

TOKENS IN AI AND BLOCKCHAIN, CRYPTOCURRENCY

1. The term “token” has various meanings across different fields, including linguistics, cryptography, computing, natural language processing (NLP), and blockchain or cryptocurrency.
2. In NLP, a token is a basic unit of text, usually a word, part of a word, or punctuation. For example, the sentence “ChatGPT is smart” consists of four tokens.
3. Types of Tokenisation –
 - Word Tokenisation – breaks down text into individual words.
 - Subword Tokenisation (used by models like GPT) breaks down unknown or rare words into smaller known units.
 - Byte Pair Encoding (BPE) works by merging the most common pairs of characters or subwords. It's widely used in tokenization for GPT-3 and GPT-4, proving effective for managing rare words and supporting multilingual data.
 - Character tokenisation considers each character a separate token.
4. Tokens are crucial in NLP models like GPT-4 and BERT because they process input as individual tokens rather than raw text. Additionally, input and output length limits are measured in tokens rather than characters or words. For example, GPT-4 turbo has a context window of 128,000 tokens.
5. Token Count v/s Word Count
 - One Word ≠ One Token
 - Artificial = 1 token
 - “Unbelievable” = might be 2 -3 tokens depending on the tokenizer
6. In Cryptocurrency/Blockchain
 - What is a Token? In blockchain technology, a token refers to a digital asset created on a blockchain platform.

7. Types of Crypto Tokens

- Utility tokens, such as BAT and Filecoin, are used to access services within a platform.
- Security Tokens signify ownership or a financial interest, such as tokenized stocks or assets.
- Governance Tokens grant voting rights in decentralized applications, such as Uniswap (UNI) and Compound (COMP).
- Stablecoins tied to fiat currencies include USDT (Tether) and USDC.
- NFTs (Non-fungible tokens) are one-of-a-kind digital assets or collectibles, like Bored Ape and CryptoPunks.

8. The Blockchain Coins include native coins for their chains, such as Bitcoin (BTC) and Ether (ETH).

The Blockchain Tokens include those built on existing chains, such as USDT (on Ethereum), LINK, and UNI.

9. Token Standards include –

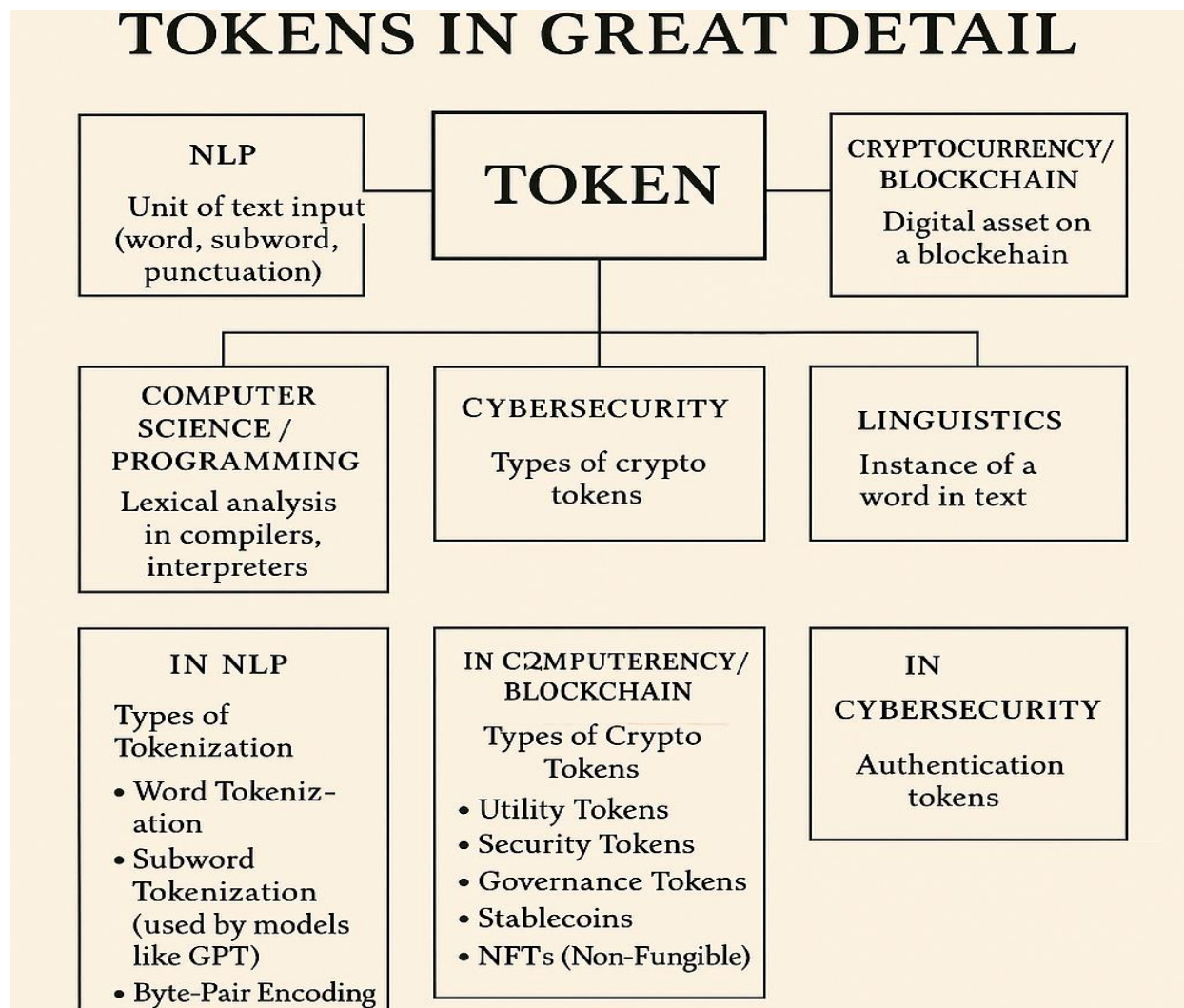
- ERC-20 includes standards for fungible tokens on the Ethereum blockchain.
- ERC-721 includes standards for NFTs.
- BEP-20, which includes the token standard for Binance Smart Chain.

10. As for Computer Science/Programming includes –

In lexical analysis, a token is a categorized block of text with various types, including keywords, identifiers, operators, literals, and punctuation.

11. In cybersecurity, authentication tokens verify user identity during login sessions. For instance, OAuth tokens are used in web logins, and JWT (JSON Web Tokens) securely transmit user data in JSON format. In linguistics, a token is a specific occurrence of a word.

In summary, NLP encompasses text units like words, subwords, and punctuation marks. Blockchain relates to digital assets recorded on a blockchain. Programming focuses on syntactic elements such as keywords and identifiers. Cybersecurity involves authentication markers for identity verification. Lastly, linguistics studies instances of words within texts.



-XXXX-

Author: Darshita Neeraj
Senior Researcher,
Mudita Learning and Knowledge Private Limited
Email: info@muditabooks.com
Date of post: 09.07.2026