Supplemental Digital Content.

The Supplemental Digital Content was not copyedited by *Archives of Pathology & Laboratory Medicine*.

Supplemental Table 1. Glossary of terms relevant to Artificial Intelligence (AI) and Generative AI (GAI)

Term	Definition
Attention mechanism	The attention mechanism is predominantly vital in driving compelling sequence modeling and transduction models. It enables the modeling of dependencies between elements in input or output sequences, irrespective of their distance [1-3]. In addition, it has been combined with recurrent networks in various research endeavors [4].
Autoregressive Model (ARM)	GAI model that uses a single neural network to predict the next element in a sequence by taking measurements from prior inputs in the sequence; it assumes there is a dependency on past values. The neural network can be recurrent, convolutional, or transformer-based, depending on the data type. Examples of applications include text generation, speech synthesis and music composition.
Computer vision	A branch of AI that enables computers to derive information from visual inputs such as whole slide digital images.
Convolutional Neural Network (CNN)	Class of deep neural network based on the human visual cortex that automatically and adaptively identify and generalize patterns from visual data. CNNs are commonly used for computer vision tasks including medical image analysis.

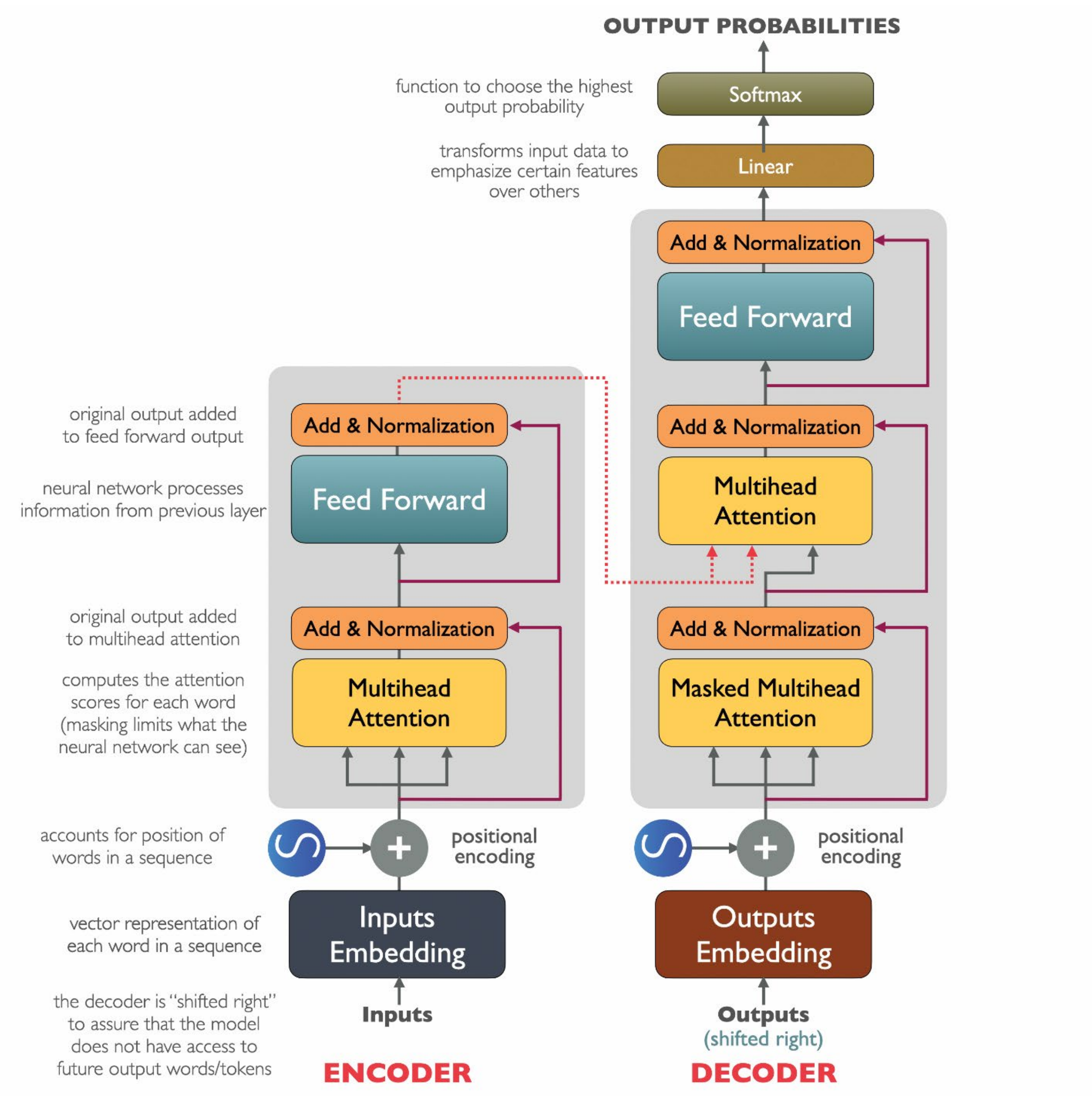
Data poisoning or contamination	Inappropriate and deliberate addition of non-representative examples in the training or fine-tuning dataset that leads to a bias in GAI output.
Decoder	Part of the transformer architecture. It is type of deep neural network architecture that uses cues and learned features from the encoder to convert numerical representations back to desired output
Encoder	Part of the transformer architecture. It is type of deep neural network architecture that extracts features from input data and converts it into a numerical representation
Fine tuning	The process of training a pre-trained large language model on specific tasks or datasets to help it generate optimal outputs specifically for domain-specific tasks.
Foundation Model (FM)	Deep learning neural network trained on terabytes of generalized and unannotated data with broad capabilities that can be adapted to a range of different, more specific purposes.
Generalizability	Ability of a model to perform similarly well with unknown data sets from different sites and settings beyond the training source (e.g., ability to handle variations in case mix and data compatibility)
Generative Adversarial Network (GAN)	GAI model that trains two neural networks, a generator and a discriminator, to compete against one another. The generator tries to create new data, and the discriminator attempts to predict if the output is real or fake. Through this iterative rivalry, the generator is forced to create more realistic content, and training is complete when the discriminator can no longer recognize the

	synthesized data. Examples of applications include image synthesis and style transfer.
Generative Pretraining Transformer (GPT)	Specific instance of a LLM that uses the transformer architecture. It is used by ChatGPT (OpenAI), T5 (Google), Megatron-LM (NVIDIA), and Jurassic-1 (AI21 Labs), among others.
Hyperparameters	Parameters that define the model architecture and control the learning process; these are specified in advance of the training process by the users and are not estimated from training data; e.g., number of hidden layers and number of nodes in each layer of a neural network
Large Language Model (LLM)	GAI model that uses a neural network architecture (often a transformer) to interpret, manipulate, or generate natural language. It has been trained on a massive dataset of text and code, allowing it to generate text, translate languages, write different kinds of creative content, and answer questions in an informative way.
Multi Modal Models	These are GAI models similar to LLMs with capabilities to work with text, images, sound, video, etc., at one time. Multimodality learning can also enable new forms of human-computer interaction, such as conversational agents, virtual assistants, and interactive systems.
Natural Language Processing (NLP)	Branch of AI that enables computers to understand, interpret, and manipulate human language.
Parameters	Weights and numerical values in LLMs that define the model's structure and behavior and help in optimizing desired output; these will be adjusted during the training process.

Pretraining	The initial phase of training large language models on a massive corpus of text data to capture general language patterns and features.
Prompt	Text or image used to as input query to generate optimized output
Reinforcement learning with human feedback	Reward mechanism used in fine tuning LLM based on output evaluation of LLM. Model output can be evaluated by humans who can give feedback in the form of suggestions and corrections. The final best output uses a reward mechanism to be selected as the desired output.
Retrieval-augmented generation (RAG)	AI framework designed to augment an LLM by integrating it with information retrieved from an external knowledge base.
Token	Smallest unit of data representing text or images, for example, that can be used by LMMs to generate new synthetic data that looks similar
Transformer	Specific kind of deep neural network that uses a self-attention mechanism to focus on particular inputs that are important in understanding the meaning of the entire input (e.g., sentence, image, etc.). This is particularly important in NLP tasks where interpreting the relationship between words and their context is essential.
Variational Autoencoders (VAEs)	GAI model that consists of two neural networks, an encoder and decoder. The encoder compresses the input data to a lower-dimensional representation (latent vector), where each dimension captures some aspect of the data variance in a probabilistic way. The decoder takes the latent vector and reconstructs the input data. Both are trained to minimize errors in the reconstruction process. VAEs generate new data by tweaking or

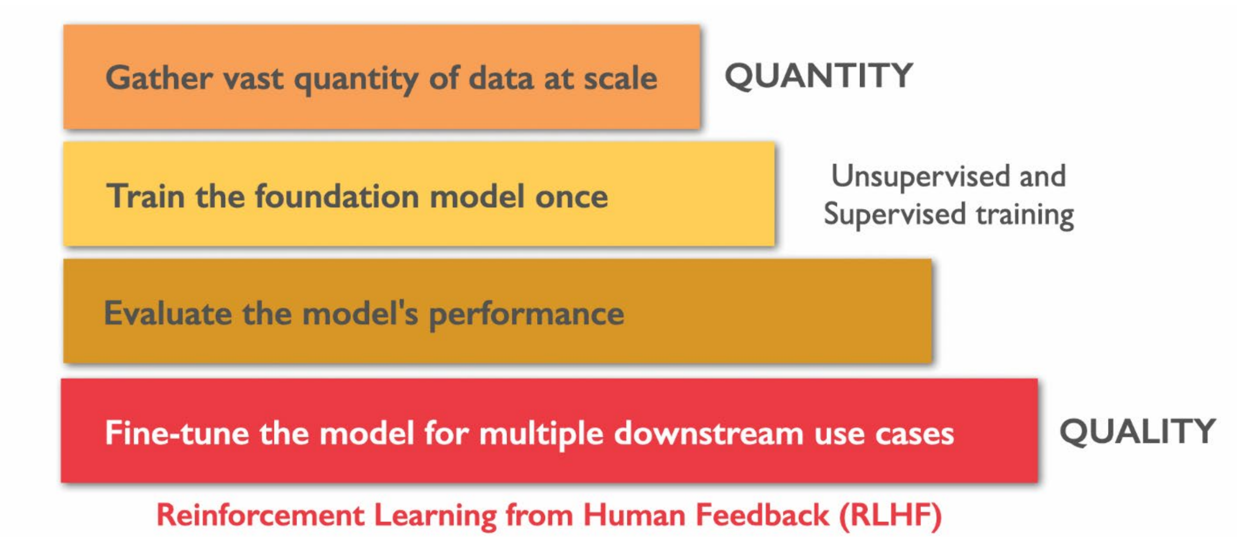
	randomly sampling the latent space. Examples of applications include image generation and anomaly detection.
Vision Image Transformers (ViT)	Deep learning models for computer vision that use transformer architecture. They can work with images and use patches as input tokens in transformers.

Supplemental Figure 1. Example of transformer model architecture designed for language. The encoder (shown on the left) takes input — text or an image — and breaks it into smaller pieces called tokens, which are then converted into numbers (called embeddings) that the computer can understand. The embeddings also contain information about the position and sequence of each token, which helps the model understand the order and context. This is done by several layers in the encoder ("Multi-Head Attention" and "Feed Forward") that analyze the input and extract important features. Once the computer has this translated information, the decoder steps in to make a response. It looks at the pieces and tries to predict what comes next by using everything the encoder learned about the order and connection of the tokens. This helps the decoder make a more accurate guess. When it's time to make a final guess, the decoder generates many possible answers, each with a score. The answer with the best score is the one it chooses for the final output. At each step, the decoder considers multiple possible outputs and chooses the one that seems most likely based on the input and context. (Adapted from "Yuening Jia, CC BY-SA 3.0, via Wikimedia Commons"; accessed 3/31/2024 [5])

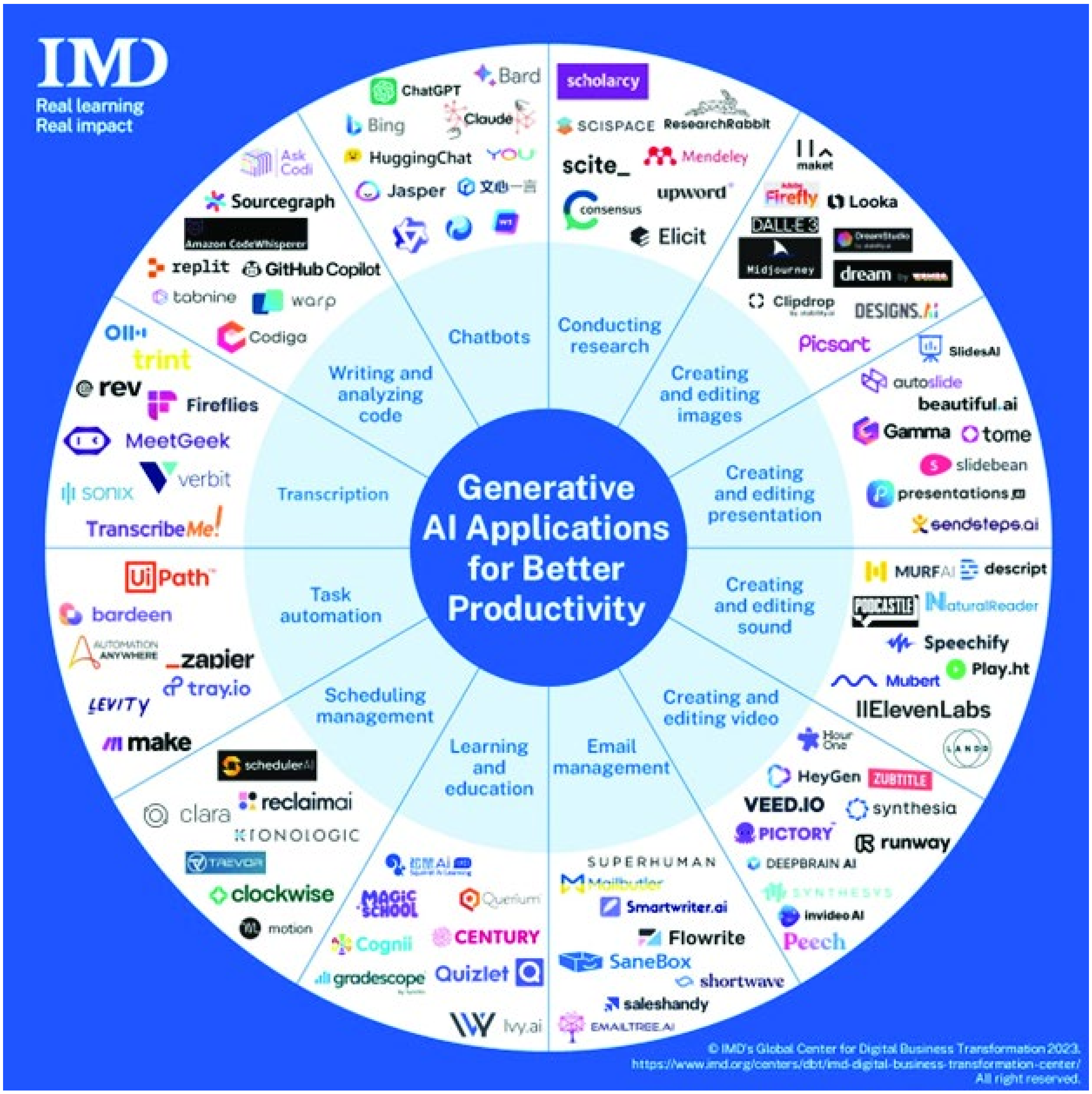


OUTPUT PROBABILITIES
function to choose the highest output probability
Softmax
transforms input data to emphasize certain features over others
Linear
Add & Normalization
Feed Forward
Add & Normalization
Multihead Attention
Add & Normalization
Masked Multihead Attention
original output added to feed forward output
Add & Normalization
neural network processes information from previous layer
Feed Forward
original output added to multihead attention
Add & Normalization
computes the attention scores for each word (masking limits what the neural network can see)
Multihead Attention
accounts for position of words in a sequence
positional encoding
positional encoding
vector representation of each word in a sequence
Inputs Embedding
Outputs Embedding
the decoder is "shifted right" to assure that the model does not have access to future output words/tokens
Inputs
Outputs (shifted right)
ENCODER
DECODER

Supplemental Figure 2. Foundational workflow for training generative artificial intelligence (GAI) models. There are four steps to the process, which focus on the balance between quantity and quality. The process begins with gathering a massive dataset of text, code, or images. This data is then used to train a foundation model using a combination of supervised and unsupervised learning techniques. Once trained, the model's performance is evaluated on a specific task. Finally, to optimize the quality of a response, the model can be fine-tuned for various downstream applications through reinforcement learning from human feedback.



Supplemental Figure 3. Examples of commercially available generative artificial intelligence (GAI) products grouped by task. Image used with permission from International Institute for Management Development (IMD) [6]. The field of GAI is rapidly evolving, so that new GAI products will continue to emerge or evolve.



Supplemental References:

1. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Guyon I, Von Luxburg U, Bengio S, et al, eds. *31st Conference on Neural Information Processing Systems (NIPS)*. Curran Associates, Inc.; 2017:6000-6010.

2. Kim Y, Denton C, Hoang L, Rush AM. Structured attention networks. *arXiv Preprint*. Feb 3 2017:arXiv:1702.00887. doi:10.48550/arXiv.1702.00887 Accessed May 20, 2024. https://api.semanticscholar.org/CorpusID:6961760

3. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv Preprint*. Sept 1 2014:arXiv:1409.0473. doi:10.48550/arXiv.1409.0473 Accessed May 1, 2024.

4. Parikh AP, Täckström O, Das D, Uszkoreit J. A decomposable attention model for natural language inference. *arXiv Preprint* Sept 25 2016:arXiv:1606.01933. doi:10.48550/arXiv.1606.01933 Accessed May 10, 2024.

5. Jia Y. Attention Mechanism in Machine Translation. *J Phys Conf Ser*. 2019;1314:012186. doi:10.1088/1742-6596/1314/1/012186

6. Wade M. Generative AI productivity map. Linkedin. Updated December 19, 2023. Accessed May 2, 2024, https://www.linkedin.com/posts/michael-wade-a435a8_here-it-is-an-update-of-our-generative-ai-activity-7142946521234210816-CV34