Moveworks

Conversational AI Trustworthiness

Conversational AI is increasingly popular for consumer and enterprise uses. Businesses across industries recognize it as a tool to improve the customer and employee experience as it operates on many platforms and allows users to engage with natural language. Additional advantages, such as 24/7 always-on service, personalized interaction, and no wait times, position conversational AI as the logical evolution for service management. However, the trustworthiness of conversational AI — specifically, ensuring that customer data is properly safeguarded — is key.

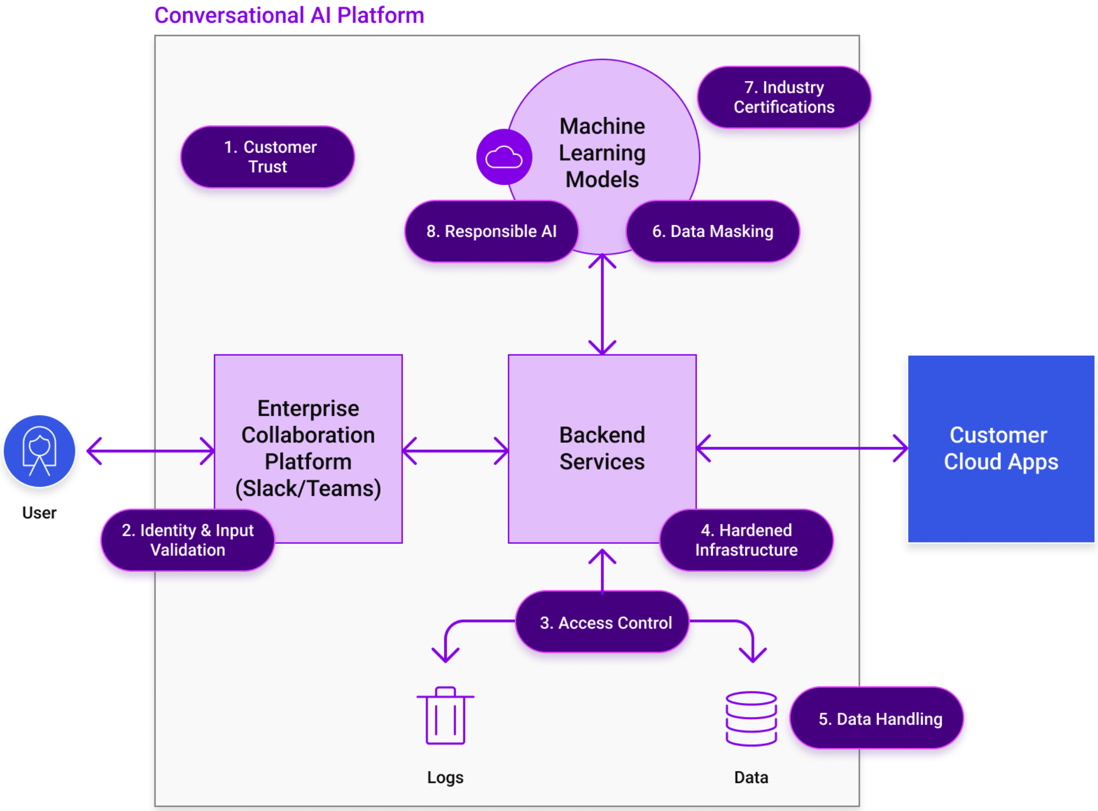

Conversational AI Platform
1. Customer Trust
Machine Learning Models
7. Industry Certifications
8. Responsible AI
6. Data Masking
User
Enterprise Collaboration Platform (Slack/Teams)
Backend Services
Customer Cloud Apps
2. Identity & Input Validation
4. Hardened Infrastructure
3. Access Control
Logs
Data
5. Data Handling

Fig. Conversational AI overview

Users typically interact with a conversational AI through the enterprise collaboration platform used by an organization, for example, Slack or Microsoft Teams. Conversational AI leverages backend services to cater to users' concerns and connect with various cloud applications deployed by the customer. These backend services also leverage machine learning to train models to generate user-specific responses.

What are the foundational blocks of trustworthy conversational AI?

Regarding user adoption, there are several valid concerns that conversational AI should address to build trust with users. The outline below defines the terminology followed by the minimum questions an organization developing conversational AI should address to build customer trust.

1. **Customer Trust:** A company earns customer trust by delivering on its commitments and promises. Building and maintaining trust requires well-established information security and privacy processes, mechanisms, and teams. Furthermore, there is a commitment from the executive leadership team and designated resources to promptly work on these important initiatives.

   Questions to keep in mind:
   - Is there a dedicated security team? Who is the CISO? What's the experience of the team?
   - What formal security and privacy processes exist to safeguard data?
   - What's the commitment to security and privacy efforts?
   - Does the product have any security and privacy features?

2. **Identity and Input Validation:** The information exchanged between the conversational AI, and the customer needs to be properly validated and the identity properly safeguarded.

   Questions to keep in mind:
   - How is the user's identity validated?
   - How does the bot ensure that the information can only be accessed by that user?
   - What validation exists for untrusted user input?

3. **Access Control:** Customer data should have proper access controls in place and be stored separately (i.e., not co-mingle with other customers' data). The conversational AI service should only have access to the services and data needed.

   Questions to keep in mind:
   - What type of data does the conversational AI have access to?
   - How does the company limit access to customers' data?
   - What permissions are in place?
   - How is customer data securely stored?
   - How are access controls reviewed?

4. **Hardened Infrastructure:** Conversational AI must operate on a properly configured and secured infrastructure with secure default configurations.

   Questions to keep in mind:
   - Where is the data hosted, and how is it safeguarded?
   - What processes and mechanisms are implemented to ensure the infrastructure is properly configured?
   - What monitoring, alerting, and auditing tools are in place to identify potential threats to the Infrastructure?
   - What evaluations (i.e., pen tests, scanning) are performed against the Infrastructure, and on what cadence?

5. **Data Handling:** Conversational AI must access and process data to service requests and generate user-specific responses. It should also have the capability to delete and mask sensitive data. A proper conversational AI relies on machine learning (ML) to improve over time, so proper privacy-enhancing

technologies should be implemented. Lastly, it is important for conversational AI to ensure the data from one customer is not intertwined with another customer.

Questions to keep in mind:

- How is the data securely handled in transit and at rest?
- What capabilities does the conversational AI have to delete data?
- What ML capabilities exist to properly safeguard data?
- How is ML data from one customer not mixed with another?

6. **Data Masking:** Data masking ensures that sensitive customer data is concealed from people who might need to review or access data as part of their job (such as a data annotator). A trustworthy conversational AI has this functionality to ensure that sensitive data is not exposed.

Questions to keep in mind:

- What data masking capabilities are in place?

7. **Industry Certification:** Several industry standards and certifications highlight security and privacy. They ensure that security and privacy controls are properly implemented, including processes and mechanisms to protect its customers and their data.

Questions to keep in mind:

- What security and privacy certification does the organization have?
- How does the organization go above & beyond certifications? How does the organization think and care about security and privacy?

8. **Responsible AI:** Unbiased machine learning (ML) models require actively minimizing or eliminating potential sources of bias with respect to protected classes, including race, age, disability, religion, color, national origin, sexual orientation, gender identification, and genetic information.

Questions to keep in mind:

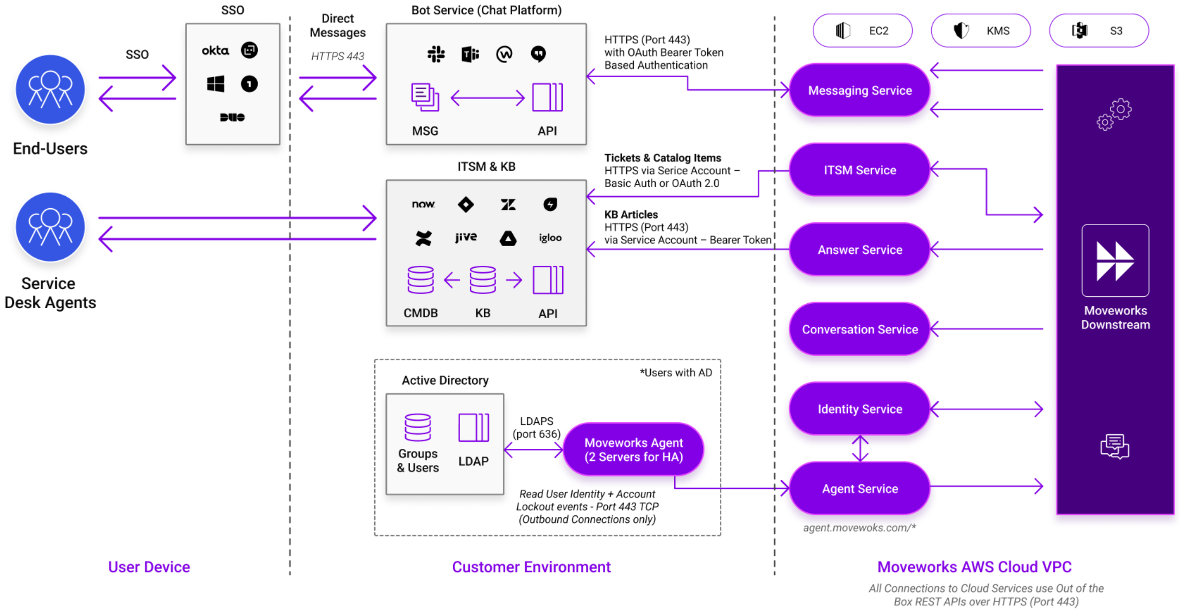
- How are the ML models trained?
- What techniques do these ML models use for training?
- What's the approach performed by annotators to avoid biases?

How does Moveworks stack up?

Moveworks is an authenticated app that lives in customers' enterprise collaboration platforms such as Slack, Microsoft Teams, etc. It also delivers support across many popular enterprise portals, including ServiceNow, SharePoint, and Epic. Moveworks takes advantage of integrations with native platforms but is not in and of itself a native platform directly accessed by an employee. For example, an end-user would access Teams or Slack to access Moveworks.

Since its inception, Moveworks has taken security and privacy seriously by following industry best practices and innovating in this space.

For example, Moveworks has added a data segmentation feature on top of AWS S3 to better safeguard customers' data. The diagram below provides a high-level architecture overview, followed by a table summarizing how Moveworks performs in each of the areas outlined above.



Foundational Trustworthy Areas	Protections?	Moveworks' Approach
Customer Trust	✔	Moveworks has a dedicated team of security and privacy professionals led by a Chief Information Security Officer (CISO). Moveworks has established security mechanisms & privacy processes, which are regularly updated with best-in-class practices. Moveworks has also developed security and privacy features to properly safeguard customer data.
Identity and Input Validation	✔	Moveworks integrates with enterprise collaboration platforms and identity management systems, ensuring that customers are properly authenticated. User data is validated, encoded, and encrypted.
Access Control	✔	Access to customer data is restricted on a need-to-know basis and controlled via the least privilege mechanism. Moveworks recognizes the sensitive nature of your data, and to mitigate risk, each Moveworks customer has a logically isolated instance of our cloud services suite that is deployed and secured on AWS US West (Oregon), AWS US East (Ohio), and AWS GovCloud (US-East).
Hardened Infrastructure	✔	The Moveworks service runs as a set of containers in the Amazon Web Services (AWS) cloud with all Moveworks runtime services and runtime customer data resident in the Moveworks virtual private cloud (VPC). Moveworks maintains controls to ensure the confidentiality, integrity, and availability of your data. The security of the infrastructure is scanned, monitored, and audited on a regular basis.
Data Handling	✔	Customer data is stored in a dedicated and encrypted AWS S3 bucket via AWS KMS. Data is encrypted at rest (AES-GCM 256)

		and in transit (TLS 1.2). All keys and credentials are encrypted and managed by AWS KMS (Rotated at least annually). Our collective learning model is designed to anonymize and irreversibly transform customer data, including personally identifiable information (PII), before ingesting it into our model for training. Moveworks deletes data upon request or if your organization stops using the product (following NIST 800-88)
Data Masking	✔	Moveworks utilizes a machine learning data masking library to mask sensitive personal information such as names, emails, phone numbers, credit card numbers, etc. Moveworks maintains strict levels of privacy by complying with industry-leading standards such as GDPR, CCPA, HIPAA, and ISO 27001.
Industry Certifications	✔	Moveworks' security and privacy program has achieved the following industry certifications: SOC2, ISO 27001, ISO 27017, ISO 27018, CSA Star Level 2 Gold, and Privacy Shield.
Responsible AI	✔	Moveworks' ML models are trained primarily on data sampled from production usage. Using a technique known as Collective Learning, many of its models are trained on anonymized data drawn from multiple customers, allowing them to learn the universal structure of requests from employees with different backgrounds and characteristics. Moveworks annotates this data without exposing any user characteristics to the annotator: no names, photos, or other category-identifying features are included in the annotation interface. When annotating the intention of a request, for example, the annotator only sees the text of the message and the organization's name. During training, Moveworks does not include protected attributes, such as gender or race, in the inputs from which the models learn to derive signals. For the vast majority of requests sent to the Moveworks bot, it would be challenging, if not impossible, for a human annotator to guess any protected categories about the knowledge worker from the text of the request—and harder still for its ML models. This means the models are much less likely to learn an intermediate representation associated with a protected class.