

Amazon Comprehend

Table of Contents



1. What is Natural Language Processing?

2. What is Amazon Comprehend?

3. Amazon Comprehend insights

4. Amazon Comprehend Custom

5. How it works

6. Topic Modeling

What is Natural Language Processing?

Natural Language Processing (NLP) is a way for computers to analyze, understand, and derive meaning from textual information in a smart and useful way.

Key aspects of NLP include:

1. **Text Analysis:** Ability to analyze textual information 🔍
2. **Comprehension:** Understanding the content and context of text 💡
3. **Meaning Extraction:** Deriving meaningful insights from text 🔑

What is Natural Language Processing?

Natural Language Processing (NLP) is a way for computers to analyze, understand, and derive meaning from textual information in a smart and useful way.

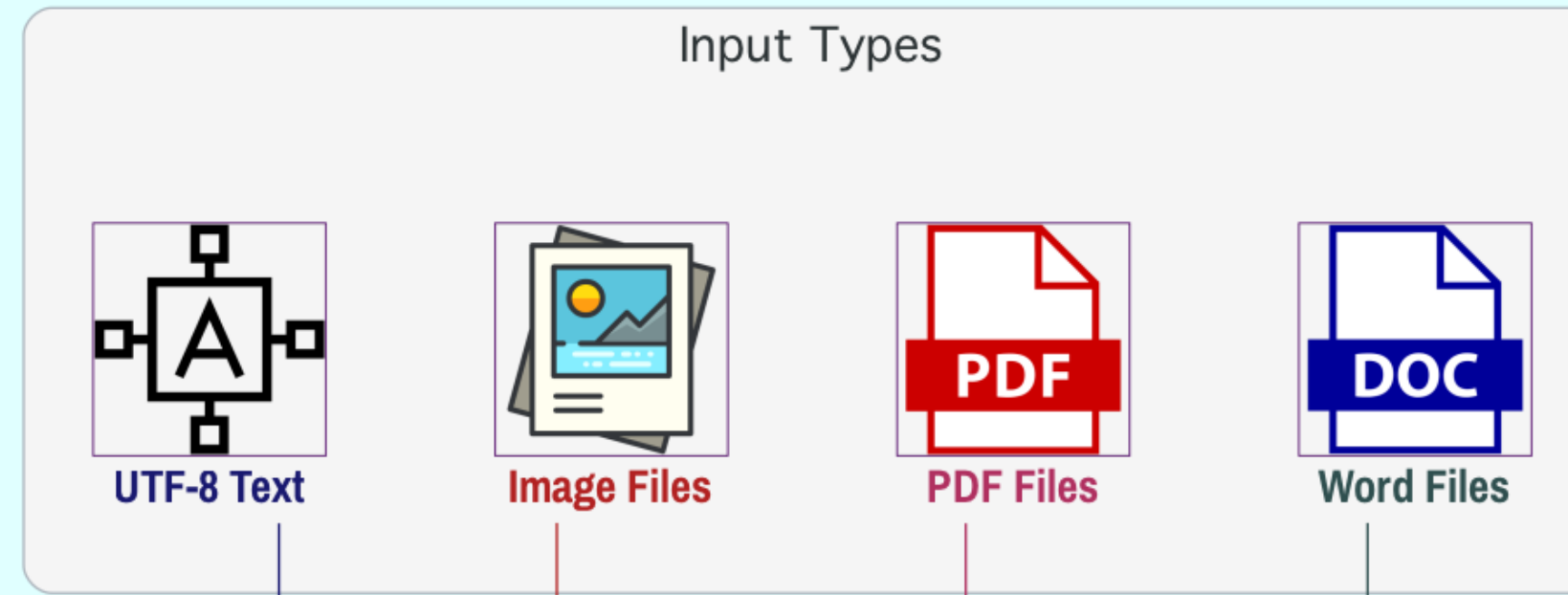
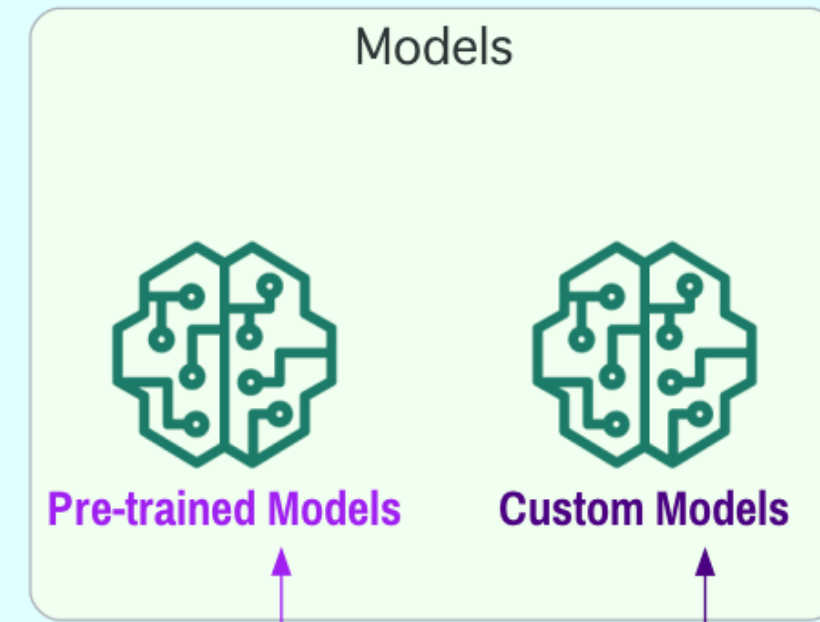
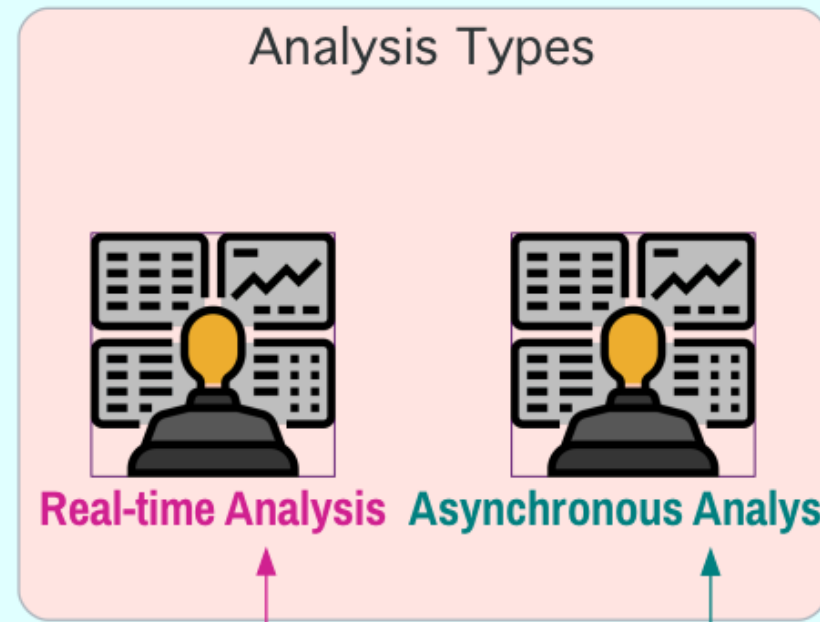
Key aspects of NLP include:

1. **Text Analysis:** Ability to analyze textual information 🔍
2. **Comprehension:** Understanding the content and context of text 💡
3. **Meaning Extraction:** Deriving meaningful insights from text 🔑

NLP enables the extraction of various elements from text:

1. Important phrases 🏷️
2. Sentiment 😊
3. Syntax 🛠️
4. Key entities, such as:
 - a. Brand 🛍️
 - b. Date 📅
 - c. Location 📍
 - d. Person 👤
5. Language of the text 💬

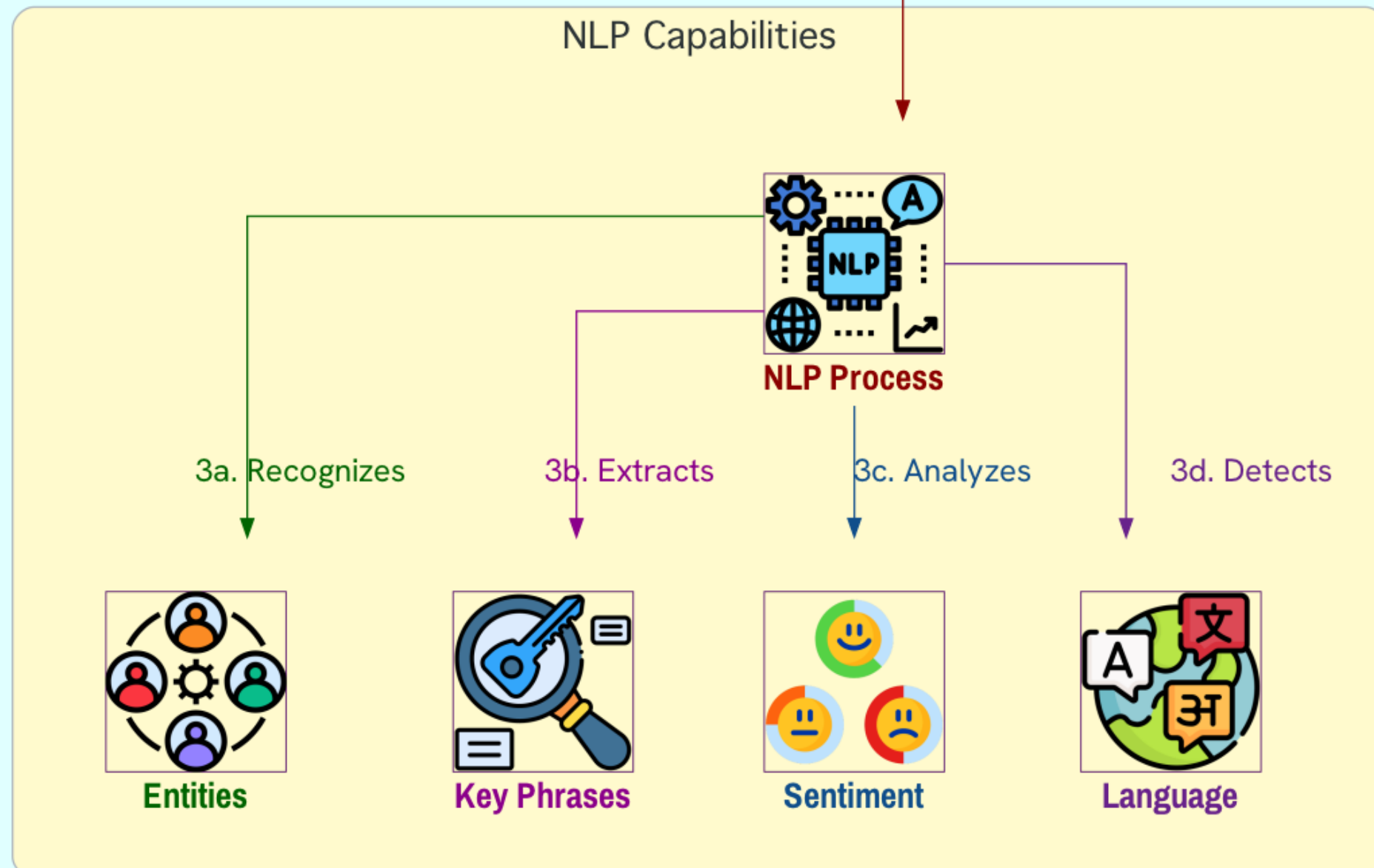
Amazon Comprehend



S3 Input



Amazon Comprehend



Document Analysis



Dominant Language



Lambda Function



Analytics



CloudWatch



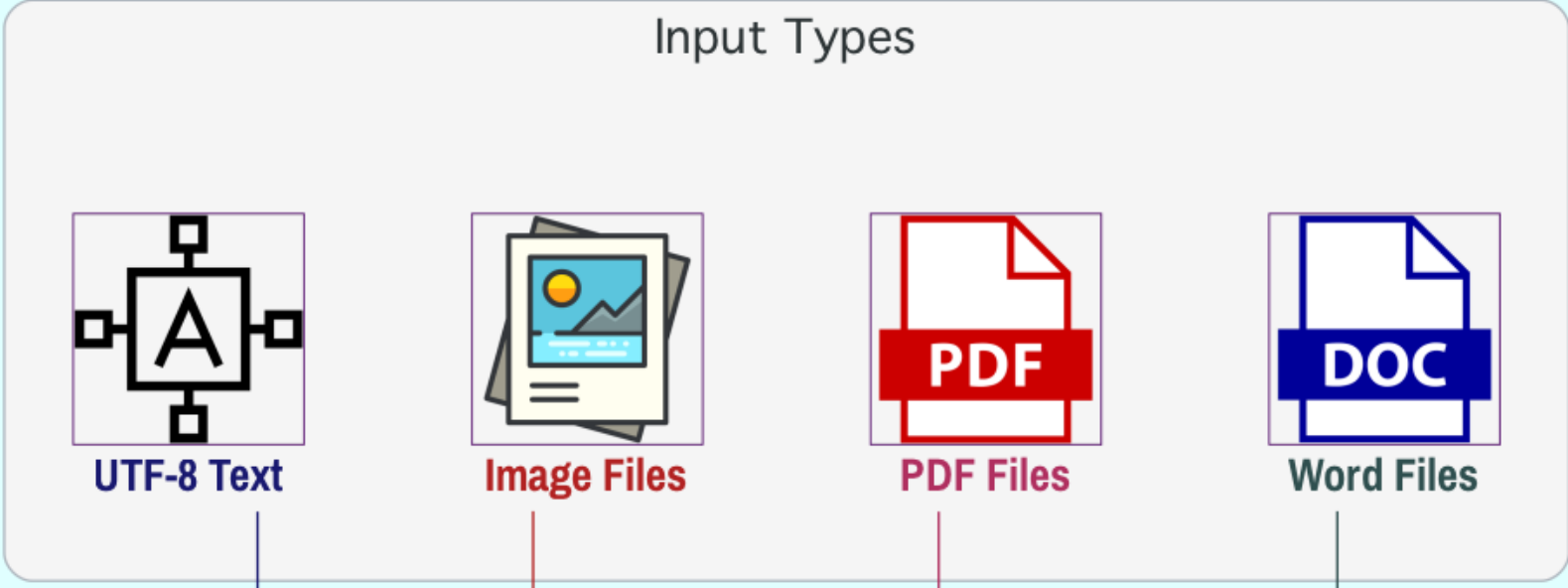
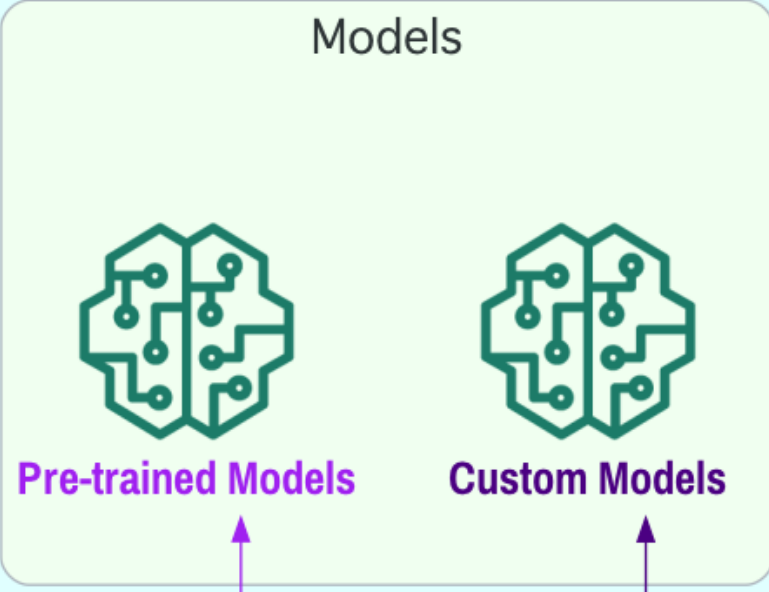
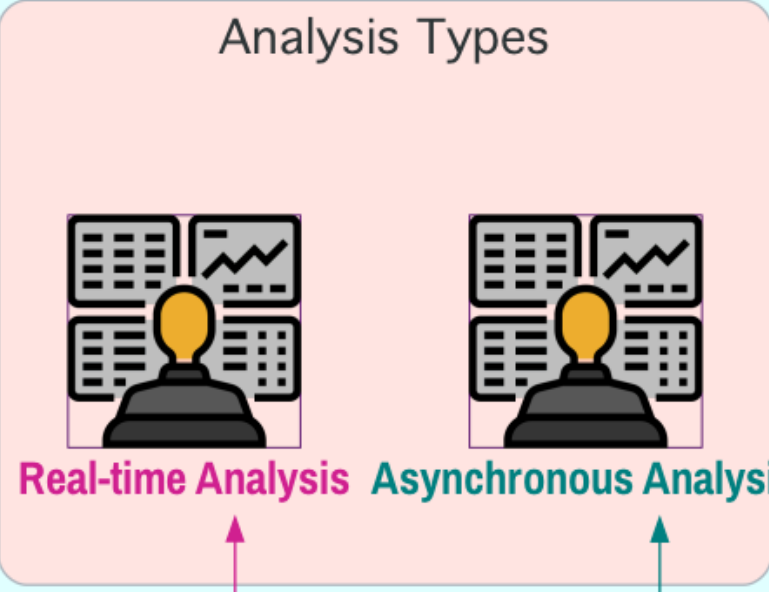
EventBridge

1. NLP-powered document analysis

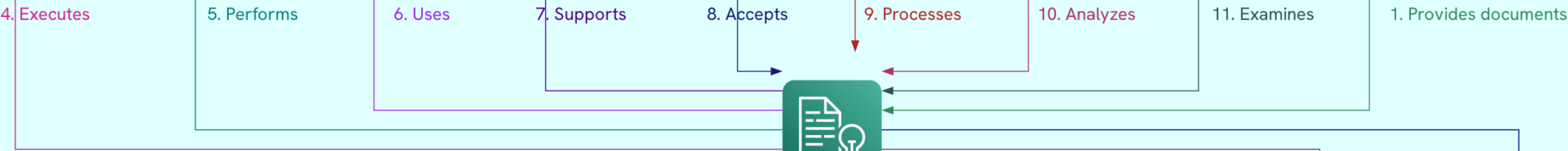
Extracts insights from documents

Deep understanding of text data

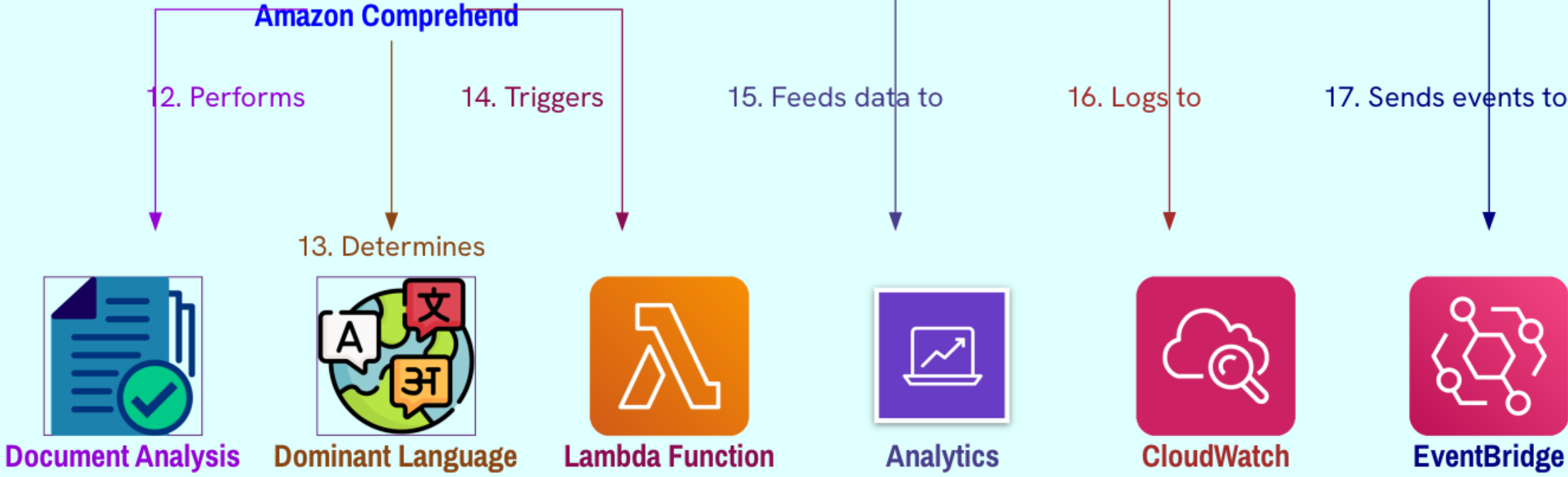
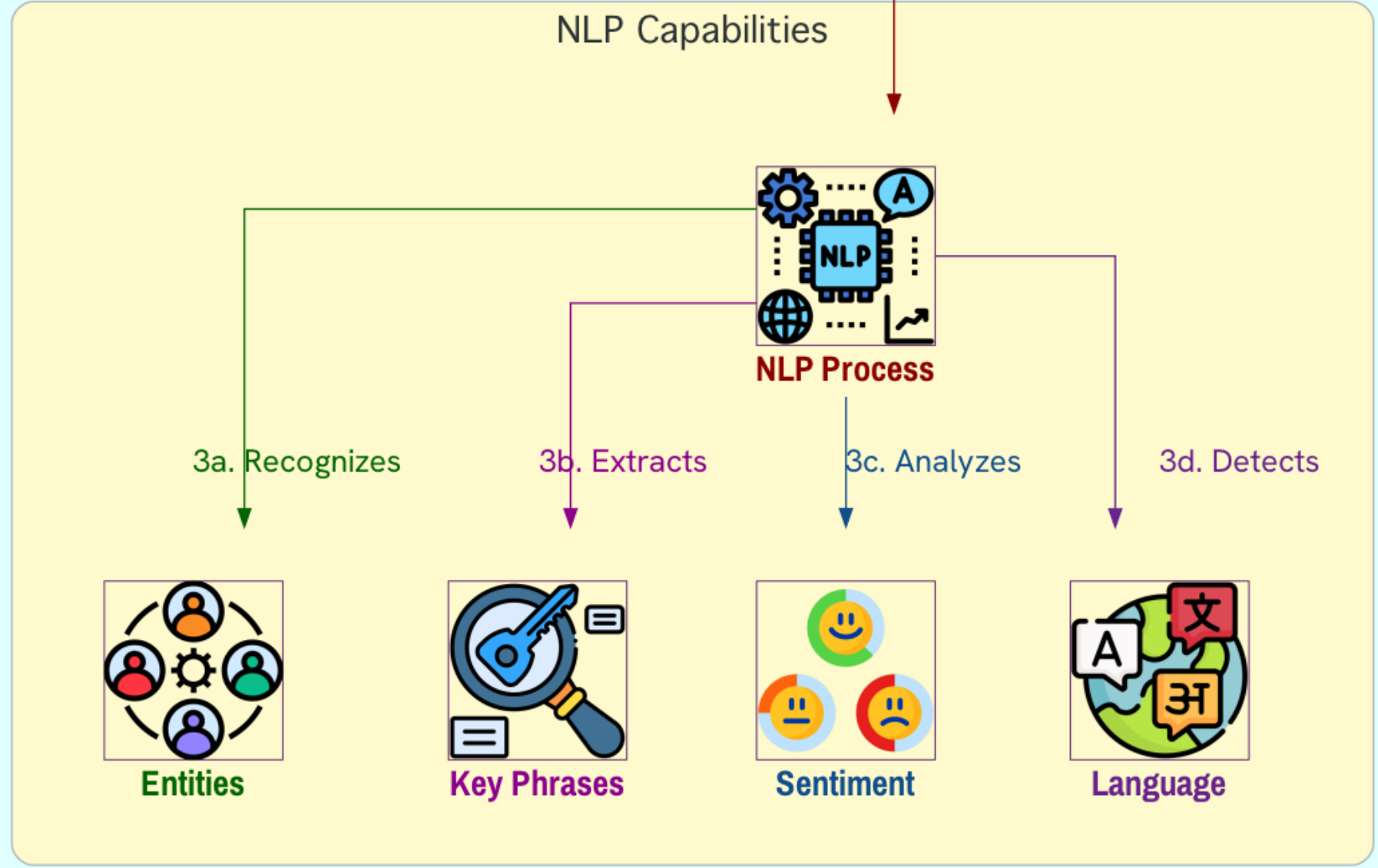
Amazon Comprehend



S3 Input

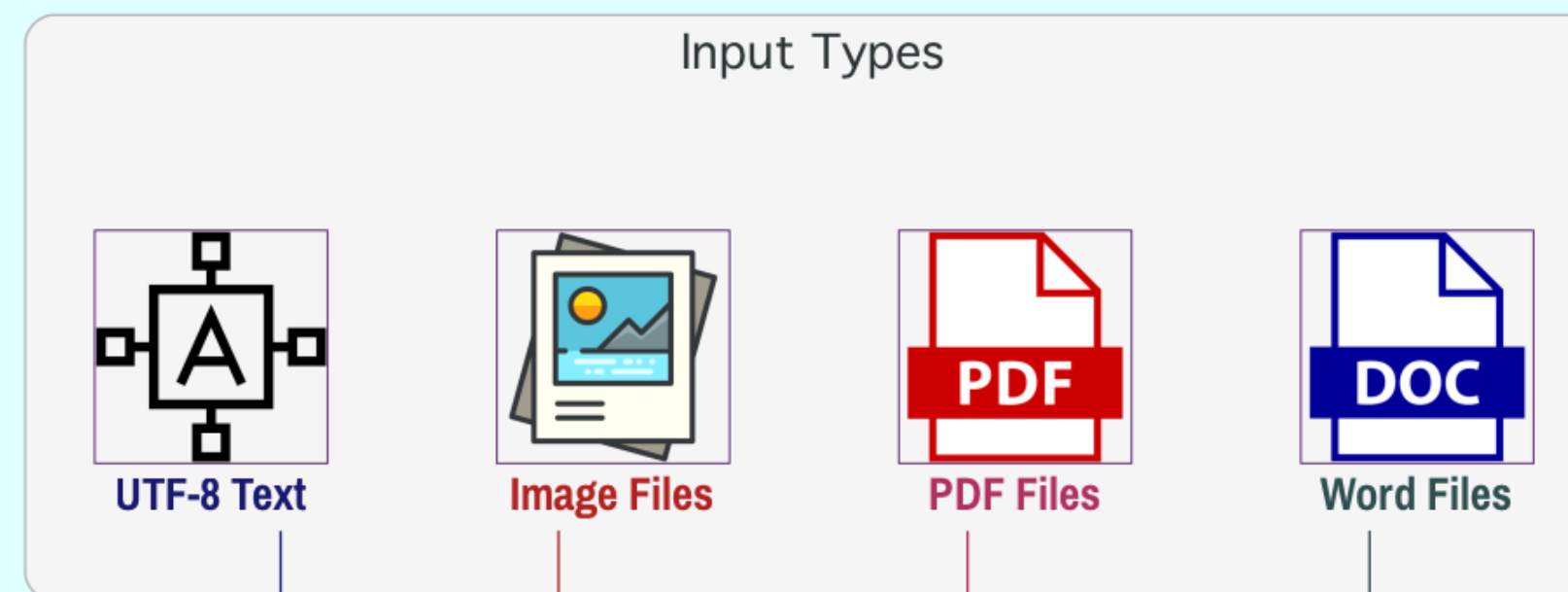
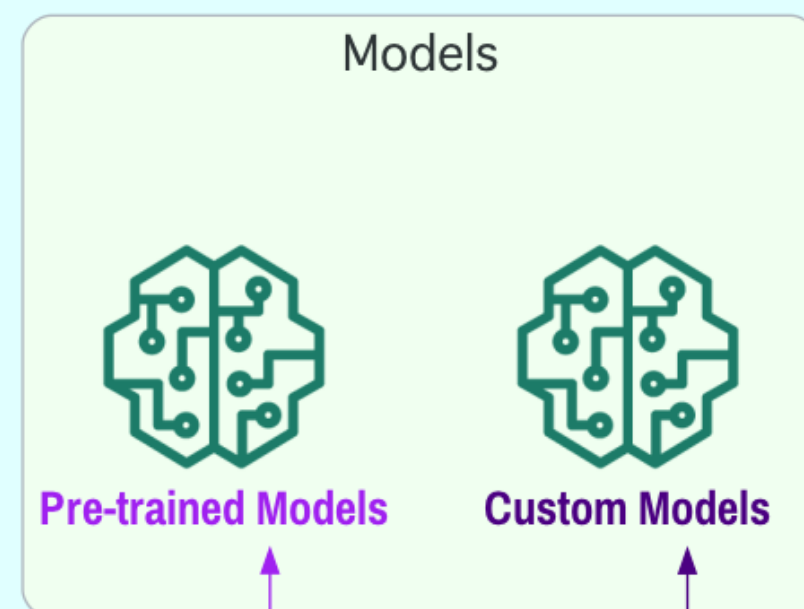
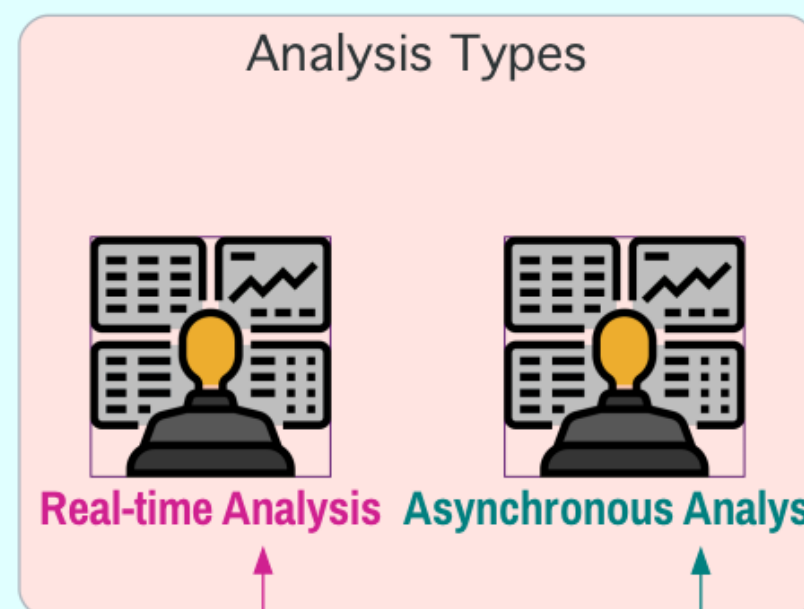


Amazon Comprehend

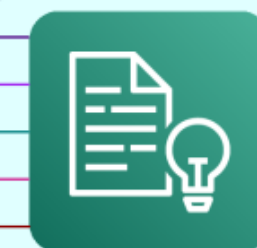


2. Recognizes key elements
Entities
Key phrases
Language
Sentiment
Other common elements

Amazon Comprehend

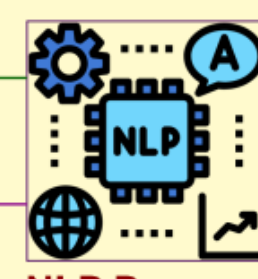


S3 Input



Amazon Comprehend

NLP Capabilities



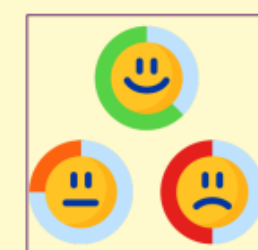
NLP Process



Entities



Key Phrases



Sentiment



Language



Document Analysis



Dominant Language



Lambda Function



Analytics



CloudWatch



EventBridge

3. Creates products from document structure

Search social networks for product mentions

Scan document repositories for key phrases

4. Executes

5. Performs

6. Uses

7. Supports

8. Accepts

9. Processes

10. Analyzes

11. Examines

1. Provides documents

2. Utilizes

12. Performs

14. Triggers

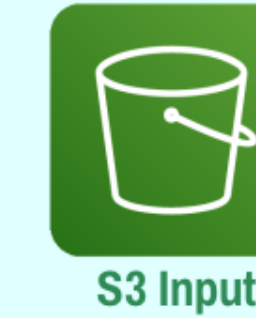
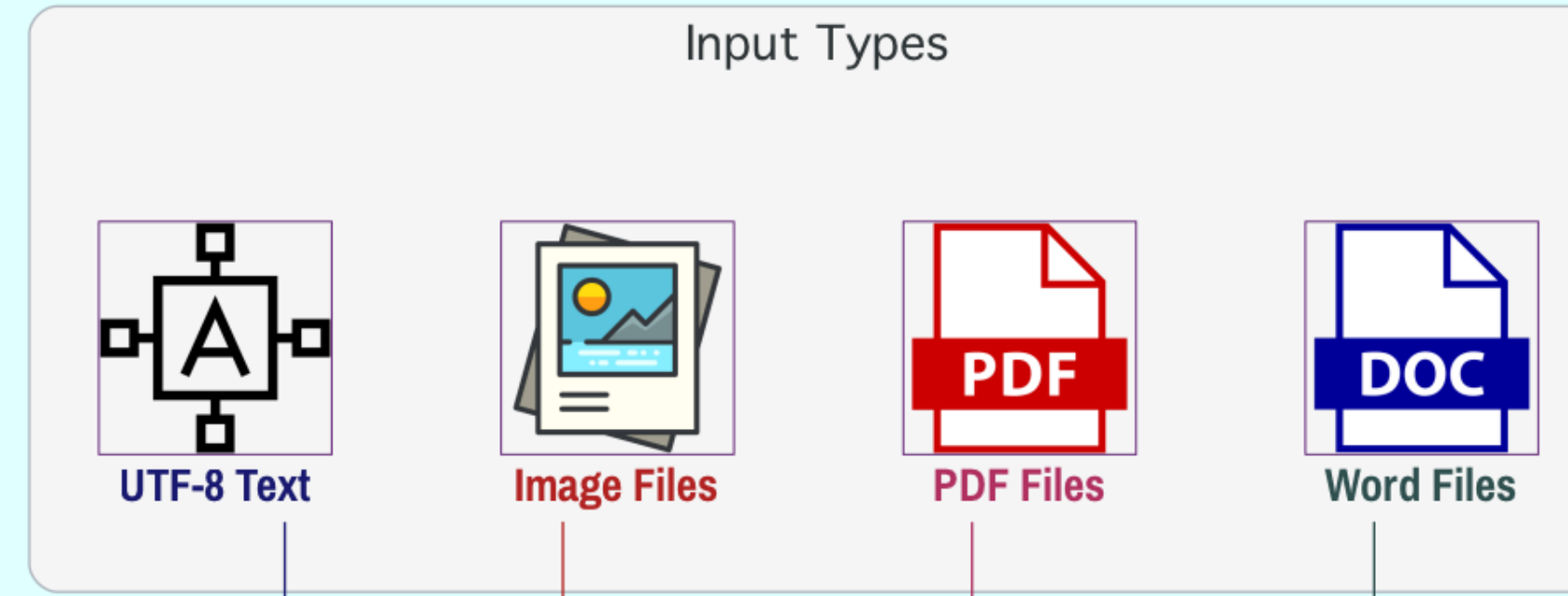
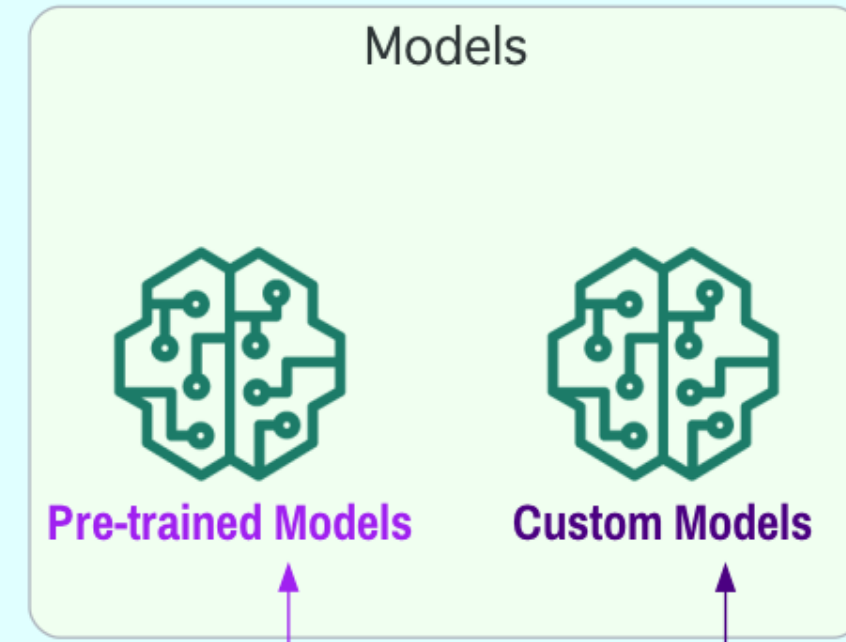
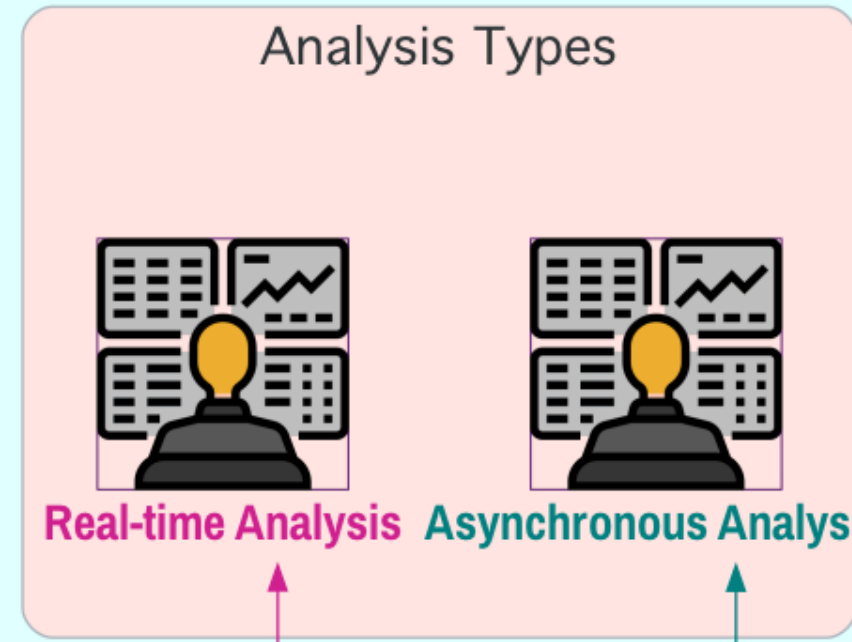
15. Feeds data to

16. Logs to

17. Sends events to

13. Determines

Amazon Comprehend



S3 Input

4. Executes

5. Performs

6. Uses

7. Supports

8. Accepts

9. Processes

10. Analyzes

11. Examines

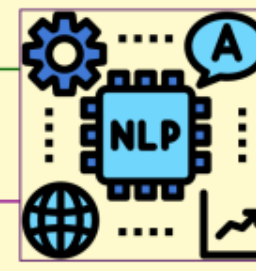
1. Provides documents



Amazon Comprehend

2. Utilizes

NLP Capabilities



NLP Process

3a. Recognizes

3b. Extracts

3c. Analyzes

3d. Detects



Entities



Key Phrases



Sentiment



Language



Document Analysis



Dominant Language



Lambda Function



Analytics



CloudWatch



EventBridge

12. Performs

14. Triggers

15. Feeds data to

16. Logs to

17. Sends events to

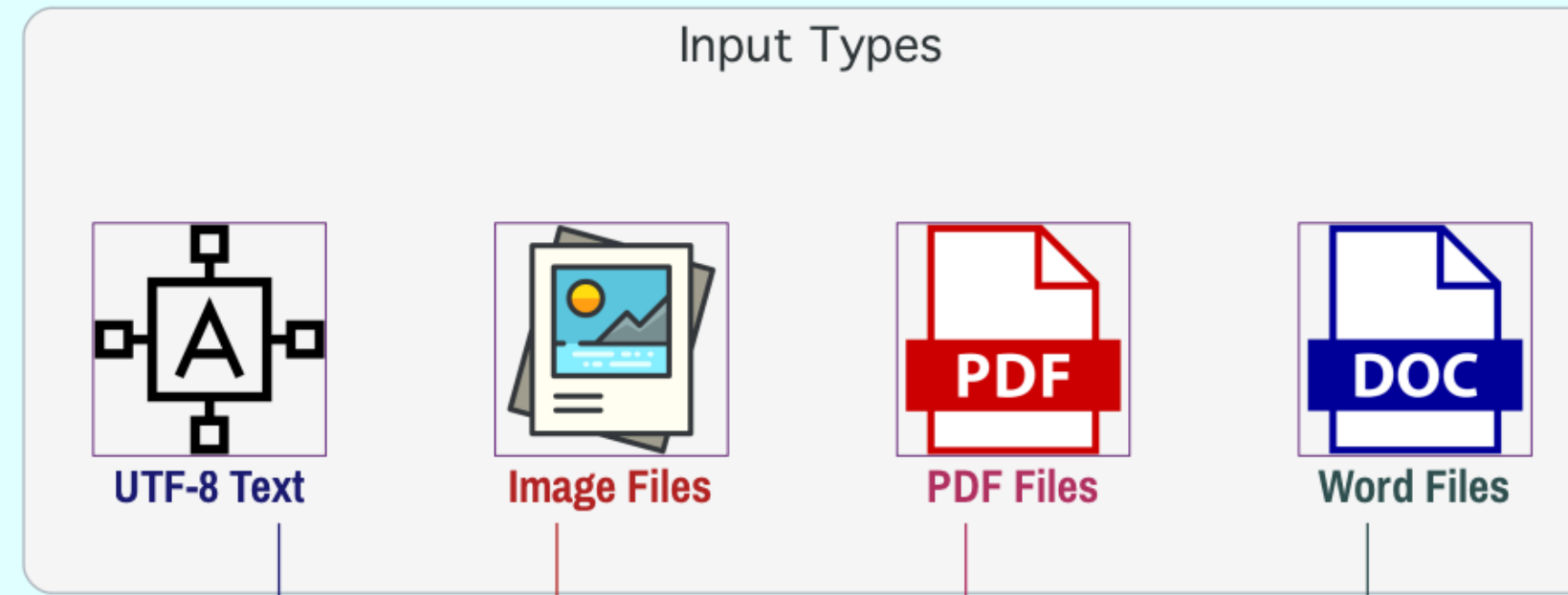
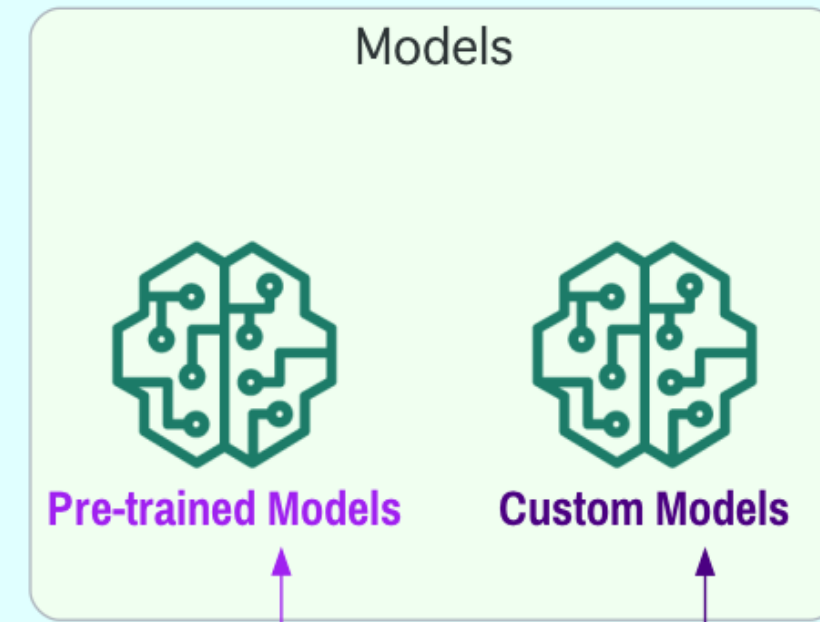
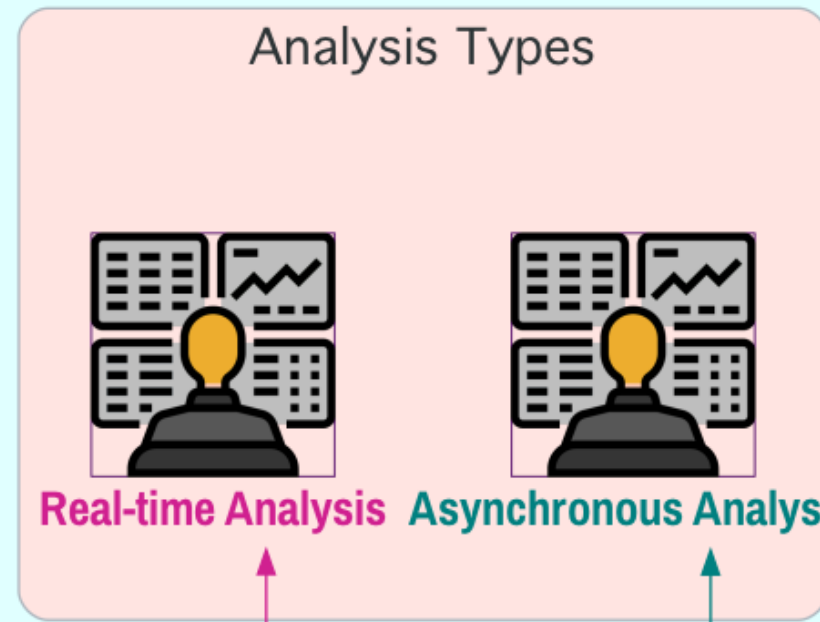
13. Determines

4. Access methods

User-friendly console

APIs for programmatic access

Amazon Comprehend

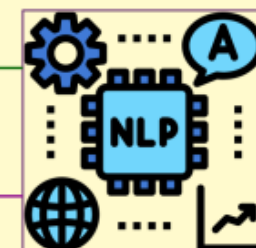


S3 Input



Amazon Comprehend

NLP Capabilities



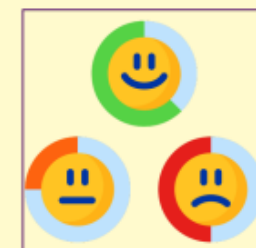
NLP Process



Entities



Key Phrases



Sentiment



Language



Document Analysis



Dominant Language



Lambda Function



Analytics



CloudWatch



EventBridge

5. ⚡ Analysis types

🕒 Real-time for small workloads

🕒 Asynchronous for large document sets

4. Executes

5. Performs

6. Uses

7. Supports

8. Accepts

9. Processes

10. Analyzes

11. Examines

1. Provides documents

2. Utilizes

12. Performs

14. Triggers

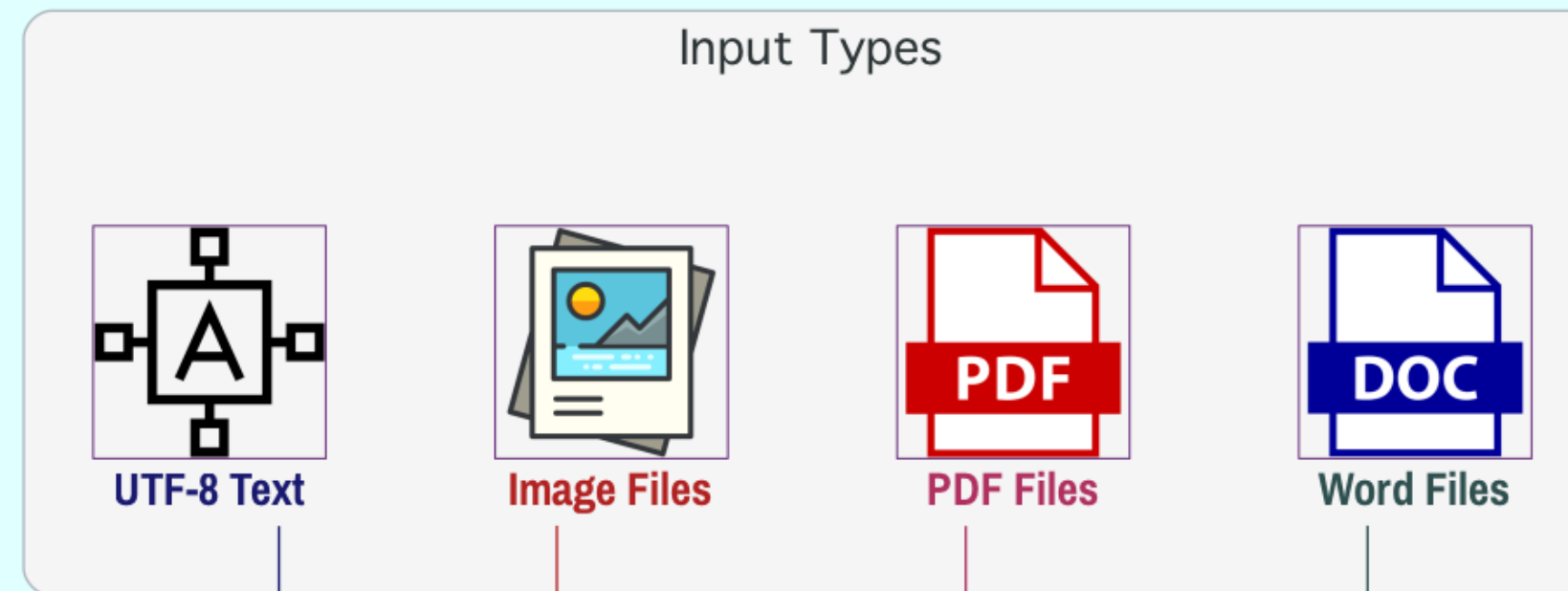
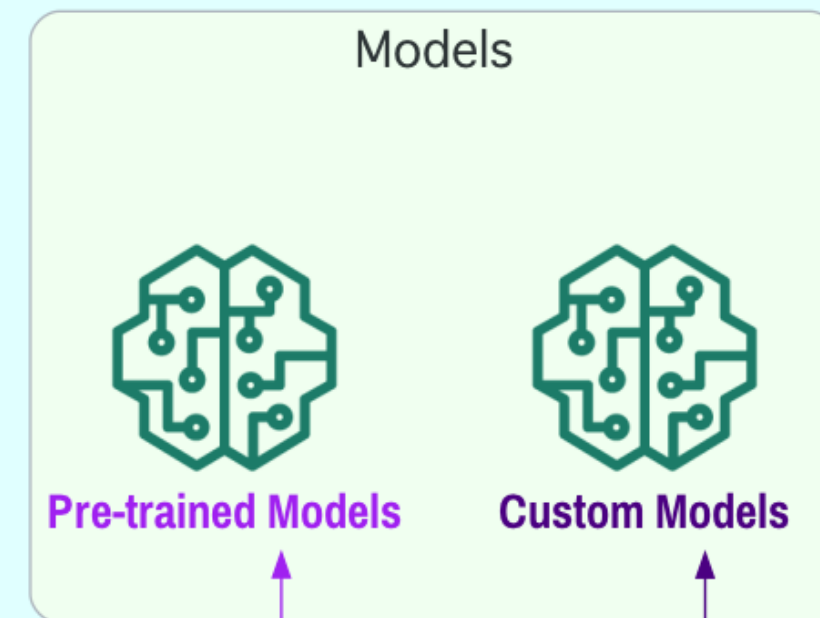
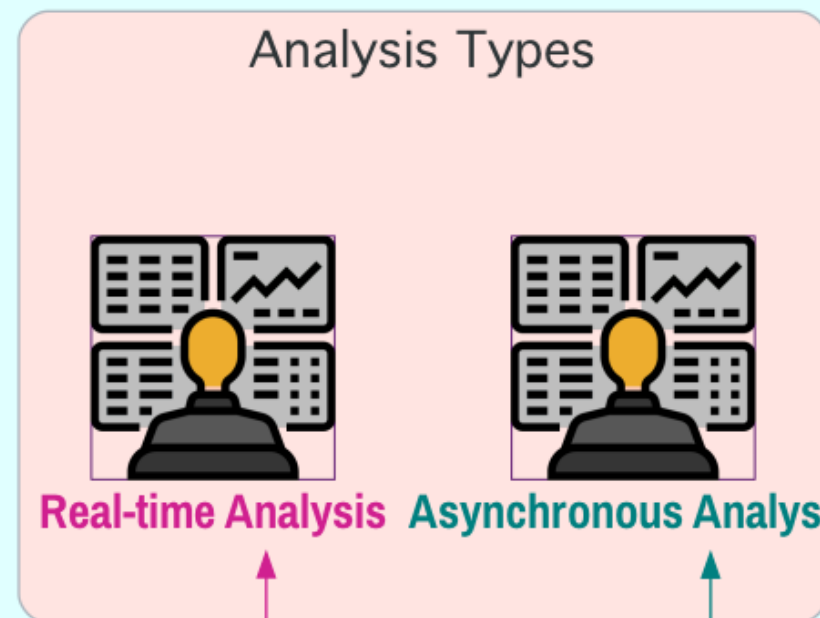
15. Feeds data to

16. Logs to

17. Sends events to

13. Determines

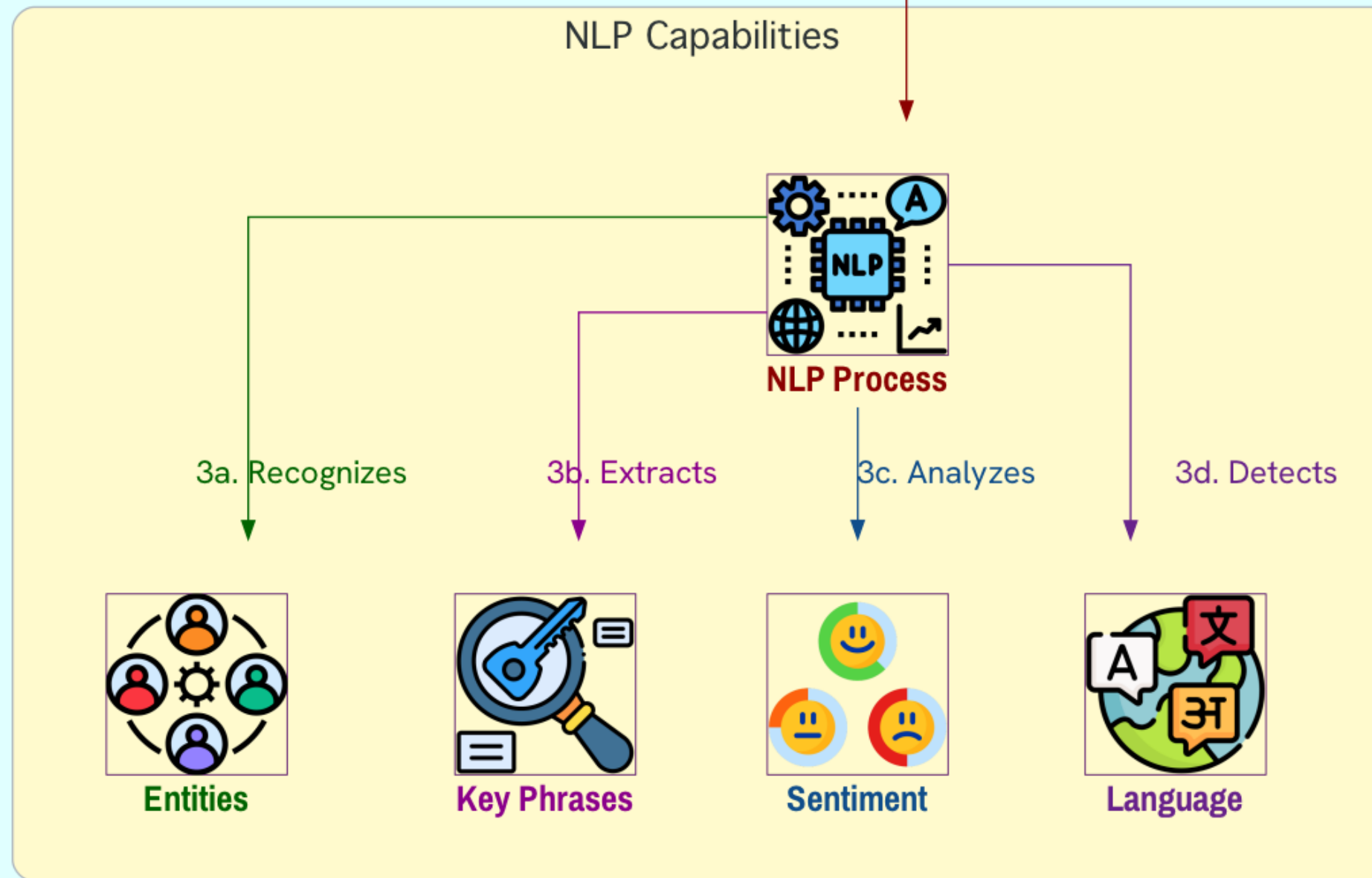
Amazon Comprehend



S3 Input



Amazon Comprehend



Document Analysis



Dominant Language



Lambda Function



Analytics



CloudWatch



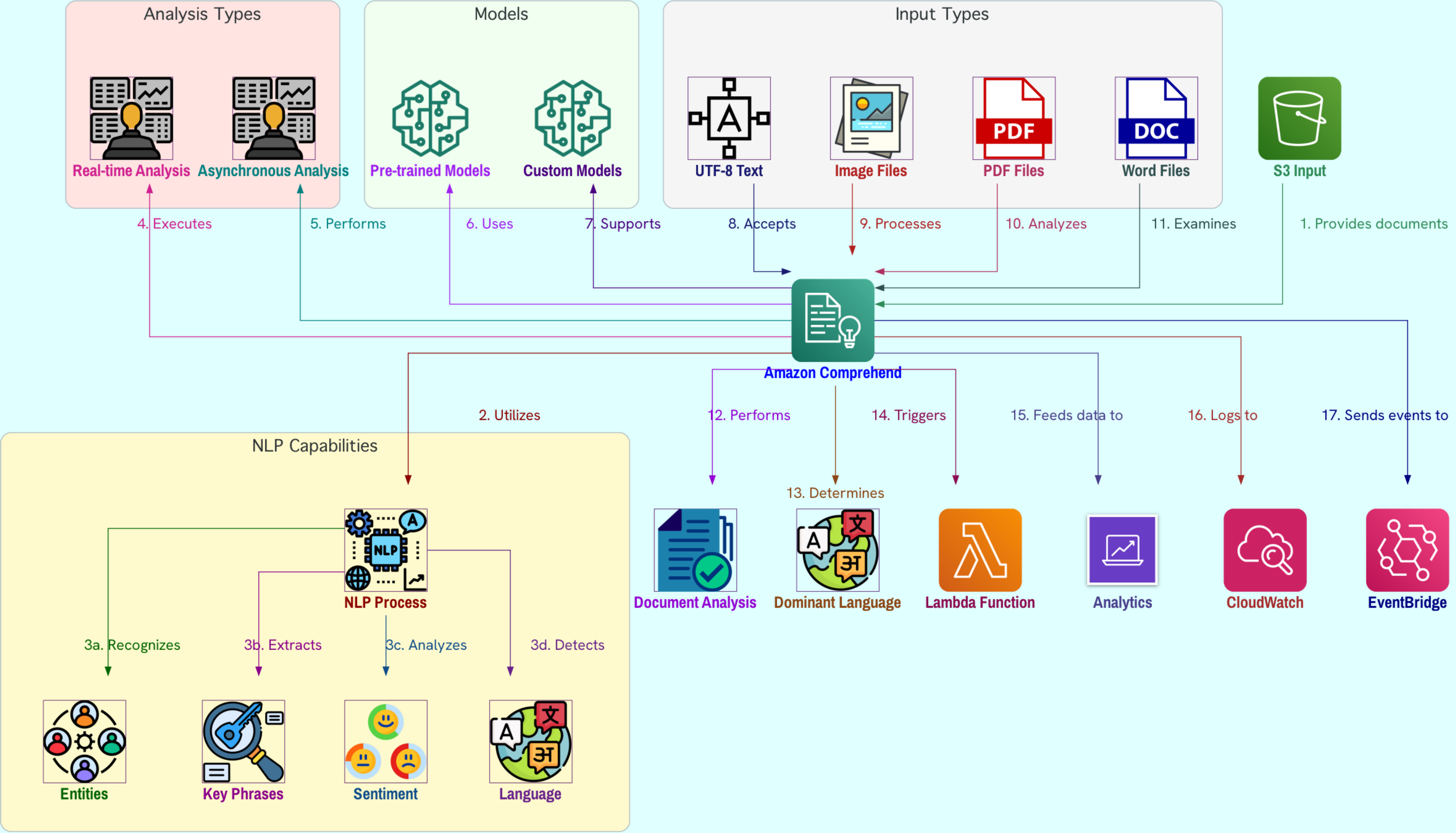
EventBridge

6. 🤖 Model options

🏭 Pre-trained models

🔧 Custom models	🏷️ Classification
	📄 Entity recognition

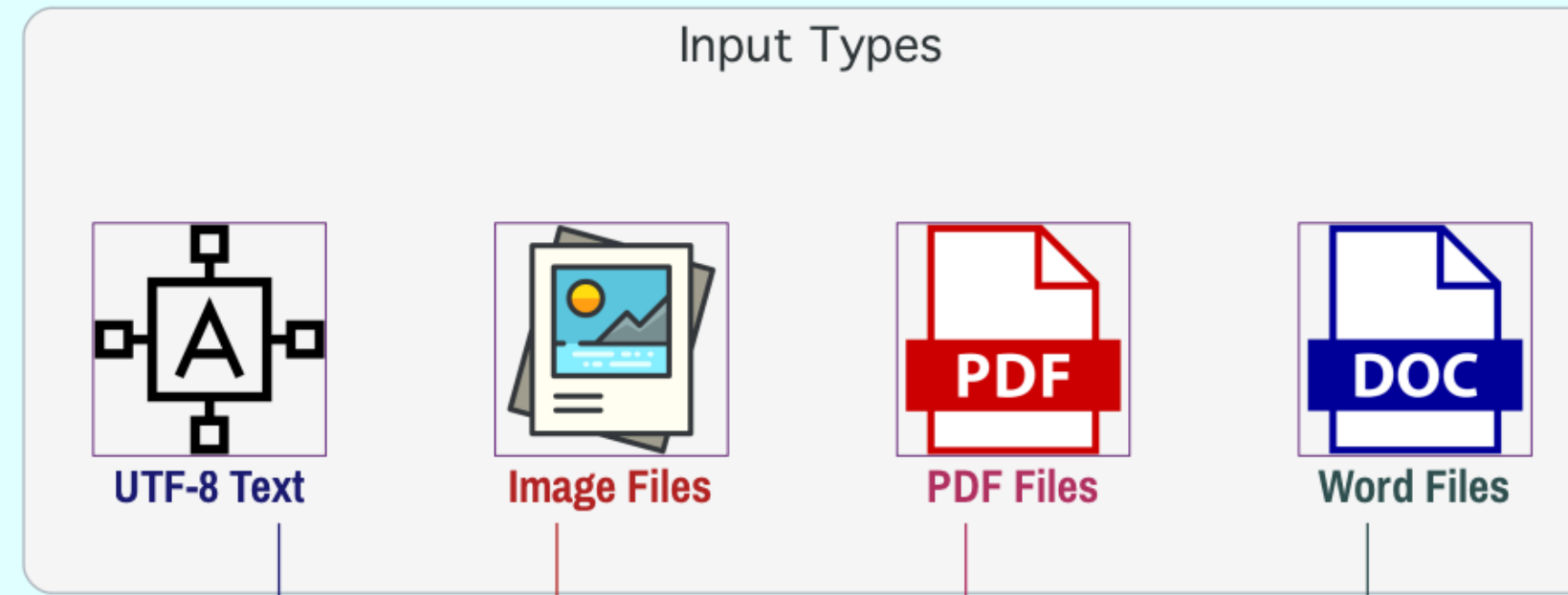
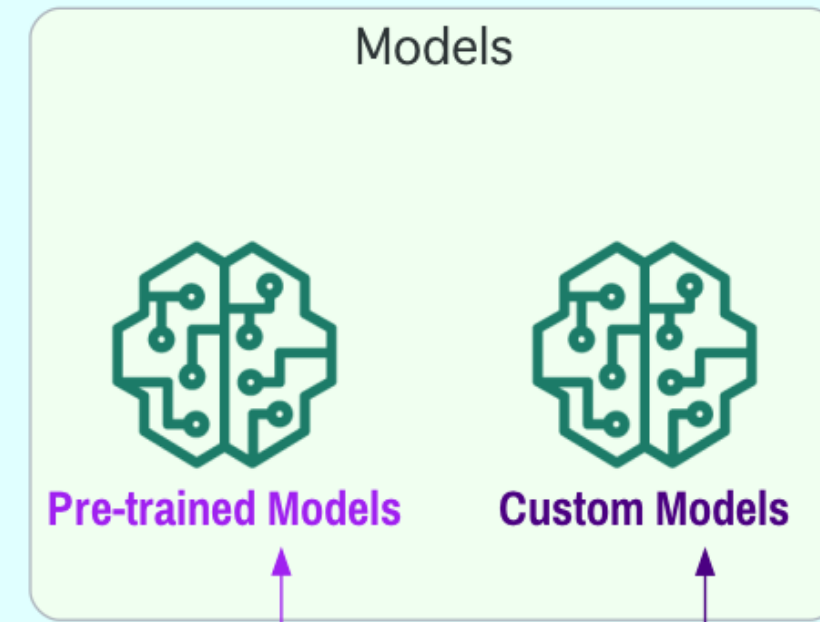
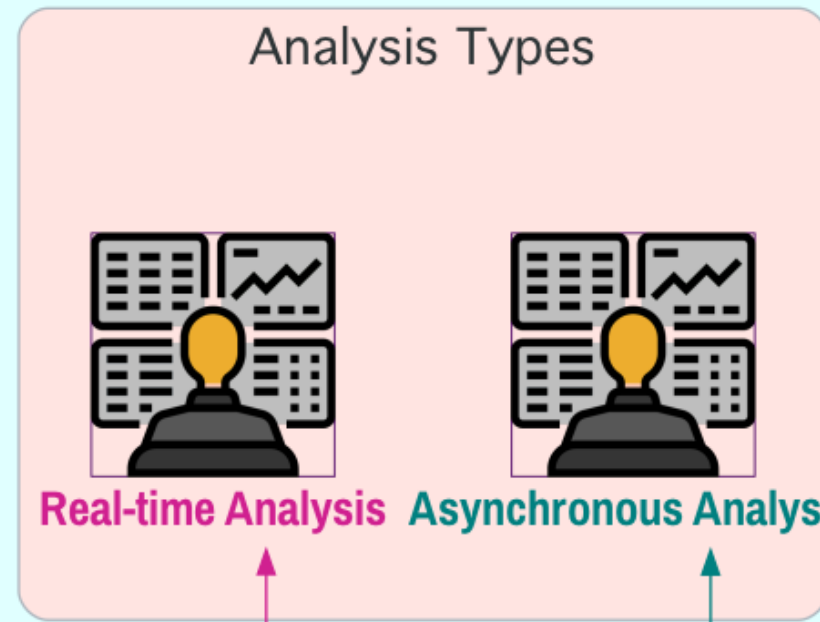
Amazon Comprehend



7. Supported input formats

- UTF-8 text documents
- Image files
- PDF files
- Word files

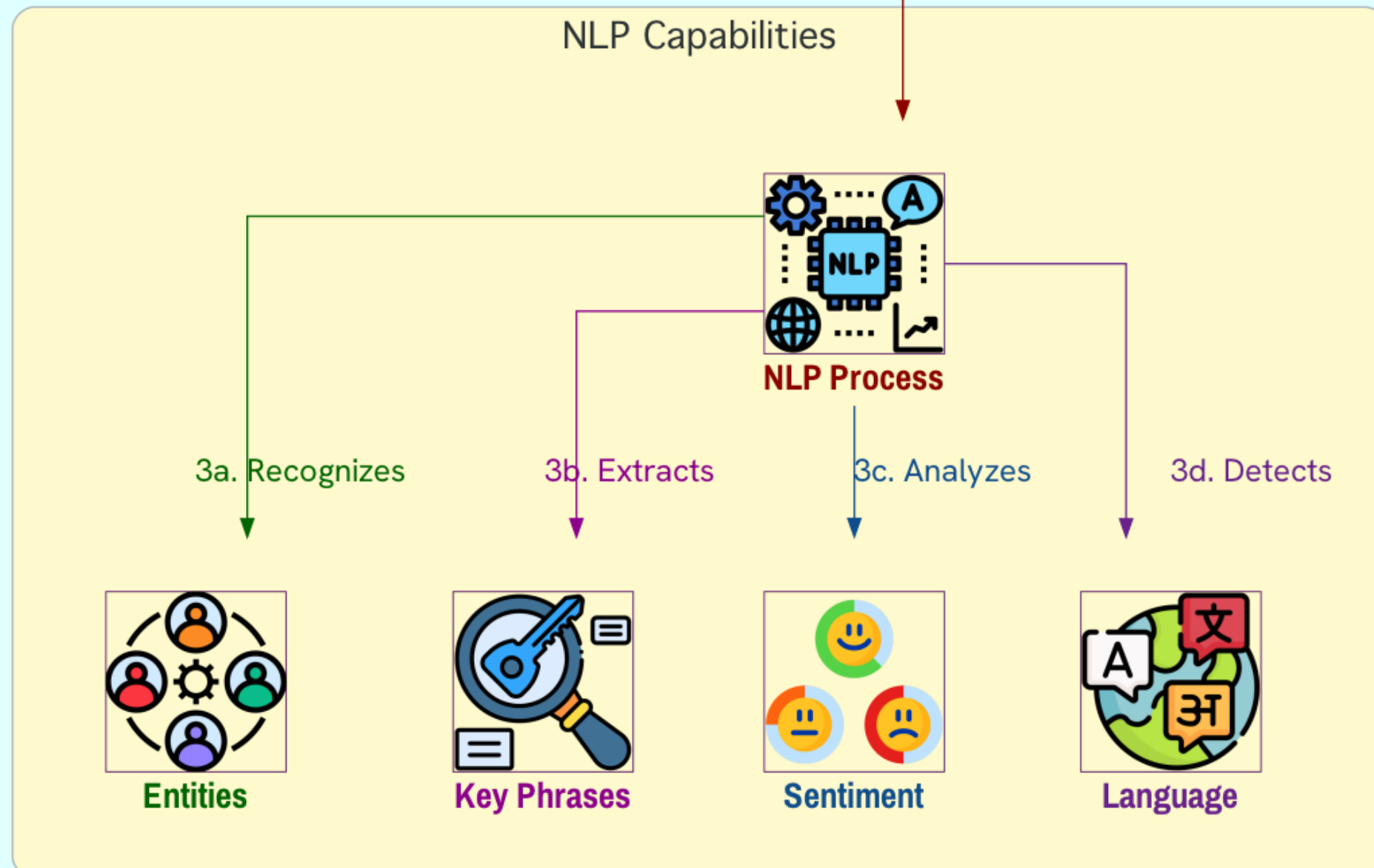
Amazon Comprehend



S3 Input



Amazon Comprehend



Document Analysis



Dominant Language



Lambda Function



Analytics



CloudWatch



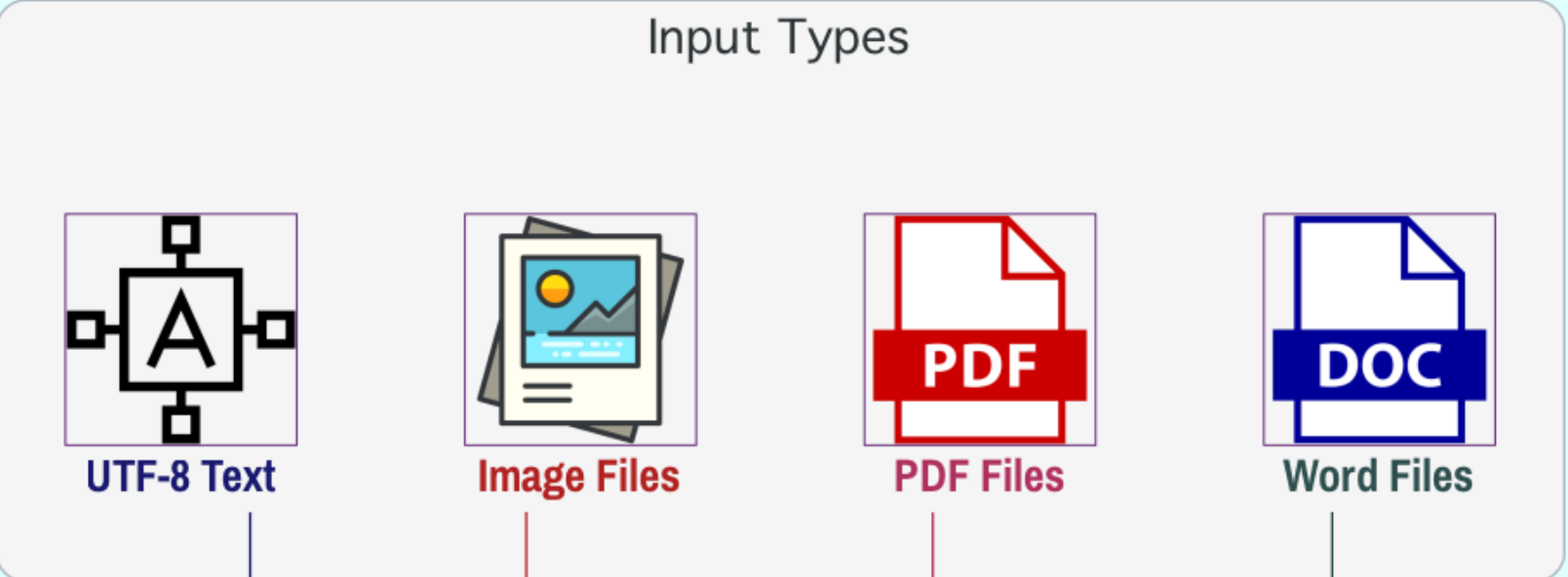
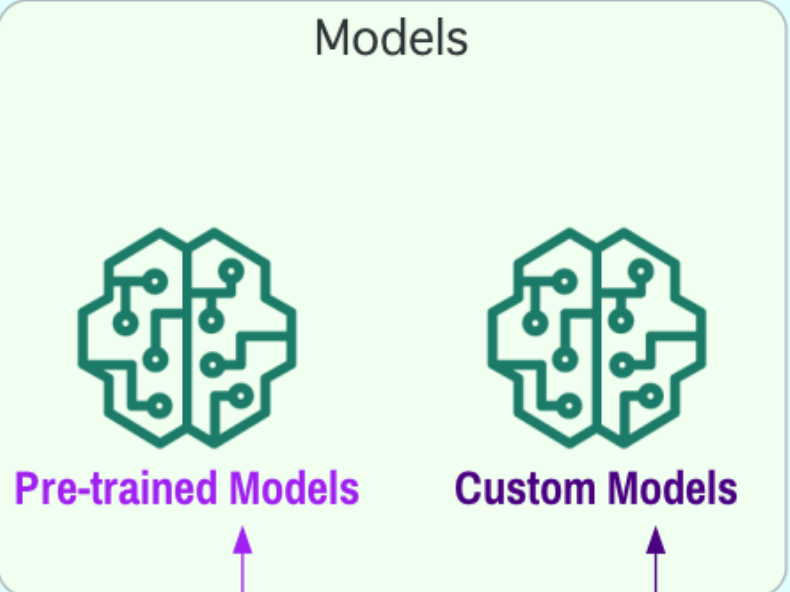
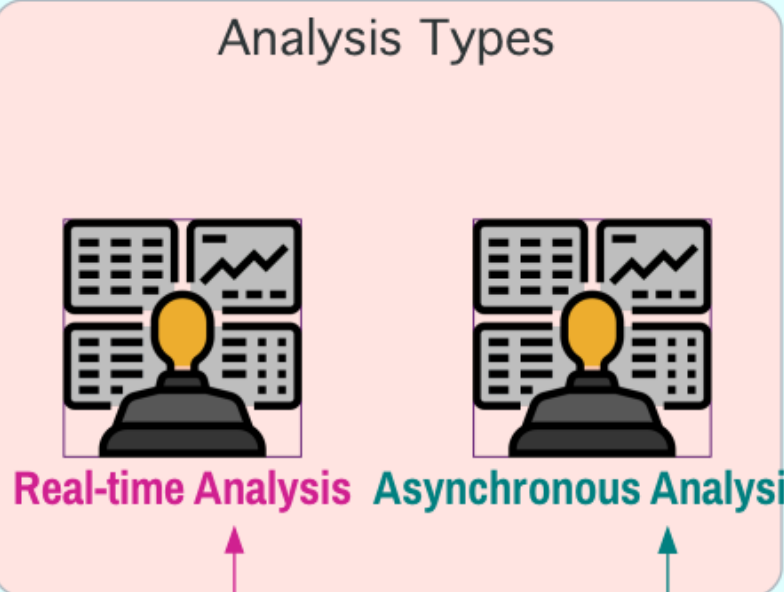
EventBridge

8. Multi-language support

Variety of languages

Global application versatility

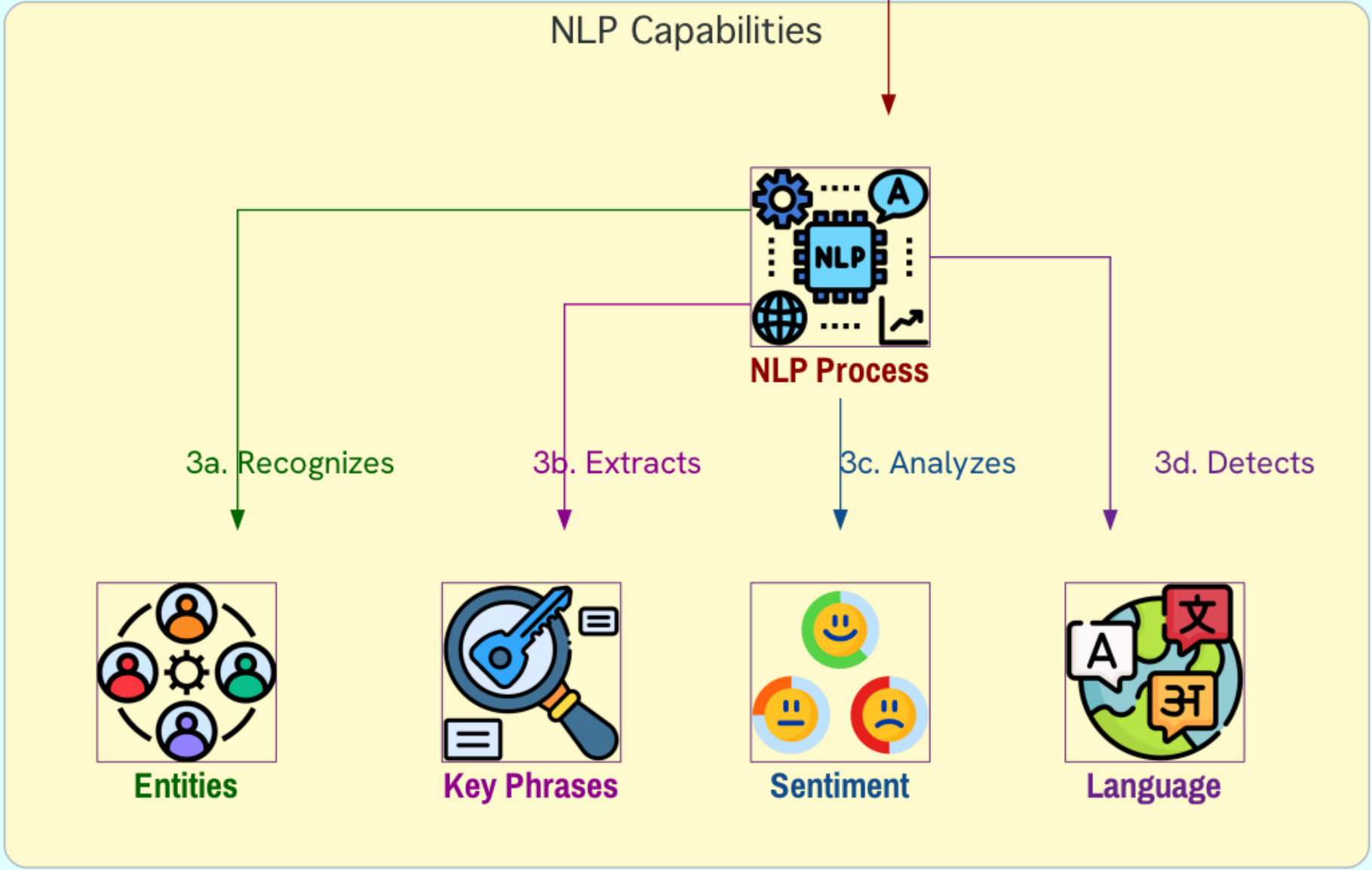
Amazon Comprehend



S3 Input



Amazon Comprehend



Document Analysis



Dominant Language



Lambda Function



Analytics



CloudWatch

























EventBridge

9. Continuous model improvement

Content storage

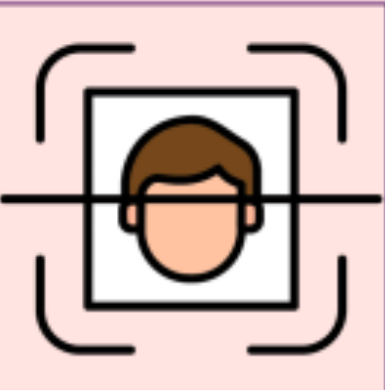
Ongoing refinement

Increased accuracy over time

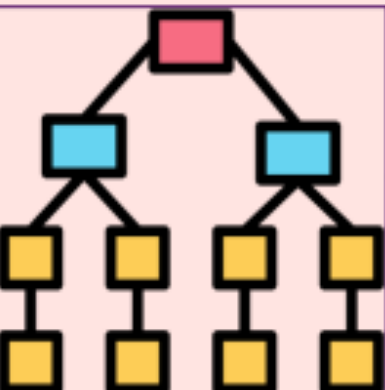
 Types of Insights	
 Entities	 Names of people, places, items, locations
 Key phrases	 Example: Basketball game document
	 Teams, venue, final score
 Personally Identifiable Information (PII)	 Address
	 Bank account number
	 Phone number
 Language	 Dominant language of document
 Sentiment	 Positive
	 Neutral
	 Negative
	 Mixed
 Targeted sentiment	 Associated with specific entities
	 Sentiment per entity occurrence
 Syntax	 Parts of speech for each word

Amazon Comprehend Custom

Custom NLP Models



Custom Entity Recognition



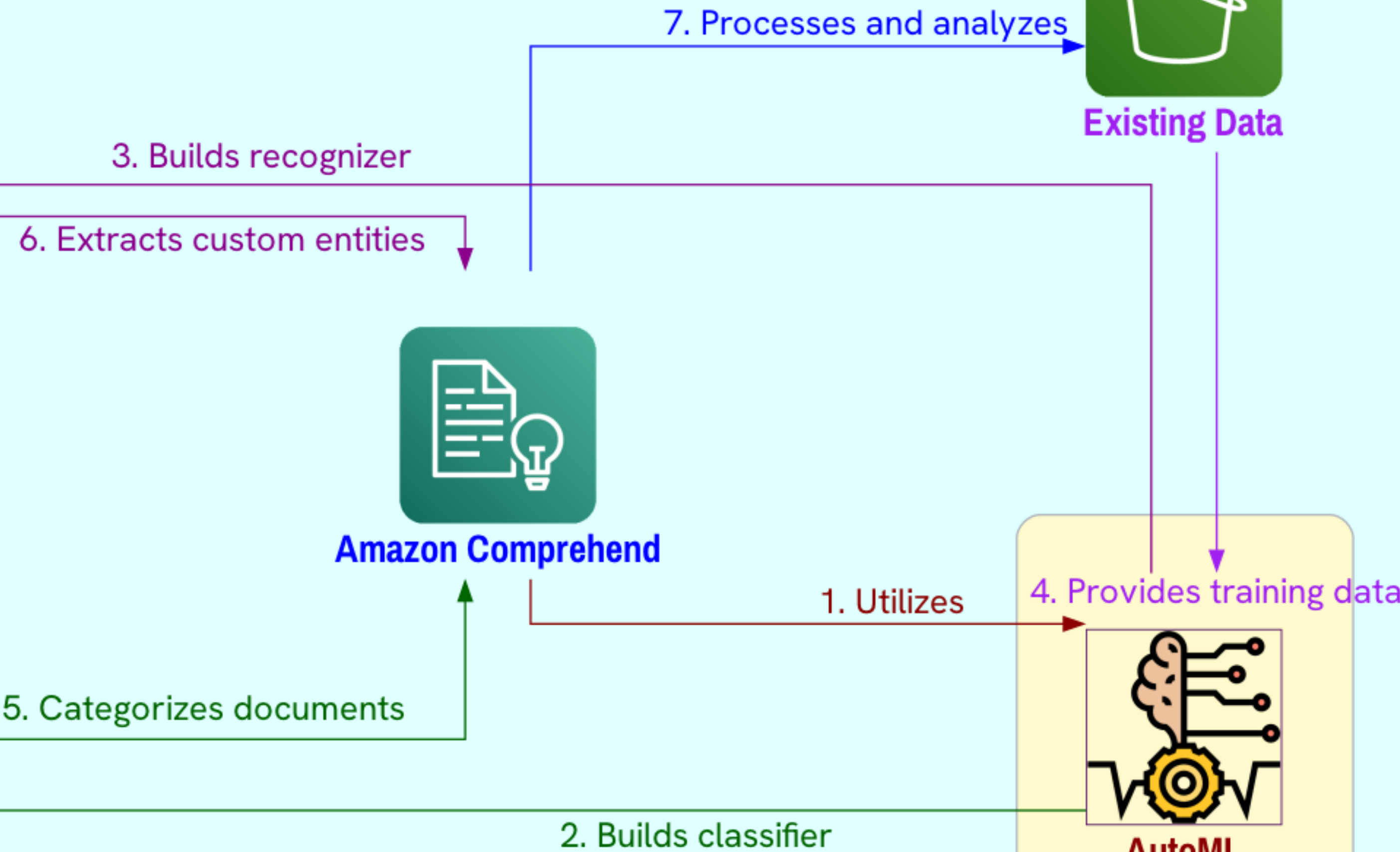
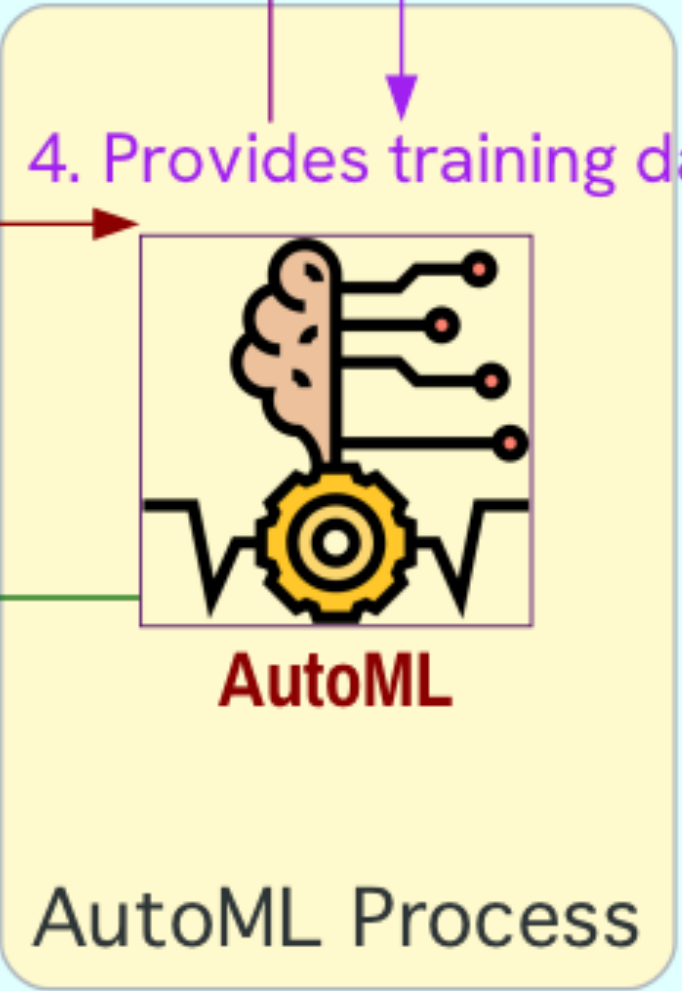
Custom Classification



Amazon Comprehend



Existing Data



1. Customization without ML expertise

Uses AutoML

Builds custom NLP models

Utilizes existing data

2. Custom classification

Creates classifiers

Organizes documents

User-defined categories

3. Custom entity recognition

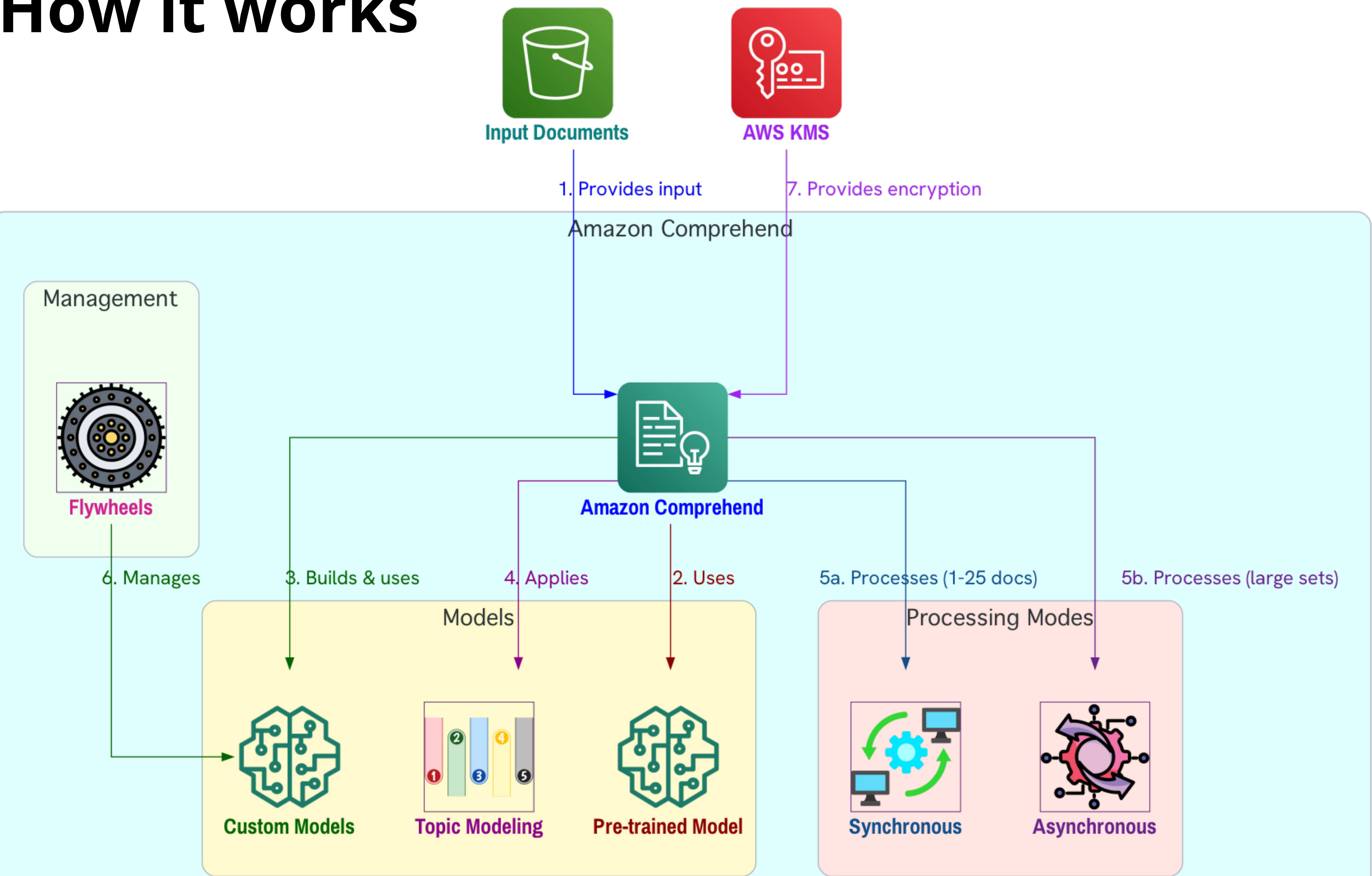
Creates recognizers

Analyzes text

Identifies specific terms

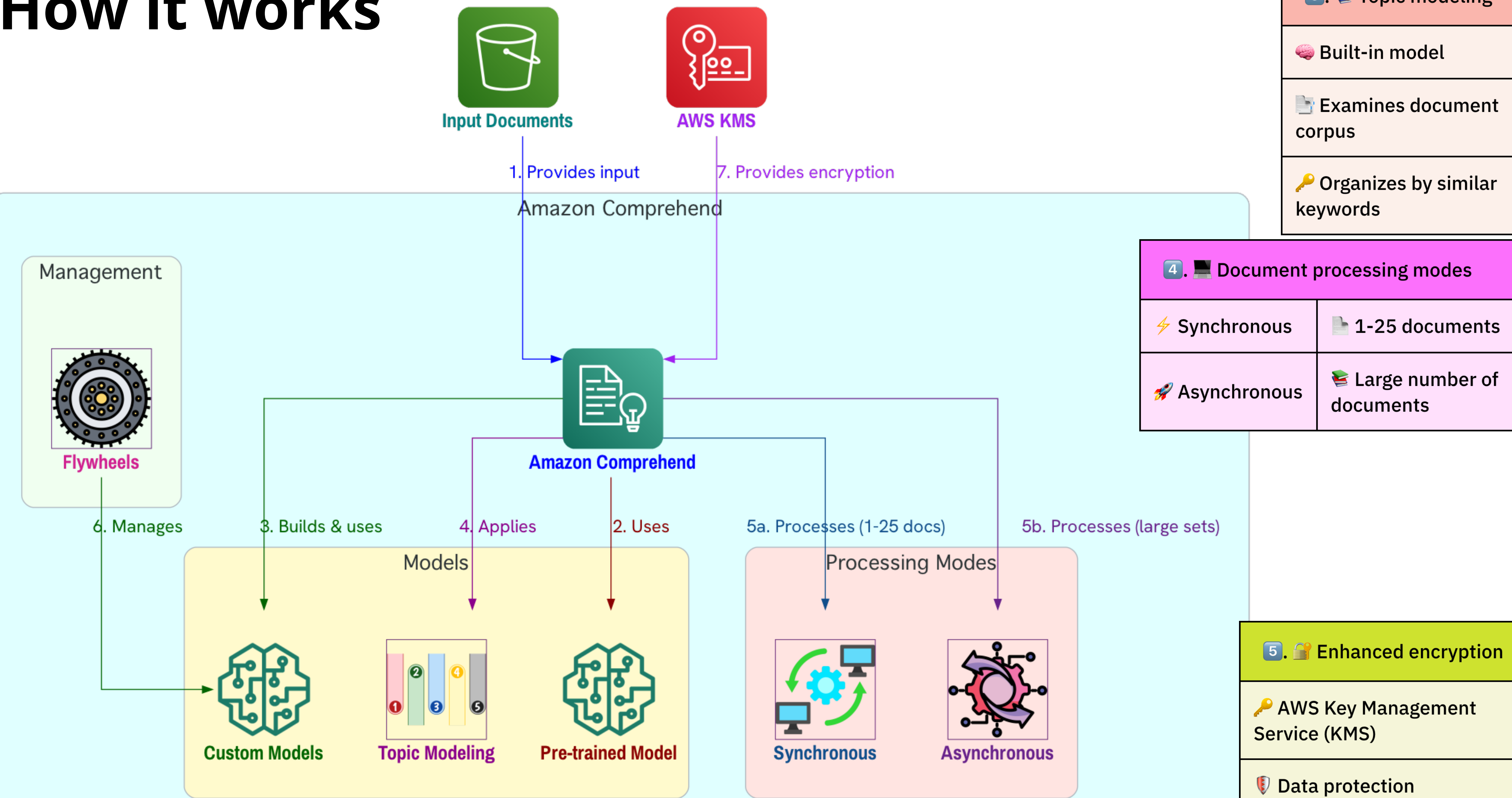
Detects noun-based phrases

How it works



- | |
|--|
| 1. 🤖 Pre-trained model |
| 📊 Gathers insights from documents |
| 🔄 Continuous training on large text corpus |
| 🚫 No user training data required |
-
- | |
|----------------------------------|
| 2. 🛠️ Custom models |
| 📁 Custom classification |
| 🔍 Custom entity recognition |
| 🔄 Flywheels for model management |

How it works



Topic Modeling

1. 🔍 Topic modeling with Amazon Comprehend

📄 Examines document collections

🔑 Determines common themes

📰 Example: news articles

⚽ Sports

🏛️ Politics

🎭 Entertainment

🚫 No annotation required

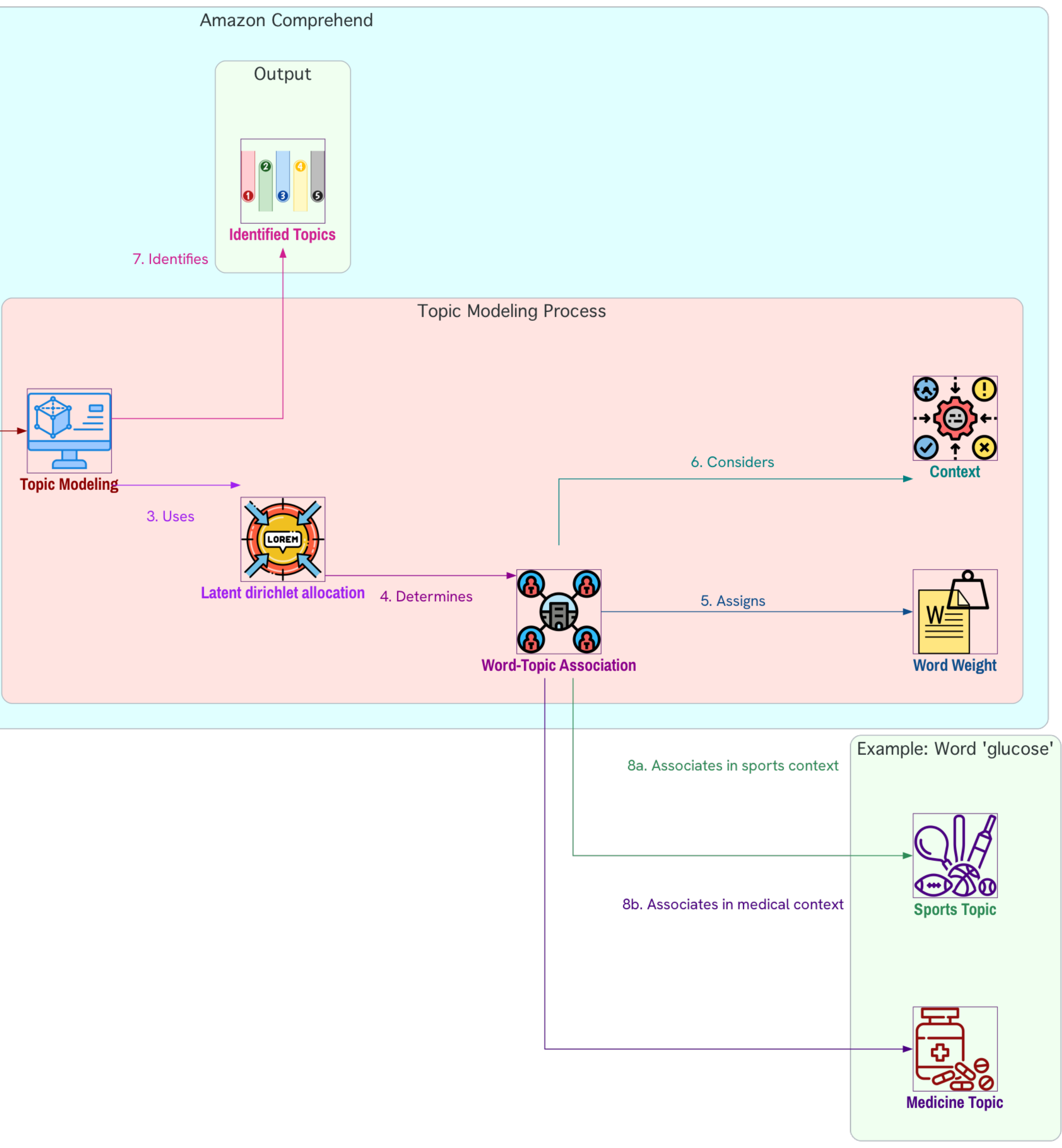
2. 🧠 Latent Dirichlet Allocation (LDA)

🤖 Learning model

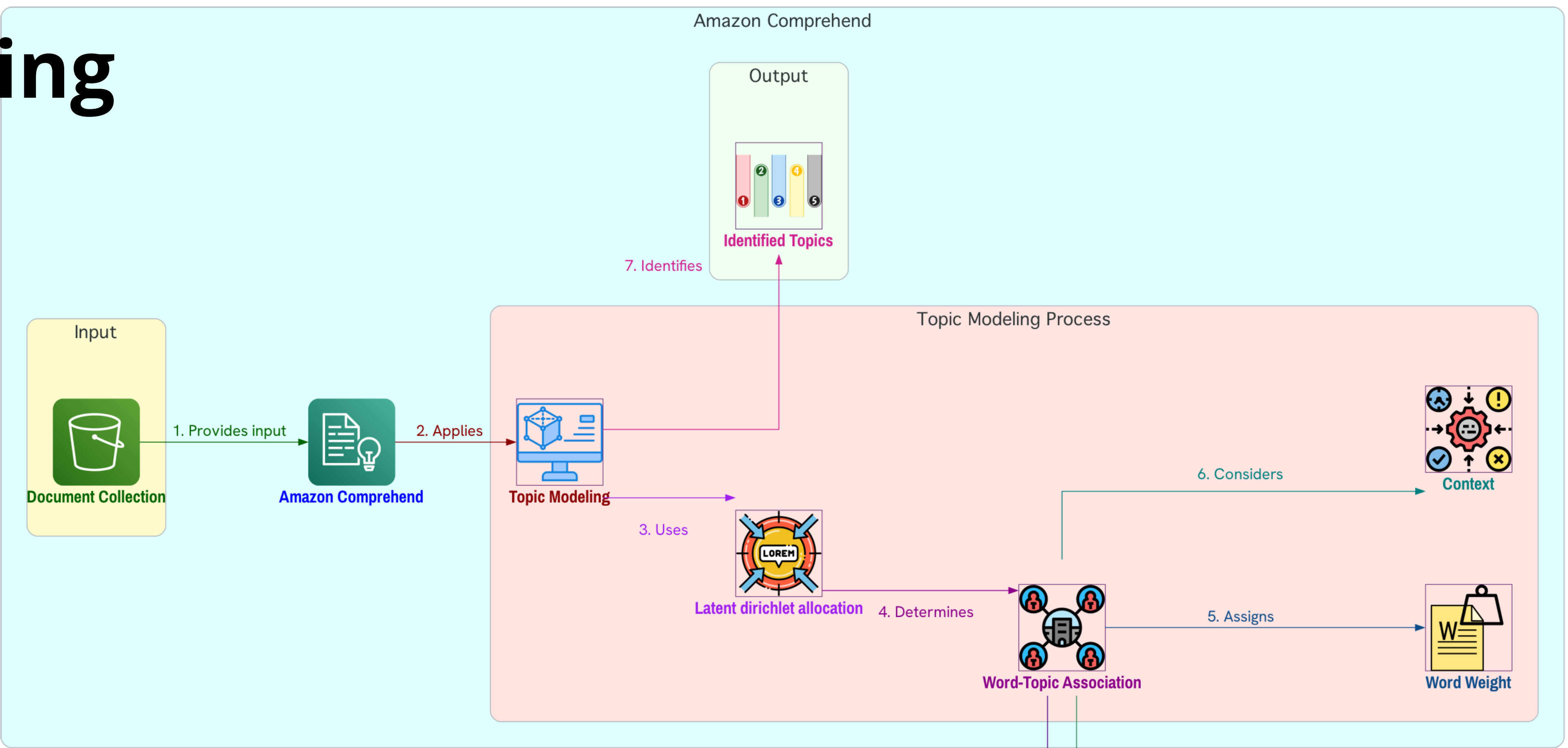
📊 Determines topics in document set

📖 Examines context and meaning of words

🔗 Words in same context form topics



Topic Modeling



3. 🔗 Word-topic association

📊 Based on topic prevalence

🔬 Considers word affinity to topic

🔄 Same word, different topics

🏆 "Glucose" in sports

💊 "Glucose" in medicine

4. ⚖️ Word weight

📊 Indicates word's importance to topic

📊 Based on word frequency

📊 Compared across entire document set

8a. Associates in sports context

8b. Associates in medical context

Example: Word 'glucose'



Sports Topic



Medicine Topic



**Thanks
for
Watching**