Nvidia llamaindex contest

Chi-Yuan Lee
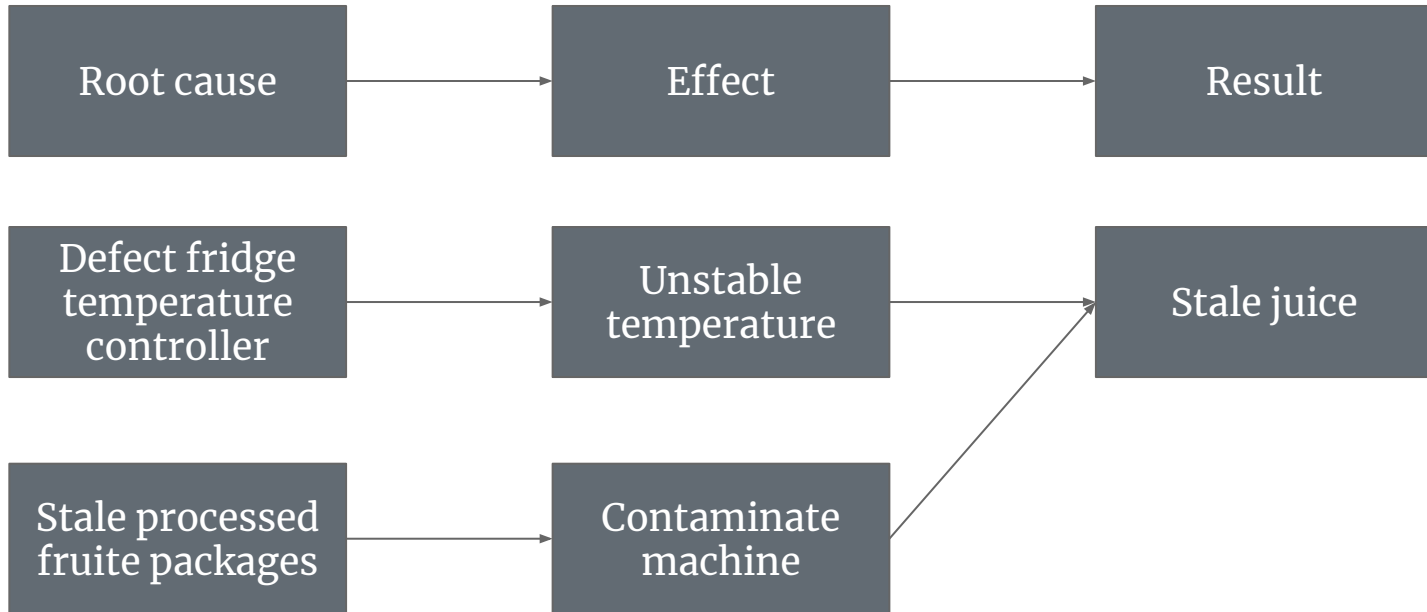
Motivation

Root cause → Effect → Result

1. The causality can be complex.
2. Each elements can be more than single element.
3. The causality can be relevant to the current events and repeated over time.
4. It might be inefficient for the user to search through the historical data.

Case study

| Root cause | → | Effect | → | Result |

| Defect fridge temperature controller | → | Unstable temperature | | Stale juice |

| Stale processed fruite packages | → | Contaminate machine | | |

Architecture

```
┌─────────────────┐      ┌─────────────────┐      ┌─────────────────────┐
│                 │      │                 │      │    Nvidia NIM       │
│    Raw data     │─────▶│ Nvidia Curator  │─────▶│  (preprocessing and │
│                 │      │ (preprocessing) │      │   summarization)    │
└─────────────────┘      └─────────────────┘      └─────────────────────┘
                                                             │
                                                             ▼
┌─────────────────┐      ┌─────────────────┐      ┌─────────────────────┐
│                 │      │   Nvidia NIM    │      │                     │
│     Users       │◀─────│  (Ranking and   │◀─────│   Vector database   │
│                 │      │    Chatbot)     │      │                     │
└─────────────────┘      └─────────────────┘      └─────────────────────┘
```
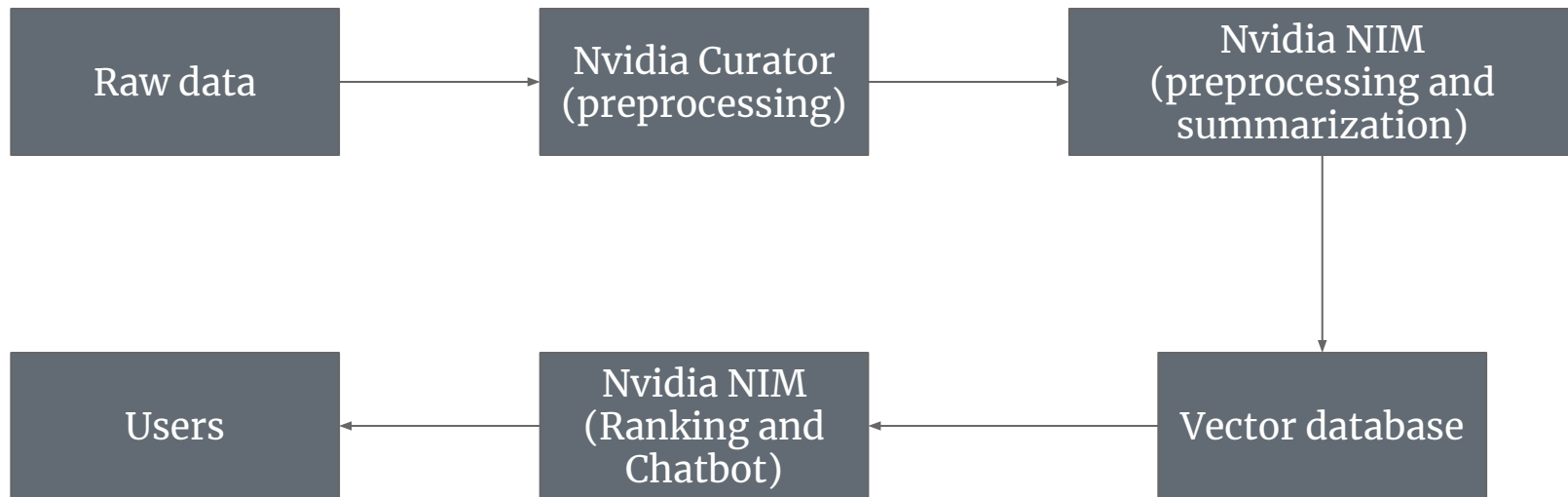
Challenges and future development

Challenges:
1. Real world practical scenario.
2. Libraries do not support to some specific environment (ex. MacOS).

Future Development:
1. Integrate more Nvidia eco-system libraries.
2. Fine-tune the LLM for more accurate and specific response.
3. Better UI design.