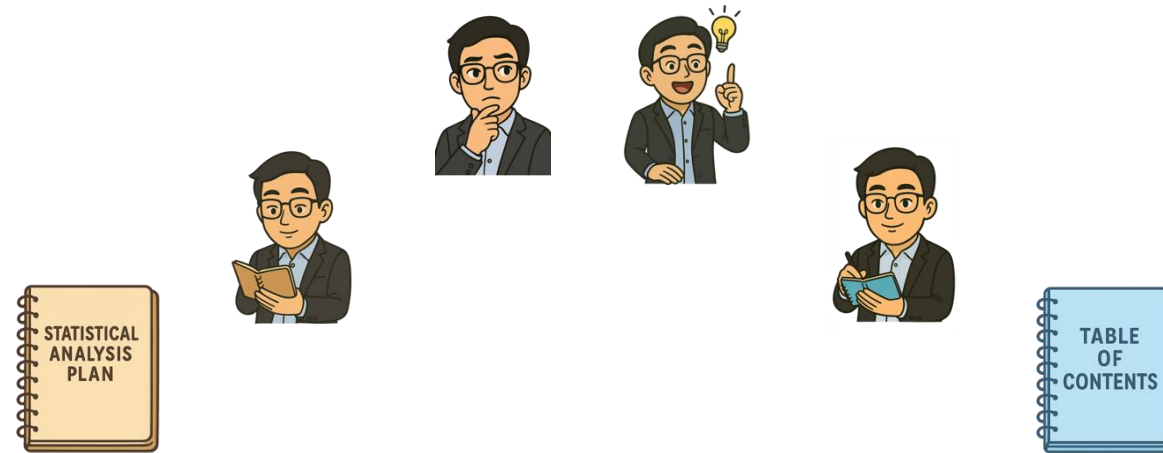


What if ...

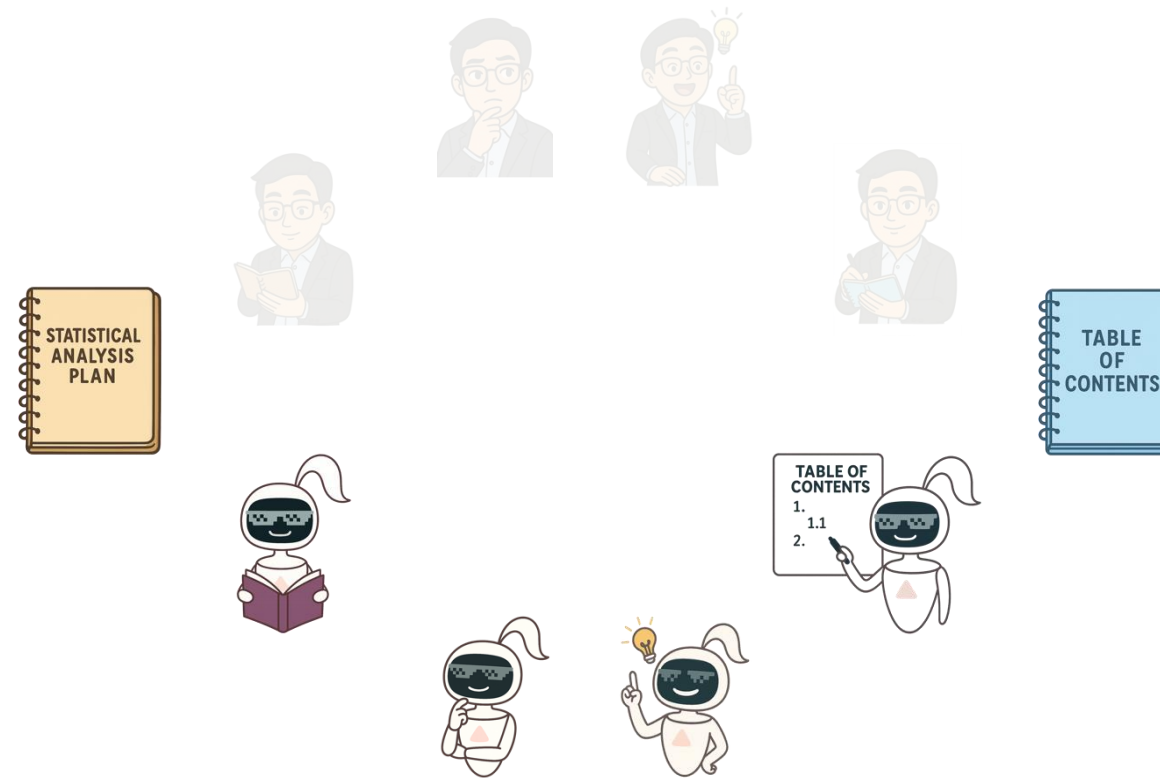
AI TOC Generation

Current Process



AI TOC Generation

Future Process



Artificial Intelligence

Reading and Understanding the SAP to Generate the TFL TOC

Toshio Kimura Arcsine Analytics
Siqi Wang Arcsine Analytics
Weiming Du Alnylam Pharmaceuticals
Songgu Xie Regeneron Pharmaceuticals

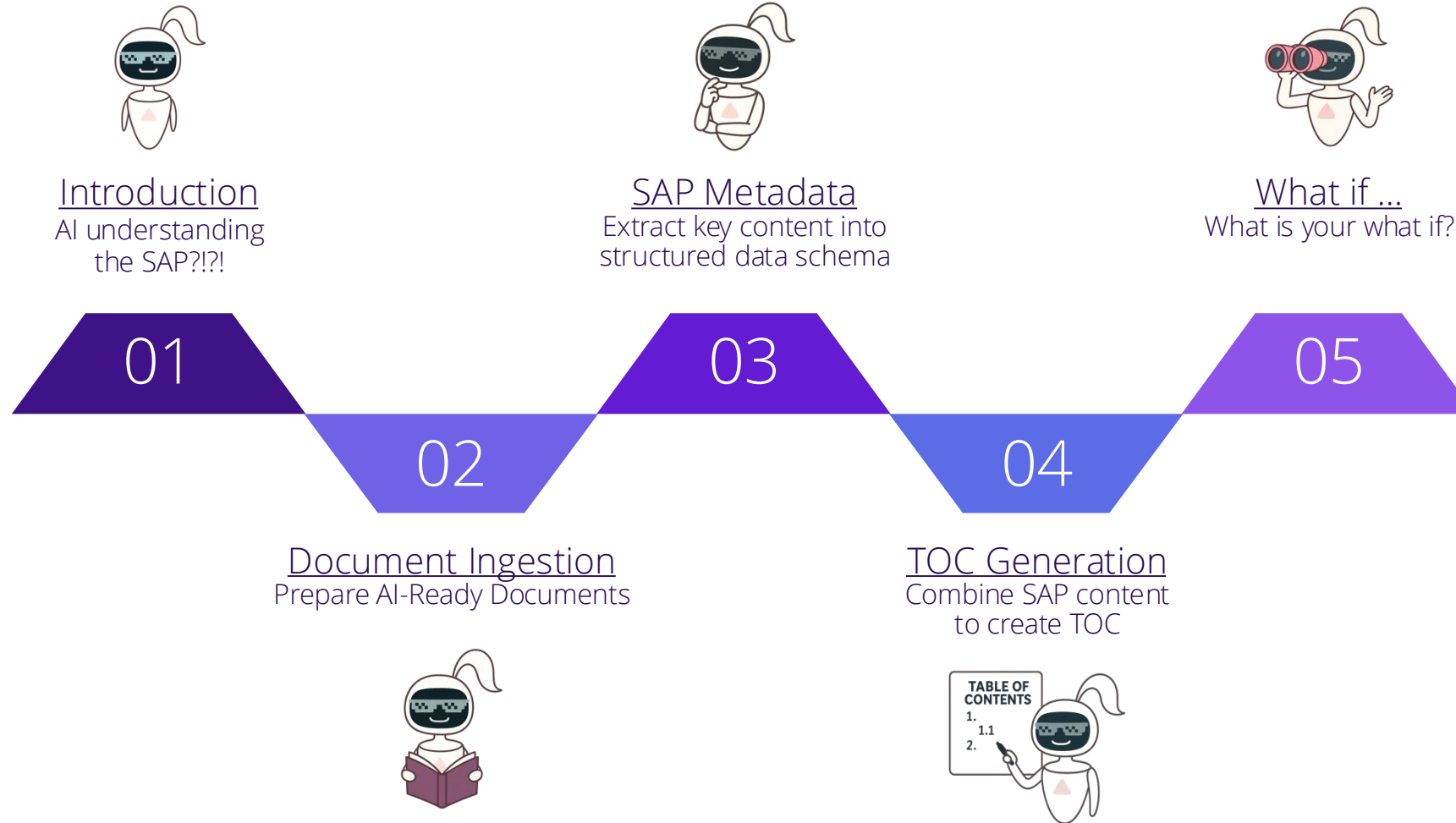
PHUSE US Connect 2026

ET-14

03.25.2026

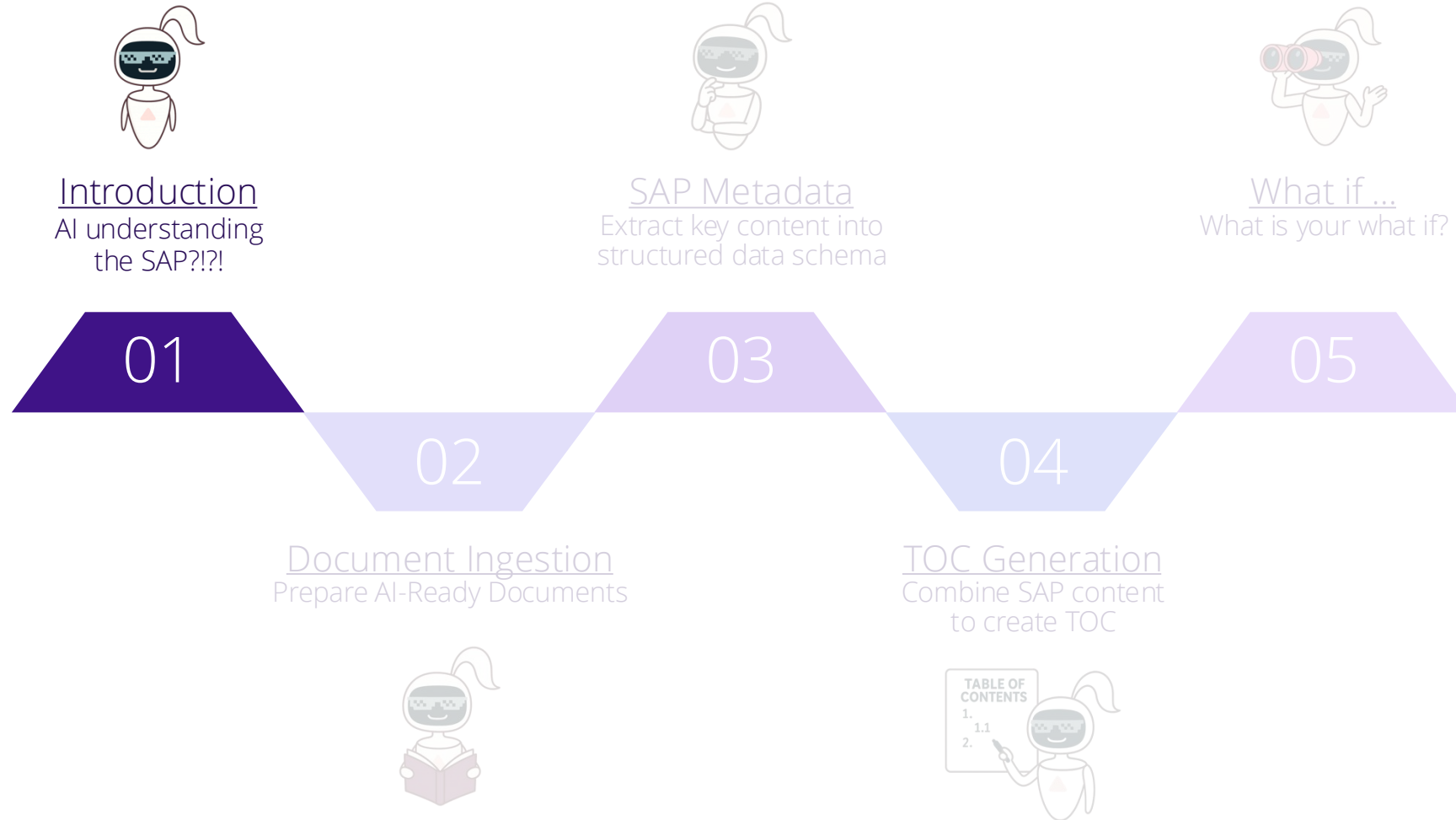
AI TOC Generation

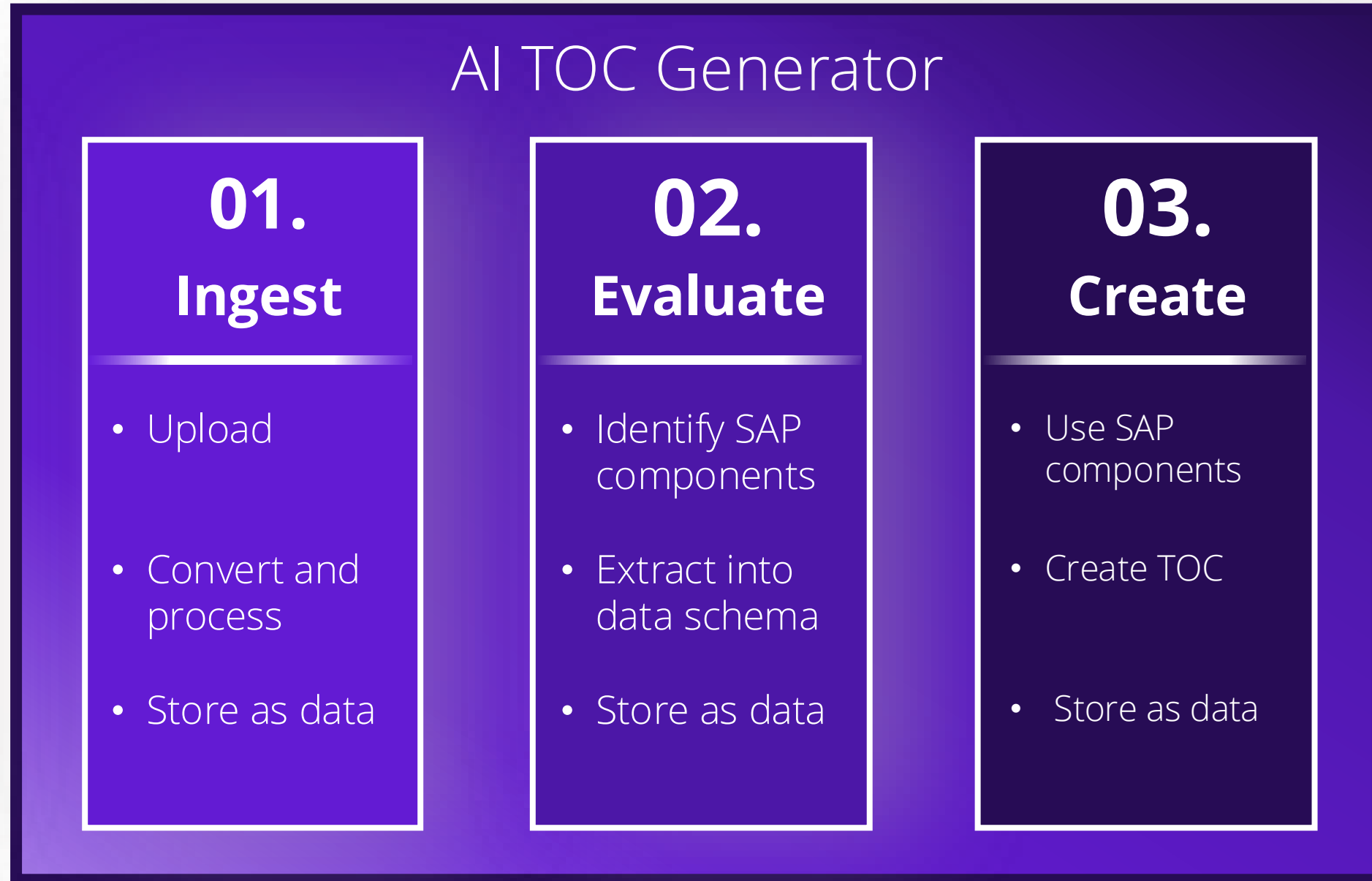
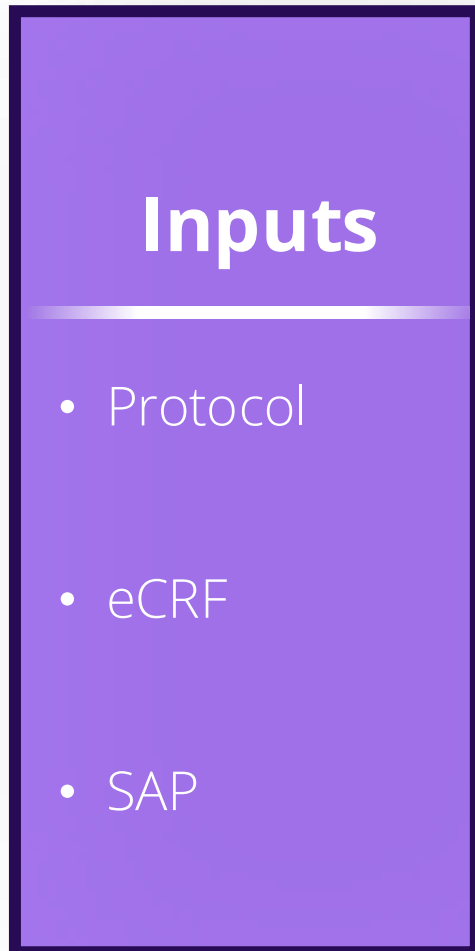
Agenda



AI TOC Generation

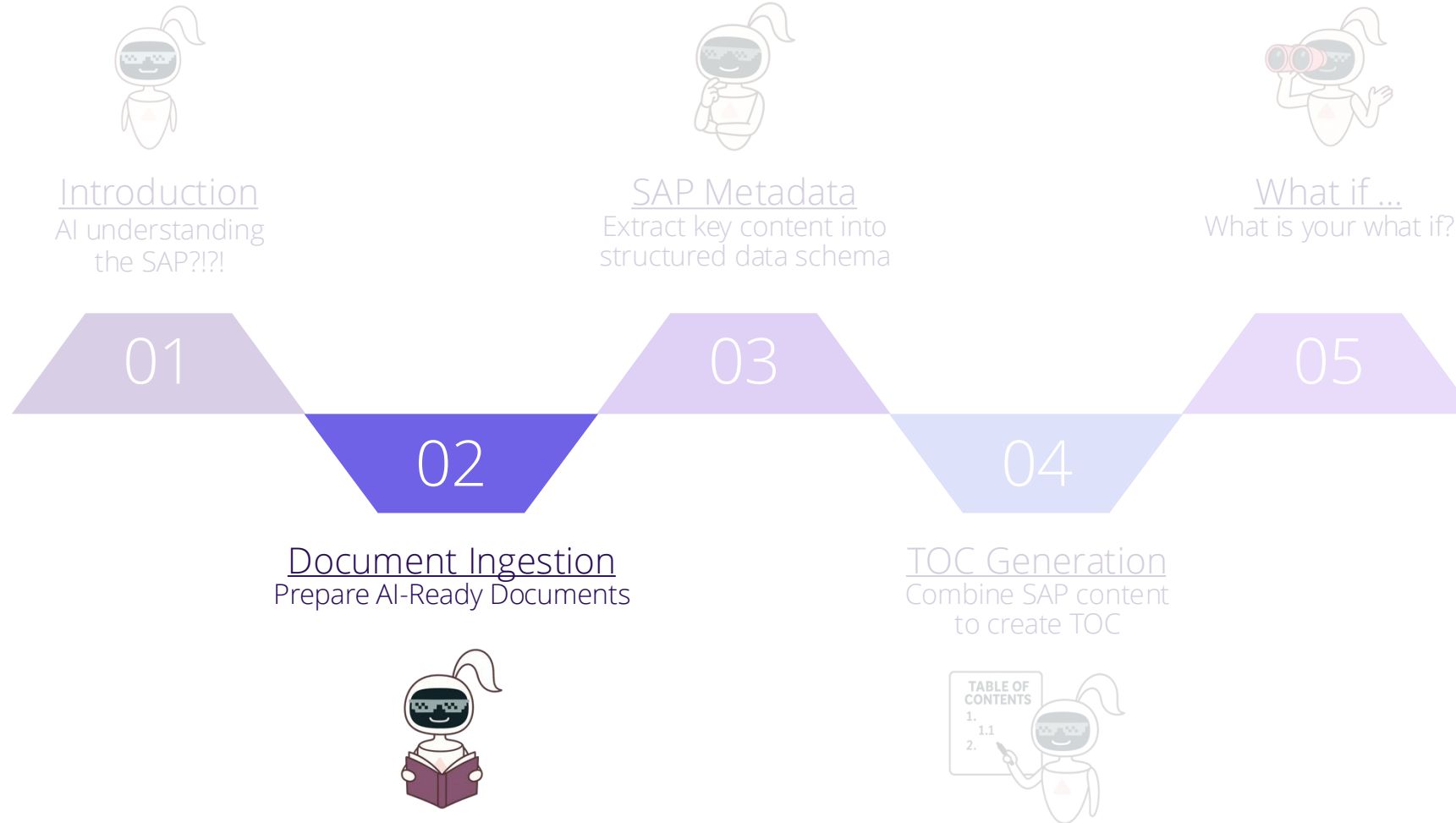
Agenda





AI TOC Generation

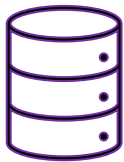
Agenda



Document Ingestion



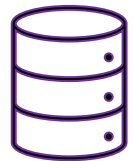
SectionNumber	SectionName	Content
5	Statistical Methods	For continuous variables, descriptive statistics will include the following: the number of patients reflected in the calculation (n), mean, median, standard deviation, minimum, and maximum. For categorical or ordinal data, frequencies and percentages will be displayed for each category.
5.1	Demo and Baseline	Demographic data and baseline characteristics variables will be summarized, by treatment group and overall, using descriptive statistics.
...	...	...





1) Adobe Extract API

2) Filter, process, clean-up



3) Store in database

01.

Upload

- Upload PDF, Word, RTF files

02.

Process

- Convert document to defined data schema

03.

Store

- Store in a database

Document Ingestion

```
{
  "Bounds": [
    72.02400207519531,
    696.8948822021484,
    298.83936364746094,
    716.3262329181562
  ],
  "clipBounds": [
    72.02400207519531,
    696.8948822021484,
    298.83936364746094,
    716.3262329181562
  ],
  "Font": {
    "alt_family_name": "Times New Roman",
    "embedded": true,
    "encoding": "Identity-H",
    "family_name": "Times New Roman",
    "font_type": "CIDFontType2",
    "italic": false,
    "monospaced": false,
    "name": "XIUJUZ+TimesNewRoman,Bold",
    "subset": true,
    "weight": 700
  },
  "HasClip": true,
  "Lang": "en",
  "ObjectID": 2305,
  "Page": 24,
  "Path": "//Document/H[11]",
  "Text": "S. STATISTICAL METHODS ",
  "TextSize": 14.039993286132812,
  "attributes": {
    "LineHeight": 16.875,
    "SpaceAfter": 18
  }
},
{
  "Bounds": [
    72.02400207519531,
    663.3760070800781,
    537.8040924072266,
    693.5800018310547
  ],
  "clipBounds": [
    72.02400207519531,
    663.3760070800781,
    537.8040924072266,
    693.5800018310547
  ],
  "Font": {
    "alt_family_name": "Times New Roman",
    "embedded": true,
    "encoding": "Identity-H",
    "family_name": "Times New Roman",
    "font_type": "CIDFontType2",
    "italic": false,
    "monospaced": false,
    "name": "QPMYAH+TimesNewRoman",
    "subset": true,
    "weight": 400
  },
  "HasClip": true,
  "Lang": "en",
  "ObjectID": 2306,
  "Page": 24,
  "Path": "//Document/P[84]",
  "Text": "For continuous variables, descriptive statistics will include the following: the number of patients reflected in the calculation (n), mean, median, standard deviation, minimum, and maximum. ",
  "TextSize": 12.0,
  "attributes": {
    "LineHeight": 14,
    "SpaceAfter": 5.5
  }
},
}
```

1 Adobe Developer



Extract PDF

Extract content from PDF documents and output it in a structured JSON format, along with tables and figures

Extract content from PDF documents and output it in a structured JSON format, along with tables and figures

- Extract PDF Content, Tables content and Tables/Figures renditions from a PDF document. Various available options are: (mutually exclusive)
1. Extract text with structure (headings, paragraphs, lists and footnotes)
 2. Extract tables data as CSV or XLSX
 3. Extract tables as images
 4. Extract bounding boxes for characters present in text block (paragraphs, list, headings)
 5. Extract figures or images as PNG

Extract API

2 Keep relevant fields

```
{
  "SectionID": "S",
  "SectionName": "STATISTICAL METHODS",
  "H1": "S STATISTICAL METHODS",
  "H2": "",
  "H3": "",
  "Text": "<p>For continuous variables, descriptive statistics will include the following: the number of patients reflected in the calculation (n), mean, median, standard deviation, minimum, and maximum.",
  "Figures/Tables": ""
},
}
```

Process

document to
defined data schema

Store

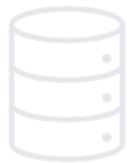
03.

- Store document content in a database

Document Ingestion



SectionNumber	SectionName	Content
5	Statistical Methods	For continuous variables, descriptive statistics will include the following: the number of patients reflected in the calculation (n), mean, median, standard deviation, minimum, and maximum. For categorical or ordinal data, frequencies and percentages will be displayed for each category.
5.1	Demo and Baseline	Demographic data and baseline characteristics variables will be summarized, by treatment group and overall, using descriptive statistics.
...	...	...



01. **Upload**

- Upload file

02. **Process**

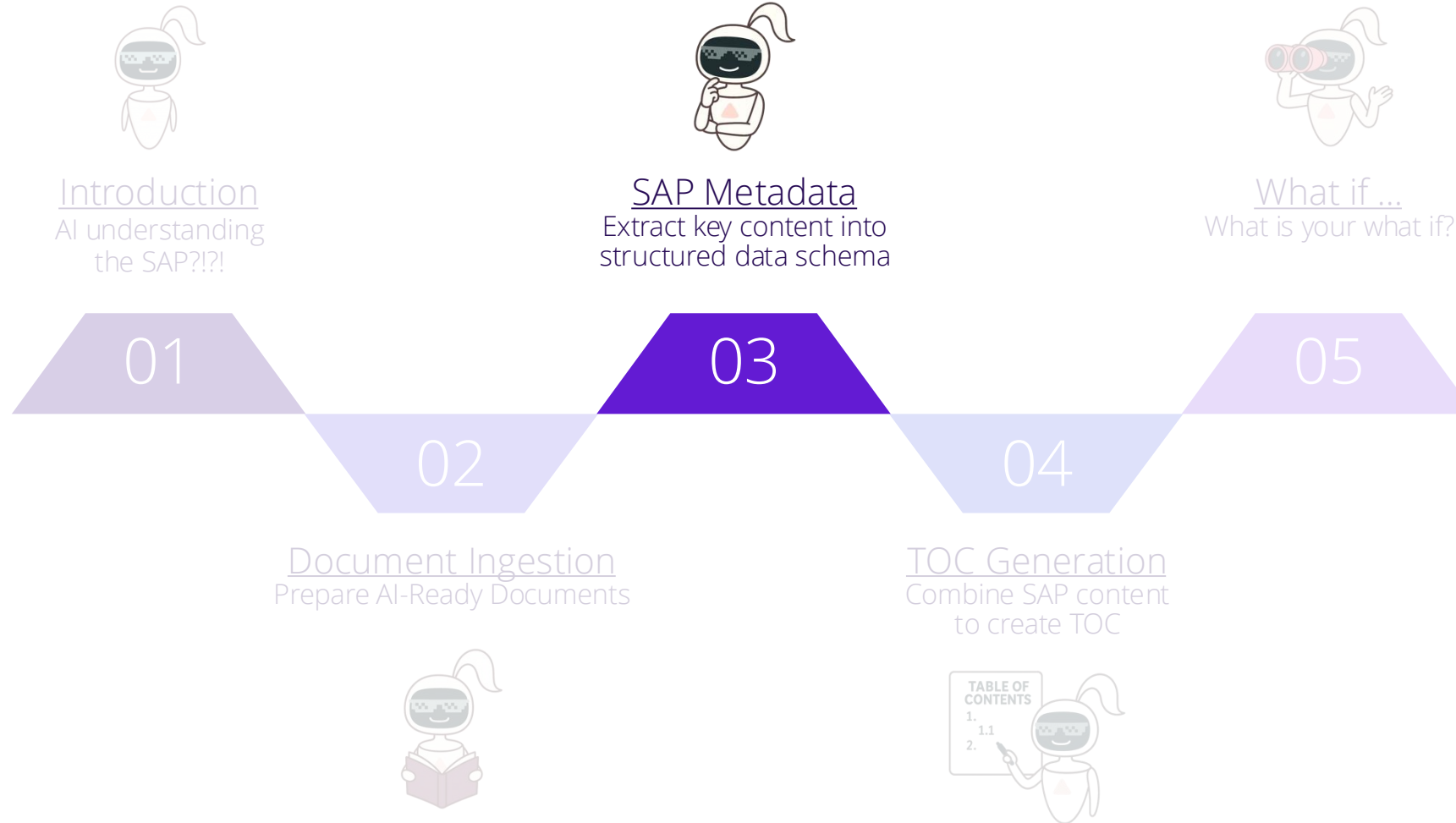
- Convert document to defined data schema

03. **Store**

- Store in a database

AI TOC Generation

Agenda



Evaluate SAP

5.6.1.1. Primary Analysis for Primary Efficacy Variable

For the analysis of the primary efficacy variable, a mixed model for repeated measurements (MMRM) will be used with baseline BCVA measurement as a covariate and treatment group (HDq16 vs. 2q8 and HDq12 vs. 2q8) and baseline CRT category (<400 μm , ≥ 400 μm), prior DME treatment (yes, no), geographical region (Rest of world, Japan), and visit as fixed effects as well as terms for the interaction between baseline BCVA and visit and for the interaction between treatment and visit. A Kenward-Roger approximation will be used for the denominator degrees of freedom. Further, an unstructured covariance structure will be used. If the model does not converge, an appropriate covariance structure will be used instead. Only data up to Week 48 will be used in this analysis.



Analysis: MMRM

Terms in model:

baseline BCVA
treatment group
baseline CRT category
prior DME treatment
geographical region
visit

interaction terms:

baseline BCVA and visit
treatment and visit

DDFM: KR

Covariance matrix: UN

Timepoints: Up to week 48

01.

Inputs

- Identify SAP components
- Few shot examples
- Rules

02.

Extract

- Extract into data schema

03.

Store

- Store in a database

3. ANALYSIS POPULATIONS

3.1. The Full Analysis Set (FAS)

The full analysis set (FAS) includes all randomized patients who received at least 1 dose of study medication; it is based on the treatment assigned to the patient at baseline (as randomized). FAS is the primary analysis set for efficacy endpoints.



Population: Full Analysis Set

Abbreviation: FAS

Treatment: as-randomized

Usage: Efficacy

Definition: All randomized patients who received at least 1 dose of study medication

01.

Inputs

- Identify SAP components
- Few shot examples
- Rules

02.

Extract

- Extract into data schema

03.

Store

- Store in a database

Evaluate SAP

1) SAP Components

2) Few Shot Examples

3) Rules, rules, rules



1) Extracted Values

2) JSON Data Schema

01.

Inputs

- Identify SAP components
- Few shot examples
- Rules

02.

Extract

- Extract into data schema

03.

Store

- Store in a database

Evaluate SAP

Few shot approach prompting method:

- ❶ Provide several examples of input and correct output pairs
 - a) Input: Example SAP text
 - b) Output: Structured data schema with extracted content
- ❷ Provide LLM with the examples and have the LLM apply the training from the examples to a new input

Implementation:

- ❶ Prepare several examples of input/output pairs from example SAP text along with the expected response (structured data schema with values filled in as per the example input)
- ❷ Pull the relevant text from the new SAP, merge with the examples and rules, and submit the combined prompt to the LLM

❶ Examples

Example Input

3.1. The Full Analysis Set (FAS)
The full analysis set (FAS) includes all randomized patients who received at least 1 dose of study medication; it is based on the treatment assigned to the patient at baseline (as randomized). FAS is the primary analysis set for efficacy endpoints.

Example Output

Analysis Set: Full Analysis Set
Abbreviation: FAS
Treatment: as-randomized
Usage: Efficacy
Definition: All randomized patients who received at least 1 dose of study medication

❷ New Input

SectionId	SectionName	Content
3.1	Intent-to-treat Population	The intent-to-treat (ITT) population includes all randomized patients. Efficacy analyses will be based on the treatment allocated by the interactive response technology (IRT) at randomization (as randomized). This is the primary analysis population for efficacy analyses.



Response with New Output

Analysis Set:	Intent-to-Treat
Abbreviation:	ITT
Treatment:	as-randomized
Usage:	Efficacy
Definition:	All randomized patients

01. Inputs

- Identify SAP components
- Few shot examples
- Rules

02. Extract

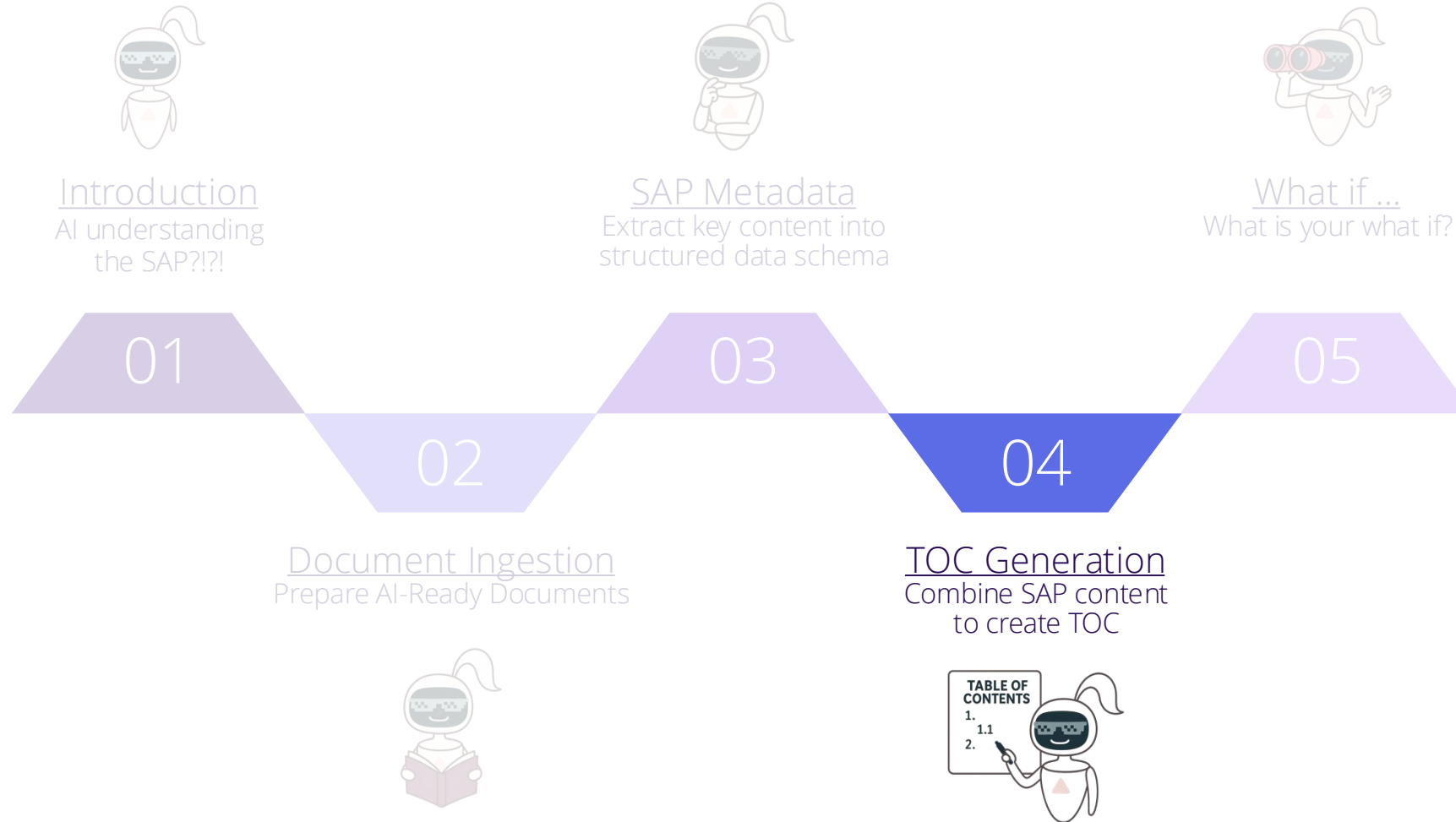
- Extract into data schema

03. Store

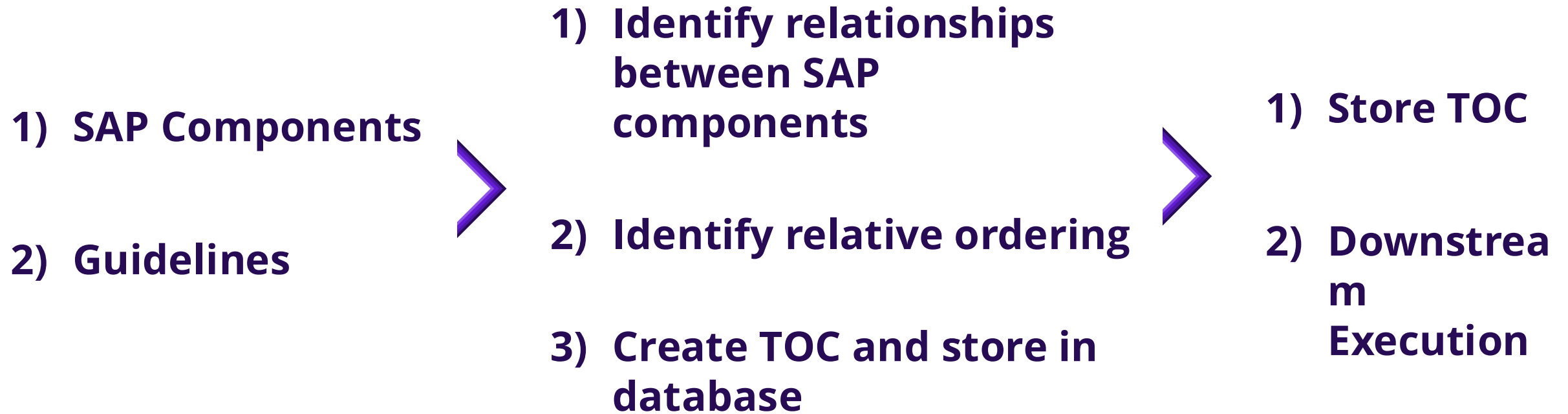
- Store in a database

AI TOC Generation

Agenda



TOC Generation



01.

SAP Data

- Use SAP components as the input data source
- Use guidance documents to guide TOC creation

02.

Create TOC

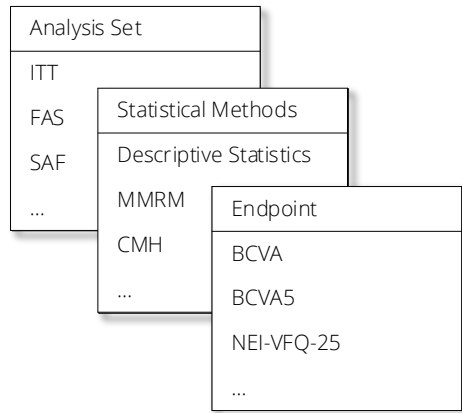
- Identify relationships and relative ordering between components
- Mix and match to create TOC

03.

Store

- Store in database

TOC Generation



E3: Structure and Content of Clinical Study Reports

- 14. Tables, Figures and Graphs Referred to but not included in the Text
 - 14.1 Demographic Data
 - 14.2 Efficacy Data
 - 14.3 Safety Data
 - 14.3.1 Displays of Adverse Events
 - 14.3.2 Listings of Deaths, Other Serious and Significant Adverse Events
 - 14.3.3 Narratives of Deaths, Other Serious and Certain Other Significant Adverse Events
 - 14.3.4 Abnormal Laboratory Value Listing (Each Patient)

1) Identify relationships between SAP components

2) Identify relative ordering

3) Create TOC

1) Store TOC

2) Downstream Execution

SAP Data

01.

- Use SAP components as the input data source
- Use guidance documents to guide TOC creation

Create TOC

02.

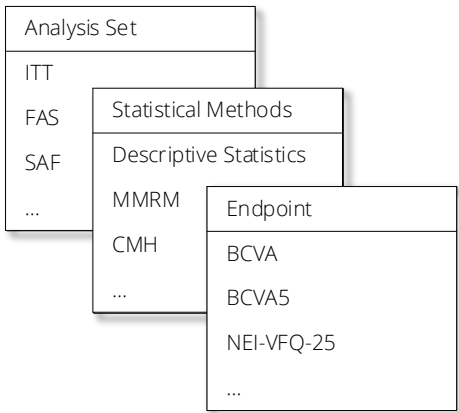
- Identify relationships and relative ordering between components
- Mix and match to create TOC

Store

03.

- Store in database

TOC Generation



E3: Structure and Content of Clinical Study Reports

- 14. Tables, Figures and Graphs Referred to but not included in the Text
 - 14.1 Demographic Data
 - 14.2 Efficacy Data
 - 14.3 Safety Data
 - 14.3.1 Displays of Adverse Events
 - 14.3.2 Listings of Deaths, Other Serious and Significant Adverse Events
 - 14.3.3 Narratives of Deaths, Other Serious and Certain Other Significant Adverse Events
 - 14.3.4 Abnormal Laboratory Value Listing (Each Patient)



Order	Abbreviation	Usage
1	FAS	All efficacy analyses
2	PPS	Supplementary efficacy analyses
3	SAF	Treatment compliance/administration and safety
4		
5		
6		

Order	LVL Order	Endpoint Level	Endpoint
1	1	Primary	Change from baseline in BCVA
2	2	Key Secondary	≥2-step improvement in DRSS
3	3	Secondary	Gain of ≥15 letters in BCVA from baseline
4	3	Secondary	BCVA ≥69 letters
5	3	Secondary	No fluid at foveal center
6	3	Secondary	Change from baseline in CRT
7	3	Secondary	No leakage on fluorescein angiography
8	3	Secondary	Change from baseline in NEI-VFQ-25 total



1) Store TOC

2) Downstream Execution

SAP Data

01.

- Use SAP components as the input data source
- Use guidance documents to guide TOC creation

Create TOC

02.

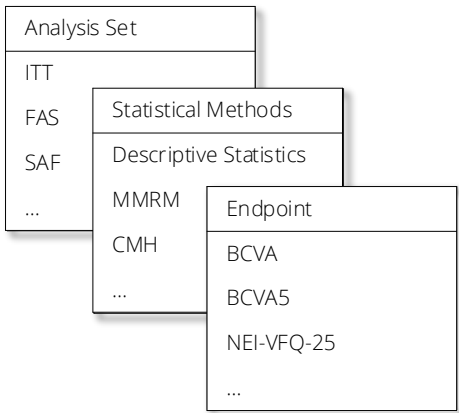
- Identify relationships and relative ordering between components
- Mix and match to create TOC

Store

03.

- Store in database

TOC Generation



E3: Structure and Content of Clinical Study Reports

- 14. Tables, Figures and Graphs Referred to but not included in the Text
 - 14.1 Demographic Data
 - 14.2 Efficacy Data
 - 14.3 Safety Data
 - 14.3.1 Displays of Adverse Events
 - 14.3.2 Listings of Deaths, Other Serious and Significant Adverse Events
 - 14.3.3 Narratives of Deaths, Other Serious and Certain Other Significant Adverse Events
 - 14.3.4 Abnormal Laboratory Value Listing (Each Patient)

Order	Abbreviation	Usage
1	FAS	All efficacy analyses
2	PPS	Supplementary efficacy analyses
3	SAF	Treatment compliance/administration and safety
4		
5		
6		

Order	LVL	Order	Endpoint Level	Endpoint
1	1	1	Primary	Change from baseline in BCVA
2	2	2	Key Secondary	≥2-step improvement in DRSS
3	3	3	Secondary	Gain of ≥15 letters in BCVA from baseline
4	3	3	Secondary	BCVA ≥69 letters
5	3	3	Secondary	No fluid at foveal center
6	3	3	Secondary	Change from baseline in CRT
7	3	3	Secondary	No leakage on fluorescein angiography
8	3	3	Secondary	Change from baseline in NEI-VFQ-25 total

Title1	Title2
Table 14.2: 1 - MMRM Analysis: Change from baseline in BCVA (ETDRS letters) at Week 48	FAS
Table 14.2: 2 - CMH Analysis: ≥2-step improvement in DRSS at Week 48	FAS
Table 14.2: 3 - MMRM Analysis: Change from baseline in BCVA at Week 60	FAS
Table 14.2: 4 - CMH Analysis: Gain ≥15 letters in BCVA from baseline at Week 48	FAS
Table 14.2: 5 - CMH Analysis: BCVA ≥69 letters at Week 48	FAS
Table 14.2: 6 - CMH Analysis: No fluid at foveal center at Week 48	FAS
Table 14.2: 7 - MMRM Analysis: Change from baseline in CRT at Week 48	FAS
Table 14.2: 8 - CMH Analysis: No leakage on fluorescein angiography (FA) at Week 48	FAS
Table 14.2: 9 - MMRM Analysis: Change from baseline in NEI-VFQ-25 total score at Week 48	FAS
Table 14.2: 10 - MMRM Analysis: Change from baseline in BCVA at Week 96	FAS
Table 14.2: 11 - CMH Analysis: Gain ≥15 letters in BCVA from baseline at Week 60	FAS
Table 14.2: 12 - CMH Analysis: Gain ≥15 letters in BCVA from baseline at Week 96	FAS
Table 14.2: 13 - CMH Analysis: BCVA ≥69 letters at Week 60	FAS
Table 14.2: 14 - CMH Analysis: BCVA ≥69 letters at Week 96	FAS
Table 14.2: 15 - MMRM Analysis: Change from baseline in CRT at Week 60	FAS
Table 14.2: 16 - MMRM Analysis: Change from baseline in CRT at Week 96	FAS
Table 14.2: 17 - CMH Analysis: No leakage on FA at Week 60	FAS
Table 14.2: 18 - CMH Analysis: No leakage on FA at Week 96	FAS
Table 14.2: 19 - MMRM Analysis: Change from baseline in NEI-VFQ-25 total score at Week 60	FAS
Table 14.2: 20 - MMRM Analysis: Change from baseline in NEI-VFQ-25 total score at Week 96	FAS
Table 14.2: 21 - CMH Analysis: No retinal fluid (total fluid, IRF, SRF) at foveal center and in center subfield at Week 48	FAS
Table 14.2: 22 - CMH Analysis: No retinal fluid (total fluid, IRF, SRF) at foveal center and in center subfield at Week 60	FAS
Table 14.2: 23 - CMH Analysis: No retinal fluid (total fluid, IRF, SRF) at foveal center and in center subfield at Week 96	FAS
Table 14.2: 24 - Cox proportional hazards model Analysis: Time to fluid-free retina over 48 weeks (total fluid, IRF, SRF at foveal center and in center subfield)	FAS
Table 14.2: 25 - Cox proportional hazards model Analysis: Time to fluid-free retina over 60 weeks (total fluid, IRF, SRF at foveal center and in center subfield)	FAS
Table 14.2: 26 - Cox proportional hazards model Analysis: Time to fluid-free retina over 96 weeks (total fluid, IRF, SRF at foveal center and in center subfield)	FAS
Table 14.2: 27 - CMH Analysis: Sustained fluid-free retina over 48 weeks (total fluid, IRF, SRF at foveal center and in center subfield)	FAS
Table 14.2: 28 - CMH Analysis: Sustained fluid-free retina over 60 weeks (total fluid, IRF, SRF at foveal center and in center subfield)	FAS
Table 14.2: 29 - CMH Analysis: Sustained fluid-free retina over 96 weeks (total fluid, IRF, SRF at foveal center and in center subfield)	FAS
Table 14.2: 30 - Cox proportional hazards model Analysis: Time to sustained fluid-free retina over 48 weeks (total fluid, IRF, SRF at foveal center and in center subfield)	FAS

SAP Data

01.

- Use SAP components as the input data source
- Use guidance documents to guide TOC creation

Create TOC

02.

- Identify relationships and relative ordering between components
- Mix and match to create TOC

Store

03.

- Store in database

AI TOC Generation



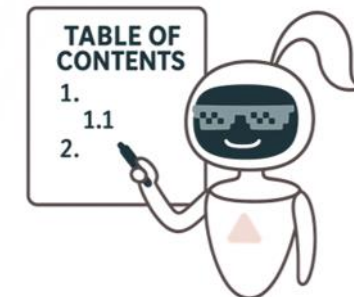
01.
Ingest



02.
Evaluate



03.
Create



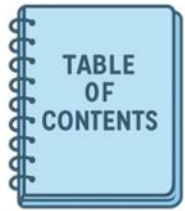
AI TOC Generation



01.
Ingest

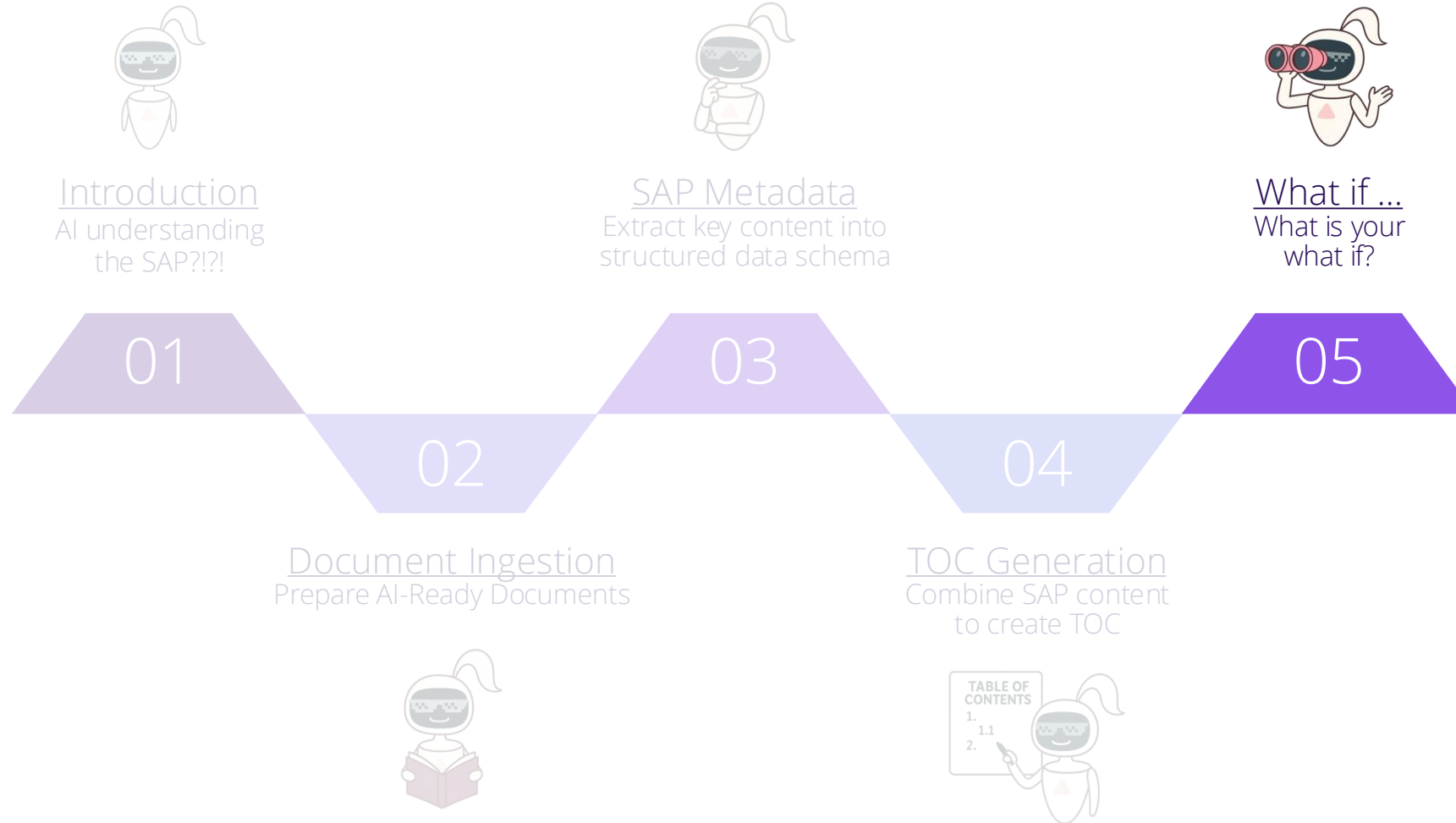
02.
Evaluate

03.
Create



AI TOC Generation

Agenda



AI TOC Generation

Boost Productivity by 60%

Reimagining life science enterprises with agentic AI

September 8, 2025 | Article

McKinsey
& Company

“

... Once the database is locked, the programming agent produces analysis datasets and results tables ... with a human biostatistician providing input for nonroutine datasets. This process could boost productivity ... by 60 percent ...

”

What if ...

AI TOC Generation



Introduction
AI understanding
the SAP?!?!?

01



SAP Metadata
Extract key content into
structured data schema

03



What if...
What is your what if?

05

toshio.kimura@arcsineanalytics.com

www.arcsineanalytics.com

02

Document Ingestion
Prepare AI-Ready Documents



04

TOC Generation
Combine SAP content
to create TOC

