Ten questions you should ask your AI vendors to ensure maximum data security

Assessing your potential AI vendors for security is crucial. Security is an organization-wide endeavor — and this includes evaluating everyone within the ecosystem, and ensuring there are no security gaps so your customers are protected.

Here's a handy list of the top questions you can ask your vendors to make sure you're implementing secure AI.

What's your approach to model training, retraining, and maintenance?

Training your models is a crucial step in implementing AI solutions. In fact, at Language I/O, we'd go so far as to say this is the most important step.

The effectiveness of your AI depends on the quality and integrity of the data it's trained on. As such, it's crucial to ensure that the data consumed is diverse. AI needs large volumes of all kinds of data, and this means there are many ways human biases and inaccuracies could creep in.

For the model to be fully reliable, you need to eliminate such biases and issues. Rigorous data curation and ethical oversight are essential to harness the full potential of these models while ensuring fairness. This is especially crucial in customer service: Your AI needs to make all your customers feel welcome, included and supported.

At Language I/O, one of the approaches that we use to ensure that our multilingual AI solution is consistently accurate and secure for your business is by employing a diverse red team. The red team is a group of diverse experts whose main objective is to "break" the AI. Many times, the AI will surprise us with output that we didn't anticipate, like swearing like a sailor or coming up with completely fabricated company policies, which are then addressed by the development team.

How do you handle data encryption?

There are several vulnerabilities along the way when you implement AI solutions, and one way to tackle these is by data encryption.

AI solutions like ChatGPT have been known to retain and leak personal information. Online AI solutions like DeepL also make it clear that their free service should not be used for professional purposes. Consider this text on their site: "We reserve the right to process the content you upload (e.g., your texts, documents) and its translation/improvement for a limited period of time to train and improve our neural networks and algorithms."

To top it all, employees who are not trained in keeping data secure are liable to put confidential customer information into such online tools. You won't believe the number of teams that tell us that they plug entire customer conversations into Google Translate or ChatGPT, personally identifiable information (PII) and all.

And this is indicative of an alarming trend: 1 in 4 employees see no issue sharing personally identifiable information with AI.

At Language I/O, we guard our customers against such vulnerabilities by masking PII in transit: Even before the real-time translation is performed, we automatically detect and encrypt PII to maintain maximum security.

How do you ensure that the AI is accurate?

AI solutions are prone to hallucinations and misinformation. In one of a string of law-related incidents, a lawyer used ChatGPT for research and ended up with many false cases and dockets that the AI had conjured out of thin air.

Especially in translation, this kind of behavior from AI quickly becomes a problem. For example, if your brand is called Cloud, how would the AI know to keep it as Cloud or translate it to the word "cloud" in another language?

This is where training and domain adaptation come in. Domain adaptation is how the AI becomes aware of the context and is prepared for accurate processing of industry-specific and brand-related terminology so that there's no miscommunication and errors conveyed to customers.

One way to ensure high accuracy in AI is to go for a multi-engine approach. Language I/O, for example, aggregates the world's leading models and intelligently selects the best model for each conversation depending on the domain, language pair, etc. This means that depending on the context, the strongest model is chosen, and you get the most accurate response.

Language I/O also has a Self-Improving Glossary (SIGLO) that it imposes in real time during customer conversations. LLMs are trained on all kinds of data and don't always have the exact context. This is where we bring in the context and ensure reliably accurate translation.

Do you store data? If so, what data, for what purpose, and where?

Asking AI vendors about their data storage practices is essential for ensuring the security and privacy of your sensitive information, mitigating risks and preventing abuse of data.

Many AI solutions might promise security and data deletion, but still store all the customer information, presumably to train their software. If this information includes PII, any business working with these translation solutions is not safe from the risk of data breaches. No one wants their confidential data languishing in some server somewhere in the world, waiting to be breached.

It's also important to know where and how your data is stored: This helps with ensuring compliance with relevant data protection laws and industry regulations. This also allows you to ensure that your data isn't shared with unauthorized parties or competitors.

How do you ensure compliance with data protection regulations? What security certifications do you hold?

Navigating AI complexities in the age of tightening regulations requires businesses to be agile but also adhere to global security laws and standards. To keep up with your legal obligations and to eliminate risk of noncompliance, you must ensure your vendors are complying with global data protection guidelines.

Several organizations like ISO and PCI clearly lay out rules to ensure customers have more control over what data they're providing and how it's stored. ISO 27001, for example, ensures that data confidentiality and integrity are protected. These global guardrails ensure that you stay ahead in the regulatory landscape while also being prepared to pivot in case new regulations emerge.

International and local authorities in different geographies also have their own set of rules. Consider GDPR, for example: Known as the strictest data security regulation, GDPR is aimed at European organizations that process personal data of EU residents and organizations outside the EU that target EU residents. It sets out rules to ensure that organizations don't collect data more than necessary and governs how that data is collected, handled and stored.

Fun fact: Language I/O is the only translation platform that is compliant with the latest ISO 27001:2022 standards. See our full certifications and compliance information.

What data access controls and authentication measures are in place?

Not everyone in your organization needs access to all your data. Only the relevant, authorized personnel should be able to access your data. This helps in many ways.

- Prevents unauthorized users from viewing, modifying, or deleting confidential information
- Protects trade secrets, customer details, and other proprietary data from leaks or misuse
- Minimizes the potential for data breaches and cyberattacks by limiting who can access sensitive systems and information
- Prevent insider threats, whether malicious or accidental, by restricting data access based on users' roles

Many regulatory standards also require companies to have access control in place to be compliant.
At Language I/O, data provided for translation is not accessible by personnel or systems not intended for processing. We follow a least privileges policy and only permit authorized personnel who need access to areas containing relevant data.

Do you use third-party solutions, vendors and subprocessors? How do you vet and monitor them?

Not everyone in your organization needs access to all your data. Only the relevant, authorized personnel should be able to access your data. This helps in many ways.

- Prevents unauthorized users from viewing, modifying, or deleting confidential information
- Protects trade secrets, customer details, and other proprietary data from leaks or misuse
- Minimizes the potential for data breaches and cyberattacks by limiting who can access sensitive systems and information
- Prevent insider threats, whether malicious or accidental, by restricting data access based on users' roles

Many regulatory standards also require companies to have access control in place to be compliant.
At Language I/O, data provided for translation is not accessible by personnel or systems not intended for processing. We follow a least privileges policy and only permit authorized personnel who need access to areas containing relevant data.

Do you offer any tools or dashboards for us to monitor our data usage and security?

Having access to tools and dashboards that allow you to measure your data usage is the first step of proactive security management.

This also ensures that the vendor you're using is auditable: You should be able to generate audit trails and real-time reports whenever you need them to ensure regulatory compliance at any point in time. Having this kind of visibility into their data practices, usage and management will also ensure that you can mitigate any security risks faster. This gives you greater control over your data.

The ability to monitor your data usage and security also ensures that you know where you need to optimize. It helps you identify threats and anomalies as well as data overuse or misuse and inefficiencies.

What are your disaster recovery and business continuity procedures?

The global average cost of a data breach is nearly $4.9 million this year (a 10% spike compared with 2023). Very few businesses recover completely from these breaches: On average, it takes more than 100 days for most of the small number (12%) of breached organizations that were able to fully recover.

In spite of all the preparation you do, sometimes things go wrong. In that case, the sooner you discover and act, the more damage reduction you can do.

Vendors should have comprehensive strategies in place to handle any disruptions. Such procedures also show you how resilient a vendor is and if they're equipped to maintain service during crises. Any issues on the vendor's end could affect your own business continuity, so this is immensely crucial.

Their answer to this question also reveals what steps the vendor takes to safeguard your data in the event of a disaster and allows you to measure their ability to recover and restore your data quickly and completely.

How quickly can you inform us of a data breach involving our data?

One of the biggest expectations your customers have of brands is that they're notified immediately if their privacy is compromised.

In fact, legally and ethically, this kind of breach reporting is required as part of several regulations, and there are firm deadlines in place as to how soon you must let your customers know. The general rule is to avoid unreasonable delays and inform your customers within 72 hours so that they can take the necessary measures to secure their data as soon as possible.

Ideally, your AI vendor would be on top of the situation and notify you within 24 hours of identifying a data breach. They would also have strong disaster recovery protocols in place that would enable them to act fast.

Ultimately, this is all a matter of trust. What you're looking for is your vendor's willingness to share such critical information. Open communication about these procedures builds trust and confidence in the partnership.

Here at Language I/O, we strive to give our customers and their customers industry-leading security. That's how we earned the reputation of being the most secure AI-powered translation platform on the market. Learn more about our commitment to building secure AI.

For a more detailed list of questions you could ask your AI vendors, download our guide Data security in an AI-driven landscape: A practical guide for CX leaders.

For more information, reach out to your Language I/O CSM or request a demo now.