CLAIMS:

1. A computer-implemented method for managing interactions with an artificial intelligence (AI) assistant on a hybrid framework, the method being executed by one or more processors and comprising:

receiving, by the one or more processors, communication data from a device, the communication data comprising data input by a user of the device;

determining, by the one or more processors, a context based on an extended finite state machine that defines contexts and transitions between contexts by querying an on-premise session manager for a current context, the current context being returned in response to the query;

determining whether at least one cloud-hosted service is to be queried based on the current context;

if data is to be sent to at least one cloud-hosted service, determining whether a data set that is to be transmitted to the at least one cloud-hosted service in a service request includes personally identifiable information;

if the data set to be transmitted to the at least one cloud-hosted service in a service request includes personally identifiable information, masking the personally identifiable information using an on-premise masking component prior to transmitting the service request to the at least one cloud-hosted service;

transmitting, by the one or more processors, a service request to at least one cloud-hosted service, the service request being provided at least partially based on the masking of the personally identifiable information;

receiving, by the one or more processors, a service response from the at least one cloud-hosted service, the service response comprising one or more of an intent, and an entity;

determining, by the one or more processors, at least one action that is to be performed by at least one back-end source system based on the service response;

providing, by the one or more processors, a response at least partially based on an action results received from the at least one back-end source system; and

transmitting, by the one or more processors, the result data to the device.

2. The method of claim 1, wherein the at least one action is determined by identifying a transition from the current context using the extended finite state machine based on one or more of the intent and the entity, the transition indicating the at least one action.

3. The method of claim 1 or 2, wherein the communication data is received from the device through a channel connector corresponding to a plurality of channels, through which communication data can be received.

4. The method of claim 1, further comprising transforming a format of the communication data to a standard format using an on-premise data unification layer.

5. The method of claim 1, wherein the hybrid framework comprises an on-premise portion that selectively transmits requests to, and receives responses from at least one cloud-hosted service.

6. The method of claim 1, wherein the at least one cloud-hosted service comprises one or more of a natural language processing (NLP) service, a sentiment analysis service, a speech-to-text service, and a translation service.

7. The method of claim 1, wherein the at least one cloud-hosted service is configured to determine one or more of the intent, and the entity.

8. The method of claim 1, wherein the extended finite state machine is configurable by a user based on one or more requirements.

9. One or more non-transitory computer-readable storage media coupled to one or more processors and having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform operations for managing interactions with an artificial intelligence (AI) assistant on a hybrid framework, the operations comprising:

receiving communication data from a device, the communication data comprising data input by a user of the device;

determining a context based on an extended finite state machine that defines contexts and transitions between contexts by querying an on-premise session manager for a current context, the current context being returned in response to the query;

determining whether at least one cloud-hosted service is to be queried based on the current context;

if data is to be sent to at least one cloud-hosted service, determining whether a data set that is to be transmitted to the at least one cloud-hosted service in a service request includes personally identifiable information;

if the data set to be transmitted to the at least one cloud-hosted service in a service request includes personally identifiable information, masking the personally identifiable information using an on-premise masking component prior to transmitting the service request to the at least one cloud-hosted service;

transmitting a service request to at least one cloud-hosted service, the service request being provided at least partially based on the masking of the personally identifiable information;

receiving a service response from the at least one cloud-hosted service, the service response comprising one or more of an intent, and an entity;

determining at least one action that is to be performed by at least one back-end source system based on the service response;

providing a response at least partially based on an action results received from the at least one back-end source system; and

transmitting the result data to the device.

10. The computer-readable storage media of claim 9, wherein the at least one action is determined by identifying a transition from the current context using the extended finite state machine based on one or more of the intent and the entity, the transition indicating the at least one action.

11. The computer-readable storage media of claim 9 or 10, wherein the communication data is received from the device through a channel connector corresponding to a plurality of channels, through which communication data can be received.

12. The computer-readable storage media of claim 9, wherein operations further comprise transforming a format of the communication data to a standard format using an on-premise data unification layer.

13. The computer-readable storage media of claim 9, wherein the hybrid framework comprises an on-premise portion that selectively transmits requests to, and receives responses from at least one cloud-hosted service.

14. The computer-readable storage media of claim 9, wherein the at least one cloud-hosted service comprises one or more of a natural language processing (NLP) service, a sentiment analysis service, a speech-to-text service, and a translation service.

15. The computer-readable storage media of claim 9, wherein the at least one cloud-hosted service is configured to determine one or more of the intent, and the entity.

16. The computer-readable storage media of claim 9, wherein the extended finite state machine is configurable by a user based on one or more requirements.

17. A system, comprising:

one or more processors; and

a computer-readable storage device coupled to the one or more processors and having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform operations for managing interactions with an artificial intelligence (AI) assistant on a hybrid framework, the operations comprising:

receiving communication data from a device, the communication data comprising data input by a user of the device;

determining a context based on an extended finite state machine that defines contexts and transitions between contexts by querying an on-premise session manager for a current context, the current context being returned in response to the query;

determining whether at least one cloud-hosted service is to be queried based on the current context;

if data is to be sent to at least one cloud-hosted service, determining whether a data set that is to be transmitted to the at least one cloud-hosted service in a service request includes personally identifiable information;

if the data set to be transmitted to the at least one cloud-hosted service in a service request includes personally identifiable information, masking the personally identifiable information using an on-premise masking component prior to transmitting the service request to the at least one cloud-hosted service;

transmitting a service request to at least one cloud-hosted service, the service request being provided at least partially based on the masking of the personally identifiable information;

receiving a service response from the at least one cloud-hosted service, the service response comprising one or more of an intent, and an entity;

determining at least one action that is to be performed by at least one back-end source system based on the service response;

providing a response at least partially based on an action results received from the at least one back-end source system; and

transmitting the result data to the device.

18. The system of claim 17, wherein the at least one action is determined by identifying a transition from the current context using the extended finite state machine based on one or more of the intent and the entity, the transition indicating the at least one action.

19. The system of claim 17 or 18, wherein the communication data is received from the device through a channel connector corresponding to a plurality of channels, through which communication data can be received.

20. The system of claim 17, wherein operations further comprise transforming a format of the communication data to a standard format using an on-premise data unification layer.

21. The system of claim 17, wherein the hybrid framework comprises an on-premise portion that selectively transmits requests to, and receives responses from at least one cloud-hosted service.

22. The system of claim 17, wherein the at least one cloud-hosted service comprises one or more of a natural language processing (NLP) service, a sentiment analysis service, a speech-to-text service, and a translation service.

23. The system of claim 17, wherein the at least one cloud-hosted service is configured to determine one or more of the intent, and the entity.

24. The system of claim 17, wherein the extended finite state machine is configurable by a user based on one or more requirements.