

Frequently Asked Questions

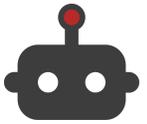
Document id: ishield.ai-Dost-FAQ-V3

Last Update Date: 31/Mar/2022

Notice of Distribution	This document is created by the ishield.ai team and published for public consumption.
Notice of Confidentiality	This document contains proprietary and confidential information of ishield.ai products and services. The recipient agrees to maintain this information in confidence and not reproduce or otherwise disclose this information to any person outside of the organization unless authorized by the document owner.
Disclaimer	This documentation will be subject to periodic revision as ishield.ai's product suite expands.

Version History

Date	Status	Author
29/July/2021	Created the first version	Product team @ ishield technologies, Inc.
09/Aug/2021	Incorporated feedback after internal review	Kesavan
23/Mar/2022	Updated with latest features	Kesavan



Frequently Asked Questions

Some Fun Questions

Q) What does the word Dost mean?

Dost means 'friend'. This word transcends many languages and cultures across South Asia and Middle-east.

Q) Why did we choose the name Dost?

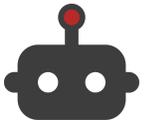
Who is (a / your) Dost? Someone you can trust! Someone who is well meaning. Someone who does not hesitate to tell you things the way they are even if you are uncomfortable hearing it. Someone who taps on your shoulders and discreetly points out your errors with a positive intent. A dost never memorizes your past errors, does not share your errors with anyone else, and both forgives and forgets.

In the context of enterprise collaboration and communications, we were looking to build a product that acts as a catalyst for promoting inclusivity in written language, so that one does not inadvertently make other people uncomfortable due to the unconscious biases and toxicity that at times seeps into the language one uses.

Seamlessly integrated with your Slack or Microsoft Teams, Dost acts as a trusted friend in the room. When a person sends a message that is deemed non inclusive or biased, Dost immediately flags the content to the user with a discreet message and provides them an opportunity to consciously review their messages and take corrective actions (delete, edit, etc).

In addition, some key qualities of Dost app include:

- **Discretion & Trust:** Dost messages are visible only to the sender. No one else.
- **Does not memorize:** Dost does not store any user-information or the flagged messages sent by the user in its systems. This way, Dost has no mechanism to trace either the message or the user.
- **No whispering to others:** Dost does not share or report data to anyone. Your messages and even the event that a message was flagged for a user is not sent to any other person. Since Dost does not memorize, there is also no mechanism in place to trace messages or the people sending it.



Q) What is the story behind the logo?



Does Dost seem familiar to you? Dost's logo is inspired by M-O from the movie Wall-E. Remember M-O? One of the most endearing and famous scenes from the movie is when M-O goes “Foreign Contaminant” when it first encounters Wall-E. The image of M-O, the purpose of the character and flagging unwanted things as “foreign contaminant” fits very well with the theme of Dost. Hence, we chose this as an inspiration to create our Logo.

Exploring the purpose of Dost

Q) Why do we need something like Dost?

Here are some excerpts from McKinsey study - *Diversity wins. How inclusion matters* (May 2020)

- The business case for Diversity and Inclusion is stronger than ever, with diverse companies more likely to financially outperform their peers.
- Hiring diverse talent isn't enough—it's the experience they have in the workplace that shapes whether they remain and thrive.
- Openness of the working environment, which encompasses bias and discrimination, was also of significant concern, with negative sentiment across industries.

McKinsey also recommends some systemic changes and bold steps. Two such bold recommendations support the case for Dost:

- Enable equality of opportunity through fairness and transparency.
- Promote openness, tackling bias and discrimination.

The purpose of Dost is to foster inclusive workplace collaboration and communications.

Q) Do people really post non-inclusive messages on enterprise communication platforms?

People post non inclusive messages due to unconscious bias or lack of context. According to SHRM, one in five workers has left their job because of a toxic workplace. This has cost employers a whopping \$223 billion over the last five year. A McKinsey survey¹ found that 84% of all respondents have experienced workplace microaggressions, which are everyday slights

¹ <https://www.mckinsey.com/business-functions/organization/our-insights/understanding-organizational-barriers-to-a-more-inclusive-workplace>



rooted in bias. In every subgroup—by gender, gender identity, minority status, or sexual orientation—more than eight in ten respondents report these indignities.

Q) So, Dost is going to read all my messages?

The following applies to any content that a user posts on enterprise communications and collaboration platforms:

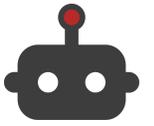
1. Users are not expected to post messages that can exclude, isolate or insult others.
2. The content a user posts is governed by their company policies and owned by the company.
3. The content that a user posts on these platforms is already public (within the company).
4. Some companies have groups to monitor and address problematic content on these platforms.
5. There is probably a grievance redressal mechanism in place to tackle workplace toxicity. If required, a company can audit all the messages a user has posted.

Dost is a technology platform that auto analyzes the content and nudges only the user sending non inclusive messages to correct their behavior to create a more inclusive workplace.

Q) What type of issues does Dost detect?

Dost detects the following types of issues:

1. Workplace microaggressions including
 - a. Identity attack
 - b. Stereotyping
 - c. Assault
 - d. Insult
 - e. Invalidation
2. Gender attack, including
 - a. Benevolent sexism
 - b. Misogyny
 - c. Bias
3. Toxicity, including
 - a. Hatred
 - b. Bullying and Harassment
 - c. Abuse
 - d. Violence
 - e. Sexual content
4. Non inclusive language that excludes a person or group of people.





Exploring how Dost works

Q) How do you decide that a message is not inclusive?

Dost is powered by a trained AI model(s) to detect non inclusive language. We rely on established research to define criteria for what constitutes non inclusive language in workplace communications.

The training data we use to train our AI models are collected through primary and secondary research. These data sets are human-labeled (with multi level reviews), based on the criteria established in the research.

Our research team is led by Linguists & Domain experts who constantly review customer feedback and errors to make systemic changes to the AI models to improve the accuracy.

We have also built explainability into our model right from start. The model responds with the message sub-text that the AI found to be non inclusive, recommends alternatives and provides material for users to learn more about the issue.

Q) What actions does Dost take when it finds non inclusive messages?

When Dost detects that a message is non inclusive, it immediately notifies the user sending such a message with a friendly nudge (customized message for the organization) that it can be interpreted as non inclusive by the people receiving it.

In addition Dost does the following:

1. Explains the issue found.
2. Recommends content for users to learn more about the issue.
3. Proposes alternative words / phrases to make the language more inclusive.
4. Celebrates a user when they send truly inclusive messages, and
5. Provides metrics for users to see how they are doing on using inclusive language.

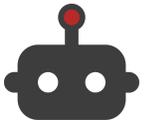
Q) Does anyone else see the message highlighted by Dost?

No. Only the user sending the message can see the flagged message.

Q) Is this a surveillance mechanism for managers to snoop around?

No.

1. On Slack, Dost does not work on 1:1 messages. It only works on channels where it is configured.
2. Channel users can add / delete the app from their channels if they like.
3. Only the sender gets the Dost response message, discreetly. No one else, including the admin, will know that a message was flagged.
4. No user information is accessed or stored.



5. No messages are stored in our systems. So, there is no mechanism to either trace the flagged message or user credentials.
6. Sender gets replacement recommendations, but the decision to edit is with the sender. The app does not decide on behalf of the sender.
7. The tool is used to educate the sender about unconscious biases that could exclude people. Today, companies spend a lot of time on one-off training on appropriate language and issue guidelines for employees.
8. We have employed a reputed third party company to do a detailed audit of the above principles, through DB review, code review and application testing. This report is available to be shared with companies on demand.

Q) What data does Dost store in their systems?

Dost takes the input message, processes it, and sends a response back. We only store the following information only for billing purposes: a) Timestamp a message was analyzed by Dost, b) the data type of the message, c) label/classification of the message when found toxic.

Dost does not store the original message in our systems. We also do not store any user data. There is no mechanism to trace the user or the message analyzed by Dost.

Q) What data do you share with the admin?

The information we share with the Slack admin is for billing purposes. This information consists of; a) the volume of messages processed, categorized by date/time and data types, b) the volume of toxic messages detected, categorized by toxicity categories, data types, and time period, and c) Any direct feedback we receive from our customers.

Q) What data formats does Dost support?

Dost will be able to detect toxicity in text, audio, video, images, emoticons, gif, memes, and documents. As of Mar/2022, we support text and images content. We are rapidly adding support for new data formats and Video support will come in Q3/2022.

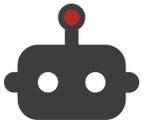
Q5) What languages does Dost support?

Currently we support the English language for the USA and other English-speaking countries.

Miscellaneous questions

Q) Okay I am convinced, how do I get started?

Administrators of Slack can go to <https://ishield.ai/dost> to download and set up Dost for their messaging platform. Alternatively they can also go to the Slack app directory, search for Dost and configure this for their organization.



Q) What kind of customizations does Dost allow for the organization?

Customizations are available for people in our paid tiers. Depending on your subscription type, we allow the following customizations:

- Company branding colors, fonts and response structure.
- Custom response messages, including
 - Default messages.
 - Issue specific messages.
- URL configurations in response messages to direct users to company specific training resources.
- Company specific rules.