



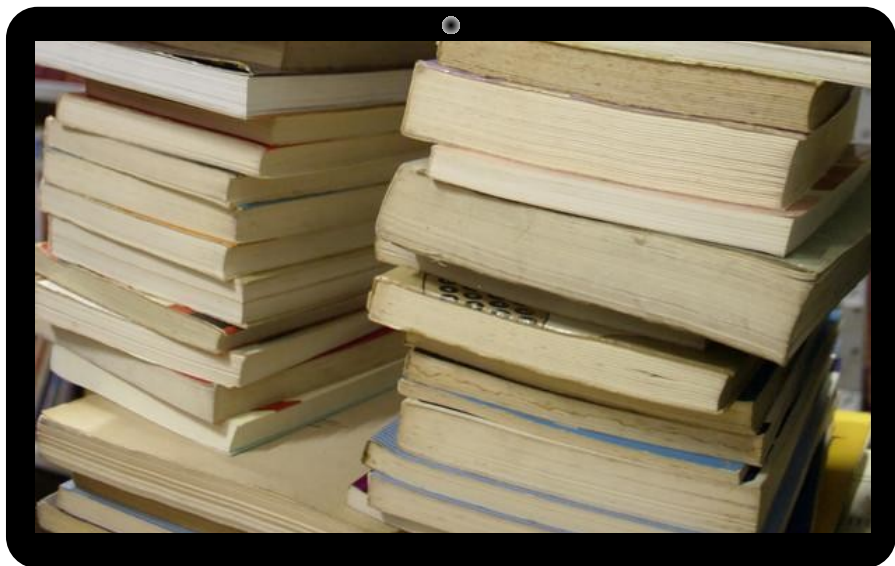
清华大学  
TSINGHUA  
UNIVERSITY

# Files to Knowledge Database

TSINGHUA UNIVERSITY

Jiang Zhicheng Wang Zehua Zhao Hanhong

# 01 Why “knowledge database” ?



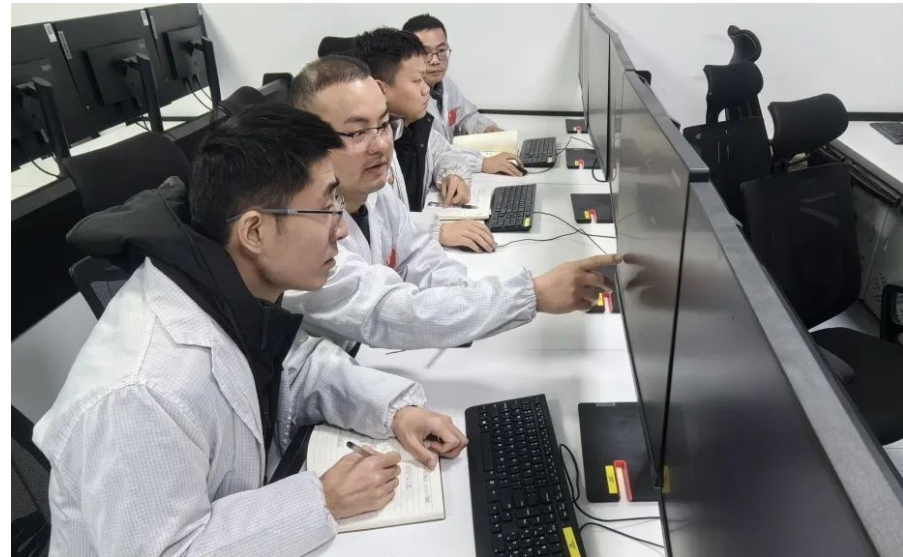
Do you have the experience of:

- remembering something you must have read it somewhere within several days ...
- The material is on your computer, but you **just cannot find it?**

!

# 01 Our Goal

- Reduce time wasting in searching materials
- Construct a database for their knowledge
- Discover new connections between topics

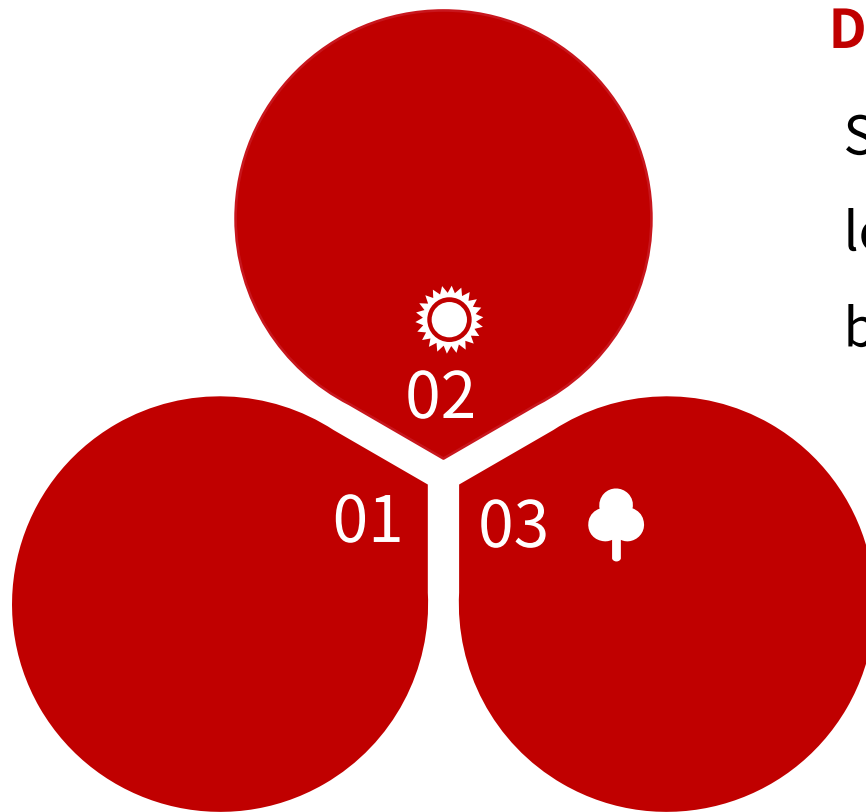


# 01 What Our Model Can Do

...

## Learning & Managing

- Learn and understand the content within your computer.
- Summarize it.
- Extract key features to assign tags to it.
- Build a “database” with hierarchy.



## Data Searching

Search for the data and the location in your computer based on your description.

## Question Answering

Answer questions based on the data in the computer.

# 01 Why non-trivial ?



## Large Amount

There' re a lot of contents in a computer: up to several GBs!



## Format Difference

They are in different formats: videos, books, audios, ...



## Trade-off

We need a model that have both high speed and accuracy!

# 01 Our Approaches

## Creating “Tags”

Using LLM to interpret and give several tags for each file.

## Querying With Templates

Designing prompt templates:

- Search through the “database” to get information
- Summarize (or answer Questions).



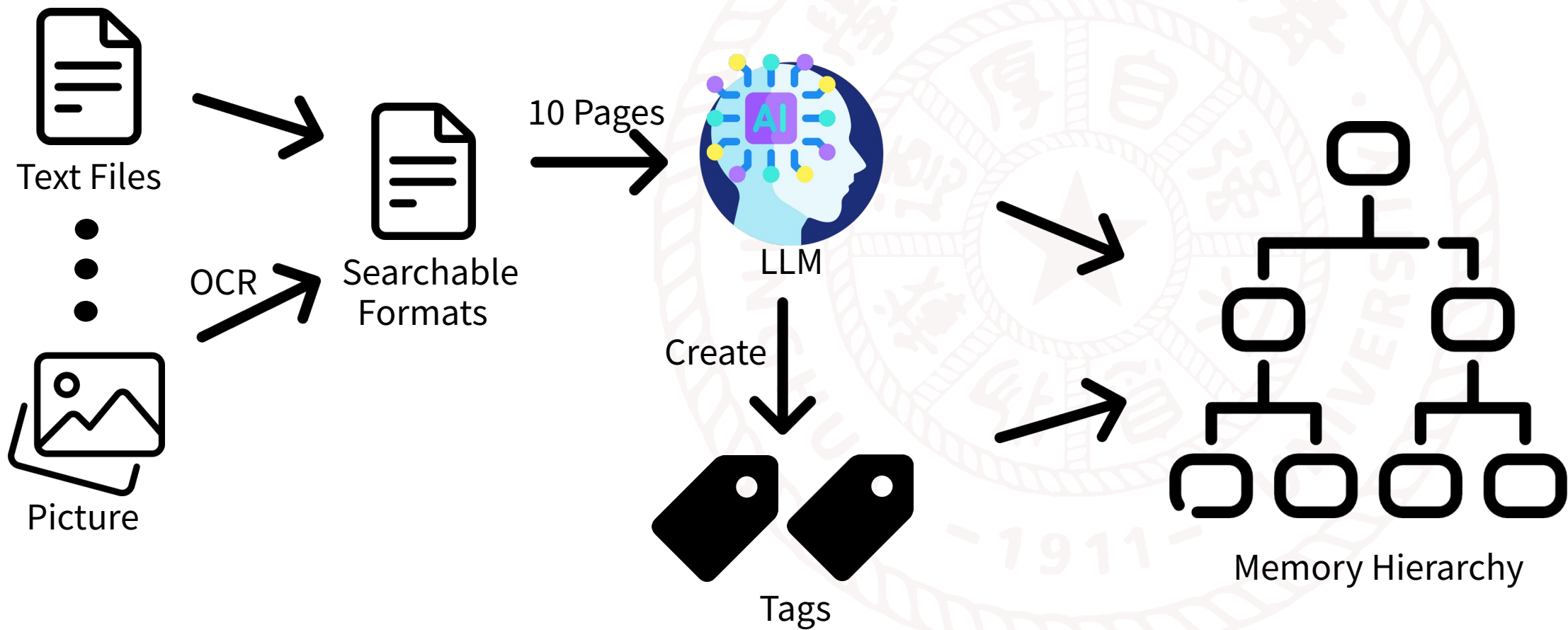
## Building a managing hierarchy

- Manage the files with tag in a multi-level hierarchy.
- Build a “database” of the files.

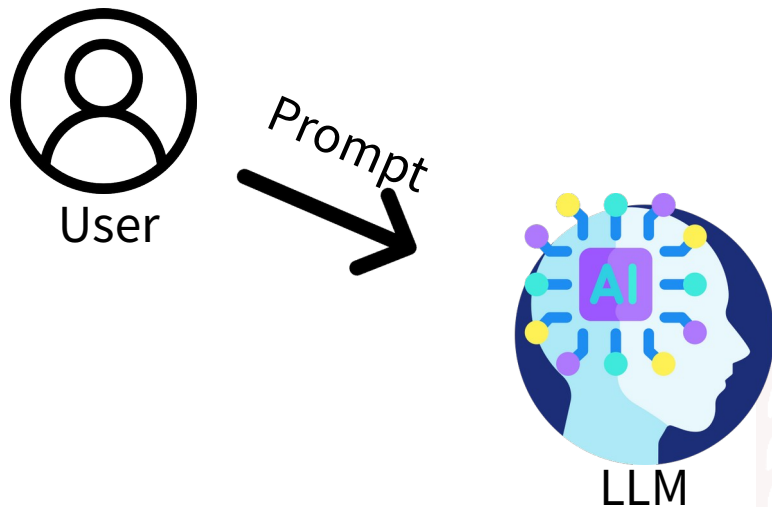
## Real-time Summarization

- Avoid reading all documents at first time.
- Only searching for relevant details each time.

# 01 Flow Chart: Pre-Processing

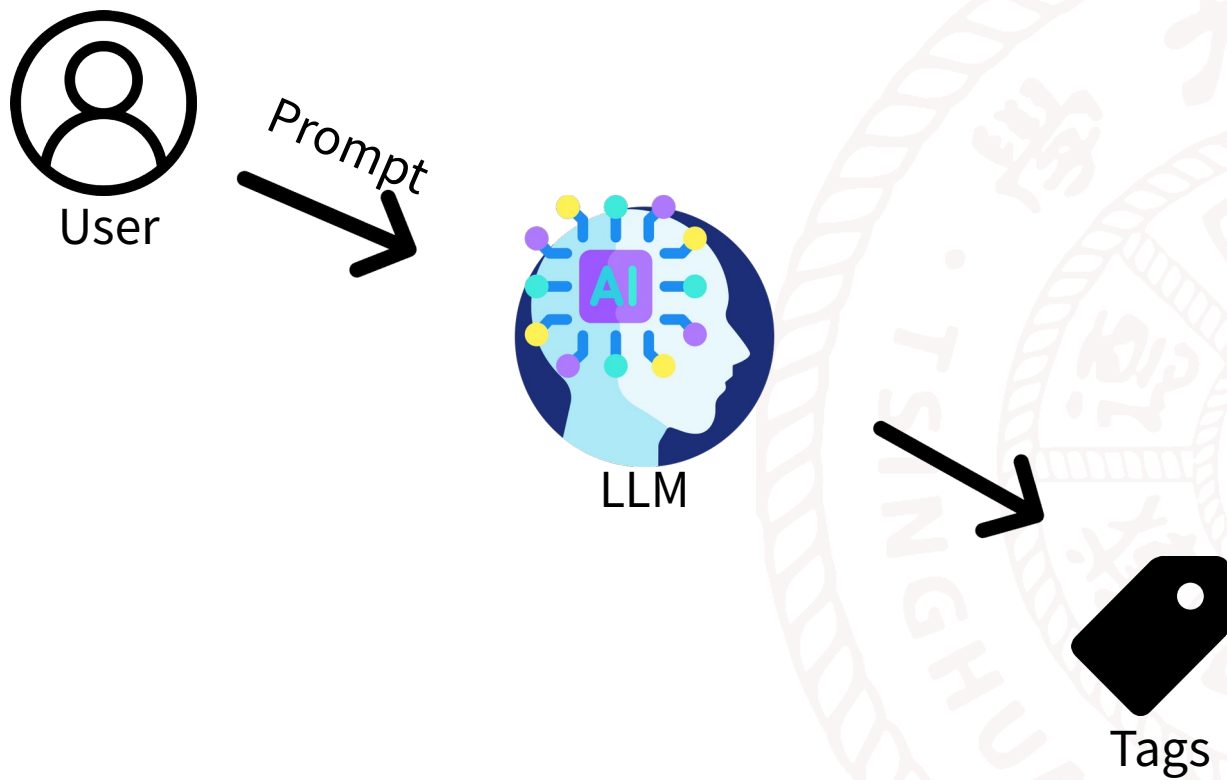


# 01 Flow Chart: Querying

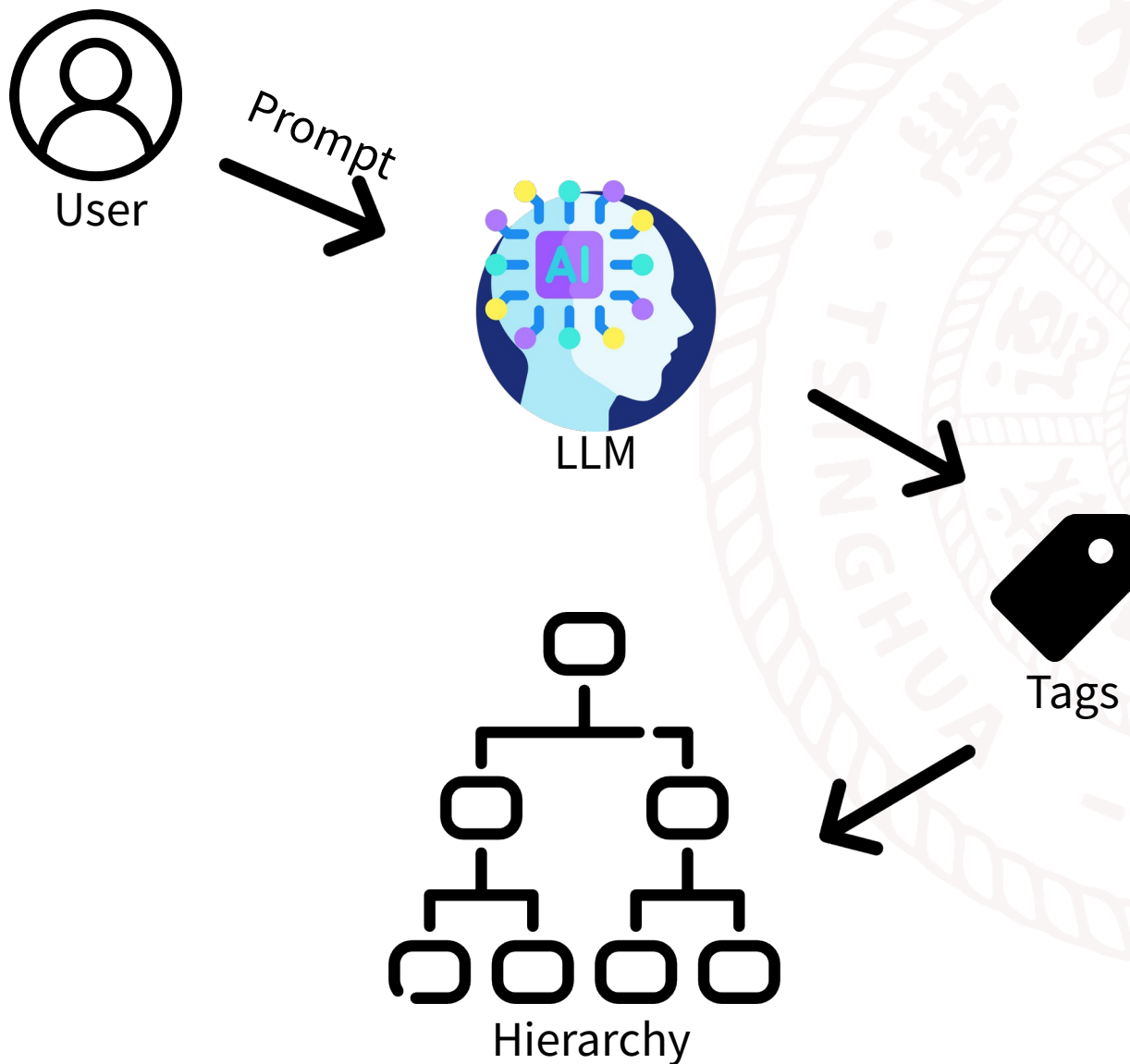




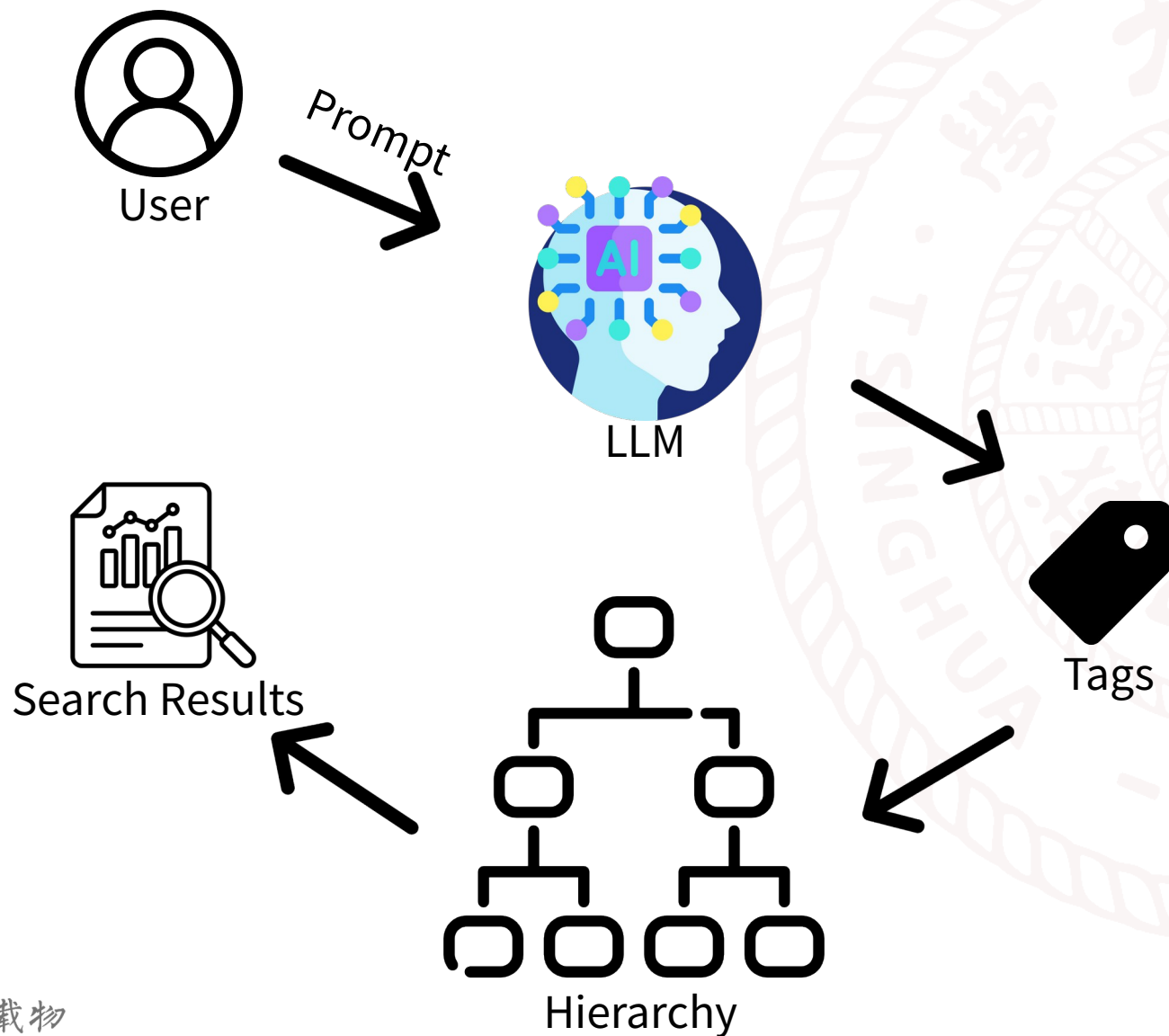
# 01 Flow Chart: Querying



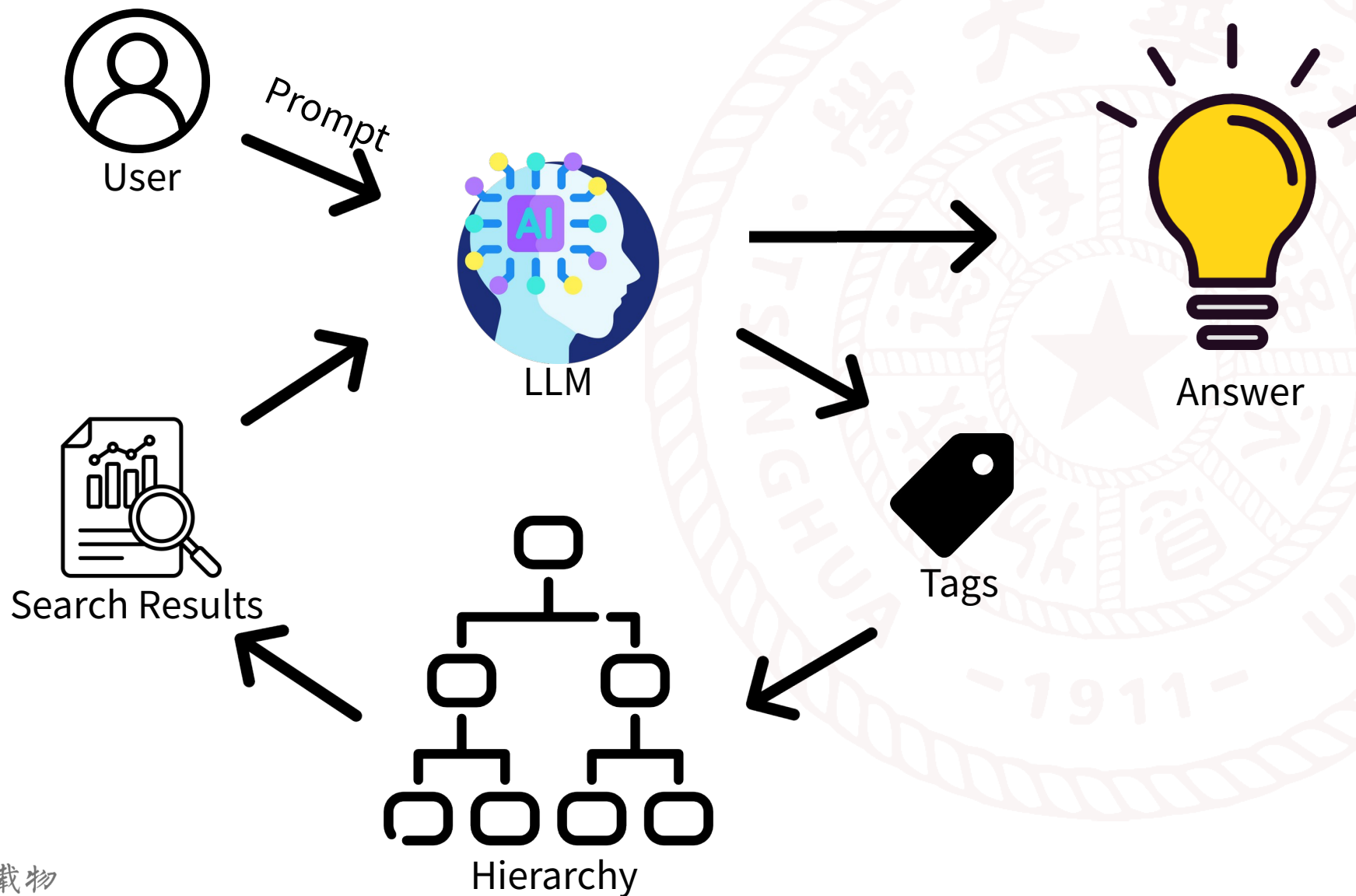
# 01 Flow Chart: Querying



# 01 Flow Chart: Querying



# 01 Flow Chart: Querying



# 01 Resource Needed

## Token

- Model: Any
- Amount: 10M

## Disk Space

- 1T storage (for user data files)







清华大学

TSINGHUA  
UNIVERSITY

# Q & A



自强不息 厚德载物





清華大學

TSINGHUA  
UNIVERSITY

Thanks!