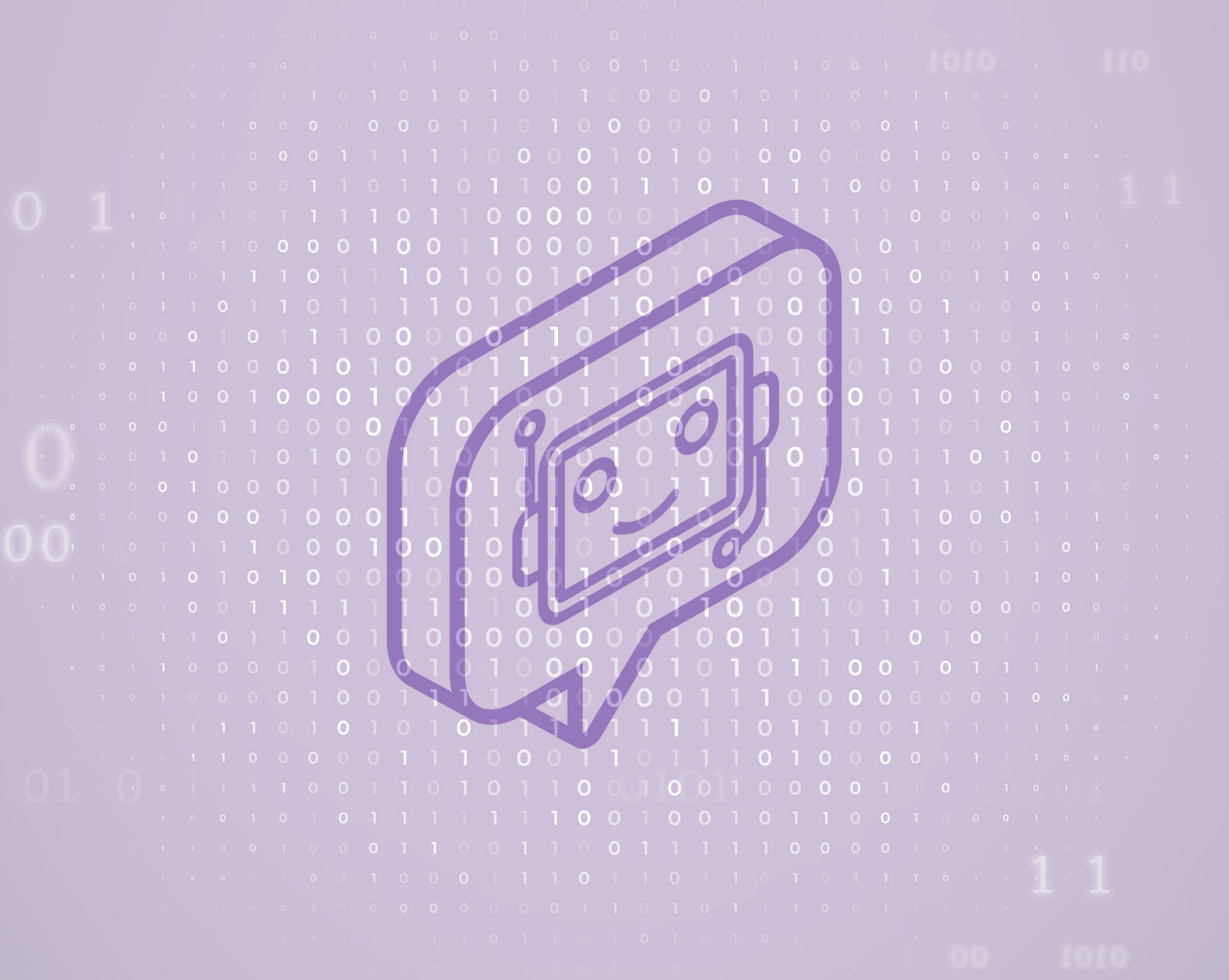
Transparency Guidelines

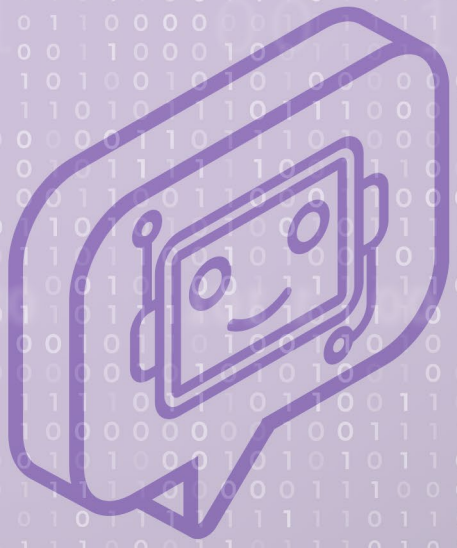
for Generative AI Chatbots

Published 20 July 2026



INFOCOMM MEDIA DEVELOPMENT AUTHORITY

CONTENTS



EXECUTIVE SUMMARY

Innovation thrives on trust, with transparency as the bridge that connects them. For generative AI to be widely adopted, users must have confidence in how it works. By sharing information about generative AI systems' capabilities and limitations, organisations can demonstrate accountability and reliability.

Yet, clear and accessible information on generative AI applications is hard to come by today. Available information is either too difficult for consumers to understand, as it is intended for technical developers; or overly simplified, as it is intended for marketing purposes. Sometimes, it can even be hard to find the right information, as it is fragmented across different sites, such as terms of service, privacy policies and blog posts.

This gap in transparency, particularly to consumers, creates real consequences. Users may inadvertently share sensitive data without understanding how it is used, overly rely on AI outputs for consequential decisions without appreciating their limitations, or unduly expose their children to unsafe content. Trust erodes as a result.

These guidelines take a first step in addressing the gap, starting with generative AI chatbots. The focus is on applications rather than the underlying models, as applications are how most consumers interact with generative AI. Chatbots serve as a natural starting point, being the application with the widest consumer reach and drawing considerable public concern.

The guidelines set out how chatbot deployers can provide meaningful transparency to their users through a chatbot info card, akin to a 'medical label'. It is a single, consolidated reference point where consumers can access essential information about the chatbot they are using. The guidelines are organised around three key principles:

1. **Relevance:** The chatbot info card addresses the four questions users care about—what the chatbot can and cannot do, how reliable and safe it is, how user data is handled and how to report issues.
2. **Accessibility:** The information is easy to understand, easy to navigate and easy to find.
3. **Timeliness:** Chatbot info cards are published by the time a chatbot is made available to users and kept up to date as the chatbot evolves.

While these guidelines are voluntary, we strongly encourage chatbot deployers to adopt at least a minimum. The chatbot info card should address each of the four questions users care about, as mentioned above, with at least one substantive disclosure in each. Beyond this, deployers can decide what and how much to disclose. Deployers should surface a high-level safety statement together with a clearly identifiable link to the chatbot info card at first use of the chatbot. The card should be easily reachable from within the chatbot thereafter. Deployers can decide how best to design and present this to users.

With greater transparency, we build a culture of accountability and trust. Chatbot deployers can be measured against their own publicly stated commitments for safety and reliability, and end-users can make informed decisions about use. The guidelines also serve as a reference for sectoral regulators to **extend transparency to other AI applications in different sectors.** As transparency cannot be achieved in isolation, Singapore will continue to work alongside international organisations and standards bodies to **broaden the global conversation on transparency from models to applications.** Together, these efforts will help ensure that as generative AI adoption grows, public trust grows with it.

INTRODUCTION

As generative AI becomes more integrated into our work and daily lives, it is critical to ensure it is deployed and used responsibly. **Meaningful transparency—information that is relevant, accessible and timely—underpins responsible AI.** It enables people to understand what an AI system does, what it is not designed to do and how risks are being addressed. Without it, governance becomes guesswork and trust is eroded.

Current State of Transparency

Information about generative AI models and applications exists today, but it is fragmented and rarely serves consumers well. Terms of service, privacy policies and technical documentation are often lengthy and complex. Consumers rarely read them and may not know where to find them. Simpler descriptions, where they exist, are often written for marketing purposes and may gloss over limitations and risks. Information across providers is also often inconsistent, which makes it difficult for consumers to understand performance and safety through meaningful comparisons.

The lack of clear and accessible information creates problems. At the individual level, users may engage with generative AI applications in uninformed ways: share confidential or proprietary information without understanding how generative AI deployers store or share their conversations, or rely on such applications for consequential decisions such as health advice, legal questions or financial planning without understanding their limitations. Vulnerable users, including children, may engage with these applications with detrimental effects. Cases have emerged of generative AI chatbots providing harmful advice[1] or inaccurate information in serious contexts.[2]

At the ecosystem level, information asymmetry creates uncertainty and erodes trust. Many consumers are concerned about the reliability of AI-generated content and how deployers handle their data. Parents worry about whether safeguards adequately protect their children from inappropriate content and potential self-harm. The result is a gap between adoption and trust: many people use generative AI applications, but they do so despite their reservations, rather than using generative AI with confidence.

Objectives of the Guidelines

The guidelines set out how generative AI deployers can provide meaningful transparency to consumers about their applications. This covers the types of information that consumers find relevant and how to ensure that the information is shared with them in a clear, accessible and timely manner.

[1] In a reported 2025 case, a user followed dietary advice from a generative AI chatbot that recommended substituting table salt with sodium bromide. Prolonged consumption led to serious health effects requiring hospitalisation. See The Guardian, Man develops rare condition after ChatGPT query over stopping eating salt.

[2] There have been several reported cases in which legal professionals submitted court documents drafted with the assistance of generative AI chatbots, only for the cited case law to be entirely fabricated. In some instances, sanctions were imposed by the courts. See Thomson Reuters, GenAI hallucinations are still pervasive in legal filings, but better lawyering is the cure.

These guidelines are voluntary and do not impose obligations.[3] Deployers know their own applications and users best, and are best placed to judge what to disclose, how to present it, and when to update it. At the same time, we **strongly encourage deployers to implement at least the minimum described in these guidelines** as a baseline of meaningful transparency for their users.[4]

Through increased transparency, we enhance accountability in the ecosystem. While setting universal safety thresholds may be impractical as risks vary widely across applications and contexts, deployers should, at least, state their own safety commitments as a clear standard against which the public can review, measure and assess them.

Transparency also enables end-user responsibility. Every player along the generative AI value chain, including end-users, plays an important role in ensuring responsible use of the technology. When consumers understand an application's capabilities and limitations, they can make informed decisions about their use and engage more responsibly. They can verify outputs where appropriate and recognise contexts where use may be inappropriate.

The guidelines also begin to standardise how information is shared with the public. A consistent disclosure framework enables users to compare applications and find relevant information quickly, rather than navigating disparate and inconsistent documentation. Ultimately, we want to create a **trusted ecosystem** where users can engage with generative AI with confidence, built on a foundation of transparency.

Generative AI Chatbots as a Start

Global attention on AI transparency has largely focused on generative AI models, with significant regulatory and industry momentum around model disclosure. **These guidelines focus on transparency at the application layer** instead, as most consumers experience generative AI through applications, rather than the underlying models. This will complement current global discussions by extending transparency beyond models.

As generative AI applications span a broad range of use cases, each with distinct considerations, these guidelines centre on **generative AI chatbots**[5] **as a start**, being the use case with the highest consumer touchpoint and growing societal concerns. Chatbots are among the most widely adopted generative AI applications, serving millions of users globally across education, work and personal life. Their scale has raised public concerns about data privacy, children's exposure to inappropriate content and safety risks for vulnerable users. Chatbot deployers should therefore clearly lay out the appropriate uses of their chatbots and how they have addressed public concerns.

[3] For the avoidance of doubt, these voluntary guidelines do not replace, modify or otherwise affect any obligations that may apply under other laws, regulations, or sector-specific requirements. For example, where a chatbot is regulated as a medical device or health product, deployers should continue to comply with the applicable regulatory requirements.

[4] In addition to transparency, generative AI deployers are encouraged to conduct risk assessments, application testing and other relevant safety practices. They may refer to other guidelines published by IMDA for further guidance on these practices. See IMDA, Artificial Intelligence in Singapore.

[5] For the purposes of these guidelines, a generative AI chatbot refers to an application, whether standalone or embedded within other products, that uses, in whole or in part, foundation models such as large language models, to produce adaptive, human-like responses in natural language to user inputs. The defining feature is that the user experiences an interactive conversational interface through text or speech. This includes chatbots that combine generative AI with rules-based or retrieval components. Chatbots may support multimodal inputs and outputs, such as the ability to process uploaded images or generate them in response to text prompts.

In particular, the guidelines are targeted at generative AI chatbots that are **external-facing** (i.e. used by external parties, such as customers or the public). Chatbots deployed solely for internal organisational use, such as employee productivity tools or internal knowledge assistants, fall outside the scope of these guidelines. These external-facing chatbots commonly fall into one or more of three broad categories:

General-Purpose Chatbots	Companion Chatbots	Domain-Specific Enterprise Chatbots
Designed for open-domain interactions across a wide variety of tasks.	**Designed for social interaction and emotional engagement.**	**Designed specifically for particular fields or tasks.**
Examples are chatbots used for writing, creative and learning tasks: ▸ ChatGPT ▸ Claude ▸ Gemini App	Examples are persona-driven and empathetic simulation chatbots: ▸ Character.AI ▸ Replika	Examples are customer service chatbots: ▸ Foodpanda's in-App Support chatbot ▸ SIA's Kris chatbot

These categories are illustrative and not mutually exclusive. A single chatbot may span more than one, and newer forms, such as agentic assistants, may not fit neatly within any of them. What matters is not the label a chatbot is given, but the risk and context of how it is used. Deployers should consider factors such as how consequential the chatbot's use is (e.g. answering informational queries versus taking actions on the user's behalf), the domain it operates in (e.g. general versus a specialised field such as health or finance), and who its users are—in particular, whether it is accessible to minors. These factors should guide which of the risks and disclosures set out in these guidelines are most relevant.

Finally, while the guidelines are targeted at generative AI chatbots, deployers can use them as a **reference when providing transparency for other generative AI applications.** Sectoral regulators can also build on them to develop more specific transparency guidelines for applications in their sectors.

Whose Responsibility is it to Provide Chatbot Transparency?

The **chatbot deployers,** which are the organisations that make the chatbot available for use, are primarily responsible for providing transparency to the end-users. They are responsible for conducting risk assessments and applying adequate mitigations, before deploying their applications. When something goes wrong, users expect answers from the deployers first, rather than from the different upstream parties involved. Deployers are therefore responsible and also best placed to provide transparency to users.

This is not to say that **upstream parties,** such as model providers, have no responsibility. They should support deployers in the responsible deployment of their applications, such as furnishing relevant and up-to-date information about the underlying models for deployers to conduct risk assessment and mitigation and be transparent to users.

CHATBOT INFO CARD

The chatbot info card is a **single consolidated reference point for consumers to access essential information about chatbots.** It can take several forms, including a disclosure document, a dedicated information page or even a pop-up that appears when consumers click on an information icon. Regardless of how a deployer chooses to communicate with consumers, what is critical is that the chatbot info card is:

RELEVANT

Contains all the essential information that consumers need to know to use the chatbot responsibly and with confidence.

ACCESSIBLE

Presents information in a way that is easy to understand, easy to navigate and easy to find.

TIMELY

Contains updated information that reflects the latest chatbot capabilities, risks and safety policies.

The guidelines are organised based on the above three principles. The first section on *Relevance* sets out key information to disclose in a chatbot info card, structured around questions users are likely to have. The second section on *Accessibility* provides design considerations addressing how to deliver and present information so that it is easy to understand, navigate and find. The third section on *Timeliness* covers the practical aspects of managing the lifecycle of chatbot info cards, including publishing and updating them over time.

The Minimum Encouraged for a Chatbot Info Card

We strongly encourage chatbot deployers to meet, at a minimum, the following:

- The chatbot info card should **cover the four key areas of information that users care about most** (detailed below), with at least one substantive disclosure in each area. Beyond this, deployers can decide what and how much to disclose.
- Deployers should also surface a **high-level safety statement together with a clearly identifiable link to the chatbot info card at first use of the chatbot.**
- The card should thereafter be **easily reachable from within the chatbot.** Deployers can decide how this is best designed and presented to users.

The sections that follow explain each of the three principles in turn. Deployers may also refer to **Annex A**, which offers additional suggested pointers they may draw on to make their chatbot info card more comprehensive.

#1 RELEVANCE

For consumers to use chatbots responsibly and with confidence, they need practical information about what the chatbot can do, what it cannot do and how their interactions are handled. **The key is not volume, but information that is relevant and actionable,** helping users make informed choices about how they engage with the chatbot.

In this section, we share how deployers can **identify the key information** to disclose and **calibrate the extent of disclosure** by balancing the need for transparency with legitimate constraints, such as proprietary information and security. **Annex A** sets out further pointers deployers can draw on.

Identify Key Information

Interviews with local users of generative AI chatbots identified four key types of information that help users engage with a chatbot responsibly and with confidence:

a) What the Chatbot Can Be Used For

When users first interact with a chatbot, it is important that they quickly understand what the chatbot can and cannot do, so that they can use it effectively and safely. They should know:

- **Capabilities** or the tasks that the chatbot can handle (e.g. answer FAQs about an airline's travel policies).
- **Limitations** or caveats related to its performance (e.g. cannot provide legal advice on aviation incidents).
- **Prohibitions** (e.g. age restrictions, out-of-scope uses).

Setting clear expectations on the chatbot's capabilities and limitations helps users get the best results from their interaction and avoid misuse. More importantly, it equips users to make sound judgments about how to engage, such as choosing appropriate tasks for the chatbot, recognising when a query falls outside its scope, and knowing when to seek human support.

b) How Reliable and Safe the Chatbot Is

To engage with chatbots in an informed manner, users need a clear view of the risks they may pose, how those risks have been mitigated, and what to keep in mind when using them. Examples of what this can include:

- **Common risks** posed by chatbots, which vary by type. For example, content safety may be a key concern for general-purpose chatbots, while emotional safety risks such as addiction or self-harm may be more salient for companion chatbots. Some risks, such as those posed to younger users, can cut across chatbot types. Annex A sets out key risk categories deployers can consider, alongside any other risks specific to their chatbots.
- **Safeguards** put in place to mitigate those risks, and the residual risks that users should be aware of.
- **User precautions** recommended for each risk (e.g. always fact-check, never share sensitive information, use parental control options for child safety).

Together, this helps users to use the chatbot with confidence, and take practical steps to protect themselves, such as configuring safety settings and calibrating how much to rely on outputs. The goal is informed engagement, rather than blind faith or excessive scepticism.

c) How User Data Will Be Used and Protected

Users may share a lot of information during chatbot conversations. While most chatbot deployers have detailed privacy policies, sharing key highlights allows users to make informed decisions about what information they should share:

- What specific data is collected (e.g. full conversations, metadata).
- Who has access (e.g. third-party sharing practices).
- Whether the data is used for model training.
- What user control options are available (e.g. ability to delete conversations, opt-out mechanism for data usage in training).

Data-related concerns were among the most common raised by the local users we spoke to, such as inadvertent personal information exposure, potential commercial exploitation (e.g. targeted advertising, especially from third parties) and professional risks (e.g. confidentiality concerns for work-related topics). Disclosing this information helps address such concerns.

d) How Users Can Report Issues

When something goes wrong or a user has concerns about a chatbot's output, they need to know how to raise them and what to expect in response. Users should be able to easily identify the right channel to report an issue, understand what types of issues they can raise, and know how their report will be acknowledged and followed up.

This information **gives users confidence that their concerns will be heard and acted upon.** It also empowers them to play an active role in improving the safety and reliability of the chatbot.

Disclose Across All Four Areas

A minimum set of disclosures matters because it helps end-users find the same core information across different chatbots, rather than disclosure that is patchy and inconsistent. **A chatbot info card should therefore address all four areas above, with at least one substantive disclosure in each.** A substantive disclosure is a **concrete, specific statement that users can act on.** For example, under 'How Reliable and Safe the Chatbot Is', a deployer might inform users that the chatbot can produce factual errors and that its responses should be verified before being relied on for important decisions. By contrast, a generic statement such as "we take safety seriously", on its own, would not meet this bar.

Deployers can reference Annex A for suggested disclosure pointers for each of the four areas. These are offered as ideas rather than requirements. **Ultimately, it is for deployers to decide what and how much to disclose.**

Calibrate Extent of Disclosure

Being transparent about a chatbot's capabilities and guardrails builds trust and helps users interact with it more thoughtfully. At the same time, there are also **practical and legitimate reasons why deployers cannot publish every detail.** In calibrating what information to disclose publicly, we recognise the following considerations:

a) Ensuring Proportionality

The guidelines balance the usefulness of information to end-users, in terms of whether it helps them make informed decisions and take meaningful action, against the efforts by deployers to obtain or produce it.

b) Protecting Proprietary Information

Deployers are not expected to reveal proprietary details such as the unique architectural choices underlying their chatbot. Where safety measures represent a competitive advantage, deployers should convey what protections are in place and how effective they have been, without disclosing implementation details.

c) Safeguarding Security

Providing information about how a chatbot application works, its capabilities and its limitations help users to make informed decisions and use it responsibly. However, excessive disclosure—such as revealing sensitive system details and vulnerabilities—can expose an application to exploitation or malicious attacks. Ultimately, disclosure should be purposeful and calibrated, ensuring that users receive sufficient information to understand and trust the chatbot, while safeguarding critical details that may compromise security.

d) Disclosing Only Reasonably Available Information

Many chatbot deployers build on third-party foundation models. In these cases, deployers need only disclose information that is reasonably available to them (e.g. through the model provider's public documentation or their contractual terms). They are not expected to attest to upstream details they cannot verify, such as the model provider's internal safety testing.

#2 ACCESSIBILITY

The preceding section sets out what key information deployers are encouraged to disclose. However, **equally important is *how* that information is presented and communicated.** Even the most relevant information will fail to serve its purpose if users cannot easily access and understand it.

This section sets out the considerations for designing the chatbot info card so that it is easy to understand, easy to navigate and easy to find. How best to achieve this is for chatbot deployers to decide. The practices set out below are recommended considerations and deployers can adapt them to suit their interface, platforms and users.

Easy to Understand

Information on the chatbot info card should be written in a way that is **easily understood by a non-technical audience.** At the same time, key information should not be oversimplified into broad statements that lack meaningful detail. Deployers should strike a balance between reducing the technicality of the statements and ensuring they remain sufficiently informative for users to make informed decisions. The following are recommended best practices:

a) Be Specific and Substantive

Broad policy statements, such as "we take safety seriously" or "we have robust safeguards in place", do not provide users with meaningful information. Where the chatbot info card describes safety measures, limitations or data practices, deployers should explain what those measures are, how they work and what users can expect.

b) Use Plain Language

The chatbot info card should be written in clear, direct language that end-users can understand without specialised AI knowledge. Technical terms should be clearly explained. For example, if a chatbot info card mentions that the chatbot may produce hallucinations, it should explain that hallucinations are instances where the chatbot generates false or misleading information that it presents confidently as fact.

Beyond terminology, deployers should also use straightforward sentence structures and avoid jargon-heavy phrasing throughout. As a practical benchmark, deployers should aim for a reading level accessible to a general adult audience—broadly equivalent to secondary school level. Readability tools such as the Flesch-Kincaid index can help deployers assess whether their language meets this standard.

Easy to Navigate

Users often scan quickly for key information that meets their needs, rather than read documents end-to-end. Therefore, the chatbot info card should also be laid out in a format that is **easily digestible.** This would not only encourage users to read the chatbot info card but also allow them to recognise critical information quickly. The following are the recommended best practices:

a) Offer Layered Disclosure

Deployers should present summary information upfront, with the option to expand drop-down sections or click through for more detailed explanations, instead of lengthy paragraphs of explanations. This allows users to understand the essential information at a glance and access more detailed information if needed.

b) Structure Content for Readability

Deployers should use layout and formatting methods, such as clear headers, bolded key terms, bullet points and short segmented sections to help users scan through the information quickly.

c) Use Visualisations

Deployers can consider using visuals, tables, icons and other design elements to make information more engaging. For example, a comparison table might show what the chatbot can and cannot do reliably, or icons might indicate which safety measures are in place.

Easy to Find

Beyond being clear and easily digestible, the chatbot info card should also be **easily discoverable.** Users should not need to know where else to look for it. The following are recommended best practices:

a) One-Stop Shop

The chatbot info card should serve as the central reference point for all key information. It may take the form of a single document or webpage, linked to related documents (e.g. Terms of Use, Privacy Policies). This unified approach reduces fragmentation of information and reinforces the card as a reliable transparency one-stop shop.

b) Surfaced at Onboarding

At first use, before users begin interacting with the chatbot, deployers should surface a high-level safety statement together with a clearly identifiable link to the chatbot info card. The design is flexible (e.g. an onboarding screen). The safety statement should be substantive rather than generic, conveying real safety caveats drawn from the chatbot info card (e.g. that the chatbot can make mistakes and outputs should be verified), rather than broad assurances such as "we take safety seriously".

c) Reachable from Within the Chatbot

Users should be able to reach the chatbot info card from within the chatbot interface. This does not require rebuilding the in-app experience or displaying all the information in the app itself. A clear, persistent access point within the interface (e.g. an information icon) is sufficient, and the card itself can be hosted on a linked web page. What matters is consistent, low-friction access from the interface, so users are not left to hunt for the information elsewhere or dig through deep menus to find it.

d) Standardised Across Different Modalities

For chatbots available on multiple platforms (e.g. website, mobile application), the chatbot info card should be equally accessible across the platforms. While exact user interface elements may vary slightly, the core navigation paths and placement should be as consistent as possible. Such consistency builds user familiarity and prevents confusion when switching between platforms.

Consolidated Cards for Chatbot Families

Deployers that offer a family of closely related chatbot experiences may publish a single, consolidated chatbot info card rather than separate cards for each variant. A consolidated card should document characteristics common to the family, such as shared safety measures or data practices, while highlighting key differences between variants, particularly regarding features, target user groups or safety settings. It should also explain how users can identify which variant they are using (e.g. through product names). This approach extends to chatbots deployed across multiple platforms or embedded within different applications.[6]

[6] For example, characters on Character.AI share the same underlying model and moderation framework but vary in behaviour and features (e.g. image generation) based on creator-defined parameters. Similarly, Meta AI is available both as a standalone chatbot application and embedded within other applications, such as WhatsApp, with features such as video generation varying by platform. In both cases, a consolidated chatbot info card could describe the shared features (e.g. safety measures) while noting the differences across variants or platforms.

#3 TIMELINESS

The chatbot info card should provide information that is current and reflective of the most up-to-date knowledge about the chatbot. A chatbot info card that reflects outdated information is of limited value. How best to manage this is for chatbot deployers to decide.

The guidance that follows sets out reference points for publishing and updating the chatbot info card, which deployers can apply in the way that best suits their context, so long as **publication remains timely and updates keep pace with developments that meaningfully affect users.**

At Deployment

Deployers should publish chatbot info cards **by the time their chatbots are made available to external users,** including in beta or pilot form. This allows users and affected stakeholders to review it before they interact with or rely on the chatbot. For example, it allows parents to research a chatbot before allowing their child to use it. For phased rollouts, deployers can publish what is known at launch and update the card as capabilities and risks settle. Publishing the card long after a chatbot is already in use defeats the purpose of transparency, as users will have engaged with the chatbot without the information they need.

For generative AI chatbots that were already publicly accessible before the issuance of these guidelines, deployers are encouraged to publish the corresponding chatbot info cards **as soon as reasonably practicable.**

Ongoing Updates

The risk landscape around generative AI chatbots is not static. New capabilities, emerging threat vectors and evolving user behaviours can all shift the risks associated with a chatbot over time. Deployers should continually assess whether the information disclosed in the chatbot info card remains accurate and current.

The chatbot info card should be **updated within a reasonable time when a new development meaningfully affects users through a change in the chatbot's capabilities or safety profile.** This applies whether the development arises from changes to the chatbot itself or from external factors such as newly identified risks from post-deployment monitoring.

Even when no such change has occurred, chatbot deployers are encouraged to **review their chatbot info cards periodically** (e.g. annually) to keep information current.

Potential Trigger Points to Consider Updates:

- **Model Changes**—Switching to a different underlying model that significantly alters the chatbot's reasoning ability or responses to safety-sensitive topics.
- **Updates to Guardrails**—Revising moderation pipelines or content filters such that the chatbot intervenes much more or much less on user inputs or model outputs.

- **New Capabilities**—Adding new features (e.g. image generation, web browsing) that expand how users can interact with the chatbot.
- **Emergence of Unexpected Use Cases**—Widespread adoption of the chatbot for high-risk use cases it was not originally designed for, altering its risk profile.

Examples of Changes That May Not Require an Update:

- **Minor Model Upgrades**—Upgrading to a newer version of the same underlying model with no significant change to capabilities or safeguards.
- **Interface Changes**—Changes to the user interface and visual design that do not alter the chatbot's core behaviour.
- **Backend and Infrastructure Changes**—Infrastructure upgrades, routine security patches or performance improvements (e.g. faster response times) that do not affect the chatbot's behaviour or safety profile.

Versioning

Chatbot deployers are encouraged to clearly version their chatbot info cards to help users verify that they are reading up-to-date information. Each chatbot info card could include:

- **Last updated date:** The date the chatbot info card was most recently revised.
- **Chatbot version:** The version of the chatbot to which the chatbot info card relates, where the chatbot uses version identifiers.

FUTURE WORK

These transparency guidelines represent an important step in Singapore's efforts to build a trusted AI ecosystem. By establishing clear expectations around how chatbot deployers should communicate their chatbots' capabilities, limitations and safety policies to users, we **lay the groundwork for a culture of accountability and trust** between AI systems and the people they serve.

As consumer-facing AI transparency is still an emerging practice, IMDA will continue to monitor how chatbot deployers implement chatbot info cards and **may refine these guidelines over time, in step with what the ecosystem can reasonably support.**

Further, as AI continues to permeate a broader range of applications—from education to legal and financial services—IMDA will continue to work with different partners, including sectoral regulators, to **extend and adapt the principles and considerations here to address the unique transparency demands of different sectors and domains.**

Finally, we recognise that transparency cannot be achieved in isolation. Singapore is committed to contributing to and aligning with global efforts to advance AI transparency. We will continue to work alongside international organisations and standards bodies to **broaden the global conversation on model transparency to application-level transparency.** Together, these efforts will help build an AI ecosystem that is both innovative and worthy of public trust.

ANNEX A: SUGGESTED DISCLOSURE POINTERS

The table below serves as a **practical resource** to help chatbot deployers identify the key information that they can share with their users. The pointers are **suggestions,** rather than requirements, and are **not exhaustive.** Deployers may wish to engage actual users or review user feedback to understand what information would be most useful to their users for enabling trusted and reliable usage of their chatbots.

Depending on the nature of their chatbot, deployers can **select the pointers most relevant to them, and may tailor the depth of disclosure to the chatbot's risk and its impact on users.** Where they do disclose, deployers are encouraged to go beyond describing features or risks in the abstract, and to **provide information that helps users take concrete steps,** whether that is adjusting their behaviour, configuring settings or knowing when to seek help elsewhere.

Where the table refers to the "effectiveness" of safeguards, deployers may express this qualitatively (e.g. a description of testing done or an assurance statement) or as evaluation results. The aim is to show how a conclusion about safety was reached, and either approach can do this. Deployers need not publish numerical scores to meet this.

Sections	Suggested Disclosure Pointers
What the Chatbot Can Be Used for *Deployers may link to their full terms of service to provide more details*	**Capabilities & Limitations:** ▸ What the chatbot is designed to do ▸ Underlying foundation model(s) used ▸ Key features and how it helps the end-user ▸ Clear statements about strengths and limitations ▸ Clear statements about accuracy, including when and how it may generate incorrect information, and guidance on verifying outputs ▸ Whether the chatbot is used in a regulated context
	Prohibitions: ▸ Out-of-scope uses (e.g. medical diagnosis in a hospital setting) ▸ Age restrictions

Sections	Suggested Disclosure Pointers
How Reliable and Safe the Chatbot Is *Deployers may link to their full content policies to provide more details*	*The following are examples of common risks posed by generative AI chatbots today, which deployers could address in the chatbot info card:* **Harmful or Inappropriate Content:** Relevant to most chatbots ▸ **Risk Addressed:** Types of harmful or inappropriate content the deployers have addressed (e.g. sexual, violent or suicide and self-harm content) ▸ **Safeguards:** Measures to prevent harmful or inappropriate content generation and summary of content management policies ▸ **Effectiveness:** How the deployer knows the safeguards are working, and any residual limitations ▸ **User Precautions:** Steps users should take to report or flag the conversation if they encounter harmful or inappropriate content **Risks to Younger Users:** Where the chatbot is available to users under 18 ▸ **Risk Addressed:** Types of harm to younger users that the deployers have addressed, such as age-inappropriate content and unhealthy usage ▸ **Safeguards:** Measures to ensure age-appropriate experiences for younger users (e.g. age assurance, restricted features, stricter content moderation), measures to reduce unhealthy usage (e.g. session limits, usage nudges) ▸ **Effectiveness:** How the deployer knows the safeguards are working, and any residual limitations ▸ **User Precautions:** Existing parental controls for parents or guardians to monitor and restrict usage by their child and steps for users to report age-inappropriate content

Sections	Suggested Disclosure Pointers
How Reliable and Safe the Chatbot Is *Deployers may link to their full content policies to provide more details*	**Content Safety in High-Risk Use Cases:** Particularly relevant to general-purpose chatbots ▸ **Risk Addressed:** Types of dangerous advice or high-risk use cases such as those pertaining to health, law or finance that the deployers have addressed ▸ **Safeguards:** Whether the chatbot identifies and responds differently to high-risk queries (e.g. providing disclaimers, redirecting users to qualified professionals) ▸ **Effectiveness:** How the deployer knows the safeguards are working, and any residual limitations ▸ **User Precautions:** Guidance on when users should seek expert human advice or verify outputs against sources of relevant and authoritative information, rather than relying on the chatbot **Emotional Safety and Healthy Usage:** Particularly relevant to companion chatbots ▸ **Risk Addressed:** Types of affect-laden uses leading to emotional overreliance or unhealthy attachment that the deployers have addressed ▸ **Safeguards:** Measures to prevent the chatbot from encouraging excessive emotional reliance, how the chatbot responds to users expressing distress or self-harm ideation, and whether the chatbot provides tools to encourage healthy usage patterns (e.g. session limits, usage nudges) ▸ **Effectiveness:** How the deployer knows the safeguards are working, and any residual limitations ▸ **User Precautions:** Guidance on maintaining healthy usage habits and where to seek professional support **Reliability in Professional Contexts:** Particularly relevant to domain-specific enterprise chatbots ▸ **Risk Addressed:** The broad scope or topics within the professional domain that the deployers have sought to improve accuracy in. ▸ **Safeguards:** Measures taken to improve accuracy or reduce hallucinations within the scope defined above ▸ **Effectiveness:** How the deployer knows the safeguards are working, and any residual limitations ▸ **User Precautions:** Guidance on when outputs should be independently verified or treated as indicative rather than definitive (e.g. for edge cases outside the chatbot's documented scope)

Sections	Suggested Disclosure Pointers
How User Data Will Be Used and Protected *Deployers may link to their full privacy policies to provide more details*	**What Data is Collected:** ▸ The types of data collected (e.g. conversation history, metadata, usage patterns, device information) ▸ Whether the chatbot accesses data from connected apps or services, and if so, what that data is ▸ How data is stored and protected ▸ The data retention period
	Who Has Access: ▸ Whether data is shared with third parties, and if so, for what purposes (e.g. service provision, analytics)
	Whether Data is Used for Model Training: ▸ Whether user interactions (e.g. prompts, outputs) or other user data are used to train or fine-tune models
	User Control Options: ▸ Whether users can opt out of having their data used for model training or fine-tuning ▸ Whether users can request the deletion of their data ▸ Other controls users have over their data (e.g. ability to export conversation history)
How Users Can Report Issues	**Support Channel:** ▸ Contact details for users to raise issues (e.g. support email) ▸ The types of issues that can be reported (e.g. harmful or inappropriate content, privacy concerns)
	Response Process: ▸ How and when user reports will be acknowledged ▸ How the deployer will investigate and resolve issues, and keep users informed of the outcome ▸ How user reports will be used to improve the chatbot, including preventing harmful or inappropriate content

ANNEX B: SAMPLE CHATBOT INFO CARD

The following sample illustrates what a chatbot info card might look like in practice. It is based on TalkPAL, a fictional general-purpose chatbot that can assist users with a variety of tasks, such as content creation, web search, translation and image generation. As a general-purpose chatbot, TalkPAL is designed for open-domain interactions and is available to the public, including younger users.

The sample chatbot info card demonstrates how the disclosure and design practices set out in these guidelines can be applied. It **illustrates the core areas of disclosure encouraged earlier and is intended as a reference, not a mandatory format.** It shows just one of many possible formats, and deployers can adapt it, reorganise sections or incorporate interactive elements to suit their audience and enhance usability.

Figure 1 - Overview of Chatbot Info Card

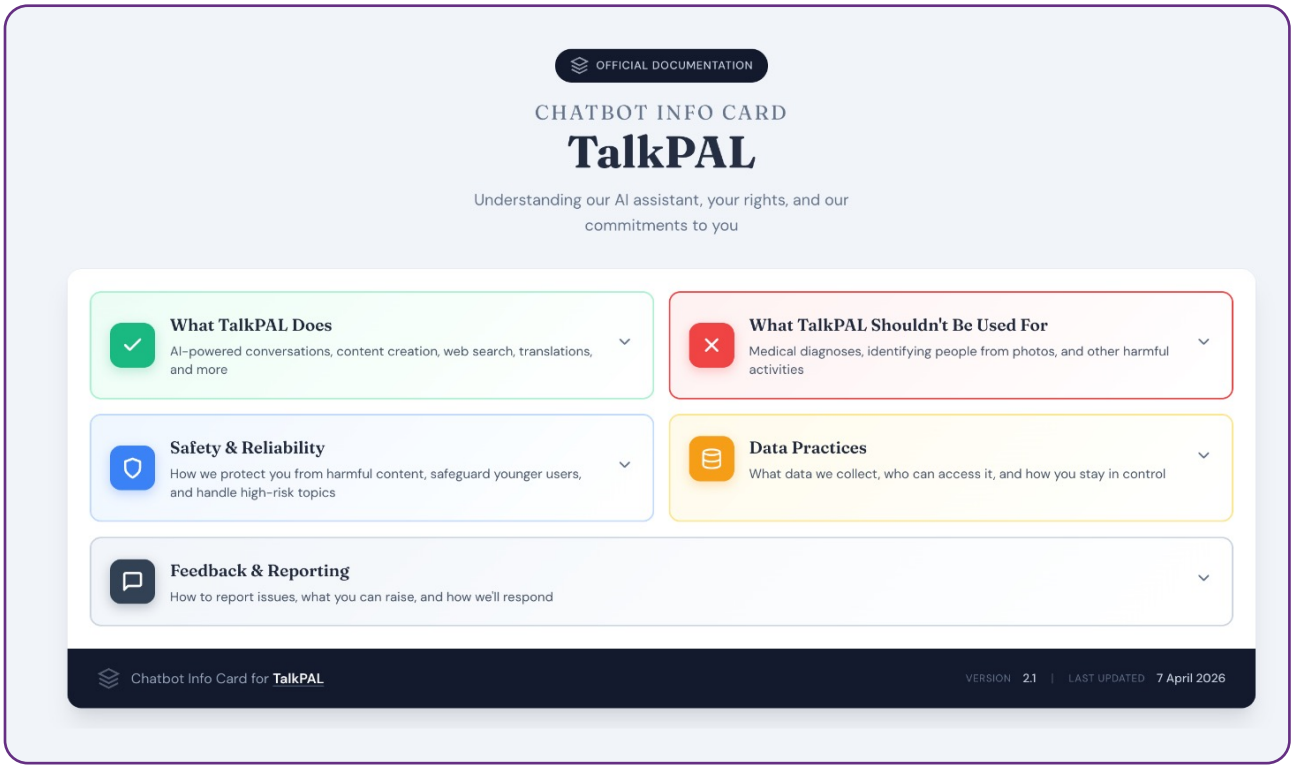


Figure 2 - What TalkPAL Does

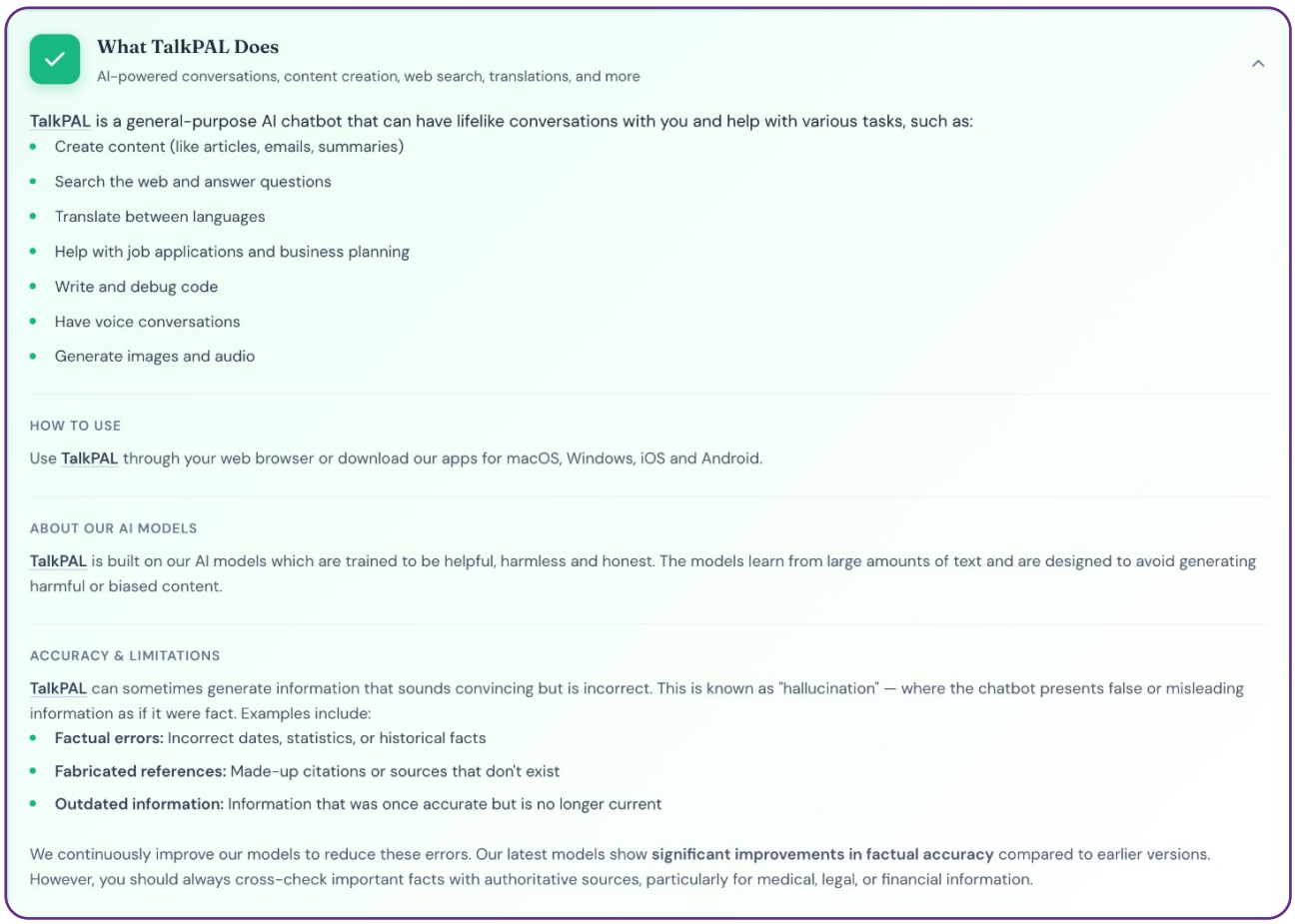


Figure 3 - Prohibited Uses

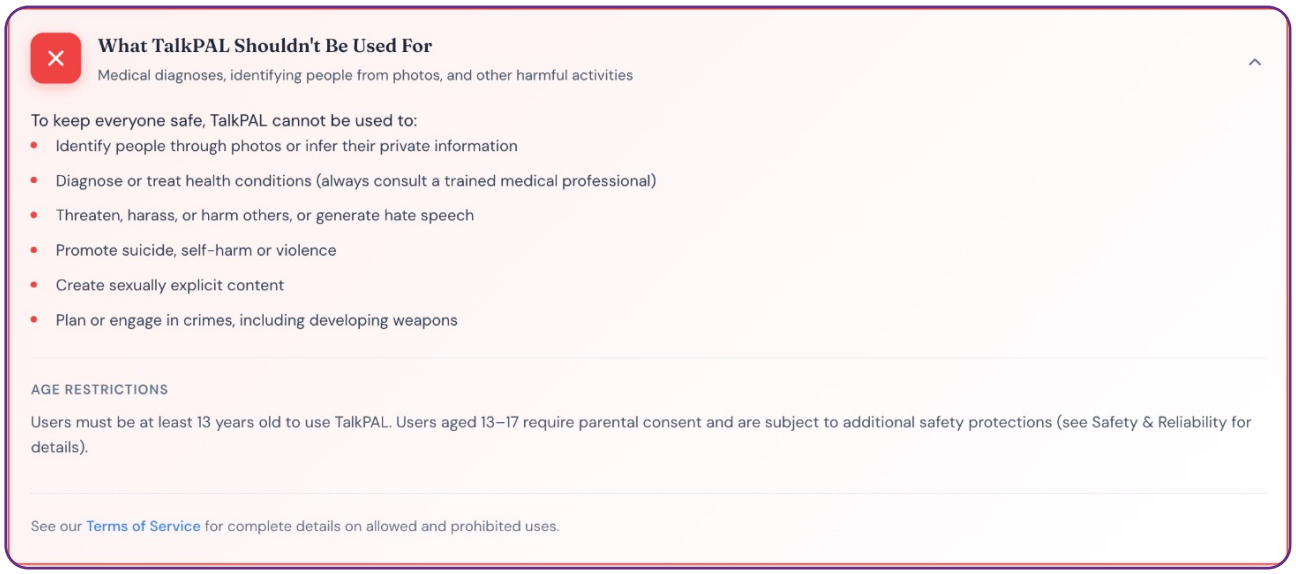


Figure 4 – Overview of Safety and Reliability Measures

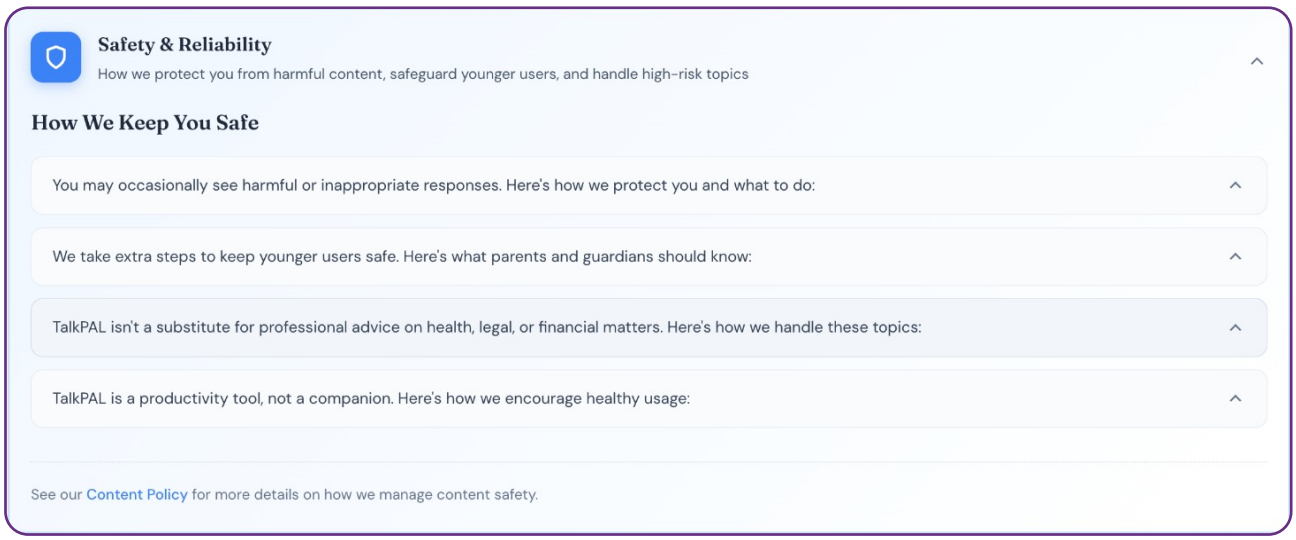


Figure 5 - Safety and Reliability Measures for Harmful and Inappropriate Responses

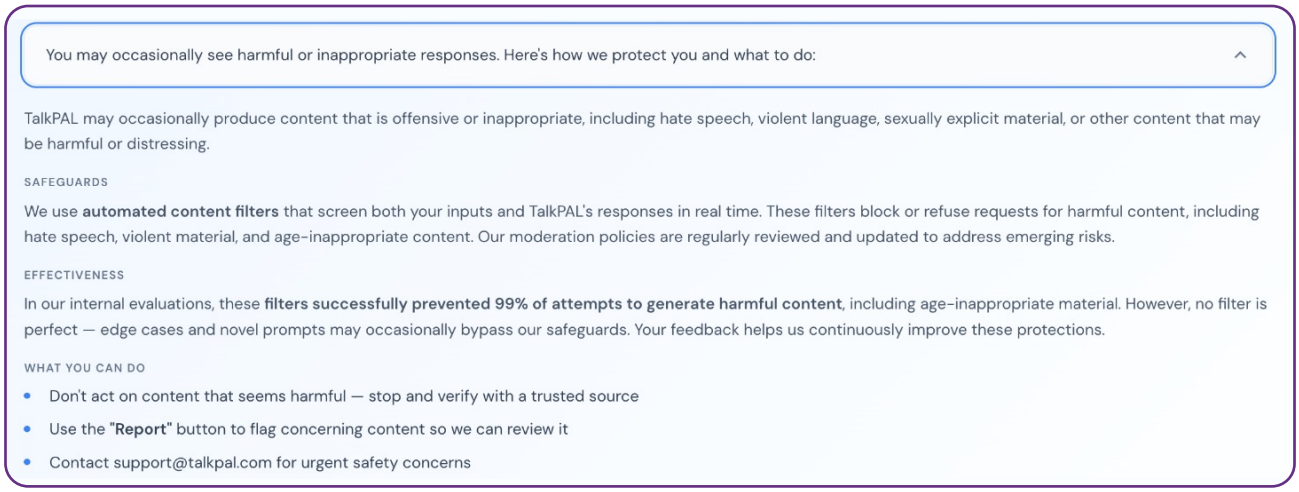


Figure 6 - Safety and Reliability Measures For Younger Users

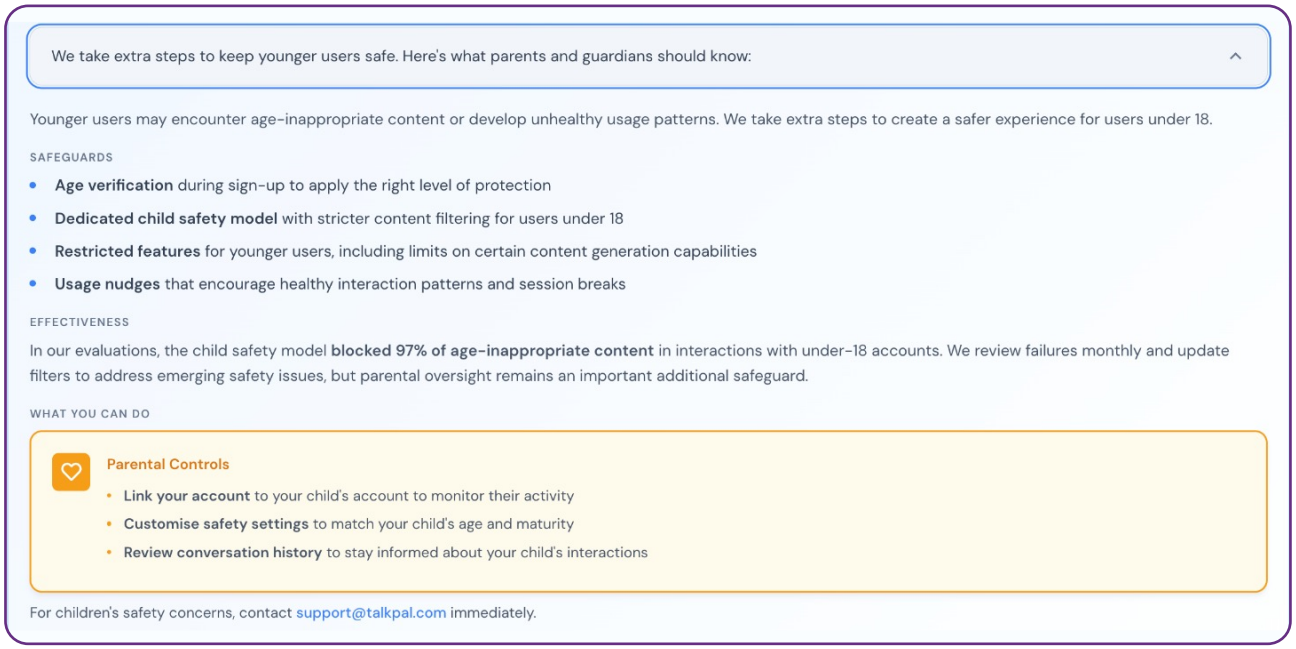


Figure 7 - Safety and Reliability Measures For High-Risk Queries

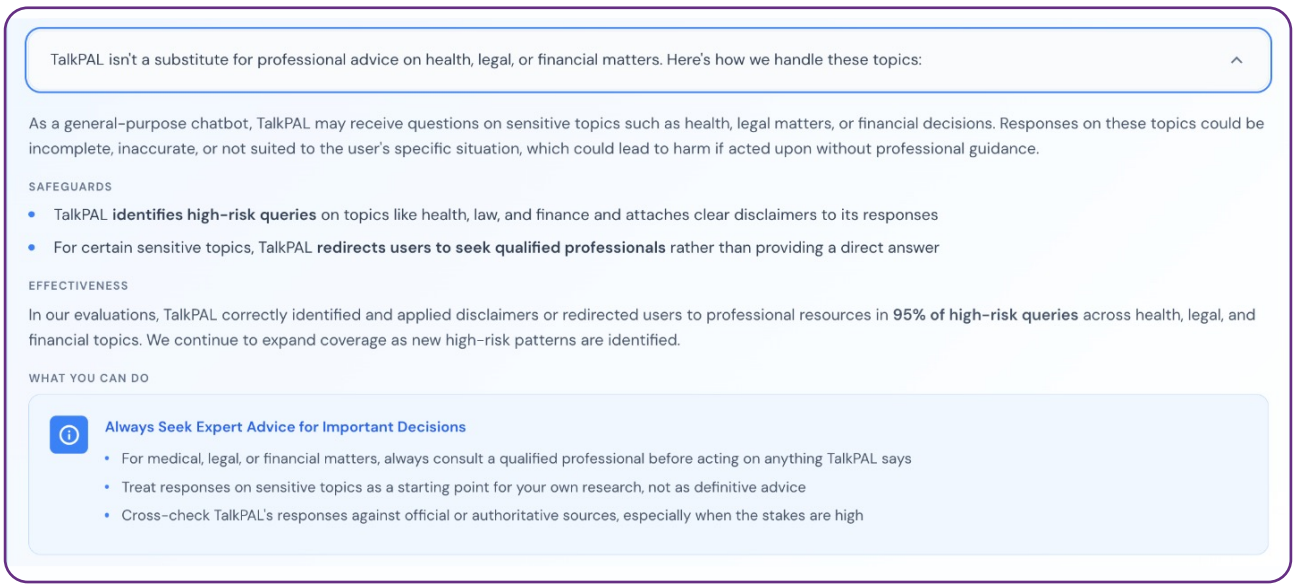


Figure 8 - Safety and Reliability Measures For Healthy Usage

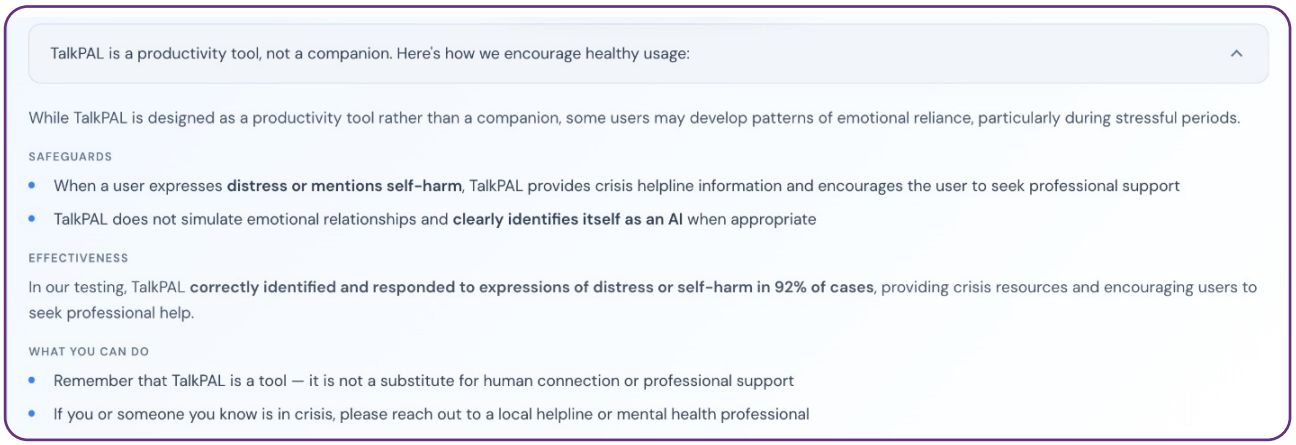


Figure 9 - Data Practices

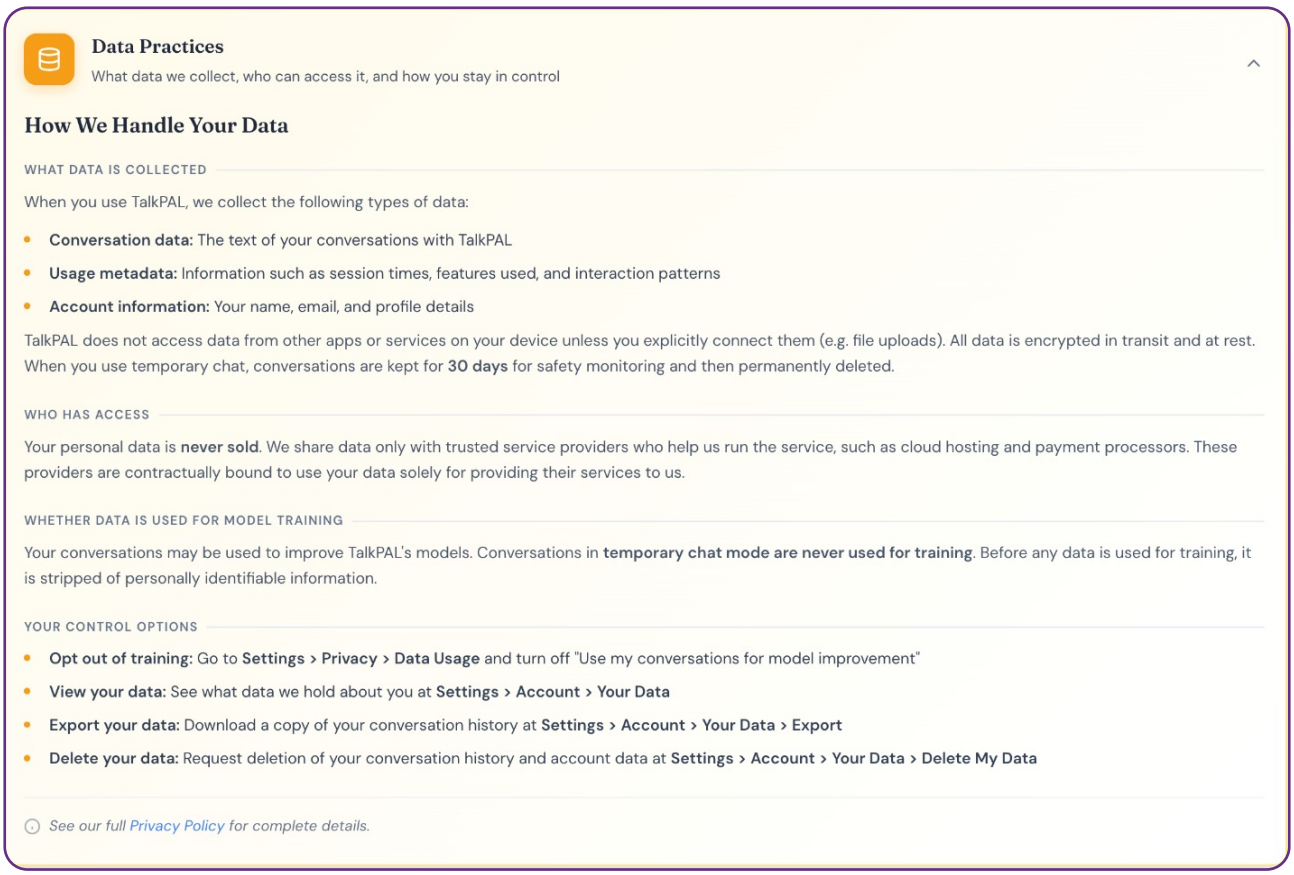
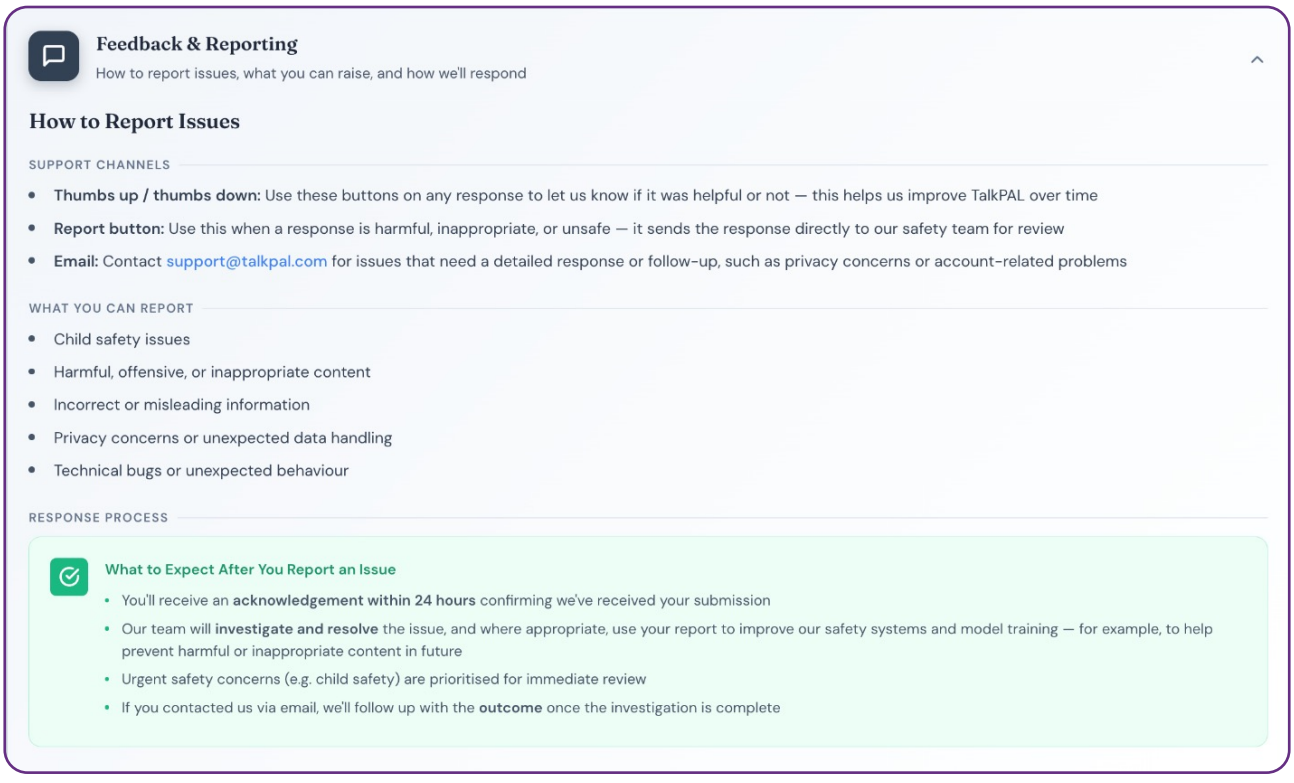


Figure 10 - Feedback & Reporting Channels



ACKNOWLEDGEMENTS

We extend our sincere gratitude to the organisations that supported the development of this document by sharing valuable feedback and participating in consultations.

Industry

Alibaba Cloud

Alibaba International

Amazon Web Services

AramGRC

AvePoint

Bridge Telco Alliance

ByteDance

Changi Airport Group

China Mobile

CyberSG Research & Development Programme Office (CRPO) - SMU

DBS

EY

Google

Grab

Ikano Pte Ltd

Impress.AI

Kindata AI

Lazada

Lumen Technologies

Mandai Wildlife Group

Meta

Microsoft

MWG

NHG Health

NTU

National University Health System

OCBC

OpenAI

OpenRouter

Resaro

Resorts World Sentosa

Salesforce

SEEK

Senoko Energy

SGS

SIA

Singapore Press Holdings

SingHealth

ST Engineering

Standard Chartered Bank

StarHub

Synapxe

TAMBI

Telstra International

Tencent

Trusted Aide

TVLK

Workday

WPH Digital

X0PA AI

YAGHMA B.V.

Government Agencies

Government Technology Agency of Singapore (GovTech)

Health Promotion Board (HPB)

Intellectual Property Office of Singapore (IPOS)

Land Transport Authority (LTA)

Ministry of Digital Development and Information (MDDI)

Ministry of Health (MOH)

Monetary Authority of Singapore (MAS)

National Library Board (NLB)

Singapore Land Authority (SLA)


INFOCOMM
MEDIA
DEVELOPMENT
AUTHORITY