McKinsey & Company

McKinsey Explainers

What is tokenization?

Tokenization is the process of creating a digital representation of a real thing. Tokenization can also be used to protect sensitive data or to efficiently process large amounts of data.

July 2024

Events of the past few years have made it clear: we're hurtling toward the next era of the internet with ever-increasing speed. Several new developments are leading the charge. Web3 is said to offer the potential of a new, decentralized internet, controlled by participants via blockchains rather than a handful of corporations.

Web3 applications rely on a process called tokenization. In this case, tokenization is a digitization process to make assets more accessible. (AI models and new modes of payments also use a process called tokenization, both of which have little to do with Web3 tokenization—or each other, for that matter. In payments, tokenization is used for cybersecurity and to obfuscate the identity of the payment itself, essentially to prevent fraud. For a detailed description of tokenization in AI, see sidebar, "How does tokenization work in AI?")

After a couple false starts, tokenized financial assets are moving from pilot to at-scale development. McKinsey analysis indicates that tokenized market capitalization could reach around $2 trillion by 2030 (excluding cryptocurrencies like Bitcoin and stablecoins like Tether). Specifically, we expect that organizations working with certain asset classes will be the quickest adopters; these include cash and deposits, bonds and exchange-traded notes, mutual funds and exchange-traded funds, as well as loans and securitization. Larry Fink, the chairman and CEO of BlackRock, said in January 2024: "We believe the next step going forward will be the tokenization of financial assets, and that means every stock, every bond … will be on one general ledger."

But before we get specific, let's get the basics down.

Learn more about McKinsey's Financial Services Practice.

How does tokenization work?

In general, tokenization is the process of issuing a digital, unique, and anonymous representation of a real thing. In Web3 applications, the token is used on a (typically private) blockchain, which allows the token to be utilized within specific protocols. Tokens can represent assets, including physical assets like real estate or art, financial assets like equities or bonds, intangible assets like intellectual property, or even identity and data.

Web3 tokenization can create several types of tokens. One example from the financial-services industry is stablecoins, a type of cryptocurrency pegged to real-world money designed to be fungible, or replicable. Another type of token is an NFT—a nonfungible token, meaning a token that is provably scarce and can't be replicated—which is a digital proof of ownership people can buy and sell.

As noted earlier, AI also uses a concept called tokenization, which is quite different from Web3 tokens (despite their shared name). A large language model (LLM) used in an AI application could tokenize the word "cat" and use it to understand relationships between "cat" and other words. (For a more detailed explanation of what tokenization means in an AI context, see sidebar, "How does tokenization work in AI?")

The benefits of Web3 tokenization for financial institutions include the following:

- *Programmability.* Programmability is the ability to embed code in the token and its capacity to engage with smart contracts, enabling higher degrees of automation. (For more on smart contracts, read our blockchain *Explainer*.)
- *Composability.* This is the ability to interact with other assets and applications on the network.
- *Operational efficiency.* Web3 tokenization can help streamline processes, automate transactions, and more, making operations more efficient.

These, in turn, can mean increased efficiency, liquidity, and new revenue opportunities.

How does tokenization work in AI?

Tokenization in AI is used to break down data for easier pattern detection. Deep learning models trained on vast quantities of unstructured, unlabeled data are called foundation models. Large language models (LLMs) are foundation models that are trained on text. Trained via a process called fine-tuning, these models can not only process massive amounts of unstructured text but also learn the relationships between sentences, words, or even portions of words. This in turn enables them to generate natural-language text or perform summarization or other knowledge-extraction tasks.

Here's how tokenization makes this possible. When an LLM is fed input text, it breaks the text down into tokens. Each token is assigned a unique numerical identifier, which is fed back into the LLM for processing. The model learns the relationships between the tokens and generates responses based on the patterns it learns.

There are a number of tokenization techniques commonly used in LLMs:

— *Word tokenization* splits text into individual words or word-like units, and each word becomes a separate token. Word tokenization might struggle with contractions or compound words.

— *Character tokenization* makes each character in text its own separate token. This method works well when dealing with languages that don't have clear word boundaries or with handwriting recognition.

— *Subword tokenization* breaks down less frequently used words into units of frequently occurring sequences of characters. Subword tokens are bigger than individual characters but smaller than entire words. By breaking words into subword tokens, a model can better handle words that were not present in the training data. Byte pair encoding (BPE) is one subword tokenization algorithm. BPE starts with a vocabulary of characters or words and merges the tokens, which most often appear together.

— *Morphological tokenization* uses morphemes, which are individual words or parts of words that carry specific meanings or grammatical functions. The word "incompetence," for example, can be broken down into three morphemes: "in-" (a prefix indicating negation), "competent" (the root), and "-ence" (a suffix indicating a state or quality). In morphological tokenization, each morpheme becomes a token, which enables LLMs to handle word variations, understand grammatical structures, and generate linguistically accurate text.

The type of tokenization used depends on what the model needs to accomplish. Different tokenization methods may also be combined to achieve the required results.

What's an example of tokenization in practice?

Financial-services incumbents like BlackRock, WisdomTree, and Franklin Templeton, as well as Web3 natives Ondo Finance, Superstate, and Maple Finance, are increasingly adopting tokenized money market funds. In first quarter 2024, these funds surpassed $1 billion in total value (not much compared with total market size, but a milestone nonetheless).

Immutable data on the shared ledger reduces data errors associated with manual reconciliation, while 24/7 instant settlement and composability provide better user experience and new revenue sources.

What types of technologies make Web3 possible?

As we've seen, Web3 is a new type of internet, built on new types of technology. Here are the three main types:

— *Blockchain.* A blockchain is a digitally distributed, decentralized ledger that exists across a computer network and facilitates the recording of transactions. As new data are added to a network, a new block is created and appended permanently to the chain. All nodes on the blockchain are then updated to reflect the change. This means the system is not subject to a single point of control or failure.

— *Smart contracts.* Smart contracts are software programs that are automatically executed when specified conditions are met, like terms agreed on by a buyer and seller. Smart contracts are established in code on a blockchain that can't be altered.

— *Digital assets and tokens.* These are items of value that only exist digitally. They can include cryptocurrencies, stablecoins, central bank digital currencies (CBDCs), and NFTs. They can also include tokenized versions of assets, including real things like art or concert tickets.

As we'll see, these technologies come together to support a variety of breakthroughs related to tokenization.

What are the potential benefits of tokenization for financial-services providers?

Some industry leaders believe tokenization stands to transform the structure of financial services and capital markets because it lets asset holders reap the benefits of blockchain, such as 24/7 operations and data availability. Blockchain also offers faster transaction settlement and a higher degree of automation (via embedded code that only gets activated if certain conditions are met).

While yet to be tested at scale, tokenization's potential benefits include the following:

— *Faster transaction settlement, fueled by 24/7 availability.* At present, most financial settlements occur two business days after the trade is executed (or T+2; in theory, this is to give each party time to get their documents and funds in order). The instant settlements made possible by tokenization could translate to significant savings for financial firms in high-interest-rate environments.

— *Operational cost savings, delivered by 24/7 data availability and asset programmability.* This is particularly useful for asset classes where servicing or issuing tends to be highly manual and hence error-prone, such as corporate bonds. Embedding operations such as interest calculation and coupon payment into the smart contract of the token would automate these functions and require less hands-on human effort.

— *Democratization of access.* By streamlining operationally intensive manual processes, servicing smaller investors can become an economically attractive proposition for financial-services providers. However, before true democratization of access is realized, tokenized asset distribution will need to scale significantly.

— *Enhanced transparency powered by smart contracts.* Smart contracts are sets of instructions coded into tokens issued on a blockchain that can self-execute under specific conditions. One example could be a smart contract for carbon credits, in which blockchain can provide an immutable and transparent record of credits, even as they're traded.

— *Cheaper and more nimble infrastructure.* Blockchains are open source, thus inherently cheaper and easier to iterate than traditional financial-services infrastructure.

There's been hype around digital-asset tokenization for years, since its introduction back in 2017. But despite the big predictions, it hasn't yet caught on in a meaningful way. We are, though, seeing some slow movement: as of mid-2023, US-based fintech infrastructure firm Broadridge was facilitating more than $1 trillion monthly on its distributed ledger platform.

Learn more about McKinsey's Financial Services Practice.

How does a Web3 asset get tokenized?

There are four typical steps involved in asset tokenization:

1. *Asset sourcing.* The first step of tokenization is figuring out how to tokenize the asset in question. Tokenizing a money market fund, for

example, will be different from tokenizing a carbon credit. This process will require knowing whether the asset will be treated as a security or a commodity and which regulatory frameworks apply.

2. *Digital-asset issuance and custody.* If the digital asset has a physical counterpart, the latter must be moved to a secure facility that's neutral to both parties. Then a token, a network, and compliance functions are selected—coming together to create a digital representation of the asset on a blockchain. Access to the digital asset is then stored pending distribution.

3. *Distribution and trading.* The investor will need to set up a digital wallet to store the digital asset. Depending on the asset, a secondary trading venue—an alternative to an official exchange that is more loosely regulated—may be created for the asset.

4. *Asset servicing and data reconciliation.* Once the asset has been distributed to the investor, it will require ongoing maintenance. This should include regulatory, tax, and accounting reporting; notice of corporate actions; and more.

Learn more about McKinsey's Financial Services Practice, and check out Web3-related job opportunities if you're interested in working at McKinsey.

Articles referenced:

— "From ripples to waves: The transformational power of tokenizing assets," June 20, 2024, Anutosh Banerjee, Julian Sevillano, and Matt Higginson

— "Tokenization: A digital-asset déjà vu," August 15, 2023, Anutosh Banerjee, Ian De Bode, Matthieu de Vergnes, Matt Higginson, and Julian Sevillano

— "Tokenizing nontraditional assets: A conversation with Ascend Bit's Brian Clark," March 17, 2023, Andrew Roth

— "Web3 beyond the hype," September 26, 2022, Anutosh Banerjee, Robert Byrne, Ian De Bode, and Matt Higginson

— "How can healthcare unlock the power of data connectivity?," December 9, 2021, Prashanth Reddy

This article was updated in July 2024; it was originally published in October 2023.

Get to know and directly engage with senior McKinsey experts on tokenization.

Matt Higginson is a partner in McKinsey's Boston office, and **Prashanth Reddy** is a senior partner in the New Jersey office.

Designed by McKinsey Global Publishing