Dr. Smith said: "Hello, everyone! Welcome to the one-day workshop on AI."

Numbers like 123 or symbols such as @, #, and $ should be ignored.

The cat (which was black) jumped over the lazy dog.

Did you notice the hyphenated word state-of-the-art? Amazing, right?

A quote inside dialogue: "She said, 'I can't believe it!'"

Finally, accents like café and naïve must be normalized.

- Example - of standalone punctuation with dashes.

Mrs. O'Connor exclaimed: "Wow! This project's going to be amazing!"

Some abbreviations like i.e., e.g., and vs. are tricky.

Remember: don't split contractions such as won't, shouldn't, and we're.

Check punctuation handling: commas, semicolons; colons: and periods.

Nested parentheses should be handled too (for example (like this) properly).

The user typed: "Login at user-name@example.com" — the email should be ignored.

Numbers mixed in words: version2, test123, alpha-1. Only alphabetic parts are tokens.

Quotes at the start of sentences: '"Amazing," she said.'

Standalone symbols: \*, &, %, $, #, ! — ignored as non-alphabetic symbols.

"Did you read the article on e-commerce?" asked Ms. Lee.

Be careful with contractions: o'clock, she'd, they'll, can't, shouldn't.

Watch numbers in context: Room 42, Level 3, Chapter 7. Ignore the digits.

Special punctuation inside sentences: Wow! Amazing; unbelievable: yes?

Nested quotes again: "He whispered, 'Don't worry, it'll be fine.'"

Standalone punctuation examples: - dash - ; semicolon ; : colon :

Emails and URLs: contact me at info@example.com or visit https://test-site.org. Ignore these.

Mixed alphanumeric: productX2, version1.5, alpha-3b — keep only letters.

Ellipses and multiple punctuation: Wait… what?! Really?!