

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)

The screenshot shows the Agent Development Kit interface with a session ID of 3c2f1b74-659b-4af5-90a2-9f180f55c3f6. The left panel displays an invocation code block:

```
aircrafts = [ {"aircraft_model": "MD-11", "MSN": "48123", "modifications": []} ]
```

The right panel shows the evaluation results in a large text box:

```
[{"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, except those with mod 24591."}, {"aircraft_model": "A320-214", "affected": "no", "reason": "The AD applies to Airbus A320-214, except those with SB A320-57-1089 Rev 04."}, {"aircraft_model": "A321-111", "affected": "no", "reason": "The AD applies to Airbus A321-111 aircraft."}]
```

A message input field at the bottom says "Type a Message..." with a microphone icon.

(Incorrect)

The screenshot shows the Agent Development Kit interface with a session ID of 3c2f1b74-659b-4af5-90a2-9f180f55c3f6. The left panel displays an invocation code block:

```
aircrafts = [ {"aircraft_model": "MD-11", "MSN": "48123", "modifications": []} ]
```

The right panel shows the evaluation results in a large text box:

```
[{"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, except those with mod 24591."}, {"aircraft_model": "A320-214", "affected": "no", "reason": "The AD applies to Airbus A320-214, except those with SB A320-57-1089 Rev 04."}, {"aircraft_model": "A321-111", "affected": "yes", "reason": "The AD applies to Airbus A321-111 aircraft."}]
```

A message input field at the bottom says "Type a Message..." with a microphone icon.

(Correct)

The screenshot shows the Agent Development Kit interface with a session ID of ffe809e3-4b3b-44c7-b9f6-7df25f1d7610. The left panel displays an invocation code block:

```
aircrafts = [ {"aircraft_model": "MD-11", "MSN": "48123", "modifications": []} ]
```

The right panel shows the evaluation results in a large text box:

```
[{"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": "no", "reason": "This AD does not apply due to having SB A320-57-1089 Rev 04 embodied."}, {"aircraft_model": "A320-232", "affected": "no", "reason": "This AD does not apply due to having mod 24591 embodied."}]
```

A message input field at the bottom says "Type a Message..." with a microphone icon.

(Correct)

The screenshot shows the Agent Development Kit interface with two sessions. The left session, labeled 'single_agent', contains an invocation: 'aircrafts = [{"aircraft model": "MD-11", "MSN": "48123", "modif...'. The right session, labeled 'EASA_EAO_US-2026-23-53_1', displays a JSON response from the AD:

```
[{"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."}, {"aircraft_model": "A320-214", "affected": "no", "reason": "The AD applies to A320-214 aircraft, except those which have accomplished Airbus SB A320-57-1089 at Revision 04."}, {"aircraft_model": "A321-111", "affected": "yes", "reason": "The AD applies to A321-111 aircraft."}]
```

A message input field and a send button are visible at the bottom of the right session window.

(Correct)

The screenshot shows the Agent Development Kit interface with two sessions. The left session, labeled 'single_agent', contains an invocation: 'aircrafts = [{"aircraft model": "MD-11", "MSN": "48123", "modif...'. The right session, labeled 'EASA_EAO_US-2026-23-53_1', displays a JSON response from the AD:

```
[{"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 and it does not have modification 24591 or SB A320-57-1089 Rev 04."}, {"aircraft_model": "A320-232", "affected": "no", "reason": "The AD applies to A320-232 aircraft, but this aircraft has modification 24591 embodied in production."}, {"aircraft_model": "A320-214", "affected": "no", "reason": "The AD applies to A320-214 aircraft, but this aircraft has accomplished SB A320-57-1089 at Revision 04 in service."}, {"aircraft_model": "A321-111", "affected": "yes", "reason": "The AD applies to A321-111 aircraft and it does not have modification 24077."}]
```

A message input field and a send button are visible at the bottom of the right session window.

- Multi Agent:
(Correct)

The screenshot shows the Agent Development Kit interface with two sessions. The left session, labeled 'single_agent', contains an invocation: 'aircrafts = [{"aircraft model": "MD-11", "MSN": "48123", "modif...'. The right session, labeled 'EASA_EAO_US-2026-23-53_1', displays a JSON response from the AD:

```
[{"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 A320-57-1089 at Revision 04."}, {"aircraft_model": "A321-111", "affected": "yes", "reason": "The AD applies to A321-111 aircraft."}]
```

A message input field and a send button are visible at the bottom of the right session window.

(Correct)

The screenshot shows the Agent Development Kit interface with a session ID of 57780d32-f568-4df6-9716-00e072a33888. The response body contains JSON data listing aircraft models and their affected status based on AD 2025-0254:

```
[{"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 A320-57-100R at Revision 04 in service."}]
```

(Correct)

The screenshot shows the Agent Development Kit interface with a session ID of 09998f80-8fc9-46c0-9143-0be0f4e1e0ed. The response body contains JSON data listing aircraft models and their affected status based on AD 2025-0254, including a note about mod 24591:

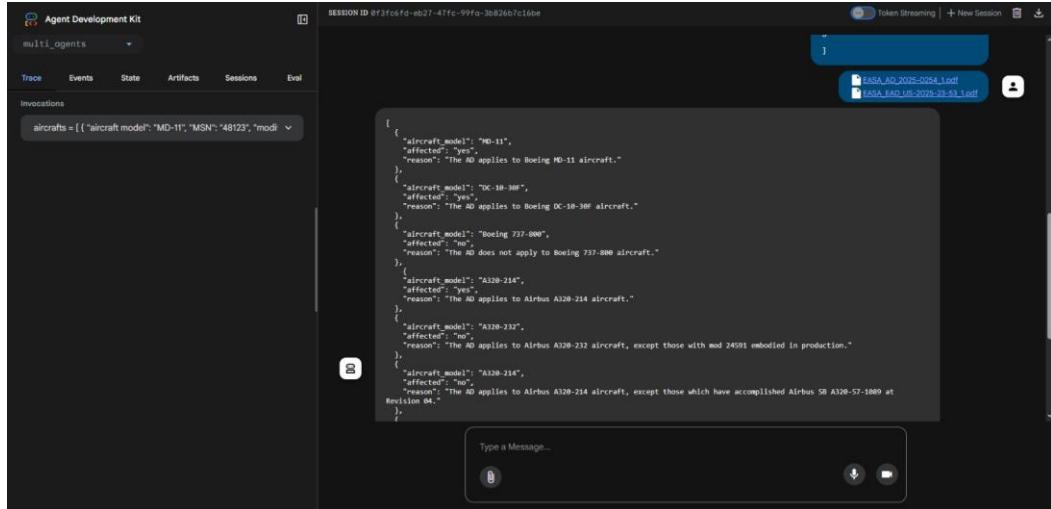
```
[{"aircraft_model": "MD-11", "affected": "yes", "reason": "The AD applies to Boeing MD-11 aircraft."}, {"aircraft_model": "DC-10-30F", "affected": "yes", "reason": "The AD applies to Boeing DC-10-30F aircraft."}, {"aircraft_model": "Boeing 737-800", "affected": "no", "reason": "The AD does not apply 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 A320-57-100R at Revision 04."}, {"aircraft_model": "A321-111"}]
```

(Correct)

The screenshot shows the Agent Development Kit interface with a session ID of 9e556392-b944-4214-8512-9cc5d5b79e66. The response body contains JSON data listing aircraft models and their affected status based on AD 2025-0254, including a note about mod 24591:

```
[{"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 A320-57-100R at Revision 04."}, {"aircraft_model": "A321-111"}]
```

(Correct)



The screenshot shows the Agent Development Kit (ADK) interface. On the left, there's a sidebar with tabs for Trace, Events, State, Artifacts, Sessions, and Eval. Under the Artifacts tab, there's a dropdown menu set to "multi_agents". Below it, there's a section titled "Invocations" with a dropdown menu set to "aircrafts = (("aircraft model": "MD-11", "MSN": "48123", "modi

```
{
    {
        "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": "A320-212",
        "affected": "no",
        "reason": "The AD applies to Airbus A320-212 aircraft, except those with mod 24591 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 A320-57-1069 at Revision 04."
    }
}
```

At the bottom of the interface, there's a message input field with placeholder text "Type a Message..." and a small microphone icon. To the right of the message input, there are icons for a video camera and a file.