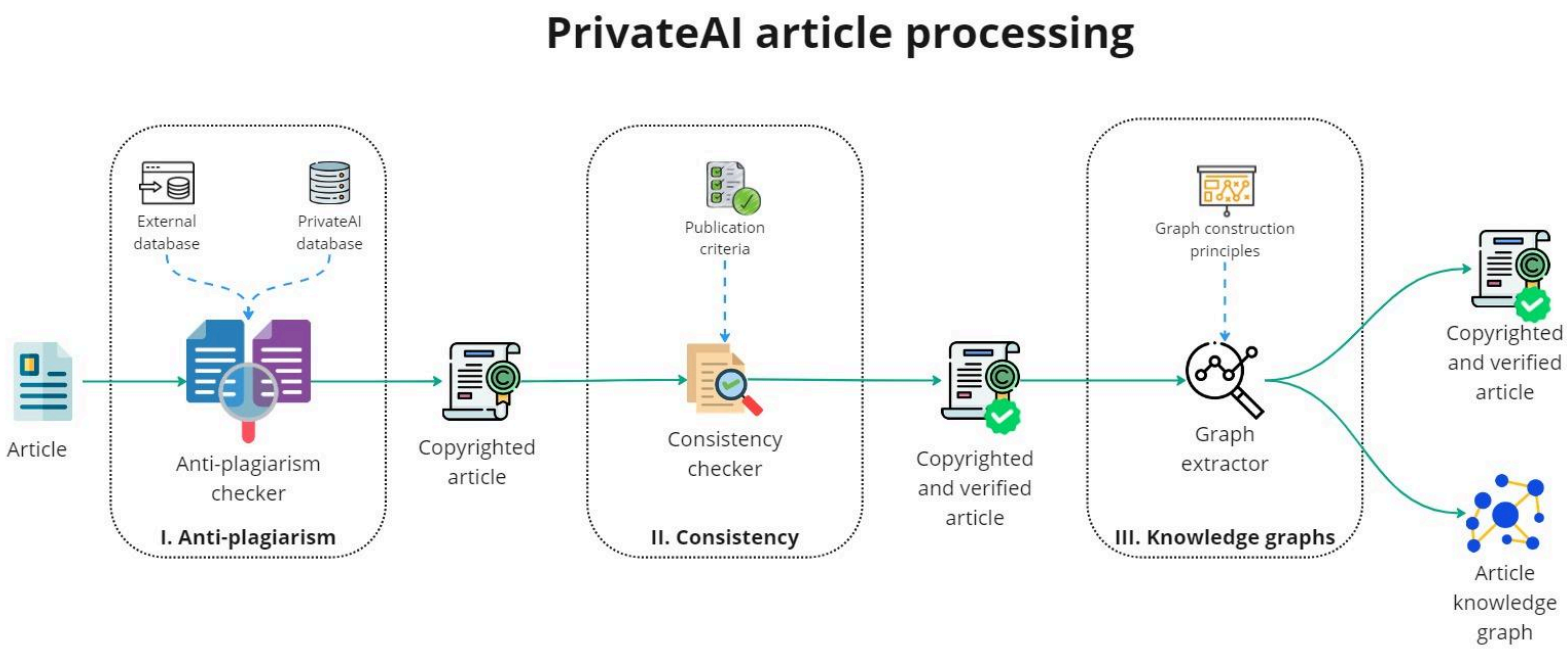
# [PrivateAI] Article processing

When uploading to the PrivateAI storage, an article should undergo three different processing stages by AI models:

1. Anti-plagiarism check
2. Consistency check
3. Knowledge graph extraction




## I. Anti-plagiarism

The reasons of implementing an anti-plagiarism system can be found here - [PrivateAI] Antiplagiarism.

Whereas such a system can be based on two different principles:

- **Comparison within PrivateAI database** - this refers to an anti-plagiarism system that operates by comparing a given text with documents stored within our own database.
- **Comparison with external open-sourced articles** - this refers to an anti-plagiarism system that compares a given text with documents available on the open internet or particular open-source platforms. This could include academic papers, articles, blog posts, and other publicly available documents.

These two approaches have their pros and cons:

| Approach | Pros | Cons |
|---|---|---|
| **PrivateAI database** | - **Customization.** The system can be configured to meet the specific needs of the platform or community, e.g. focusing on certain types of documents or formats.<br>- **Privacy.** Since it only checks within the platform, it maintains the privacy of users' data as it doesn't need to be shared with external services.<br>- **Independence.** The system is independent of external services.<br>- **Speed.** It can be faster as it only needs to search through a limited database. | - **Limited scope.** The system only checks against files within the platform, so it won't detect plagiarism from sources outside of the system. |
| **External database** | - **Broad scope.** The system checks against files available on the open internet or other open-source platforms. | - **Integration requirement.** It is much simpler to use an existing anti-plagiarism system than develop our own.<br>- **Cost.** External anti-plagiarism services can be expensive, especially for high-volume usage or for services that provide detailed reports.<br>- **Self-plagiarism.** The author can publish articles on other sources and upload them to PrivateAI as well. In this case, it is necessary to verify the user's identity, but such a requirement lowers the author's privacy. |

As seen, using an internal checker can provide quick and private results and be applied for the first stage of plagiarism detection, as it only searches through a limited database within the platform. However, to ensure a broader and more accurate check, later the system can be completed with an external checker that compares the text with documents available on the open internet.

## II. Consistency

Along with the copyright verification, it is necessary to check whether the article contains appropriate information or not.

Although the exact list of consistency criteria is not formed yet, the main principles should be as follows:

- all data should be informative, the article should not contain flood-type information or be auto generated;

- all data should contain only normal vocabulary;
- all data should meet the scientific article requirements.

## III. Knowledge graphs

As the last step of article processing, we will extract its knowledge graph to provide a quicker article overview for data searchers. The more detailed description of knowledge graph definition, importance and extraction can be found here - [PrivateAI] Whitepaper v2.0.

As a result of this processing, a user can make sure that the article content is unique and does not include any inappropriate information. The knowledge graph extraction also ensures that the article's main ideas and concepts are clearly represented, making it easier for users to understand the content.