

Benchmarking for Every Methods Including Pros and Cons

Results:

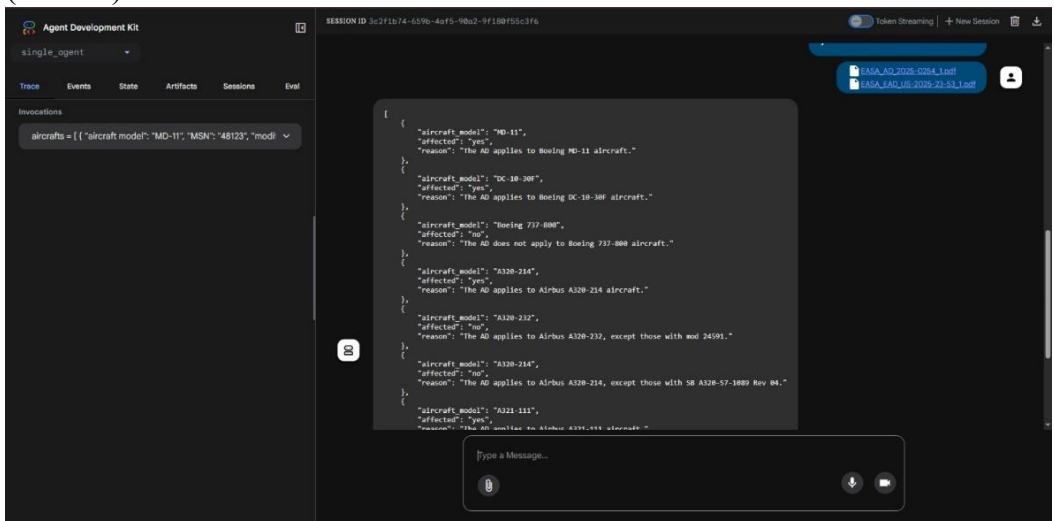
- Rule-Based:
Score: 5/5
Accuracy: 100%
Reason: Rule Based Approach manually determines the rules that applicable to the ADs. This approach is also used as a standard for this problem
Pros: Perfect accuracy (100%), Extremely fast execution, No external API dependencies or costs, Works offline without internet connection
Cons: Requires manual rule creation for each new AD, Time-consuming initial setup, Needs developer intervention for updates, Difficult to scale with hundreds of ADs, No natural language understanding, Code changes required for rule modifications, Cannot adapt to document variations automatically
- Single Agent:
Score: 4/5
Accuracy: 80% (Based On 5x testing)
Reason: The model Failed to understand the exceptions in the ADs documents
Pros: Simple architecture, Handles natural language documents, Can process new ADs without code changes, Lower token usage compared to multi-agent makes this approach more affordable
Cons: Inconsistent accuracy (80%), Struggles with complex exception logic, No built-in validation or quality control, Requires API access and authentication
- Multi Agent:,
Score: 5/5
Accuracy: 100% (Based On 5x testing)
Reason: Each of the agent have specific task, that makes each agent only focuses on each assigned task.
Pros: Perfect accuracy (100%), Self-validating with compliance reviewer, Automatically extracts rules from documents, Adapts to new ADs without code changes,
Cons: Slower execution due to multiple agent interactions, Higher token usage and API costs, Requires API access and authentication, Overkill for simple, well-defined rules

Proofs:

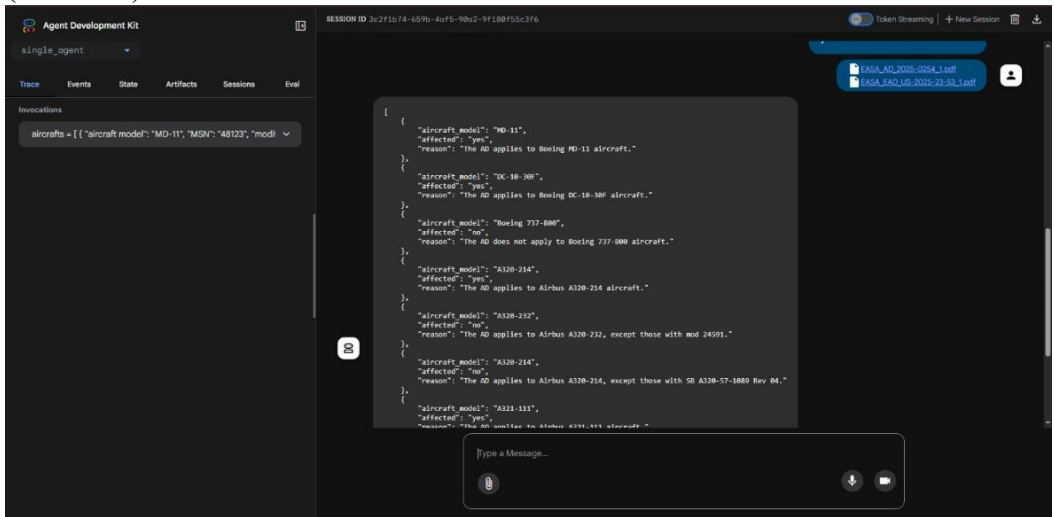
- Rule Based:

```
Aircraft MD-11 is affected by FAA AD 2025-23-53
Aircraft DC-10-30F is affected by FAA AD 2025-23-53
Aircraft Boeing 737-800 is not affected by any ADs
Aircraft A320-214 is affected by EASA AD 2025-0254
Aircraft A320-232 with modification mod 24591 (production) is not affected by EASA AD 2025-0254
Aircraft A320-214 with modification SB A320-57-1089 Rev 04 is not affected by EASA AD 2025-0254
Aircraft A321-111 is affected by EASA AD 2025-0254
Aircraft A321-112 is not affected by any ADs
Aircraft A319-100 is not affected by any ADs
Aircraft MD-10-10F is affected by FAA AD 2025-23-53
```

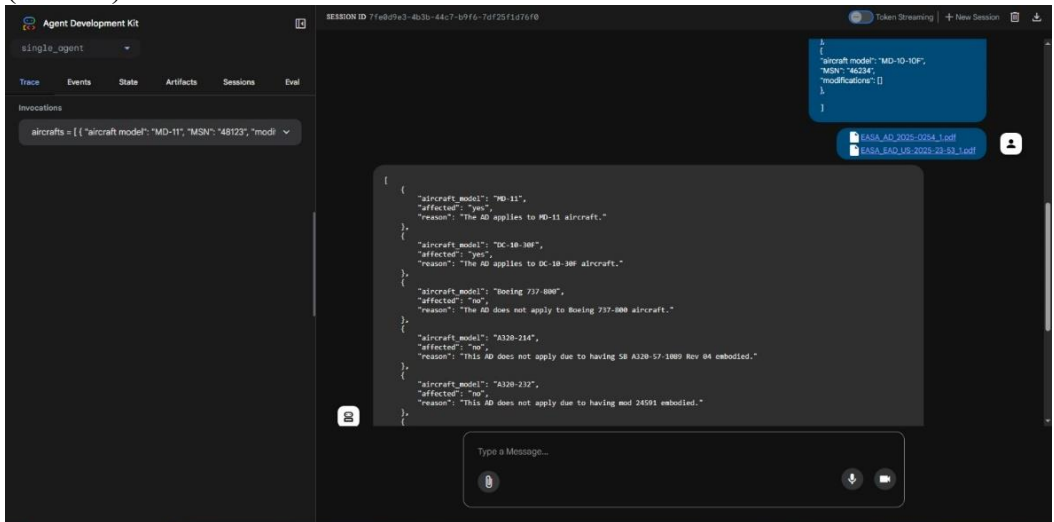
- Single Agents:
(Correct)



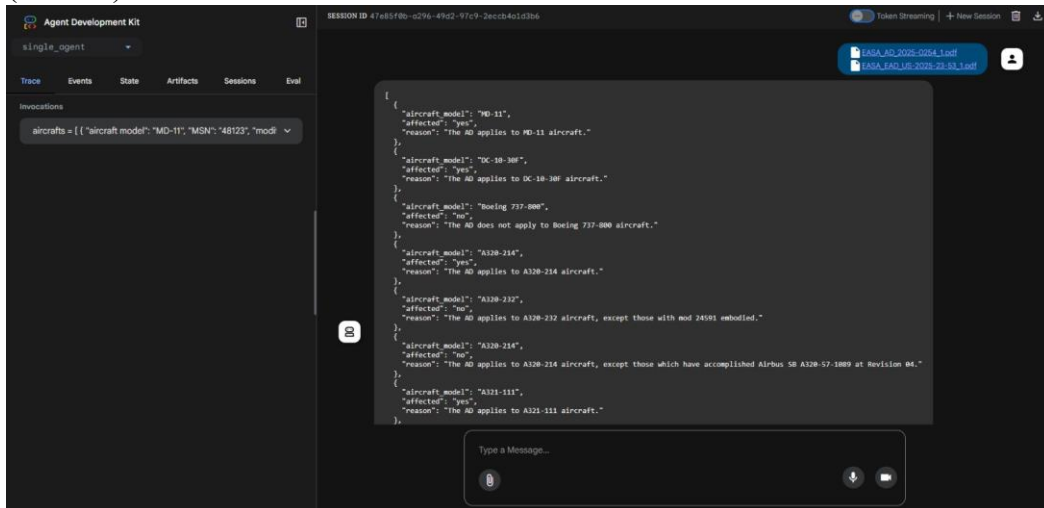
(Incorrect)



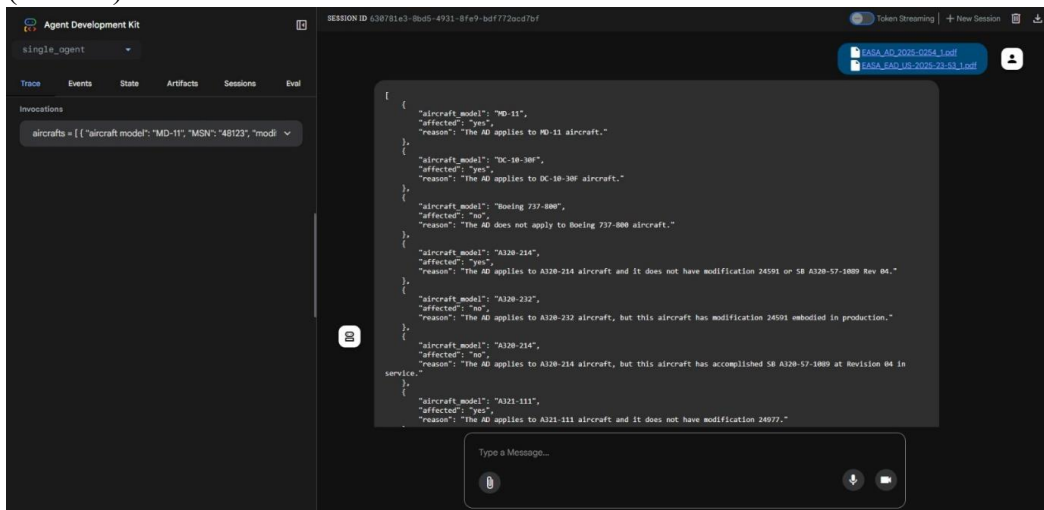
(Correct)



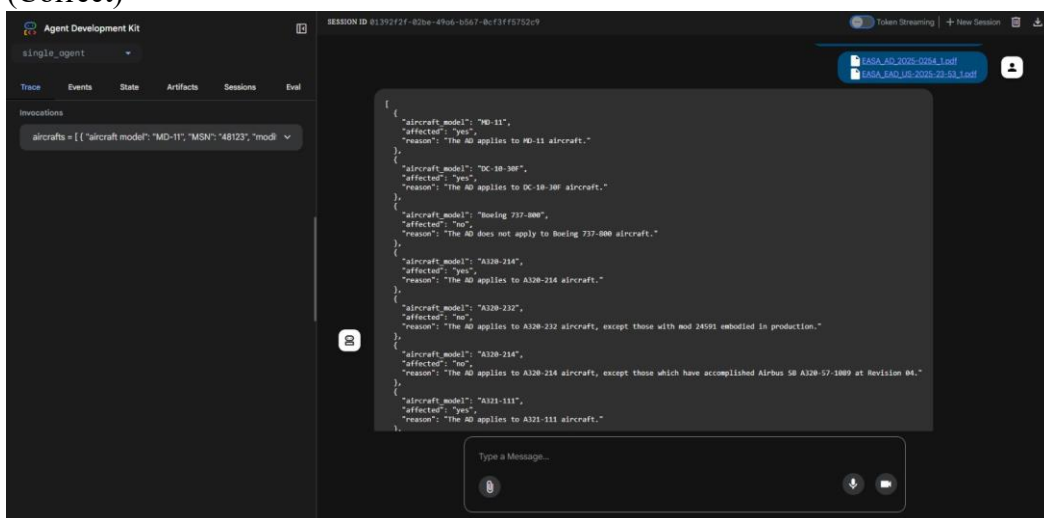
(Correct)



(Correct)



- Multi Agent:
(Correct)



(Correct)

The screenshot shows the Agent Development Kit interface with the following details:

- Left Panel:** "multi_agents" dropdown, tabs for Trace, Events, State, Artifacts, Sessions, Eval. The "Invocations" section shows a dropdown menu with the text: `aircrafts = [{ "aircraft_model": "MD-11", "MSN": "48123", "mod":`
- Right Panel:** Session ID `57780a32-f5a8-4d76-9716-00e072d38980`. It displays a JSON array of objects, each representing an aircraft model and its compliance status with various ADs. The JSON is as follows:

```
{
  "aircraft_model": "MD-11",
  "affected": "yes",
  "reason": "The AD applies to Boeing MD-11 aircraft."
},
{
  "aircraft_model": "DC-10-30F",
  "affected": "yes",
  "reason": "The AD applies to Boeing DC-10-30F aircraft."
},
{
  "aircraft_model": "Boeing 737-800",
  "affected": "no",
  "reason": "The AD does not apply to Boeing 737-800 aircraft."
},
{
  "aircraft_model": "A320-214",
  "affected": "yes",
  "reason": "The EASA AD 2025-0254 applies to A320-214 aircraft."
},
{
  "aircraft_model": "A320-232",
  "affected": "no",
  "reason": "The EASA AD 2025-0254 applies to A320-232 aircraft except those with mod 24591 embodied in production."
},
{
  "aircraft_model": "A320-214",
  "affected": "no",
  "reason": "The EASA AD 2025-0254 applies to A320-214 aircraft except those which have accomplished Airbus SB A320-57-1889 at Revision 04 in service."
},
{
  "aircraft_model": "A321-111",
  "affected": "no",
  "reason": ""
}
```

(Correct)

The screenshot shows the Agent Development Kit interface with the following details:

- Left Panel:** "multi_agents" dropdown, tabs for Trace, Events, State, Artifacts, Sessions, Eval. The "Invocations" section shows a dropdown menu with the text: `aircrafts = [{ "aircraft_model": "MD-11", "MSN": "48123", "mod":`
- Right Panel:** Session ID `8999878d-87c5-46d3-9143-0be1b4c1edcd`. It displays a JSON array of objects, each representing an aircraft model and its compliance status with various ADs. The JSON is as follows:

```
{
  "aircraft_model": "MD-11",
  "affected": "yes",
  "reason": "The AD applies to Boeing MD-11 aircraft."
},
{
  "aircraft_model": "DC-10-30F",
  "affected": "yes",
  "reason": "The AD applies to Boeing DC-10-30F aircraft."
},
{
  "aircraft_model": "Boeing 737-800",
  "affected": "no",
  "reason": "The AD does not apply to Boeing 737-800 aircraft."
},
{
  "aircraft_model": "A320-214",
  "affected": "yes",
  "reason": "The EASA AD 2025-0254 applies to A320-214 aircraft."
},
{
  "aircraft_model": "A320-232",
  "affected": "no",
  "reason": "The EASA AD 2025-0254 does not apply to aircraft with mod 24591."
},
{
  "aircraft_model": "A320-214",
  "affected": "no",
  "reason": "The EASA AD 2025-0254 does not apply to aircraft which have accomplished Airbus SB A320-57-1889 at Revision 04."
},
{
  "aircraft_model": "A321-111",
  "affected": "no",
  "reason": ""
}
```

(Correct)

The screenshot shows the Agent Development Kit interface with the following details:

- Left Panel:** "multi_agents" dropdown, tabs for Trace, Events, State, Artifacts, Sessions, Eval. The "Invocations" section shows a dropdown menu with the text: `aircrafts = [{ "aircraft_model": "MD-11", "MSN": "48123", "mod":`
- Right Panel:** Session ID `9655a382-b0a4-4214-8512-9cc5d5679a66`. It displays a JSON array of objects, each representing an aircraft model and its compliance status with various ADs. The JSON is as follows:

```
{
  "aircraft_model": "MD-11",
  "affected": "yes",
  "reason": "The AD applies to MD-11 aircraft."
},
{
  "aircraft_model": "DC-10-30F",
  "affected": "yes",
  "reason": "The AD applies to DC-10-30F aircraft."
},
{
  "aircraft_model": "Boeing 737-800",
  "affected": "no",
  "reason": "The AD does not apply to Boeing 737-800 aircraft."
},
{
  "aircraft_model": "A320-214",
  "affected": "yes",
  "reason": "The AD applies to A320-214 aircraft."
},
{
  "aircraft_model": "A320-232",
  "affected": "no",
  "reason": "The AD applies to A320-232 aircraft, except those with mod 24591 embodied in production."
},
{
  "aircraft_model": "A320-214",
  "affected": "no",
  "reason": "The AD applies to A320-214 aircraft, except those which have accomplished Airbus SB A320-57-1889 at Revision 04."
},
{
  "aircraft_model": "A321-111",
  "affected": "no",
  "reason": ""
}
```

(Correct)

Agent Development Kit

multi_agents

Trace Events State Artifacts Sessions Eval

Invocations

aircrafts = [{"aircraft_model": "MD-11", "MSN": "48123", "mod": "mod1"}]

SESSION ID: 872f6f0-eb27-47e-99fa-3d2267c10be

Token Breaching | + New Session

1

EASA_AD_2025-2024_1.pdf
EASA_AD_148-2021-23-31_1.pdf

```
{
  "aircraft_model": "MD-11",
  "affected": "yes",
  "reason": "The AD applies to Boeing MD-11 aircraft."
},
{
  "aircraft_model": "DC-10-30F",
  "affected": "yes",
  "reason": "The AD applies to Boeing DC-10-30F aircraft."
},
{
  "aircraft_model": "Boeing 737-800",
  "affected": "no",
  "reason": "The AD does not apply to Boeing 737-800 aircraft."
},
{
  "aircraft_model": "A320-214",
  "affected": "yes",
  "reason": "The AD applies to Airbus A320-214 aircraft."
},
{
  "aircraft_model": "A320-232",
  "affected": "no",
  "reason": "The AD applies to Airbus A320-232 aircraft, except those with mod 24091 embodied in production."
},
{
  "aircraft_model": "A320-214",
  "affected": "no",
  "reason": "The AD applies to Airbus A320-214 aircraft, except those which have accomplished Airbus SB A320-57-1089 at Revision 04."
}
}
```

Type a Message...

🗨️ 📎 📄