BANK OF AMERICA INSTITUTE

BANK OF AMERICA

Transformation

AI dictionary, part 1: The basics

30 June 2025

Key takeaways

- Artificial intelligence (AI) enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity, and autonomy. But just when we think we've wrapped our heads around what it is, AI rewrites the rules.
- The pace of innovation is accelerating so quickly that many of us struggle to keep up. And for those non-experts that try, the language and concepts can be bamboozling. This three-part "dictionary" series was created to help define key terms and provide everything you need to know about AI, the biggest revolution yet.
- This first installation covers the AI basics - from exploring the common types of AI widely used today, to discussing key AI technologies, including machine learning, deep learning, and natural language processing.

Artificial intelligence

What is it?

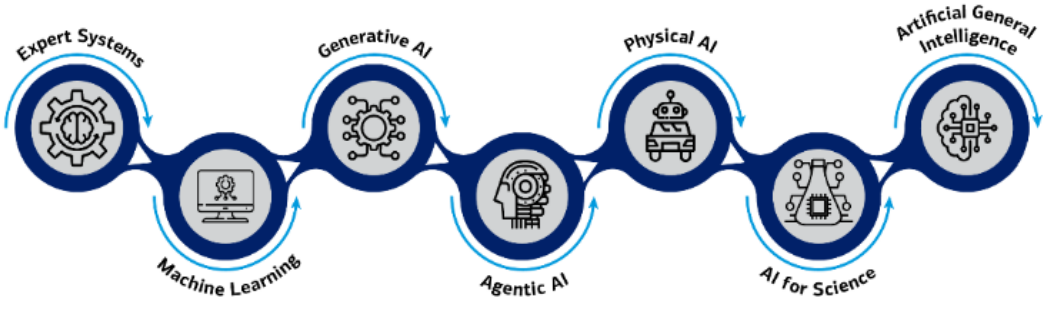
Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity, and autonomy.[1] It uses complex algorithms to process, learn from, and formulate solutions to problems using data, mimicking the human brain's process of intaking data, processing it, and drawing from it to make decisions. In building an AI compute tool, massive datasets must first be curated, then an AI algorithm or model is designed to fit a particular use case (a step called training). Finally, the compiled data is fed to the model to draw insights and conclusions (a step called inference).

Historic development of AI: Generative, agentic, embodied...

The field of AI was recognized as an academic discipline in 1956. However, it took decades for the applications based on AI to be commercially available, primarily due to limitations in the computational power required to run those models. Today, some of the key AI technologies include machine learning (ML), deep learning (DL), and natural language processing (NLP).

Generative AI (see below) has gained significant popularity since the release of ChatGPT. But the phases of AI are now evolving beyond generative to agents, physical, scientific, and general, requiring increasing infrastructure to enable them (Exhibit 1). Today, AI can recognize objects, understand natural language, and perform physical tasks (via robots).

Exhibit 1: The phases of AI are evolving beyond generative to agents, physical, scientific, and artificial general intelligence, all of which require increasing infrastructure to enable them
Illustrating seven phases of AI



Source: Steve Brown; BofA Global Research

BANK OF AMERICA INSTITUTE

[1] Kavlakoglu, E., & Stryker, C. (2024, August 9). *What Is Artificial Intelligence (AI)?* IBM. https://www.ibm.com/think/topics/artificial-intelligence

Generative AI

What is it?

Generative AI is a type of AI that can create content such as text, images, videos, audio, or software, in response to a user's prompt or request.[2] It uses deep learning (DL) models that simulate the learning of the human brain. The models identify patterns and relationships in datasets used to "train" them and use this information to create new content. In contrast, traditional AI systems are designed to recognize patterns and make predictions.

The shift to multimodality

Initial research and deployment of language models has been for text-based applications; however, the technology is now being applied to a much broader set of form factors, including numbers, images, video, music, and coding. As the capabilities and proficiency of AI improve, a combination of technologies can be applied to gradually more challenging domains, such as programming, art, and science (e.g., drug discovery or genetic analysis).

Use cases: Driving revenue enhancements and/or operational efficiencies

Given that generative AI has a wide variety of use cases, it can be embedded across many industries, and its increasing adoption can help drive revenue enhancements and/or operational efficiencies. To drive revenue enhancements, AI can be integrated into legacy and new products, such as robots, autonomous vehicles, or cybersecurity. AI strategies can also improve corporate efficiency and productivity by optimizing processes and reducing labor costs. For example, generative AI can be used for notetaking, coding, creating marketing material, human resource application filtering, customer feedback analysis, supply chain optimization, and database queries.

BofA Global Research views the launch of generative AI as the beginning of the next major tech cycle, following the introduction of the PC (personal computer) in 1981 and the internet in 1994. It emerged following 80 years of advances that produced more powerful chips and new neural network architectures, which reduced the time and cost to train increasingly large foundation models. As a result, generative AI apps could democratize access to powerful computer intelligence and drive a paradigm shift in corporate efficiency and productivity over the next one to three years.

The rise of ChatGPT

Generative AI has gained significant popularity since the release of ChatGPT in November 2022 – the first chatbot openly available to wide audiences. Previously, AI could read and write, but it could not *understand* content. However, applications like ChatGPT changed that, by leveraging machine learning (ML) and natural language processing (NLP) to create human-like conversational responses and content for a wide variety of purposes.

Since 2022, additional chatbot platforms have been introduced and have gained popularity. And today, beyond text-to-text models, there are now text-to-image and text-to-video models. Use cases include content generation (e.g., writing essays, news articles, social media posts, marketing content, stories, music, emails, etc.), data extraction, summarizing text, optimizing web browsers, language translation, and computer programming. In fact, programmers are already using this technology for program generation or to explain code or concepts.

Deep learning

What is it?

Deep learning (DL) is a subset of machine learning (ML) that uses algorithms inspired by the function and structure of the brain called artificial neural networks. The neural network is made up of "layers" – i.e., different stages where the data is processed. These layers are made up of interconnected nodes, each building on the previous layer to refine and optimize the model's prediction.[3] In turn, the interconnected nodes (or neurons, like in the human brain) transform the input data into abstract and meaningful representations. This structure allows for higher precision in the patterns it extracts from the raw data.

DL has more layers than ML, allowing the model to acquire more concepts

Traditional ML models use simple neural networks with one or two layers, while DL models use at least three but typically hundreds or thousands of layers to train the models.[4] By inputting data into a DL system with multiple layers, data abstraction is increased with each layer, which allows computer systems to acquire concepts that were previously unattainable with ML models.

DL's strength stems from the system's ability to ascertain additional data relationships that are difficult to identify. After sufficient training, the network of algorithms constantly improves predictions or interpretations.

[2] Scapicchio, M., & Stryker, C. (2024, March 22). *What is generative AI?* IBM. https://www.ibm.com/think/topics/generative-ai
[3] Holdsworth, J., & Scapicchio, M. (2024, June 17). *What is deep learning?* IBM. https://www.ibm.com/think/topics/deep-learning
[4] Ibid.

DL models used for complex systems that require pattern recognition

In illustrating the differences in use cases between ML and DL, a ML model might learn to identify and block spam emails based on patterns in previous emails or be used for recommendation systems on consumer platforms.[5] However, DL might be used for more complex systems that require pattern recognition, like image recognition, self-driving cars, and voice-controlled virtual assistants.[6] In fact, DL is the technology behind the algorithms in automated driving programs, which in some cases can detect the speed at which a driver is blinking and deduce from this data that they are in danger of falling asleep, alerting them and/or stopping the vehicle.

The three features of neural networks

Each neural network can be described by three attributes: 1) The **architecture** specifies the variables involved in the network and the relationship between them. In the neural network, it could be the weight and activities of "neurons" (interconnected nodes); 2) the **activity rule** defines how the activities of the neurons change in response to each other; and 3) the **learning rule** defines how the neural network's weight changes with time.

A variety of neural networks form the basis for generative AI models, including the following:

- **Convolutional neural networks** (CNNs) are used for image processing problems, such as classification and object recognition. Use cases include face detection, facial emotion detection, object detection, and document classification.
- **Recurrent neural networks** (RNNs) can process and predict sequential data. Traditional neural networks are feed-forward networks that cannot handle sequential data (like sentences). Use cases include autocorrection, fraud detection, and analytics.
- **Generative adversarial networks** (GANs) use two subnetworks to auto train a model. One is a discriminator model that differentiates between the real and generated data. The other is a generator model (typically a CNN) that generates images meant to mimic the actual data, where each data feed is sent to a discriminator model. As the model scales and learns, the generated outputs become closer to the actual data. Examples of generative AI models that use GANs are image generators.
- **Transformer** models apply mathematical methods to understand relationships between sequential data (e.g., words that make up a sentence). They take a series of inputs and generate output that is understandable. These models can be useful in real-time text/speech translation as well as in understanding patterns/relationships in genetic sequences in DNA.

Foundation models

What are they?

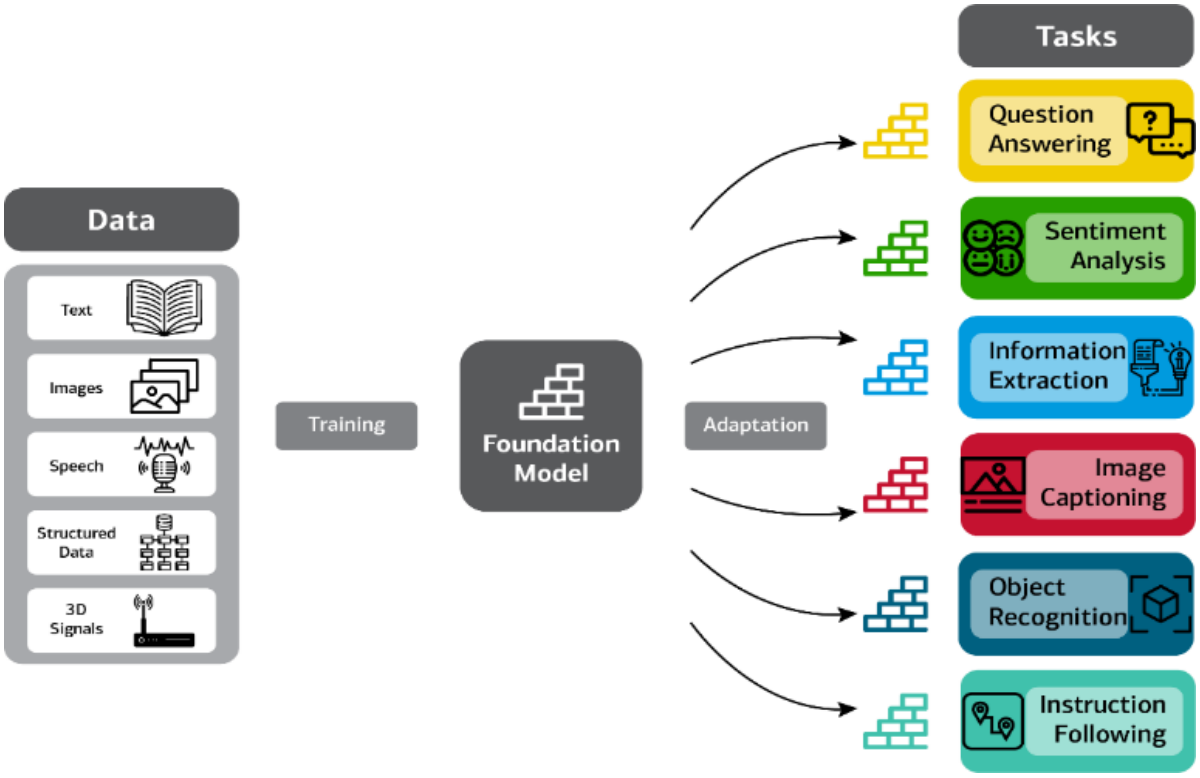
Foundation models (FMs) are a type of deep learning (DL) model, trained on a broad spectrum of generalized and unlabelled data, which can perform a wide variety of general tasks (e.g., understanding and conversing in natural language, answering a question, and generating text and images) (Exhibit 2).[7] Large language models (LLMs) are a popular type of foundation model and are focused on natural language processing (or, NLP, understanding and communicating with human language).

More general-purpose versus machine learning models created for specific tasks

FMs and traditional machine learning (ML) models differ, as traditional ML models typically perform specific tasks, such as sentiment analysis, classifying images, or forecasting trends. FMs, however, due to their size and general-purpose nature, can act as base models for more specialized downstream applications. This is done by training the FM on technical data within an industry or field.

[5] Rudderstack.
[6] Ibid.
[7] Amazon Web Services. (n.d.). *What are Foundation Models? - Foundation Models in Generative AI Explained - AWS*. Amazon Web Services, Inc. https://aws.amazon.com/what-is/foundation-models/

Exhibit 2: Foundation models centralize information from several data modalities to adapt to a wide range of tasks
Illustration of how foundation models work



Source: Center for Research on Foundation Models (CRFM), Stanford University Institute for Human-Centered Artificial Intelligence

BANK OF AMERICA INSTITUTE

Large language models

What are they?

Large language models (LLMs) are generative AI foundation models (FMs) that can recognize, understand, predict, and produce text, as they are designed for natural language processing (NLP) tasks. FMs are deep learning (DL) models trained on large datasets that can perform a wide variety of general tasks. They can act as a base for developing more specialized downstream applications. LLMs are specifically trained on language data, which makes them capable of understanding and generating natural language.[8]

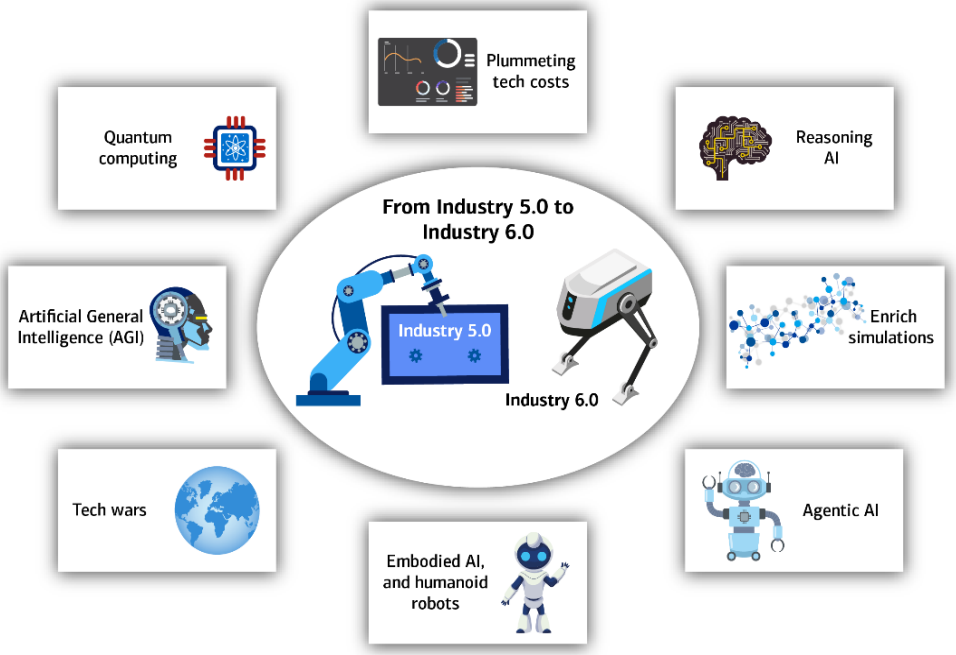
The AI LLM revolution accelerated the adoption of Industry 5.0, which emphasized collaboration between humans and advanced technologies, such as AI-driven robots, to optimize workplace processes. We are now seeing AI being integrated in every aspect of our lives and "humanizing" automated processes. This is moving society from the humanization era (Industry 5.0) to Industry 6.0, which aims to minimize human intervention by creating fully integrated, intelligent manufacturing systems based on the next generation of technologies (Exhibit 3).

More general-purpose versus traditional ML models created for specific tasks

The size and general-purpose nature of LLMs differentiate them from traditional machine learning (ML) models. Traditional ML models may typically perform specific tasks, such as sentiment analysis on text, classification of images, and forecasting of trends. LLMs have been trained on a large dataset of text and can generate text, like the training data. This is done by assigning a probability distribution over sequences of words. LLMs can generate code and new content at scale, summarize content, retrieve information, analyze sentiment, and translate languages.

[8] Daivi. (2024, October 28). *17 Best Open-Source LLMs Data Scientists Must Know*. ProjectPro. https://www.projectpro.io/article/open-source-llm-models/879

Exhibit 3: As the world moves from Industry 5.0 to 6.0, trends include falling tech costs, reasoning AI, enriching simulations, agentic AI, embodied AI, tech wars, artificial general intelligence (AGI), and quantum computing
Illustration of tech trends over the next five years



Source: BofA Global Research

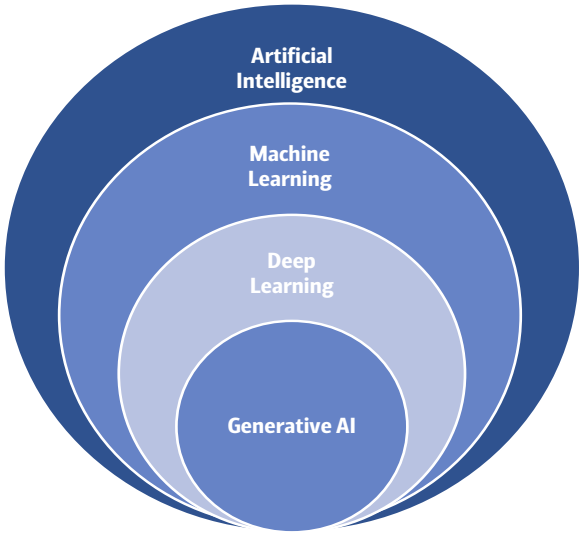
BANK OF AMERICA INSTITUTE

Machine learning

What is it?

Machine learning (ML) is a type of AI that allows programs to solve complex problems without being specifically programmed to do so (Exhibit 4). From being trained on large datasets of sentences and messages, the algorithm "learns" the patterns of human text language and then "predicts" what word is likely to be next (e.g., search engines predicting the next word a user may want to type). In general, with ML, if the data is of the same type, then increasing its amount will result in improved accuracy of output.

Exhibit 4: Machine learning is a type of AI, and deep learning is a type of machine learning
AI stack



Source: BofA Global Research

BANK OF AMERICA INSTITUTE

Choice of data inputs is key to improving model performance

There is still no single method of ML that can provide the most accurate responses to any problem or any type of data. And while there are three main stages in ML: data preparation, model training and development, and model deployment (Exhibit 5), one of the most important operations in applying ML is feature value selection. The features are the data items input into the ML process, and the accuracy of ML is heavily affected by the method of selecting and determining these features. It is also important to note that ML cannot respond to events that have not been learned yet. If the system is made to learn irregular events, then this will lower the accuracy of recognition under normal conditions.

Deep learning can independently recognize these data inputs

However, deep learning (DL), a type of ML, allows the computer to 1) independently recognize these feature values and 2) generate the program itself. These points represent a breakthrough for ML. Why? Because classic ML is more dependent on human intervention to learn – humans determine the set of features to understand the differences between data inputs, which typically requires more structured data to learn.[9]

Exhibit 5: Three main stages in ML: data preparation, model training and development, and model deployment
Machine learning workflow

Stage 1	Stage 2	Stage 3
Data preparation	**Model training and development**	**Model deployment**
Data ingestion	Train models	Deploy model
Data cleaning, processing	Tune hyperparameters (parameters to help the learning process)	Monitor model performance
Data labelling	Determine model performance (accuracy, recall, precision)	Improve model
Create variables (parameter)		

Source: BofA Global Research

BANK OF AMERICA INSTITUTE

Natural language processing

What is it?

Natural language processing (NLP) is a field within AI that enables machines to understand human language, including slang, contractions, and pronunciations, and in turn produce human-like text. NLP applies different algorithms to identify and extract natural language rules to convert the unstructured language data into a form that computers can interpret. NLP is a key tool to analyze text and speech data fully and efficiently.

Enables generative AI to understand human language to then generate new content

NLP is already part of our everyday life – powering search engines, translation apps, spam filters in email services, sentiment analysis, chatbots, and digital assistants. Generative AI uses NLP to understand and interpret human language, allowing it to generate new content.

Open source versus closed source

What are they?

Open source refers to software released under a license that grants users the right to use, study, modify, and distribute the software and its source code freely. Therefore, open-source models are publicly and freely accessible models that developers can use to create various applications. On the other hand, closed-source models are proprietary systems whose source code is not made publicly available.

Democratizes access to AI models, greater transparency and collaboration

Open-source models permit anyone to access and modify the model, allowing for more transparency and collaboration when compared to closed source, ultimately democratizing the use and access to AI models. According to AI experts, open-source models can be run locally on a company's own devices without model development or training costs, just inference cost, and can thus be much cheaper to access.[10] However, open-source models can have security risks, including the potential for malicious

[9] IBM. (2021, September 22). *What Is Machine learning?* IBM. https://www.ibm.com/think/topics/machine-learning
[10] Amit Mandelbaum.

code injection, data breaches, and vulnerabilities that can be exploited by malicious actors.[11] Closed source, on the other hand, allows the developers to ensure security and control over the source code.

Open or closed source, which will win?

Open-source models can be better than closed-source models because they allow developers to achieve the same performance with a smaller model in a more efficient way. However, per AI experts, rather than seeing these model developments as competitive as to which "wins" over time or which approach is best, the two methods *can* be complementary over time, with closed-source models used for complex instruction-based tasks and open-source models fine-tuned for specific repetitive tasks potentially with confidential/private data.[12]

Since ChatGPT's release, a variety of both closed- and open-source models have been introduced, with companies already starting to develop, adopt, or integrate AI into their products or businesses. The momentum is only likely to intensify from here, with more AI tools (e.g., simulation, knowledge graphs) and applications likely to be available soon, which could enable an abundance of opportunities beyond the digital to the physical domains of end-devices, robotics, and life sciences.

Small Language Model

What is it?

Like large language models (LLMs), small language models (SLMs) are generative AI models that can understand and generate natural language. They are smaller, with parameter size ranging from a few million to a few billion, whereas LLMs have hundreds of billions or even trillions of parameters.

More efficient than LLMs and more task-specific

Being smaller, SLMs are typically more efficient than LLMs and require less memory and computational power. This can make them more suitable for edge environments (places where computing occurs close to data origination) and mobile apps or when AI inferencing must be done offline without a data network. However, while SLMs are more efficient, they are not generalist models like LLMs. SLMs are more task-specific, like chatbots or virtual assistants. Performance-wise, SLMs are improving on metrics of common-sense reasoning, problem solving, and math, closing the gap between LLMs in general reasoning tasks.

Achieved via knowledge distillation, pruning, and quantization

SLMs' smaller size and efficiency are achieved via knowledge distillation (transfers knowledge from a pre-trained LLM to a smaller model), pruning (removes less useful parameters), and quantization. Quantization converts weights of high-precision data into low-precision data (e.g., 32-bit floating point to 8-bit integer), reducing the number of bits needed to represent information.

[11] Sidhu, A. (2025, March 17). *Securely Introducing Open Source Models into Your Organization.* Hidden Layer. https://hiddenlayer.com/innovation-hub/securely-introducing-open-source-models-into-your-organization/
[12] Amit Mandelbaum.

Contributors

Vanessa Cook

Content Strategist, Bank of America Institute

Lynelle Huskey

Analyst, Bank of America Institute

Sources

Haim Israel

Equity Strategist, BofA Global Research

Lauren-Nicole Kung

Equity Strategist, BofA Global Research

Felix Tran

Equity Strategist, BofA Global Research

Martyn Briggs

Equity Strategist, BofA Global Research

Disclosures

These materials have been prepared by Bank of America Institute and are provided to you for general information purposes only. To the extent these materials reference Bank of America data, such materials are not intended to be reflective or indicative of, and should not be relied upon as, the results of operations, financial conditions or performance of Bank of America. Bank of America Institute is a think tank dedicated to uncovering powerful insights that move business and society forward. Drawing on data and resources from across the bank and the world, the Institute delivers important, original perspectives on the economy, sustainability and global transformation. Unless otherwise specifically stated, any views or opinions expressed herein are solely those of Bank of America Institute and any individual authors listed, and are not the product of the BofA Global Research department or any other department of Bank of America Corporation or its affiliates and/or subsidiaries (collectively Bank of America). The views in these materials may differ from the views and opinions expressed by the BofA Global Research department or other departments or divisions of Bank of America. Information has been obtained from sources believed to be reliable, but Bank of America does not warrant its completeness or accuracy. These materials do not make any claim regarding the sustainability of any product or service. Any discussion of sustainability is limited as set out herein. Views and estimates constitute our judgment as of the date of these materials and are subject to change without notice. The views expressed herein should not be construed as individual investment advice for any particular person and are not intended as recommendations of particular securities, financial instruments, strategies or banking services for a particular person. This material does not constitute an offer or an invitation by or on behalf of Bank of America to any person to buy or sell any security or financial instrument or engage in any banking service. Nothing in these materials constitutes investment, legal, accounting or tax advice.