University of Stuttgart
Institute of Industrial Automation
and Software Engineering

Large Language Model Agents for AI-Assisted Document Editing for Manufacturing Process Compliance

Presenter: Xiao, Sa

Supervisor: Yuchen Xia M. Sc.

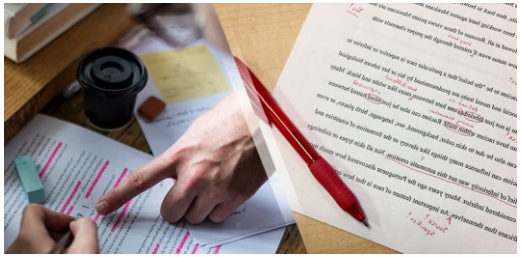
Examiner: Prof. Dr. Ing. Michael Weyrich

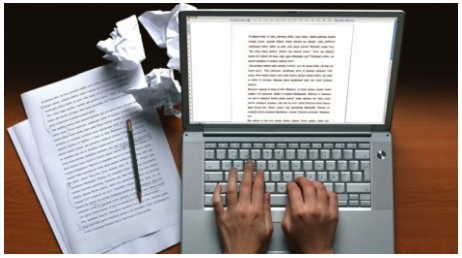
1 Introduction

- Current editing process
- Industrial use case

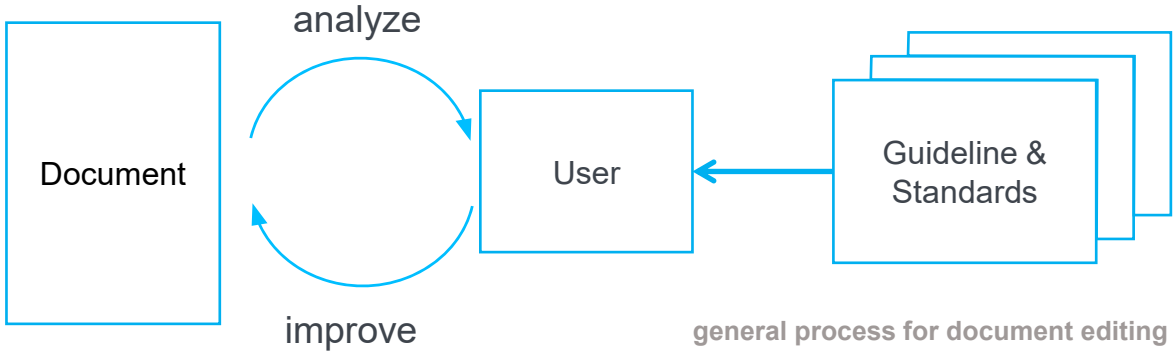
Introduction

Current editing process





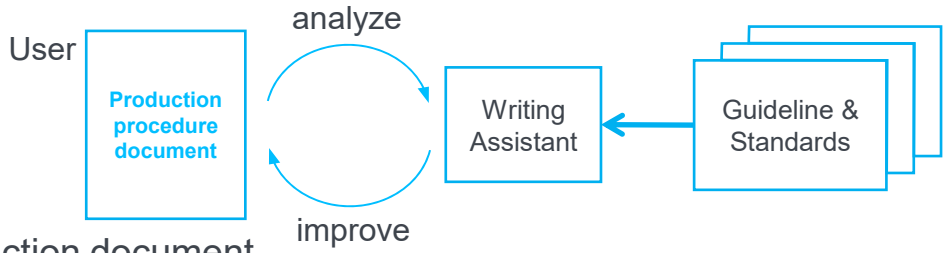
- Document Editing/ Improve Paper work
  - Analyze documents (based on **guidelines & standards**)
  - Involves **multiple rounds of review**, **correction**, and **improvement**



general process for document editing

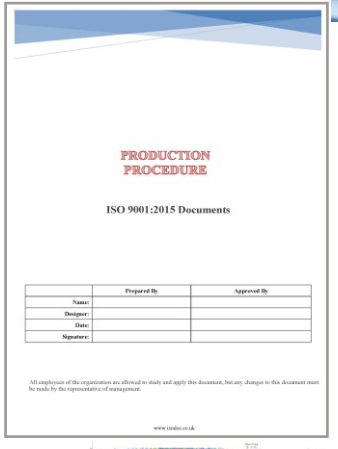
Introduction

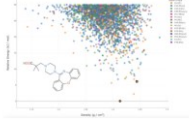
Industrial use case

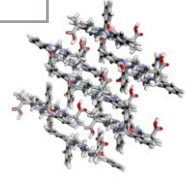


- Chemical engineer: keep track on a production document
- Pharmaceutical companies: **revise a manufacturing process document**

  ->editing Production procedure document
  (detailed & regulated)







Challenge

Traditional knowledge utilization method
- Knowledge-Intensive
  Complex structure, multiple levels, highly specialized
- Manual verification
  Extremely attentive, easily overlook details

! **Quality deviations** Drugs, **wasted time**, **higher production costs**, and even **health risks** in the final product, resulting in it being **unable to enter the market**.

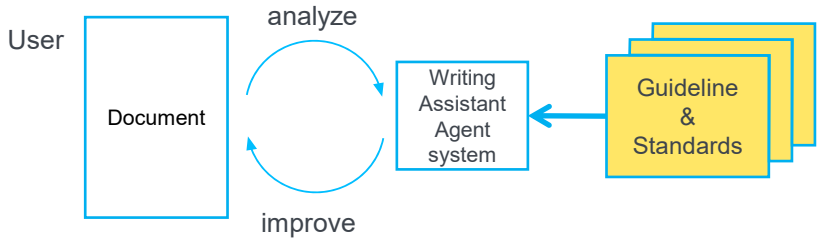
What if:
- Storage of data in structured Database
- Large Language Model(LLM) Retrieval
  - Guideline & Standards

2 Basics

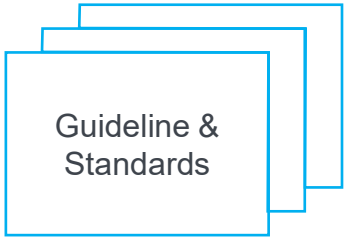
- Guideline & Standards
- State of the art

Basics

Guideline & Standards



- When editing **Production procedure document**:
  - GMP(Good Manufacturing Practices)
  - Ensures consistent quality control throughout the drug production process.



- Good Manufacturing Practices (GMP)
- Current Good Manufacturing Practice (CGMP)
- Good distribution practice(GDP)

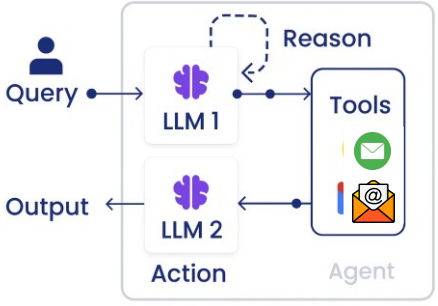
- EUROPEAN MEDICINES AGENCY (**EMA**)
- U.S. FOOD AND DRUG ADMINISTRATION (**FDA**)
- Medicines and Healthcare products Regulatory Agency (**MHRA, UK**)
- World Health Organization (**WHO**)

- 21 CFR Part 210+211 (**main regulations**) - GMP standards for **drug manufacturing**, **processing**, **packaging**, and **storage** in the U.S.
- 21 CFR Part 207 - Database registration of drug manufacturers for FDA **supply chain monitoring**
- 21 CFR Part 11 - Compliance **framework** for **electronic records** and **signatures** under FDA supervision
- 21 CFR Part 206 (suitable for **prescription drugs**) - **Distinguish prescription drugs** to prevent errors and enable quick recall
- ICH Q7 (adopted by the FDA) - GMP guidelines for **active pharmaceutical ingredient** (API) **manufacturing**

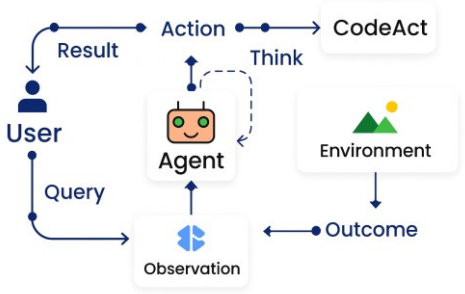
Basics

state of the art

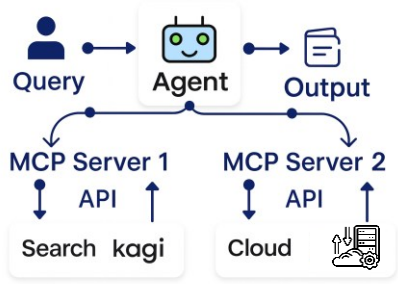
- Agent architecture



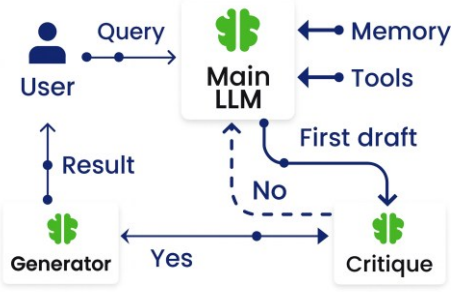
[1] ReAct: Synergizing Reasoning and Acting in Language Models



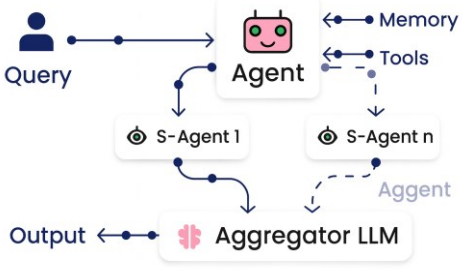
[2] Executable Code Actions Elicit Better LLM Agents



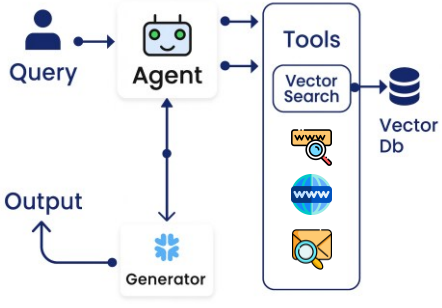
[3] A Review of Prominent Paradigms for LLM-Based Agents: Tool Use (Including RAG), Planning, and Feedback Learning (arXiv, 2024)



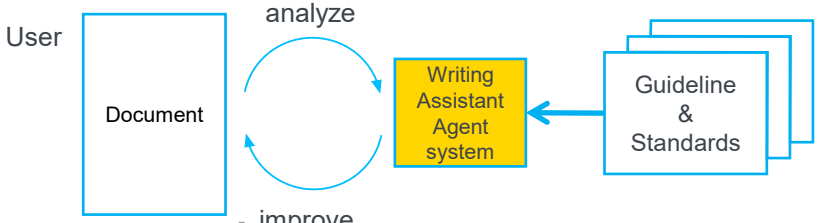
[4] Self-Reflection in LLM Agents: Effects on Problem-Solving Performance (arXiv, May 2024)



[5] A survey on LLM - based multi-agent systems: workflow, infrastructure and challenges



[6] **Agentic Retrieval- Augmented Generation: A Survey on Agentic RAG (arXiv, Jan 2025)**

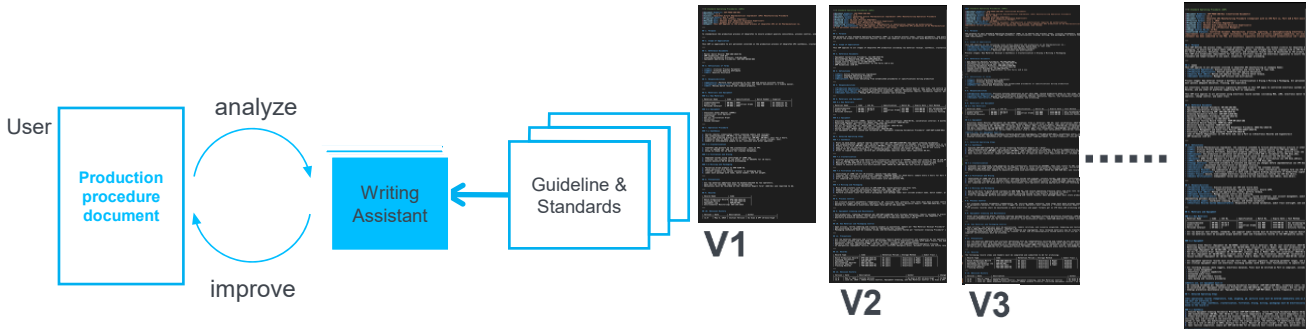


3 Conceptual Design

- Use case illustration
- System Disign
- Regulation document storage and utilization

Conceptual Design

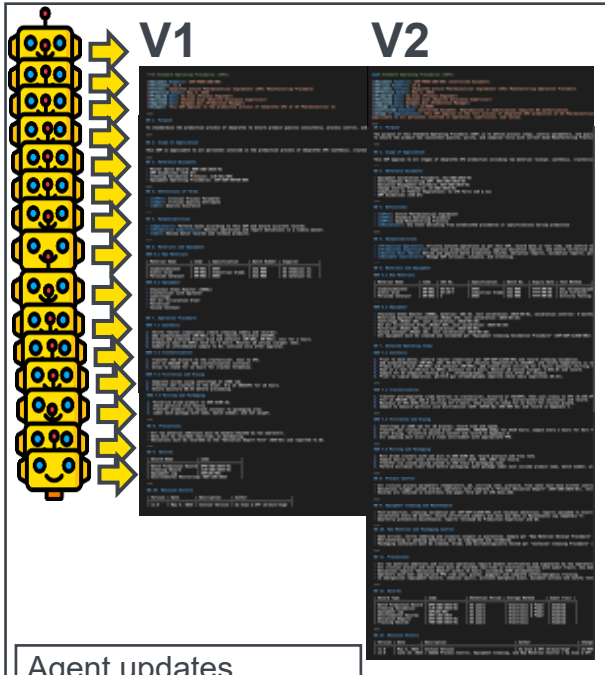
Use case illustration



diff. file

```
# 13. Revision History
```diff
- ## 10. Revision History
+ ## 13. Revision History
| Version | Date | Description | Author |
|---|---|---|---|
| v1.0 | May 5, 2025 | Initial Version | Sa Xiao & GPT o4-mini-high |
+ | v2.0 | June 16, 2025 | Added Process Control, Equipment Cleaning, and Raw Material Control | Sa Xiao & GPT o4-mini-high | CC-0002 |
+ | Environmental Records | ENV-LOG-2025 | ≥3 years | Electronic & Paper | Enabled |
+ | Deviation Reports | DEV-IBU-2025-01 | ≥5 years | Electronic | Enabled |
+ | Training Records | TRN-IBU-2025-01 | ≥5 years | Electronic | Enabled |
```
```

Agent generates a difference file



Agent updates the document file

Iterative process continue multiple times

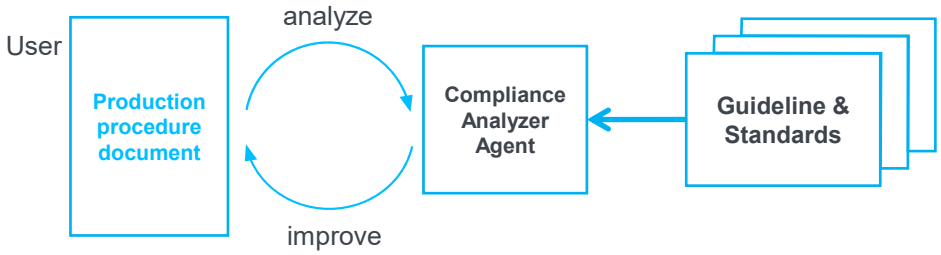
Conceptual Design

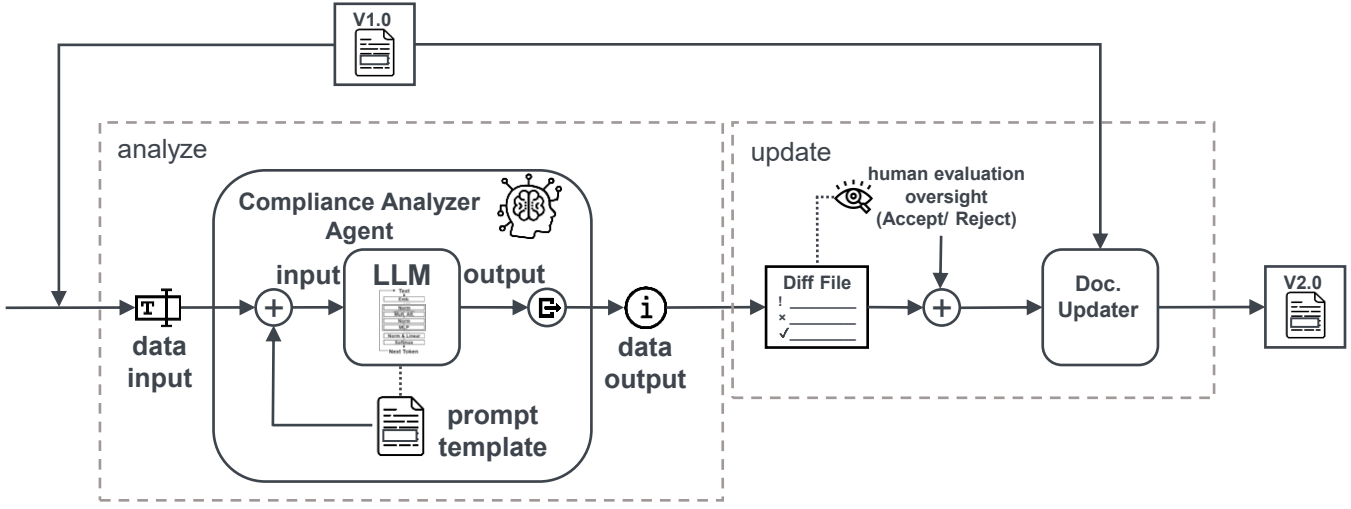
System Design

- Goal: target production line Ibuprofen

Questions

How did the system achieve **editing the Ibuprofen production procedure document?**



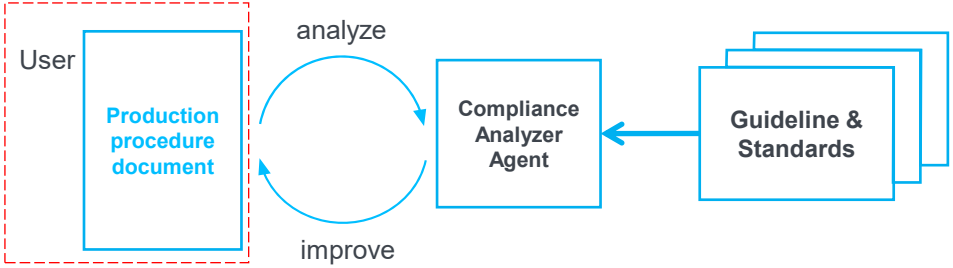


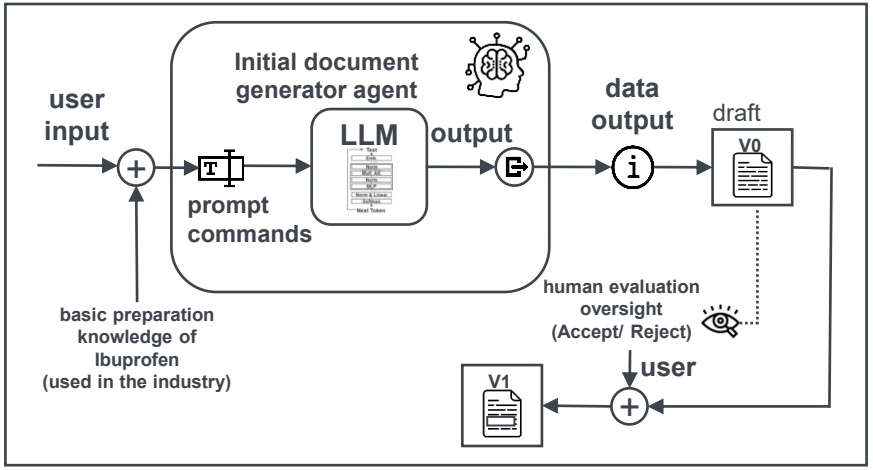
Conceptual Design

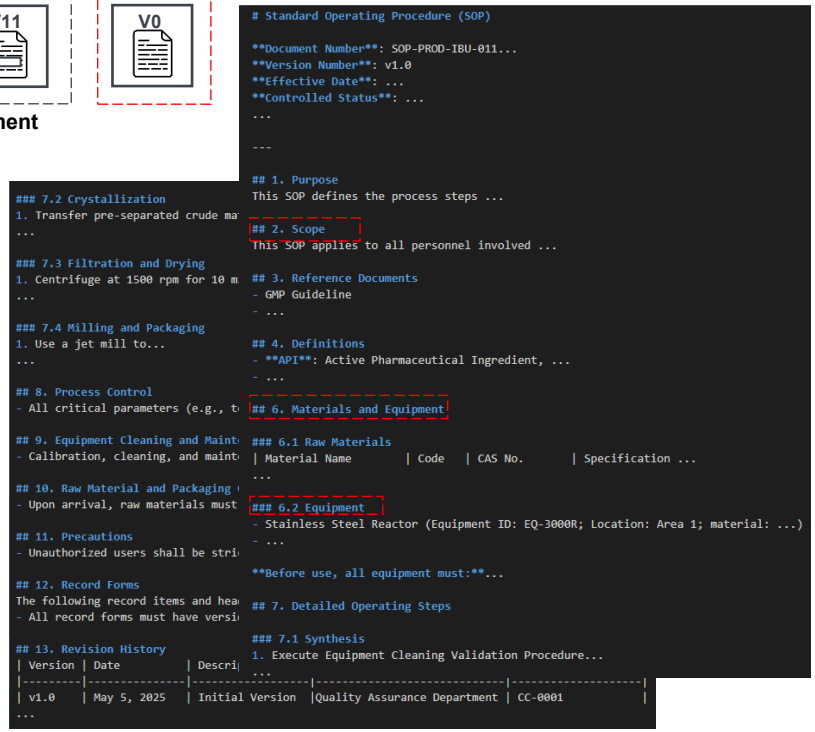
System Design

Questions

How to create the initial version of production process document?







```
# Standard Operating Procedure (SOP)

**Document Number**: SOP-PROD-IBU-011...
**Version Number**: v1.0
**Effective Date**: ...
**Controlled Status**: ...
...

---

## 1. Purpose
This SOP defines the process steps ...

## 2. Scope
This SOP applies to all personnel involved ...

## 3. Reference Documents
- GMP Guideline
- ...

## 4. Definitions
- **API**: Active Pharmaceutical Ingredient, ...
- ...

## 6. Materials and Equipment

### 6.1 Raw Materials
| Material Name        | Code   | CAS No.       | Specification ...
...

### 6.2 Equipment
- Stainless Steel Reactor (Equipment ID: EQ-3000R; Location: Area 1; material: ...)
- ...

**Before use, all equipment must:**...

## 7. Detailed Operating Steps

### 7.1 Synthesis
1. Execute Equipment Cleaning Validation Procedure...
...
```

```
### 7.2 Crystallization
1. Transfer pre-separated crude ma
...

### 7.3 Filtration and Drying
1. Centrifuge at 1500 rpm for 10 m
...

### 7.4 Milling and Packaging
1. Use a jet mill to...
...

## 8. Process Control
- All critical parameters (e.g., t

## 9. Equipment Cleaning and Maint
- Calibration, cleaning, and maint

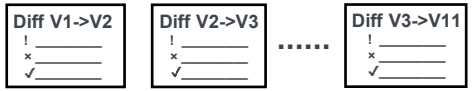
## 10. Raw Material and Packaging
- Upon arrival, raw materials must

## 11. Precautions
- Unauthorized users shall be stri

## 12. Record Forms
The following record items and hea
- All record forms must have versi

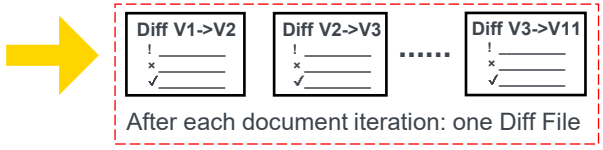
## 13. Revision History
| Version | Date          | Descri
|---------|---------------|-------------------|---------------------------|-------------------|
| v1.0    | May 5, 2025   | Initial Version  |Quality Assurance Department | CC-0001           |
...
```

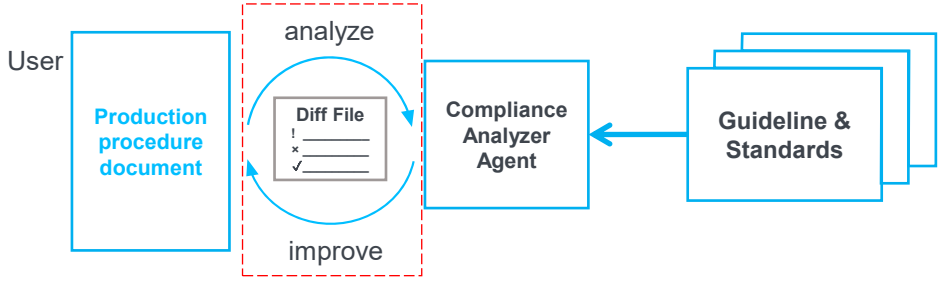
- What is **Diff File**?



Conceptual Design

System Design






```
# 0. Document Header Changes
```diff
- **Document Number**: SOP-PROD-IBU-001
+ **Document Number**: SOP-PROD-IBU-002
- **Version**: v1.0
+ **Version Number**: v2.0
- **Title**: Ibuprofen Active Pharmaceutical Ingredient (API) Manufacturing Procedure
+ **Document Title**: Ibuprofen Active Pharmaceutical Ingredient (API) Manufacturing Operation Procedure
  **Effective Date**: May 5, 2025
  **Prepared by**: Sa Xiao (Process Engineer)
  **Reviewed by**: Shouwen Xiao (Quality Assurance Supervisor)
  **Approved by**: Hongmei Qiu (Production Manager)
+ **Controlled Status**: Controlled document. Reproduction or modification requires QA authorization.
- **Scope**: This SOP applies to the production process of Ibuprofen API at XX Pharmaceutical Co.
+ **Scope of Application**: This SOP applies to the full process of Ibuprofen API production at XX Pharmaceutical Co., from raw material receipt, synthesis, crystallization, drying, milling, to packaging, applicable to all personnel involved in operation, supervision, and review.
```

change reason:
1. The scope description was incomplete: the original only stated "production process" without specific stages; added raw material receipt, synthesis, crystallization, drying, milling, and packaging.
2. The responsible positions were unclear: GMP documentation requires defining which roles the SOP applies to.
3. Lack of traceability: without clear definition, audit boundaries and responsibilities are ambiguous.

Regulatory references:
- 21 CFR §211.22(d): The quality unit shall ensure procedures are followed and apply to all relevant departments.
- 21 CFR §211.180(a): Records must be maintained and readily available, with clear responsibility, version, and effective date.
- 21 CFR §211.100(a): Procedures must have complete execution controls and defined personnel applicability.

Impact Analysis:
-Improved traceability and version control reduces the risk of using outdated or incorrect procedures, enhancing drug safety.
-clear scope and responsibilities decrease ambiguity and process deviations, supporting product stability.
-Better documentation control improves regulatory audit readiness, reducing potential economic losses from non-compliance.

Diff File

Diff code
- **Differences** between two Versions

```diff
+ add
  original test
- delete
```

! **Change reason:**
- Why **modify** the **content** of this module
- Why **update** the output **target content**

× **Regulatory reference:**
- Based on which **specific provisions** in Guideline doc.

✓ **Impact analysis:**
- what **risks** will be **avoided**
- what **costs** will be **reduced**
- what **profit** will be **increased**
```

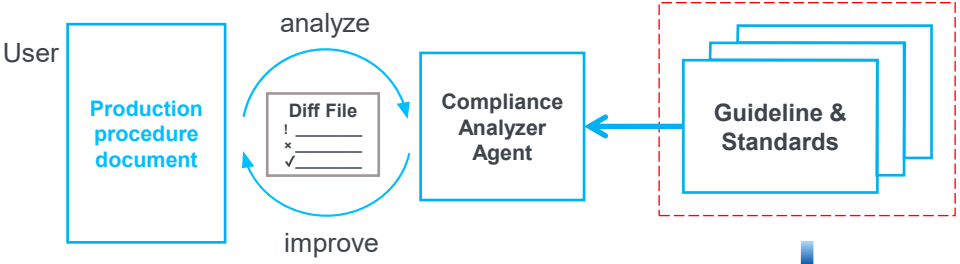
Conceptual Design

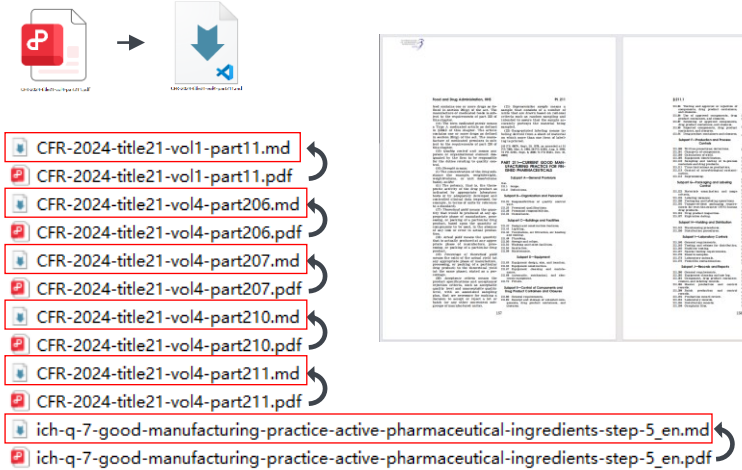
Regulation document storage and utilization

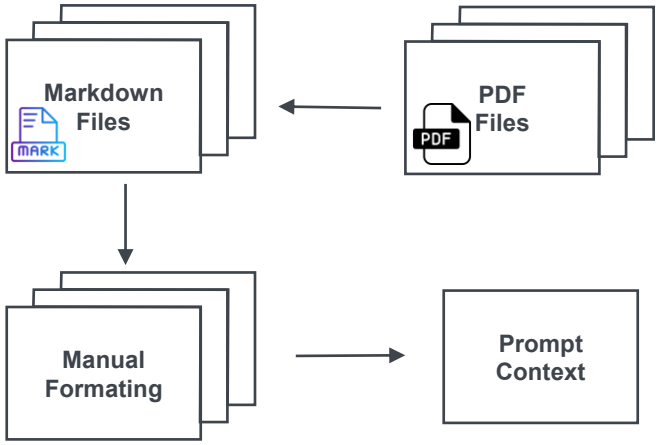
Questions

How are regulatory documents stored and utilized?

- Information **Pre- Processing**
- Guideline doc. in pdf -> Markdown
- 6 Guideline doc.





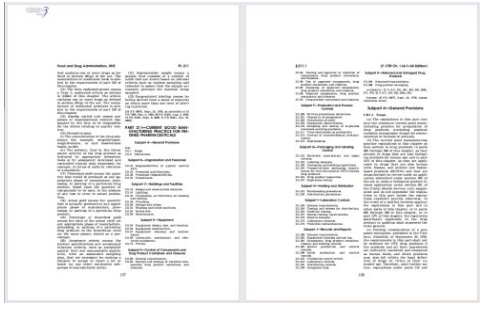


Conceptual Design

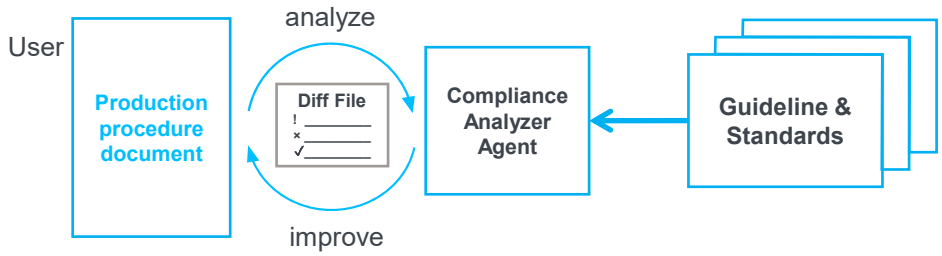
Questions

How are regulatory documents stored and utilized?

- Information **Pre- Processing**
- Guideline doc. in pdf -> Markdown
- 6 Guideline doc.



PDF file




```
# Food and Drug Administration, HHS Pt. 211 – Current Good Manufacturing Practice for Finished Pharmaceuticals

*21 CFR Ch. I (4-1-24 Edition)*

---

## Table of Cor

- [Subpart A –
  - [§ 211.1 Sc
  - [§ 211.3 De
- [Subpart B –
  - [§ 211.22 F
  - [§ 211.25 F
  - [§ 211.28 F
  - [§ 211.34 C
- [Subpart C –
  - [§ 211.42 C
  - [§ 211.44 C
  - [§ 211.46 V
  - [§ 211.48 F
  - [§ 211.50 S
  - [§ 211.52 V
  - [§ 211.56 S
  - [§ 211.58 M
- [Subpart D –
  - [§ 211.63 E
  - [§ 211.65 E
  - [§ 211.67 E
  - [§ 211.68 A
  - [§ 211.72 E
- [Subpart E –
(#subpart-e--co
  - [§ 211.80 C
  - [§ 211.82 E
(#-21182-rece
- [Subpart F –
  - [§ 211.100
  - [§ 211.101
  - [§ 211.103
  - [§ 211.105
  - [§ 211.110
(#-211110-sam
  - [§ 211.111
  - [§ 211.113
  - [§ 211.115
- [Subpart G –
  - [§ 211.122

---

## Subpart A – General Provisions

### § 211.1 Scope

(a) The regulations in this part contain the minimum current good manufacturing practice for the preparation of drug products (excluding positron emission tomography drugs) for administration to humans or animals.

(b) The current good manufacturing practice regulations in this chapter as they pertain to drug products; in parts 600 through 680, as they pertain to drugs that are also biological products for human use; and in part 1271, as they are applicable to drugs that are also human cells, tissues, and cellular and tissue-based products (HCT/Ps) and that are drugs (subject to review under an application submitted under section 505 of the act or under a biological product license application under section 351 of the Public Health Service Act) supplement and do not supersede the regulations in this part unless the regulations explicitly provide otherwise. In the event of a conflict between applicable regulations in this part and in other parts of this chapter, or in parts 600 through 680, or in part 1271, the regulation specifically applicable to the drug product in question shall supersede the more general.

(c) Pending consideration of a proposed exemption published in the Federal Register of September 29, 1978, the requirements in this part shall not be enforced for OTC drug products if the products and all their ingredients are ordinarily marketed and consumed as human foods and which products may also fall within the legal definition of drugs by virtue of their intended use.

*Source: [43 FR 45077, Sept. 29, 1978, as amended at 62 FR 66522, Dec. 19, 1997; 69 FR 29828, May 25, 2004; 74 FR 65431, Dec. 10, 2009; 80 FR 56168, Sept. 17, 2015]*

### § 211.3 Definitions

The definitions set forth in § 210.3 of this chapter apply in this part.

---

## Subpart B – Organization and Personnel

### § 211.22 Responsibilities of Quality Control Unit

(a) There shall be a quality control unit with the responsibility and authority to approve or reject all components, drug product containers, closures, in-process materials, packaging materials, labeling, and drug products, and to review production records to assure that errors have been fully investigated. This unit is responsible for approving or rejecting drug products manufactured, processed, packed, or held under contract by another company.

(b) Adequate laboratory facilities for testing and approval (or rejection) of components, containers, closures, packaging
```

processed regulatory documents - converted Markdown format

4 Test

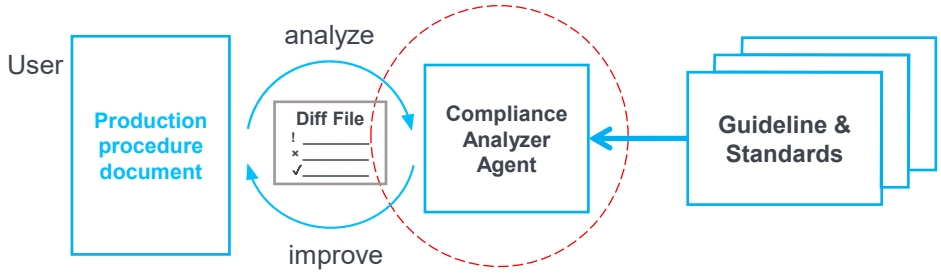
- Output discovery
- Fine-tuning model

Test

Output discovery

Questions

What can the Compliance Analyzer Agent actually do?



Data Operations

- Add new item
- Modify item
  Detailed/enriched existing entries
- Delete item
- Add new **Section**

```
+ 4. Sample to measure particle size distribution (d50: 20–50 µm; d90 ≤80 µm) and record
in Appendix C.
```

adding missing content

```
## 7.2 Crystallization
```diff
- 1. Transfer the mixture to the crystallizer. Cool to 5°C.

+ 1. Transfer pre-separated crude material to crystallizer, dissolve at 35–40°C, then cool
slowly to 5°C (0.2–0.5°C/min), recording temperature curve.
```

adjusting temperature ranges/ details

```
# 5. Responsibilities
```diff
- **Operator**: Perform tasks according to this SOP and ensure accurate records.
```

doesn't meet compliance standards

```
```diff
+ # 9. Equipment Cleaning and Maintenance
+ - Post-production, cleaning validation per SOP-EQP-CLEAN-001 with residual detection;
reports uploaded to electronic system.
```

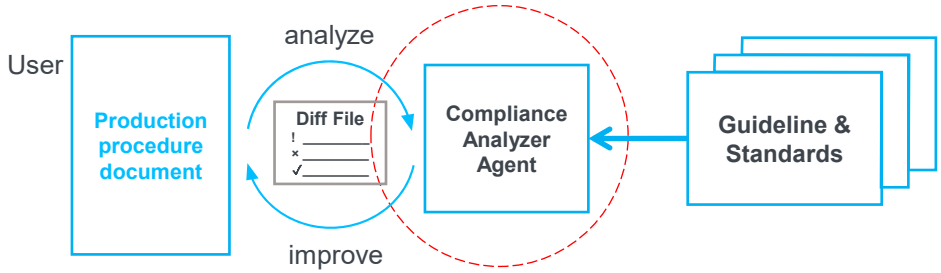
adding a missing procedure
```

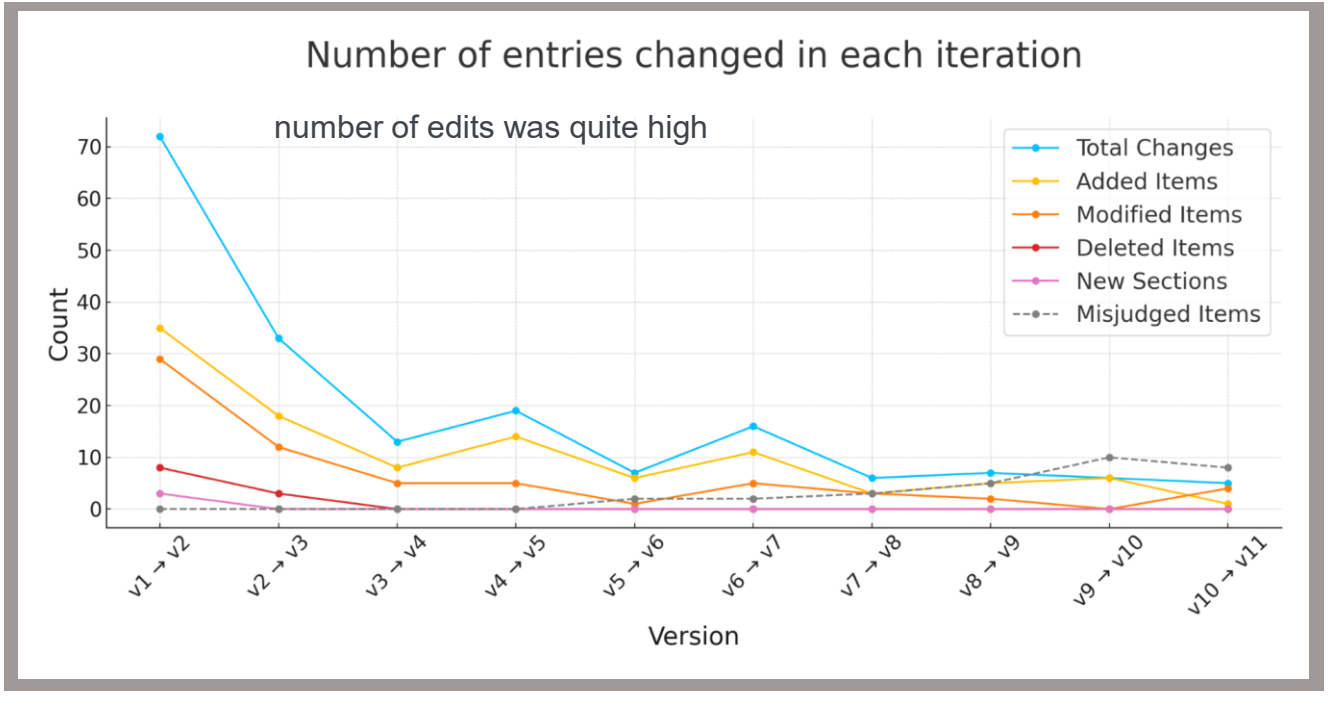
Test

Output discovery - Test Results

Questions

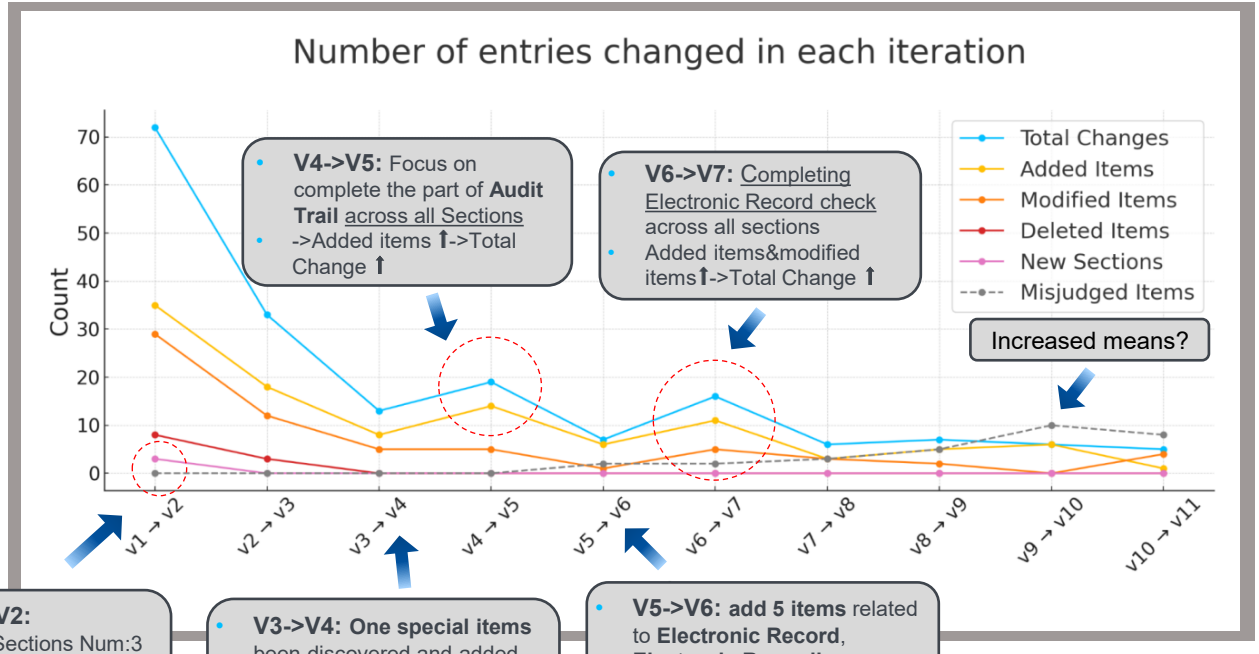
How the document evolved over time?





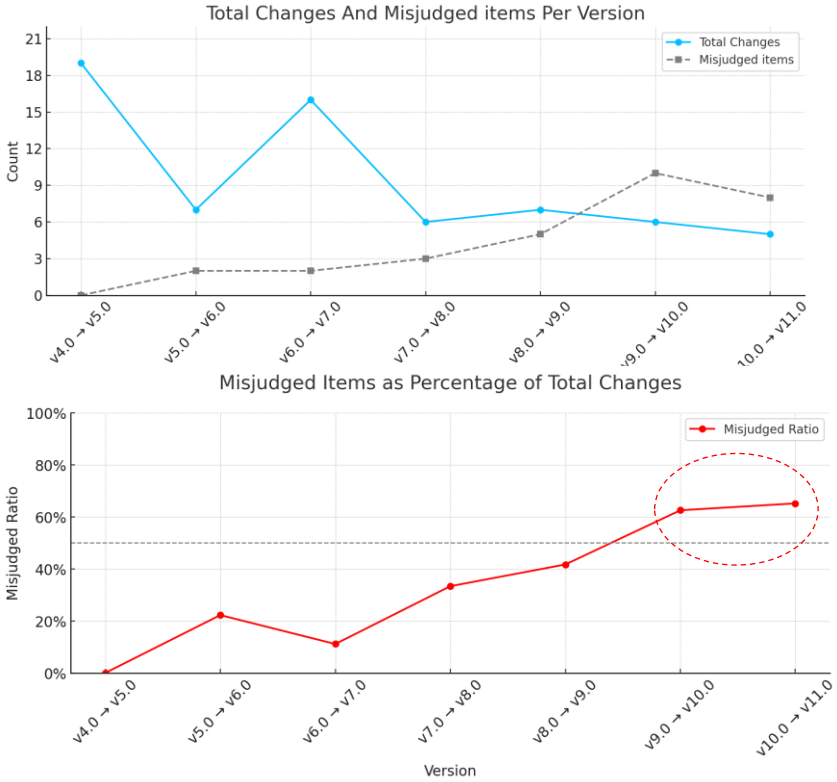
Test

Output discovery - more insights behind the changes



Test

Output discovery - Overediting



Misjudged items

- **Unnecessary**: Change a sentence but ended up adding the unnecessary content
- **Repeatedly**: Repeatedly modifying items: have already been modified; added content being completely identical to the original content.

$$Misjudged\ Ratio = \frac{Misjudged\ items}{Total\ Changes}$$

Misjudged ratio ≥ 50%

Half of the edits made were **unnecessary** after the whole process converged.

Insights

-> avoid this overediting **by early stop**

5 Model Fine-tuning

- Dataset and Model
- Training process
- Validation Result

Model Fine-tuning

Questions

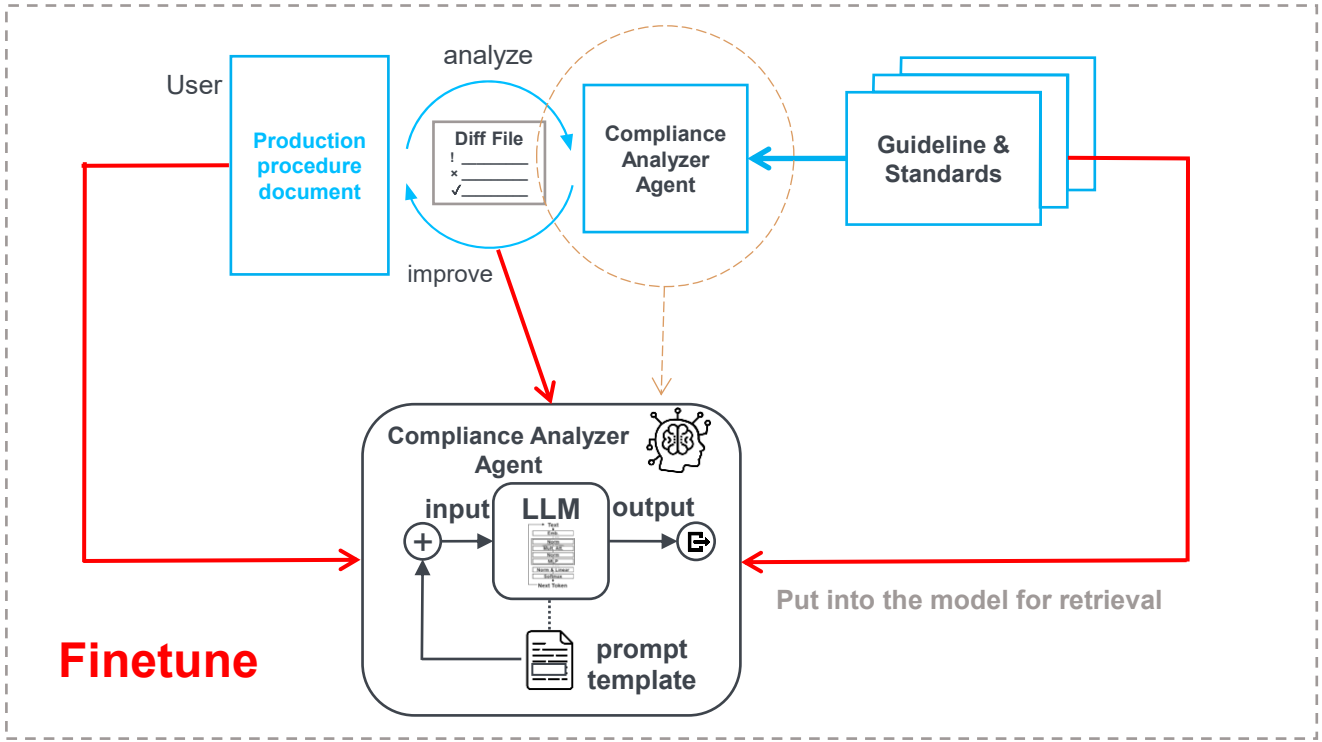
How to improve task specific capability of Agent?

Collect:

- document inputs
- the diff output

->feed them back as training data

Fine-tune the LLM directly



Model Fine-tuning

Dataset and Model

- Build my dataset

  Use data generated from my project

- Supervised fine-tuning

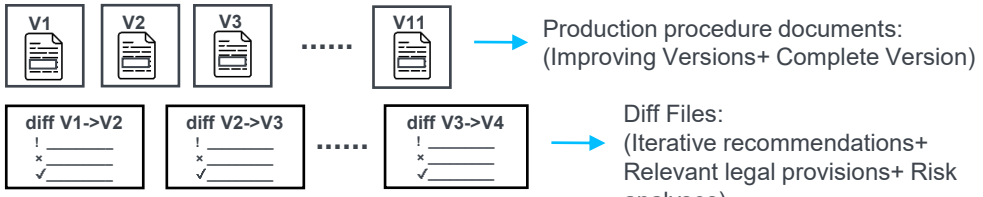
Dataset design

- JSONL format
- QA Pairs design
  - 101 QA Pairs
    - 44519 words
    - 83 pages

    (Font style: Consolas, 10.5)

Questions

What the Fine-tune file format look like?




```
{ "messages": [{"role": "user","content": "Answer the question in accordance with spe
{ "messages": [{"role": "user","content": "Answer the question in accordance with spe
{ "messages": [{"role": "user","content": "Why are the sections “Process Control,” “E
{ "messages": [{"role": "user","content": "Answer based on specific legal provisions:
```

“role”s

“User”
“content”:
(user prompt)

- **Who** are you?
- Whats your **task**?
- Scope of **legal provisions** to be searched?
- How should you provide answers?(output format description)
- **One section** production procedure document **text** (that you need to deal with)

“Assistant”
“content”:
(assistant response)

- Diff Files (for the given One section)

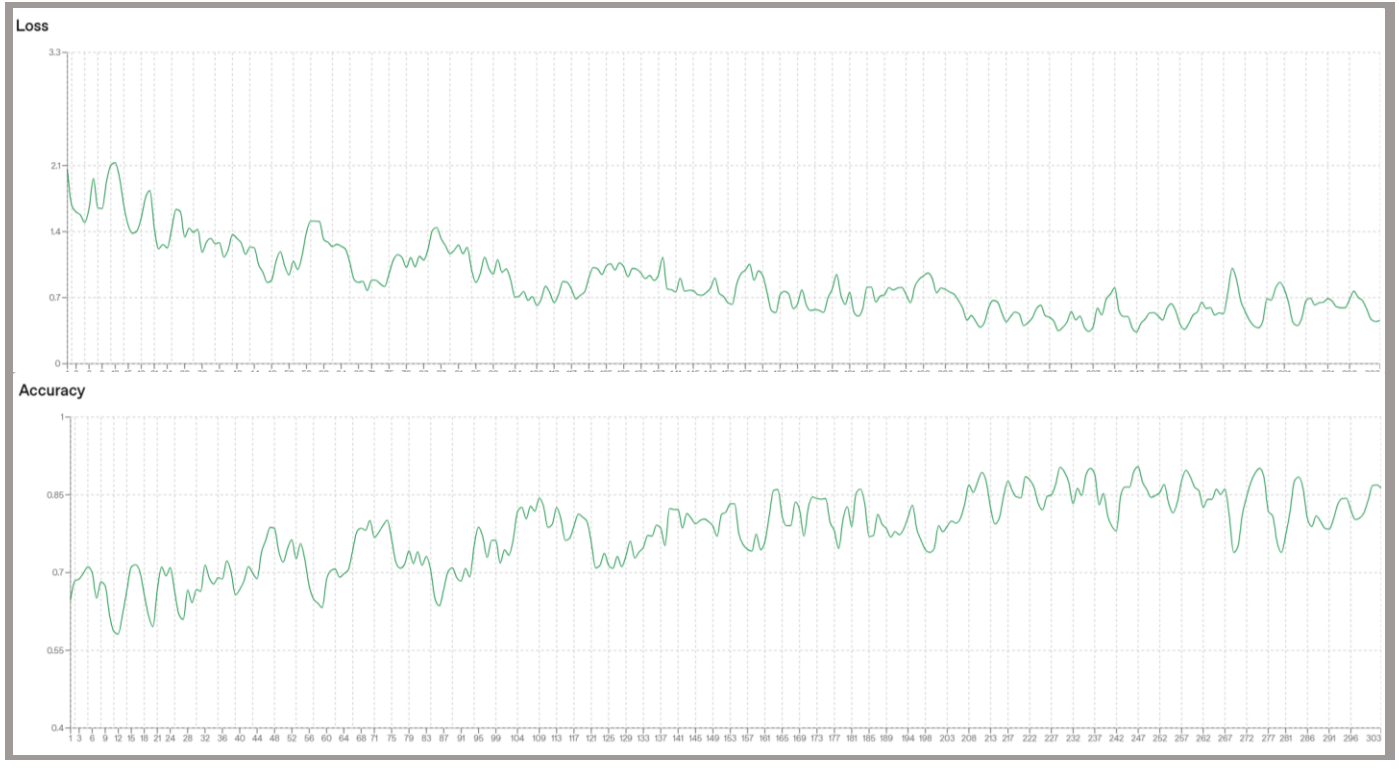
Model Selection

- gpt 4o - mini (200,000 tokens per min, TPM)
- gpt 4o: limit (30,000 tokens per min, TPM)

-> gpt 4o - mini

Model Fine-tuning

Training process



- **Loss**
  step 12 ->2.130
  step 269 ->1.011
  step 275 ->0.380

- **Accuracy**
  step 12 ->0.581
  step 269 ->0.739
  step 275 ->0.901

- **Smoothing -> 0.5**
  - Loss decreases and approaches good convergence
  - Accuracy increases, is learning effectively and performing better

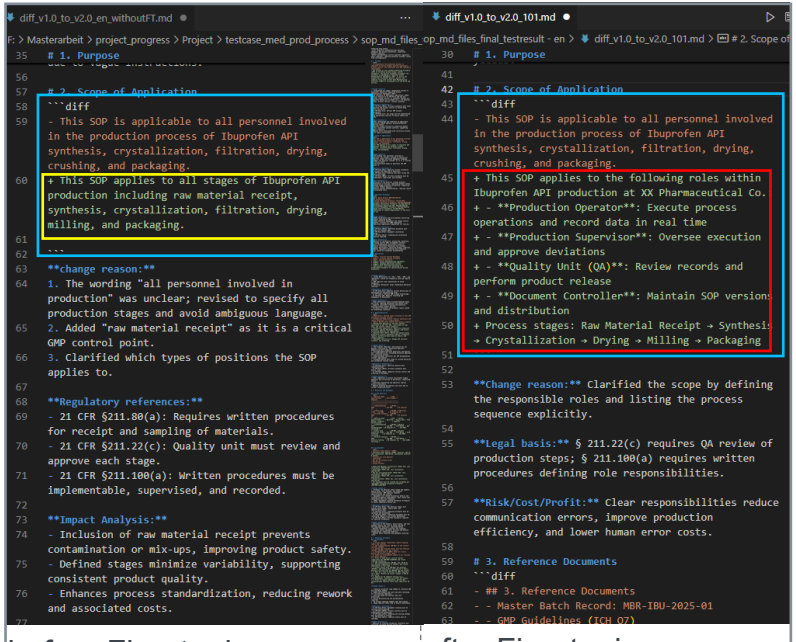
Model Fine-tuning

Validation Result - Effect of fine-tuning

Number of items changed in the first iteration
(v1->v2, before vs. after fine-tuning)

	Added items	modified items	Deleted items	Added Sections	Total changes
before Fine-tuning	36	32	8	3	76
after Fine-tuning	31	33	8	3	75

Quantity Level
×

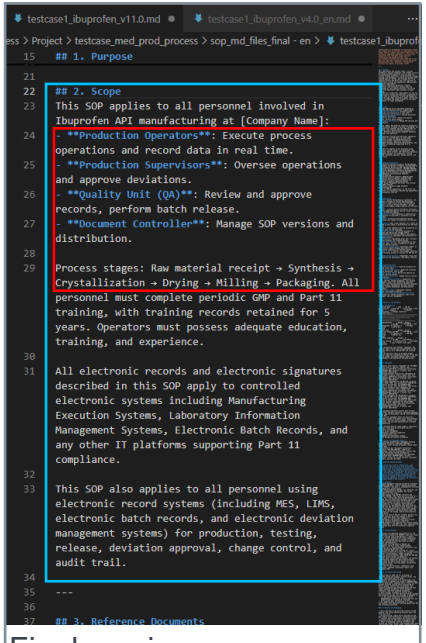


before Fine-tuning

vague and incomplete

after Fine-tuning

much more accurate and detailed result



Final version

Quality Level
✓

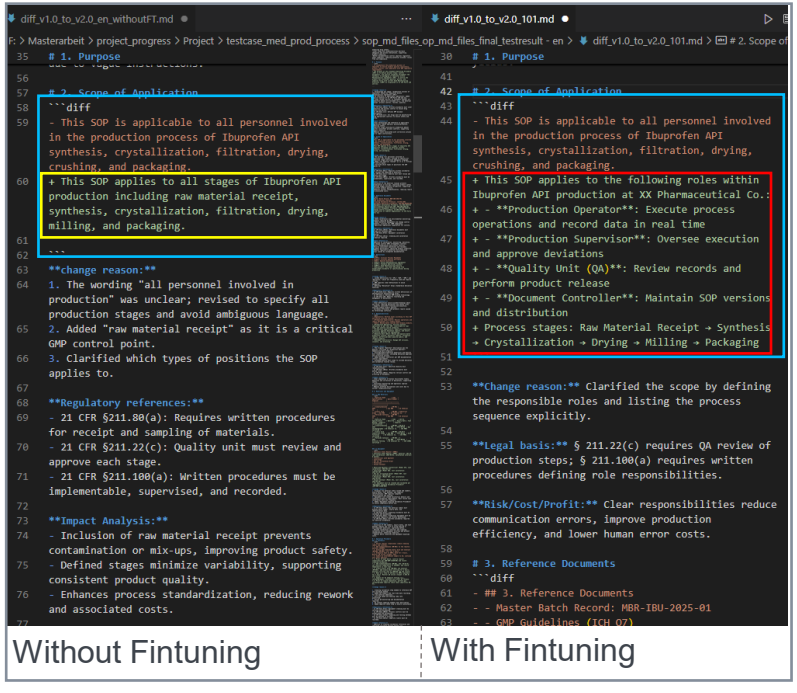
Model Fine-tuning

Validation Result - Summarizes

- **Technical** verification of LLM output level
  - Training has effect:
    - **loss decreased**

- Validation of **document** quality
  - text quality **difficult to quantify**
  - need **expert knowledge**
    - but we see effects

  -> result **more comprehensive**



6 Summary & Outlook

Summary and Outlook

Summary

- Designed a system for **improving production documents**

  (Combines regulatory documents & industrial standards)

- Ensures **version traceability** & improves **document precision**
- Reduces **manual editing** & enhances **productivity**

Limitation & Future Work

- need **expert input** to validate the results
- **Final approval** still depends on **human expertise**

  (even system improves technical quality)

University of Stuttgart
Institut of Industrial Automation
and Software Engineering

Thank you!

Sa Xiao

e-mail st181115@stud.uni-stuttgart.de
phone +49 (0) 711 685-
fax +49 (0) 711 685-

University of Stuttgart

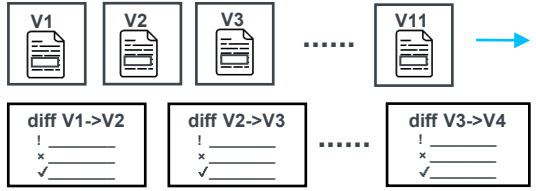
Quelle

- https://arxiv.org/abs/2210.03629
- https://arxiv.org/abs/2402.01030
- https://arxiv.org/abs/2406.05804
- https://arxiv.org/abs/2405.06682
- https://link.springer.com/article/10.1007/s44336-024-00009-2
- https://arxiv.org/abs/2501.09136
- https://platform.openai.com/settings/organization/limits

Model Fine-tuning

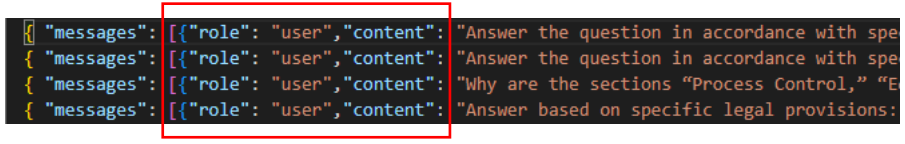
Dataset

- Use data generated from my project
- Build my dataset



Questions

What Data I obtained in my project?



“role”

- Supervised fine-tuning (为什么用这个方法？)

Dataset design

- JSONL format
- QA Pairs design

“User”
“content”:
(requirement, prompt)
- Your **role**?
- Your **task**?
- Scope of **legal provisions** to be searched?
- How should you provide answers?(output format description)
- **One section** production procedure document **text** (that you need to deal with)

“Assistant”
“content”:
(desired output of the model)
- Diff Files (for the given One section)

101 QA Pairs

31 QA Pairs -> redesign

“User”
“content”:
(requirement, prompt)
- summarize the phenomenon
- “Answer questions using legal provisions:”
- “Why is it necessary/important to add/delete those information if (this phenomenon occurs)?

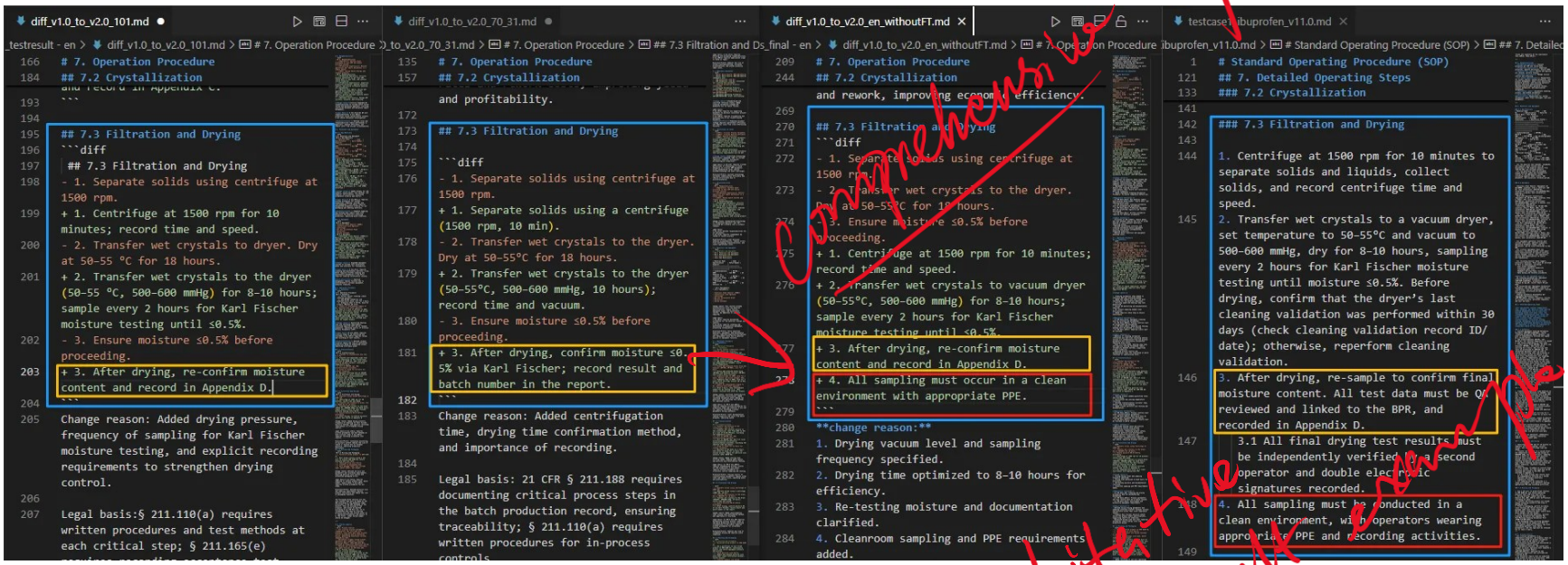
“Assistant”
“content”:
(desired output of the model)
- Relevant legal provisions

70 QA Pairs +31 redesign QA Pairs

Model Fine-tuning

Evaluation

Fine-tuning experimental phenomena - 31+70 QA pair



- 要出统计数据

Summary and future work

Experimental details and findings

Summary:

- If some items in the dataset are missing Risk/Cost/Profit, even if the playground assistant prompt has multiple cases with Risk/Cost/Profit, the output on the right side will not produce Risk/Cost/Profit. Next, requesting the output of Risk/Cost/Profit will produce the desired result.
- The results cannot supplement missing sections. If you input whether to supplement new sections, it will increase
- The model cannot distinguish between the “-” and “ -” in the ```diff``` file. Therefore, special attention should be paid to distinguishing them in the future, such as using “@@@” instead of the original database's “ -”
- The prompt is quite critical: “When designing the 30 QA pairs: Why must stirring speed and heating rate be quantified if they are not quantified in the original version?” If changed to “when,” it might cause ambiguity, leading to poor training results.
- Based on the training curve of the fine-tuned model, the dataset needs to be expanded.