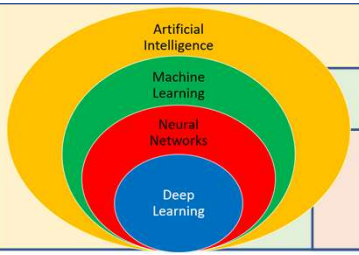


Artificial Intelligence (AI)

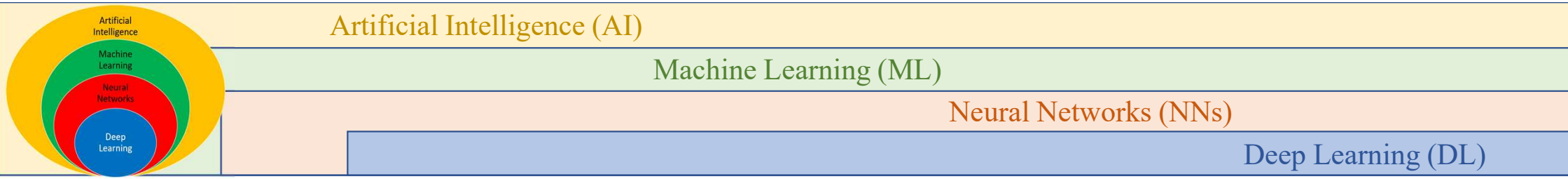
Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

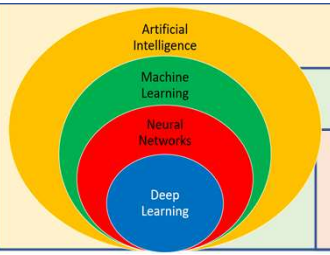
Advanced Deep Learning

Dr. Rastgoo



Diffusion Model Techniques

- Central to the diffusion model's operation are several key mechanisms that collectively drive its performance.
- Understanding these elements is vital for grasping how diffusion models function.
- These include **score-based generative modeling**, **denoising diffusion probabilistic models**, and **stochastic differential equations**, each playing a critical role in the model's ability to process and generate complex data.



Artificial Intelligence (AI)

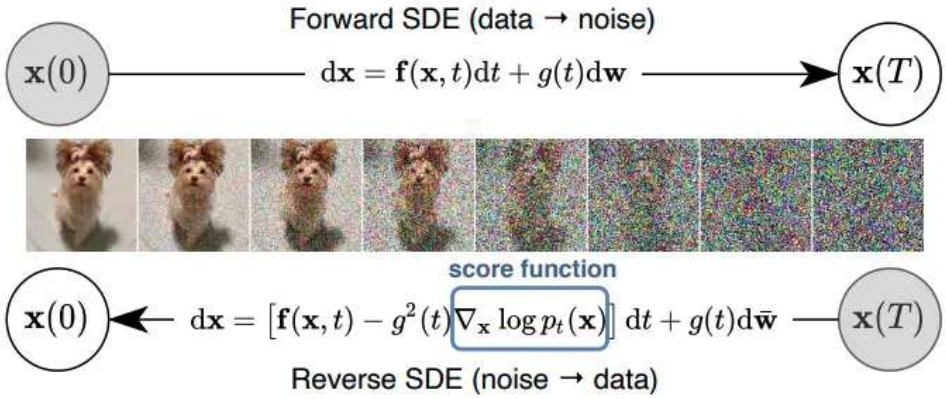
Machine Learning (ML)

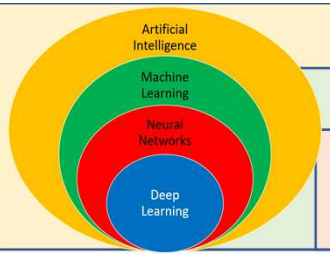
Neural Networks (NNs)

Deep Learning (DL)

Stochastic differential equations (SDEs)

- SDEs are mathematical tools that describe the noise addition process in diffusion models.
- They provide a detailed blueprint of how noise is incrementally added to the data over time.
- This framework is essential because it gives diffusion models the flexibility to work with different types of data and applications, allowing them to be tailored for various generative tasks.





Artificial Intelligence (AI)

Machine Learning (ML)

Neural Networks (NNs)

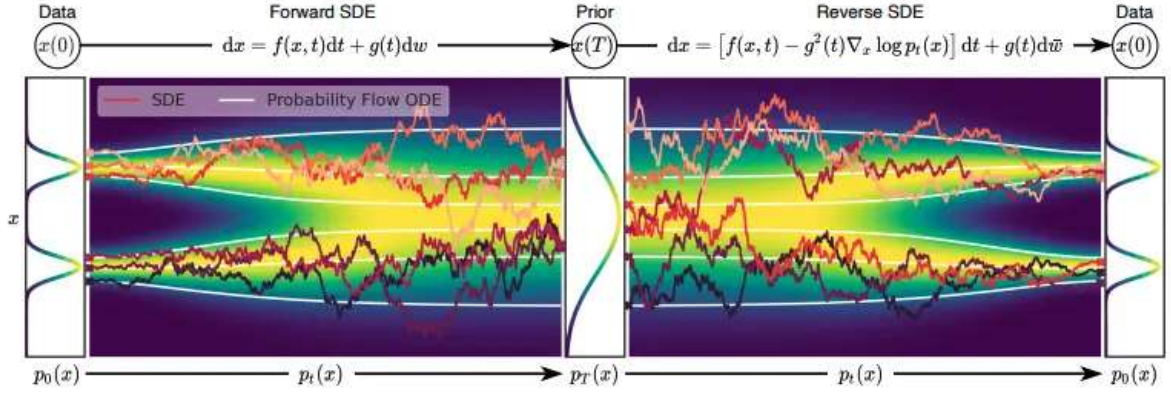
Deep Learning (DL)

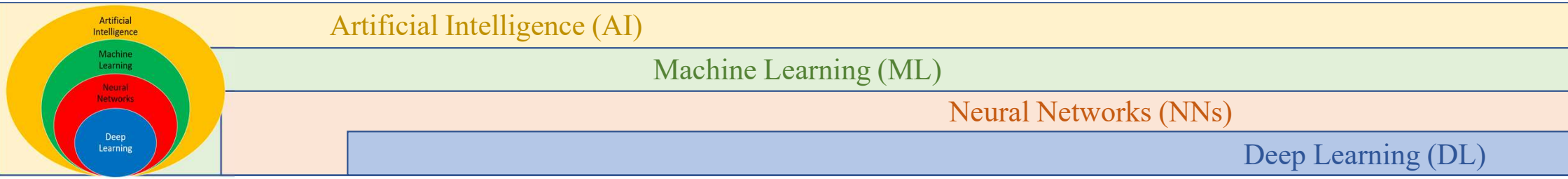
Score-based generative models (SGMs)

- This is where the model learns to understand and reverse the process of noise addition.

- Imagine adding layers of noise to an image until it's unrecognizable. Score-based generative modeling teaches the model to do the opposite - starting with noisy data and progressively removing noise to reveal clear, detailed images.

- This process is critical to creating realistic outputs from random noise.





Denoising diffusion probabilistic models (DDPMs)

- Denoising diffusion probabilistic models (DDPMs) are a specific type of diffusion model that focuses on probabilistically removing noise from data.
- During training, they learn how noise is added to data over time and how to reverse this process to recover the original data.
- This involves using probabilities to make educated guesses about what the data looked like before noise was added.
- This approach is essential for the model's capability to accurately reconstruct data, ensuring the outputs aren't just noise-free but also closely resemble the original data.

Denoising diffusion probabilistic models (DDPMs)

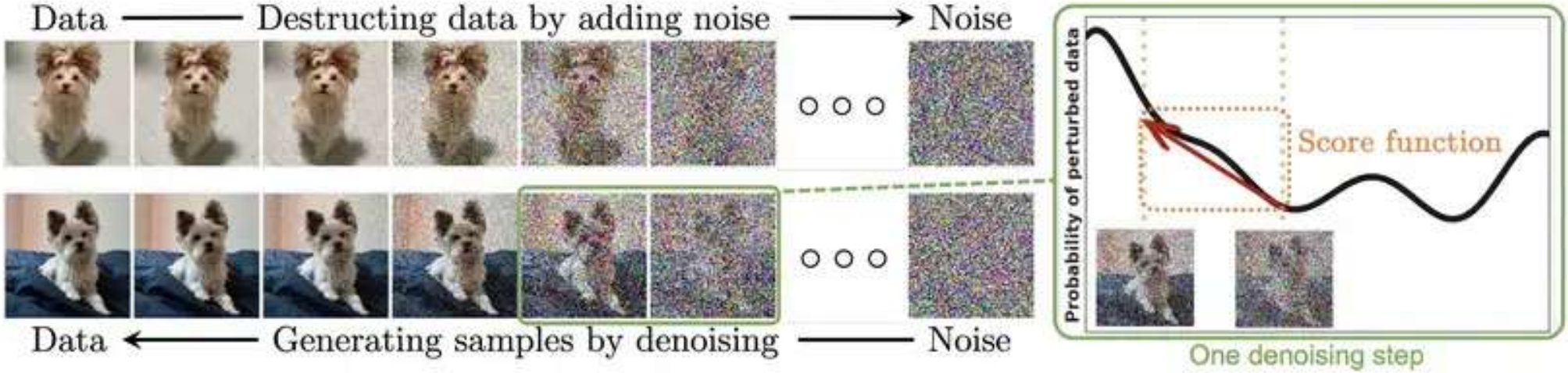
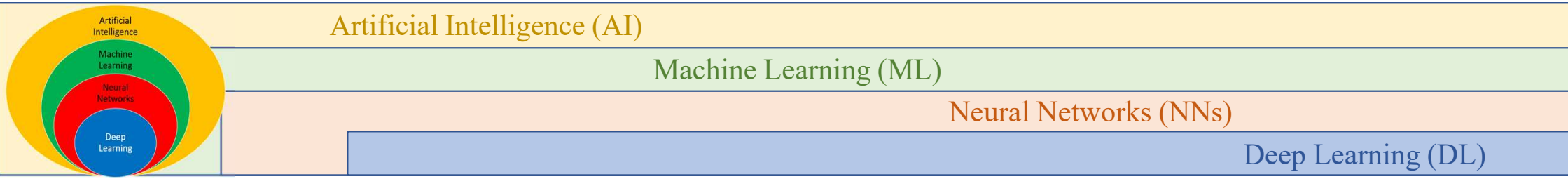


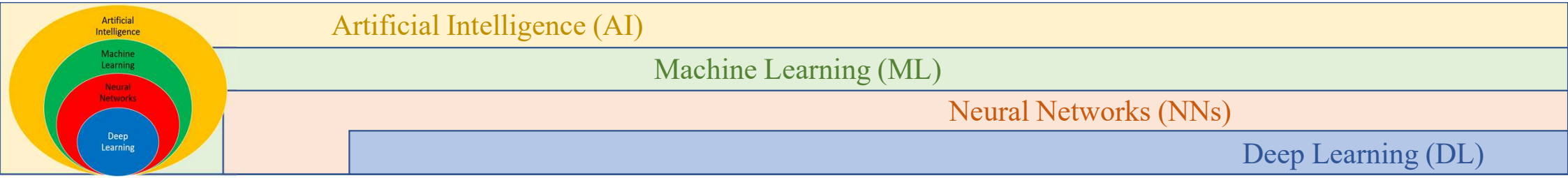
Fig. 2. Diffusion models smoothly perturb data by adding noise, then reverse this process to generate new data from noise. Each denoising step in the reverse process typically requires estimating the score function (see the illustrative figure on the right), which is a gradient pointing to the directions of data with higher likelihood and less noise.



Denoising diffusion probabilistic models (DDPMs)

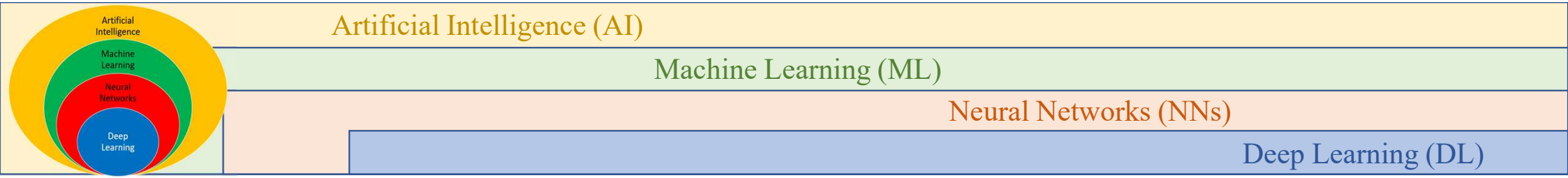
- Together, these components enable diffusion models to transform simple noise into detailed and realistic outputs, making them powerful tools in generative AI.

- Understanding these elements helps in appreciating the complex workings and capabilities of diffusion models.



Techniques for speeding up diffusion models

- Generating a sample from DDPM using the reverse diffusion process is quite slow because it involves many steps, possibly up to a thousand.
- For instance,, it takes about 20 hours to generate 50,000 small images with a DDPM, while a GAN can create the same amount in less than a minute using an Nvidia 2080 Ti GPU.
- There is an alternative method called Denoising Diffusion Implicit Model (DDIM) that stands out for its efficiency and quality.
- Unlike traditional models, DDIM needs fewer steps to create clear images from noisy data.



Techniques for speeding up diffusion models

❖ When you compare DDIM with the traditional Denoising Diffusion Probabilistic Model (DDPM), the key benefits are:

✓ **Speed**: DDIM works much faster, needing fewer steps to achieve results that used to take much longer.

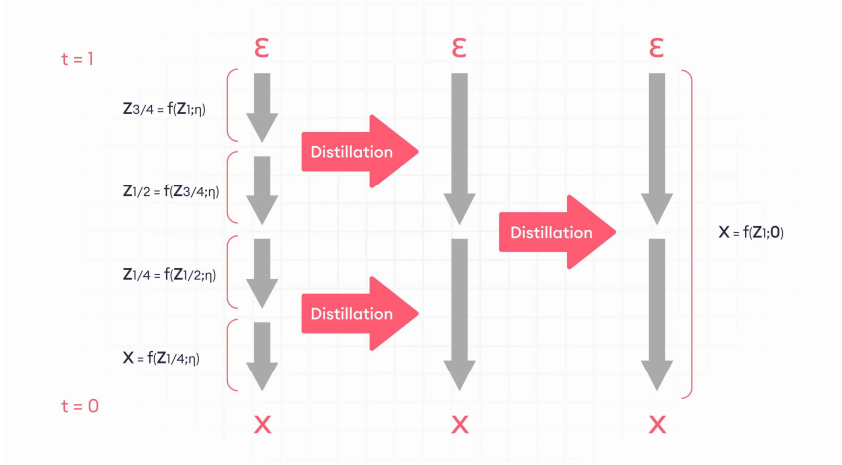
✓ **Predictability**: With DDIM, the results are more predictable. If you start with the same seed, you'll get very similar images every time. This reliability is perfect for tasks where you can't afford surprises.

New techniques that improve DDIM

- There are techniques that improve DDIM even more:

- **Progressive distillation**: Progressive distillation takes a complex model (the teacher) and simplifies it into a less complex one (the student).

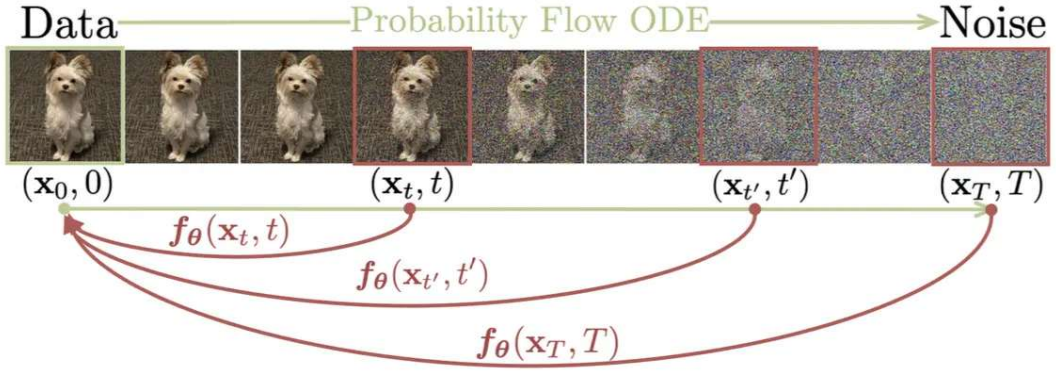
- This student model learns to achieve in one step what the teacher model does in two, speeding up the whole process without losing accuracy.

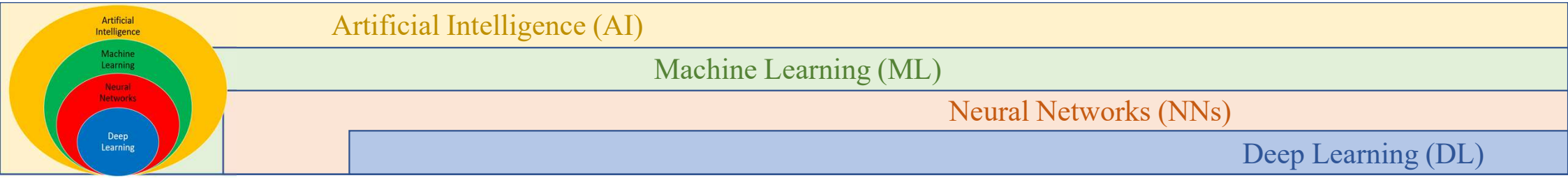


New techniques that improve DDIM

- There are techniques that improve DDIM even more:

- **Consistency models:** In 2023, Song and colleagues introduced a method that helps any point in the image generation process trace its way back to the start. This "Consistency Model" ensures that all points that follow the same path lead back to the same origin. It's like having a reliable map that guides every point back to where it began, which is crucial for maintaining the integrity and consistency of the images generated.



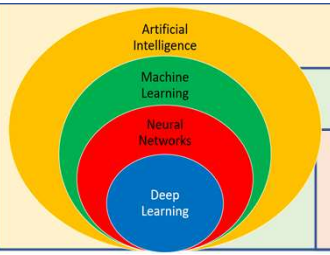


Applications of diffusion models

- There are very diverse applications of diffusion models, one of the most exciting being digital art creation.

- Artists can use these models to transform abstract concepts or textual descriptions into detailed, visually striking images.

- This capability allows for a new form of artistic expression where the boundary between technology and art blurs, enabling creators to explore new styles and ideas previously difficult or impossible to achieve.



Artificial Intelligence (AI)

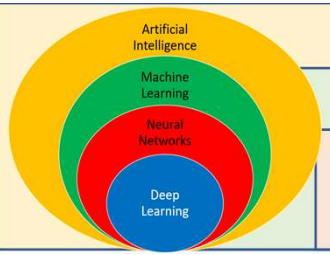
Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

Graphic design

- In graphic design and illustration, diffusion models provide a tool for rapidly generating visual content.
- Designers can input sketches, layouts, or rough ideas, and the models can flesh these out into complete, polished images.
- This can significantly speed up the design process, offering a range of possibilities from the initial concept to the final product.



Artificial Intelligence (AI)

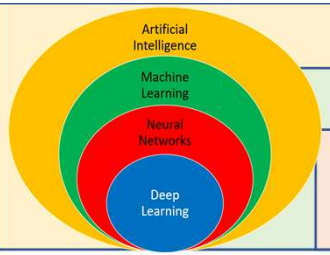
Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

Film and animation

- Another creative application is in the field of film and animation.
- Diffusion models can generate realistic backgrounds, characters, or even dynamic elements within scenes, reducing the time and effort required for traditional production methods.
- This streamlines the workflow and allows for greater experimentation and creativity in visual storytelling.
- An artist used a set of Stable Diffusion algorithms to produce the first full AI animation.



Artificial Intelligence (AI)

Machine Learning (ML)

Neural Networks (NNs)

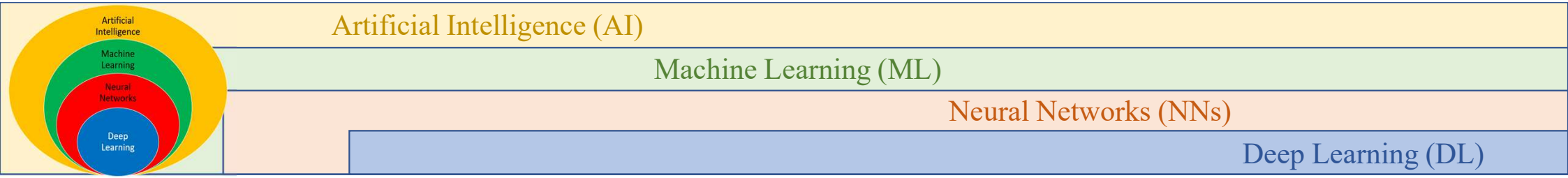
Deep Learning (DL)

Music and sound design

- In music and sound design, generative diffusion models can be adapted to generate unique soundscapes or represent music, offering new ways for artists to visualize and create auditory experiences.

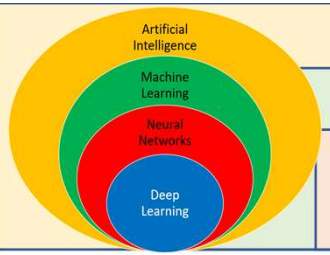
- A paper titled "Controllable Music Production with Diffusion Models and Guidance Gradients" discusses a diffusion model example used in the music industry.

- The authors demonstrate how conditional generation from diffusion models can be used to tackle a variety of realistic tasks in the production of music.



Media and gaming industry

- The interactive media and gaming industry also stands to benefit from diffusion models.
- They can be used to create detailed environments, characters, and other assets, adding realism and immersion to games and interactive experiences previously challenging to achieve.
- In essence, diffusion models are a powerful tool for anyone in the creative field, offering a blend of precision, efficiency, and artistic freedom.
- These models allow creators to push the boundaries of traditional mediums, explore new forms of expression, and bring imaginative concepts to life with unprecedented ease and detail.



Artificial Intelligence (AI)

Machine Learning (ML)

Neural Networks (NNs)

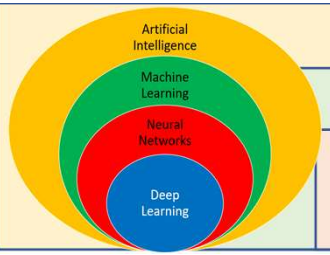
Deep Learning (DL)

Popular Diffusion tools

- Some of the most popular diffusion models, which have gained widespread attention for their impressive capabilities in image generation, include:

- **DALL-E 2**: Developed by OpenAI, DALL-E 2 is known for highly detailed and creative images from textual descriptions.

- It uses advanced diffusion techniques to produce images that are both imaginative and realistic, making it a popular tool in creative and artistic applications.



Artificial Intelligence (AI)

Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

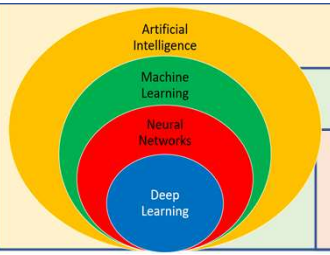
Popular Diffusion tools

- Some of the most popular diffusion models, which have gained widespread attention for their impressive capabilities in image generation, include:

- **DALL-E 3**: DALL-E 3 is the latest version of OpenAI image generation models and is a huge advancement over DALL-E 2.

- The most notable change is that this latest version isn't just an app but is integrated into ChatGPT.

- It also stands out with its image generation quality.



Artificial Intelligence (AI)

Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

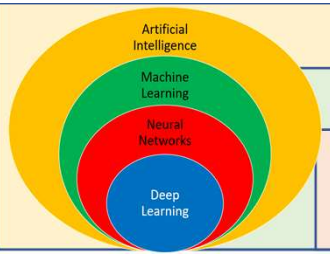
Popular Diffusion tools

- Some of the most popular diffusion models, which have gained widespread attention for their impressive capabilities in image generation, include:

- **Sora**: Sora's the latest model by OpenAI, and it's a game-changer.

- It is the first-ever text-to-video model by OpenAI.

- Sora can make 1080p videos in any resolution up to a minute long, and the videos it creates are scarily realistic.



Artificial Intelligence (AI)

Machine Learning (ML)

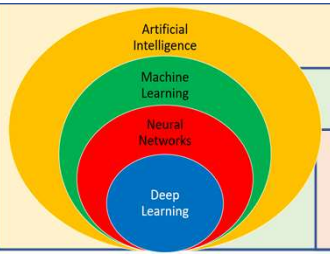
Neural Networks (NNs)

Deep Learning (DL)

Popular Diffusion tools

- Some of the most popular diffusion models, which have gained widespread attention for their impressive capabilities in image generation, include:
- **Sora**: Sora's the latest model by OpenAI, and it's a game-changer.
- It is the first-ever text-to-video model by OpenAI.
- Sora can make 1080p videos in any resolution up to a minute long, and the videos it creates are scarily realistic.

Prompt: The camera directly faces colorful buildings in Burano, Italy. An adorable dalmatian looks through a window on a building on the ground floor. Many people are walking and cycling along the canal streets in front of the buildings.



Artificial Intelligence (AI)

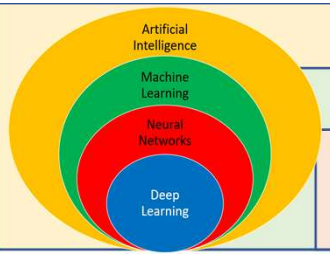
Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

Popular Diffusion tools

Prompt: The camera directly faces colorful buildings in Burano, Italy. An adorable dalmatian looks through a window on a building on the ground floor. Many people are walking and cycling along the canal streets in front of the buildings.



Artificial Intelligence (AI)

Machine Learning (ML)

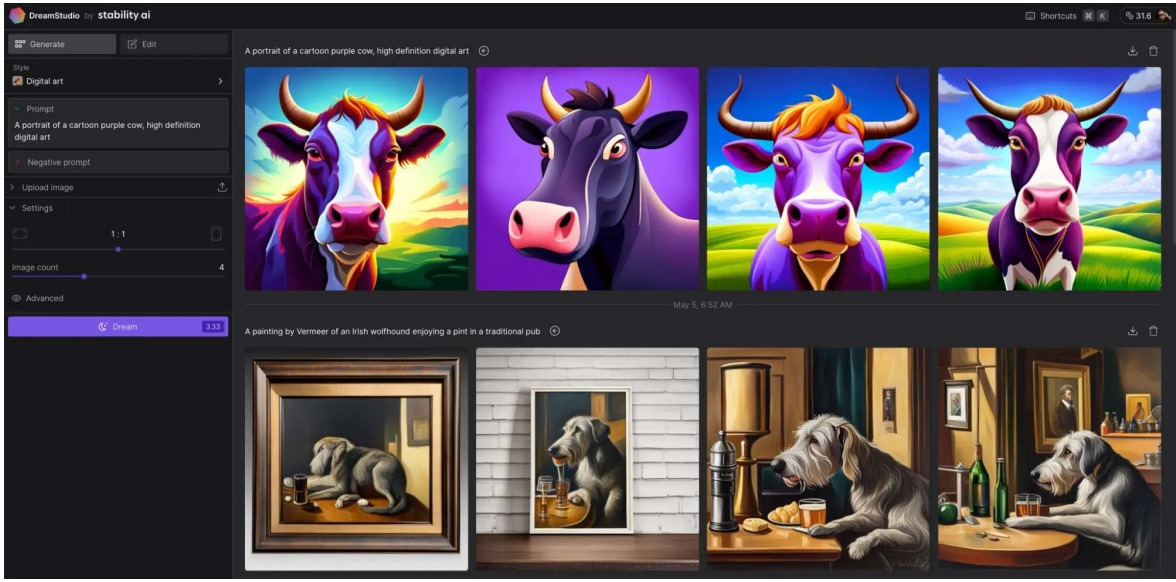
Neural Networks (NNs)

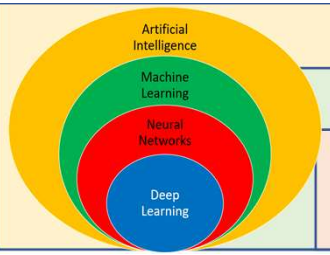
Deep Learning (DL)

Stable Diffusion

- Stable diffusion was created by researchers at Stability AI, who had previously taken part in inventing the latent diffusion model architecture used by Stable Diffusion.

- This model stands out for its efficiency and effectiveness in converting text prompts into realistic images. It has been recognized for its high-quality image generation capabilities.





Artificial Intelligence (AI)

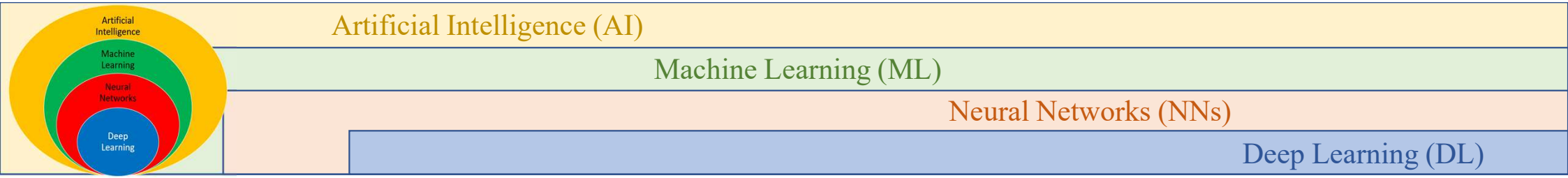
Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

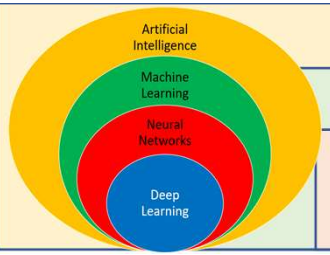
Stable Diffusion

- Stable Diffusion 3 is Stability AI's most recent release, and it's impressive.
- It's their most capable text-to-image model with greatly improved performance in multi-subject prompts, image quality, and spelling abilities.
- Look at the angled text "Stable diffusion" on the side of the bus. This used to be a dream for image generation tools.



Stable Diffusion

- The Stable Diffusion 3 collection spans models with 800M to 8B parameters, embracing Stability AI's will to widen access for everyone.
- This variety ensures users can find the perfect fit for their scalability and quality requirements, fueling their creativity.
- They've just opened the waitlist for an early preview.
- Stable diffusion also features an exciting application that can extend an image in various directions. This is called stable diffusion outpainting and is used to expand an image beyond its original borders.



Artificial Intelligence (AI)

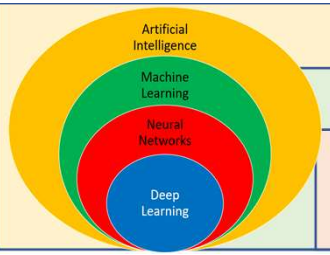
Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

Stable Diffusion

- Prompt: Resting on the kitchen table is an embroidered cloth with the text 'good night' and an embroidered baby tiger. Next to the cloth, there is a lit candle. The lighting is dim and dramatic.



Artificial Intelligence (AI)

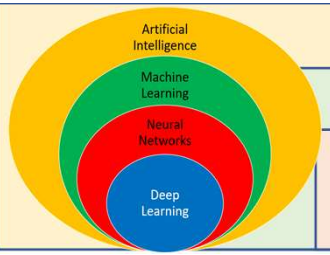
Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

Diffusion model limitations

- Deploying diffusion models like those used in DALL-E can be challenging.
- They are computationally intensive and require significant resources, which can be a hurdle for real-time or large-scale applications.
- Additionally, their ability to generalize to unseen data can be limited, and adapting them to specific domains may require extensive fine-tuning or retraining.
- Integrating these models into human workflows also presents challenges, as it's essential to ensure that the AI-generated outputs align with human intentions.
- Ethical and bias concerns are prevalent, as diffusion models can inherit biases from their training data, necessitating ongoing efforts to ensure fairness and ethical alignment.



Artificial Intelligence (AI)

Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

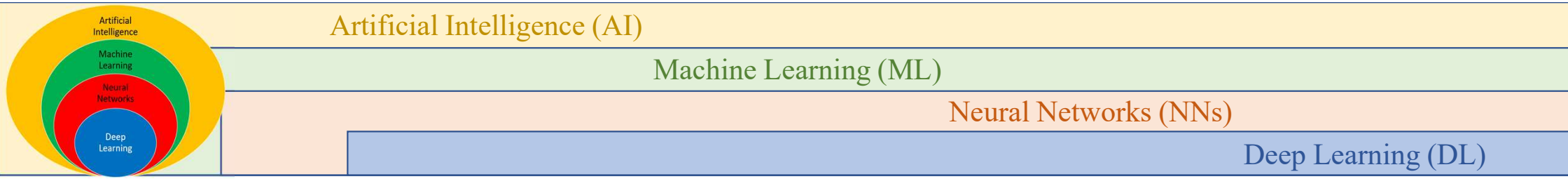
Diffusion model limitations

- Also, the complexity of diffusion models makes them difficult to interpret, posing challenges in applications where understanding the reasoning behind outputs is crucial.

- Managing user expectations and incorporating feedback to improve model performance is an ongoing process in the development and application of these models.

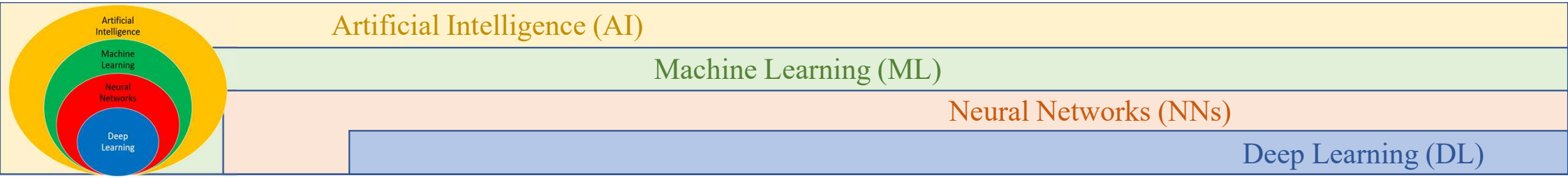
- Another big downside is their slow sampling time: generating high-quality samples takes hundreds or thousands of model evaluations.

- There are two main ways to address this issue: The first is new parameterizations of diffusion models that provide increased stability when using a few sampling steps. The second method is the distillation of guided diffusion models. Progressive distillation for fast sampling of diffusion models to distill a trained deterministic diffusion sampler results in a new diffusion model that takes half as many sampling steps.



Stable Diffusion

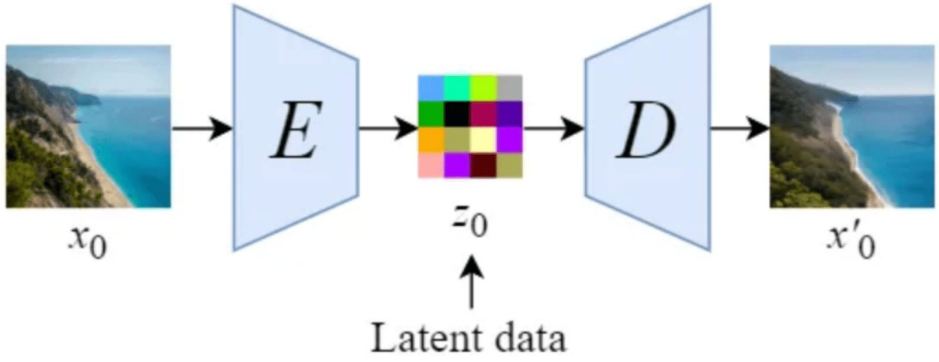
- The reverse diffusion process in conventional diffusion models involves iteratively passing a full-sized image through the U-Net architecture in order to obtain the final denoised result.
- However, this iterative nature presents challenges in terms of computational efficiency.
- This is emphasized when dealing with large image sizes and a high number of diffusion steps (T).
- The time required for denoising the image from Gaussian noise during sampling can become prohibitively long.
- To address this issue, a group of researchers proposed a novel approach called Stable Diffusion, originally known as Latent Diffusion Model (LDM) [15].

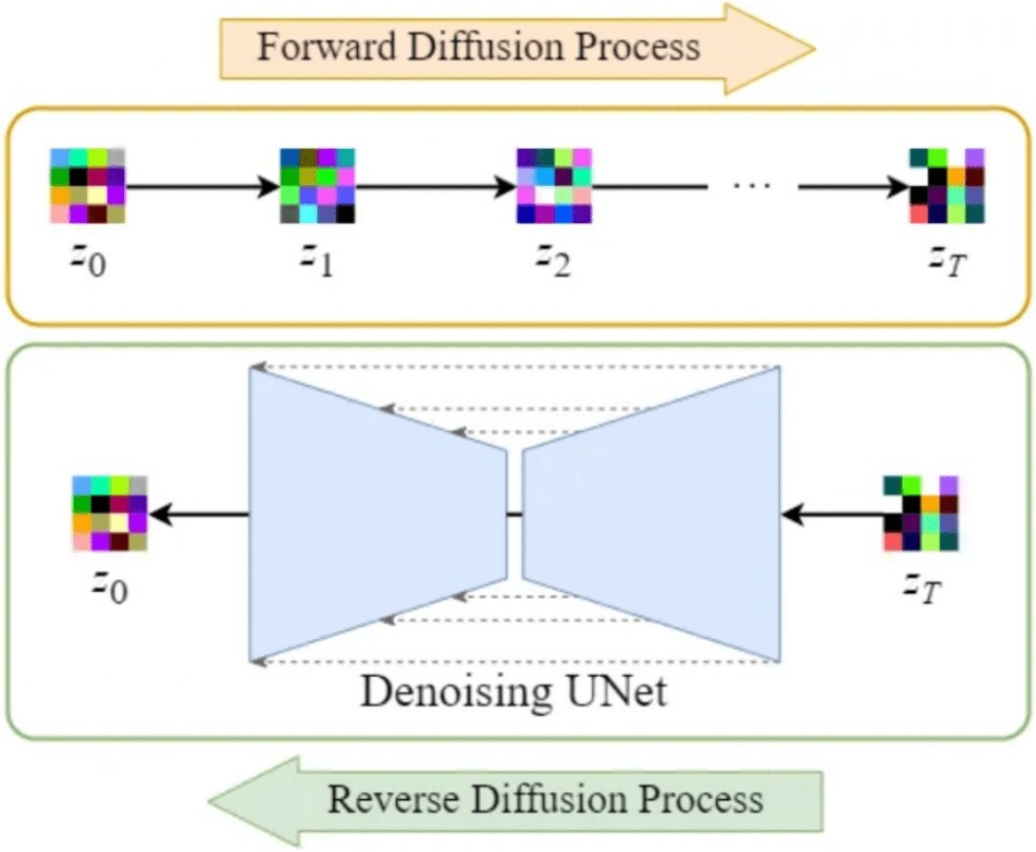


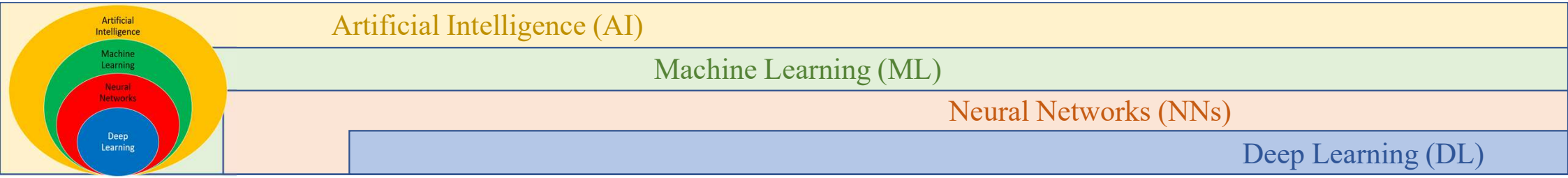
Latent Diffusion Models

- Stable Diffusion introduces a key modification by performing the diffusion process in the latent space.
- This works by using a trained Encoder E for encoding a full-size image to a lower dimension representation (latent space).
- Then making the forward diffusion process and the reverse diffusion process within the latent space.
- Later on, with a trained Decoder D, we can decode the image from its latent representation back to the pixel-space.
- For constructing the encoder and decoder, we can train some variant of a Variational AutoEncoder (VAE). This network is then decoupled for using both components separately.

Latent Diffusion Models

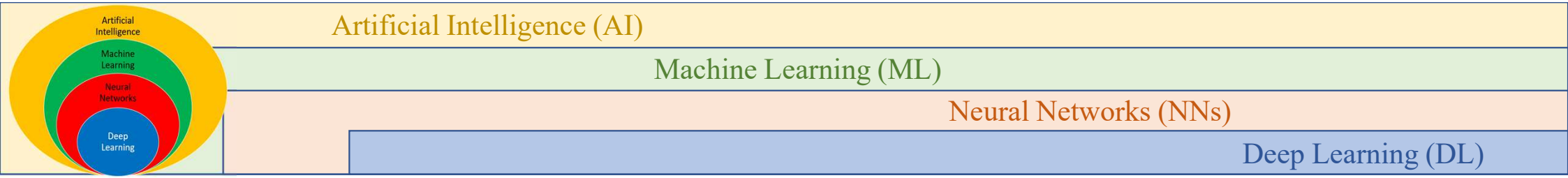






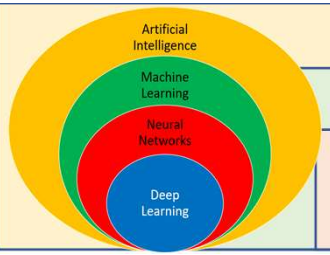
Latent Diffusion Models

- Shifting diffusion operations to the latent space in Stable Diffusion enhances speed and reduces costs.
- This advancement accelerates denoising and sampling processes, making it an efficient solution for high-quality image generation and stable training.
- By leveraging the latent space, Stable Diffusion eases the computational burden in the reverse diffusion process.
- This enables quicker denoising of images, enhancing both speed and overall model stability and robustness.



Conditioning

- Until then, generating images of a specific class was possible mainly through the addition of the class label in the input. Commonly known as Classifier Guidance.
- However, one of the standout features of the Stable Diffusion model, is its ability to generate images based on specific text prompts or other conditioning inputs.
- This is achieved by introducing conditioning mechanisms into the inner diffusion model.
- To enable conditioning, the denoising U-Net of the inner diffusion model makes use of a cross-attention mechanism.
- This allows the model to effectively incorporate conditioning information during the image generation (denoising) process.



Artificial Intelligence (AI)

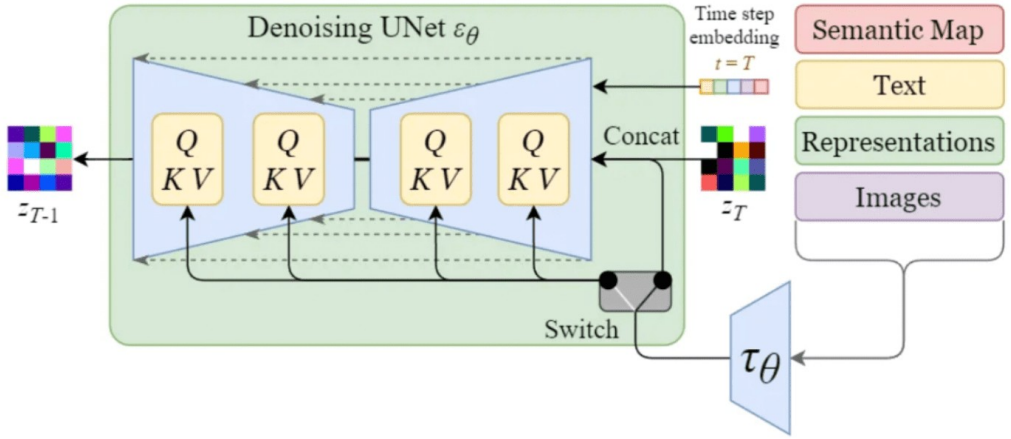
Machine Learning (ML)

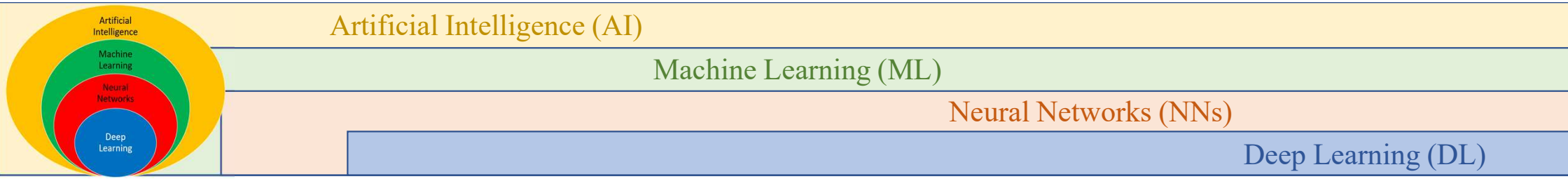
Neural Networks (NNs)

Deep Learning (DL)

Conditioning

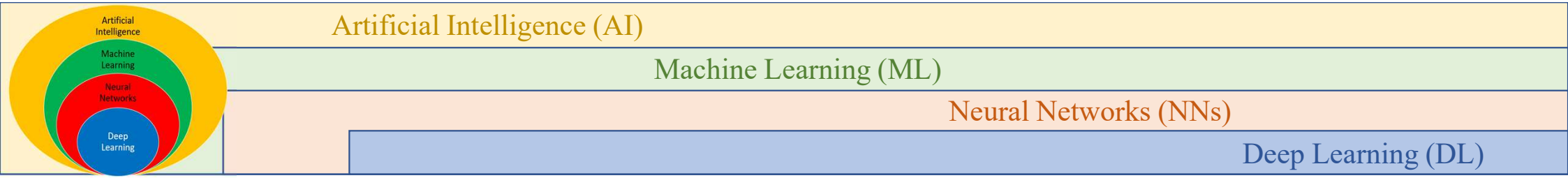
- The conditioning inputs can take various forms depending on the desired output:
- Text inputs are first transformed into embeddings through language models like BERT or CLIP. In the conditioning, we map these embeddings into the U-Net using a Multi-Head Attention layer, represented as Q, K, and V in the diagram.
- Other conditioning inputs such as spatially aligned data such as semantic maps, images, or inpainting act similarly.
- However, the integration of these conditioning mechanisms is usually achieved through concatenation.





Conditioning

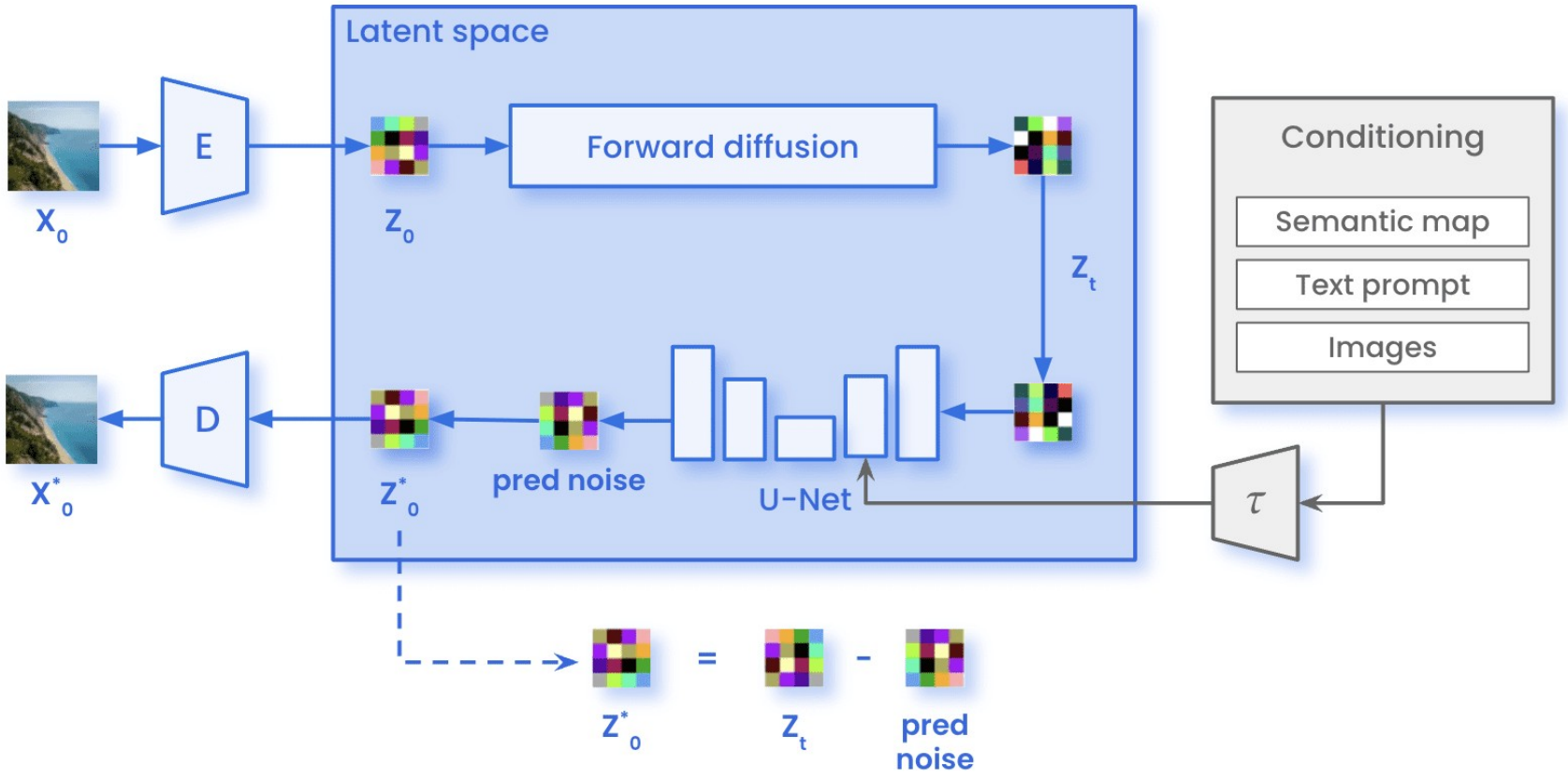
- By incorporating conditioning mechanisms, the Stable Diffusion model expands its capabilities to generate images based on specific additional inputs.
- Text prompts, semantic maps, or additional images, enable more versatile and controlled image synthesis.
- By using prompt engineering, it's possible to create even more compelling images.

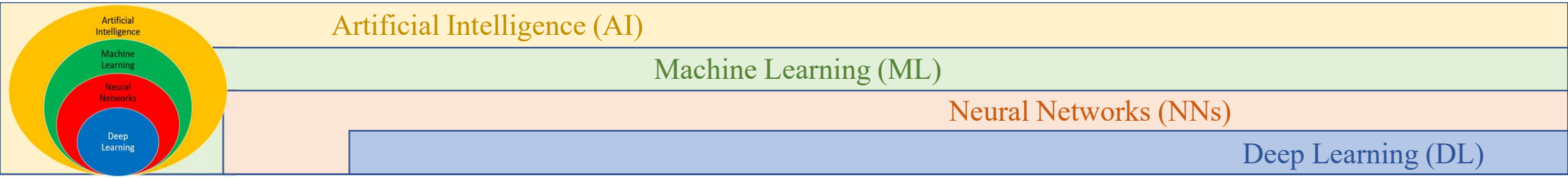


Architecture

- During training, the images (x_0) are encoded through the *Encoder E*, reaching the latent representation of the image (z_0).
- In the forward diffusion process, the image undergoes the addition of Gaussian noise, obtaining a noisy image (z_T).
- The image then passed through the U-Net, in order to predict the noise present in z_T.
- This comparison between the actual noise added in the forward diffusion and the prediction allows the calculation of the loss previously mentioned.
- With the calculated loss, we update the parameters of the U-Net through backpropagation.

Architecture

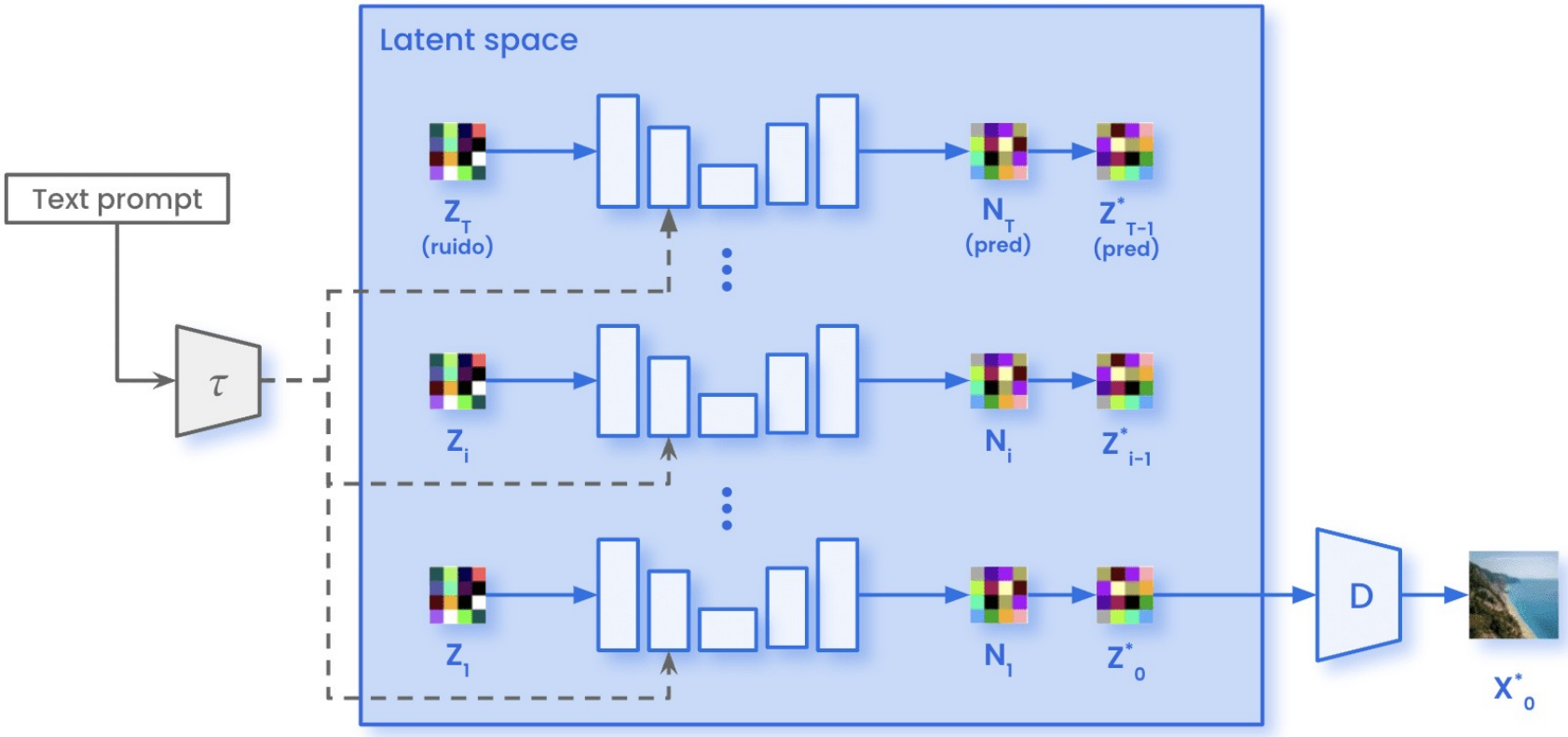


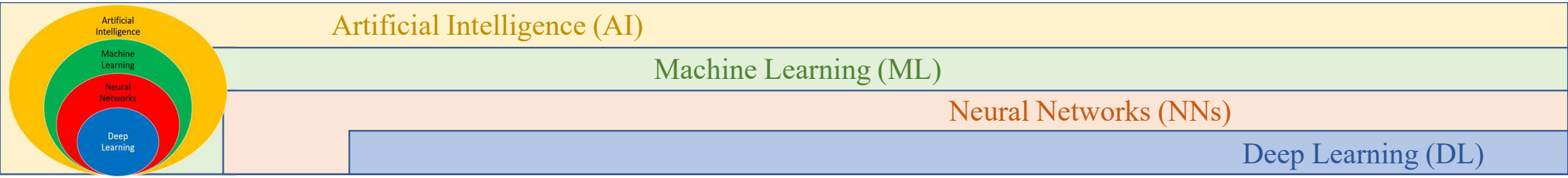


Architecture

- On the other hand, the forward diffusion process does not occur during sampling.
- We just sample Gaussian noise with the same dimensions present in the latent space (z_T).
- This noise passes through the U-Net for the specified number of inference steps T.
- At each step *t*, the U-Net predicts the whole noise present in the image.
- The model removes just a fraction of the predicted noise to obtain the representation of the image at timestep *t-1*.
- After all the T inference steps are iteratively, we obtain the representation within the latent space of the generated image ($\hat{z}_0$). Using the Decoder D, we can then transform that image from the latent space to the pixel-space ($\hat{X}_0$).

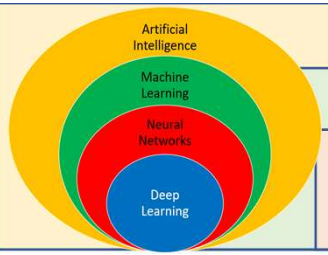
Architecture





Architecture

- On the other hand, the forward diffusion process does not occur during sampling.
- We just sample Gaussian noise with the same dimensions present in the latent space (z_T).
- This noise passes through the U-Net for the specified number of inference steps T.
- At each step *t*, the U-Net predicts the whole noise present in the image.
- The model removes just a fraction of the predicted noise to obtain the representation of the image at timestep *t-1*.
- After all the T inference steps are iteratively, we obtain the representation within the latent space of the generated image ($\hat{z}_0$). Using the Decoder D, we can then transform that image from the latent space to the pixel-space ($\hat{X}_0$).



Artificial Intelligence (AI)

Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)