



# Taking the risk out of Conversational AI

How to stop losing sleep and start getting your data on point for CAI





*A Conversational AI (CAI) solution has a lot of moving parts, but one of the most important elements to get right is the data. Inaccurate, flawed, or biased data can result in CAI that's more business risk than solution. Find out more in this eBook about how to create a winning data strategy to ensure the success of your Conversational AI initiatives.*

## Is your CAI strategy keeping you up at night?

It was 4am, and Oliver had just finished work—a day that started at 6am the morning before. His company was in the process of rolling out a new chatbot, but with his team already stretched thin, they were struggling. Complex compliance requirements and the need to support multiple languages was hanging over his head, creating additional challenges with the project. With a launch date set in stone due to an upcoming industry event, Oliver wondered how many more all-nighters his team would have to pull in order to meet the deadline.

Whether you're working on Conversational AI (CAI) in-house, or you're a technology vendor building custom CAI offerings for clients, your challenges can be significant. Not only do you have to ensure that your solution results in a great user experience, you also need to make sure that it performs efficiently, is cost-effective, and delivers competitive advantage.

And then there's the data. At the core of any CAI initiative is the data that fuels the success of AI-powered applications. It must be gathered, annotated, and tested to effectively inform the understanding of intent, the prediction of response, and deliver natural, human-like responses and recommendations.

Data provisioning for CAI is a complex function, and most companies don't have dedicated in-house staff with a track record of expertise in creating highly targeted CAI datasets. In this eBook, you'll learn how to build a data strategy that will help you get the results you need from your CAI solution.



# Conversational AI 101: The basics

Conversational AI brings together natural language processing (a subfield of artificial intelligence describing the interaction between human language and computers) and software to produce realistic, human-like interactions between people and computer applications. Today, it's most commonly experienced in the form of chatbots, voice assistants, and virtual agents.

Delivered through a spoken or typed interface, CAI uses data, algorithms, and user input to understand conversation intent, decipher language, recognize context, predict behaviors, and respond in a natural way that is familiar to the user. The best part is that when used over extended periods of time, user input serves to further refine and “train” algorithms, resulting in improved accuracy and a better user experience.





# Who's using CAI— and how?

Today, the conversational AI market is valued at around \$6.8 billion (USD) and IDC forecasts that it will grow at a compounded annual rate of nearly 29% between now and 2025\*. With that kind of dramatic growth on the horizon, many industry segments are using CAI to position themselves at the forefront of their respective markets, including:

**Customer service** uses virtual assistant solutions to decrease overhead with automated customer support

**Healthcare** uses CAI for front-office automation, filling in forms, managing patient history, updating records, and sending alerts about appointments and prescription refills

**Ecommerce** uses CAI to add value to the customer experience, capturing and qualifying leads more rapidly, providing a consistent omni-channel presence, and responding faster to customer inquiries and issues

**Manufacturing** uses chatbots to automate sales, marketing, and support functions, and to eliminate communication delays that can result in critical supply-chain issues

**Travel and hospitality** uses CAI to provide 24/7 automated assistance for reservations, insurance, passport and visa information, and upgrades, and to customers who may require urgent assistance when in a far-away time zone

\* [International Data Corporation](#)



# Rolling out a CAI solution? Here's what you need to know

As with any major product launch, your CAI initiative has to start with a solid plan to get you from idea to deployment. Beyond the obvious, such as feature set and functionality, there are certain core elements that must be included in every CAI strategy to ensure that it not only meets technical requirements, but also business goals.

## Approved datasets

Where does your data come from? Using free, easily accessible unstructured data from sources like social media, web pages, mobile devices, and so on, creates the risk of a CAI solution using or revealing hidden and sensitive information that may be embedded in data considered to be anonymized. By ensuring that your CAI solution uses only data that you have the legal right to use, you can reduce or eliminate liability that comes from using risky data to train CAI algorithms.

## Compliance

There are many potential points of exposure for user data, and with government and industry regulators shifting the compliance goal posts on a regular basis, a great rule of thumb is to always prioritize compliance. From GDPR and HIPAA to SOC2 and beyond, the use and storage of sensitive information related to identity, health, online activity, etc., is being watched more closely than ever. Have a compliance strategy that's robust and ready for today and tomorrow; retrofitting infrastructure to accommodate unforeseen compliance issues can be tricky—and expensive.

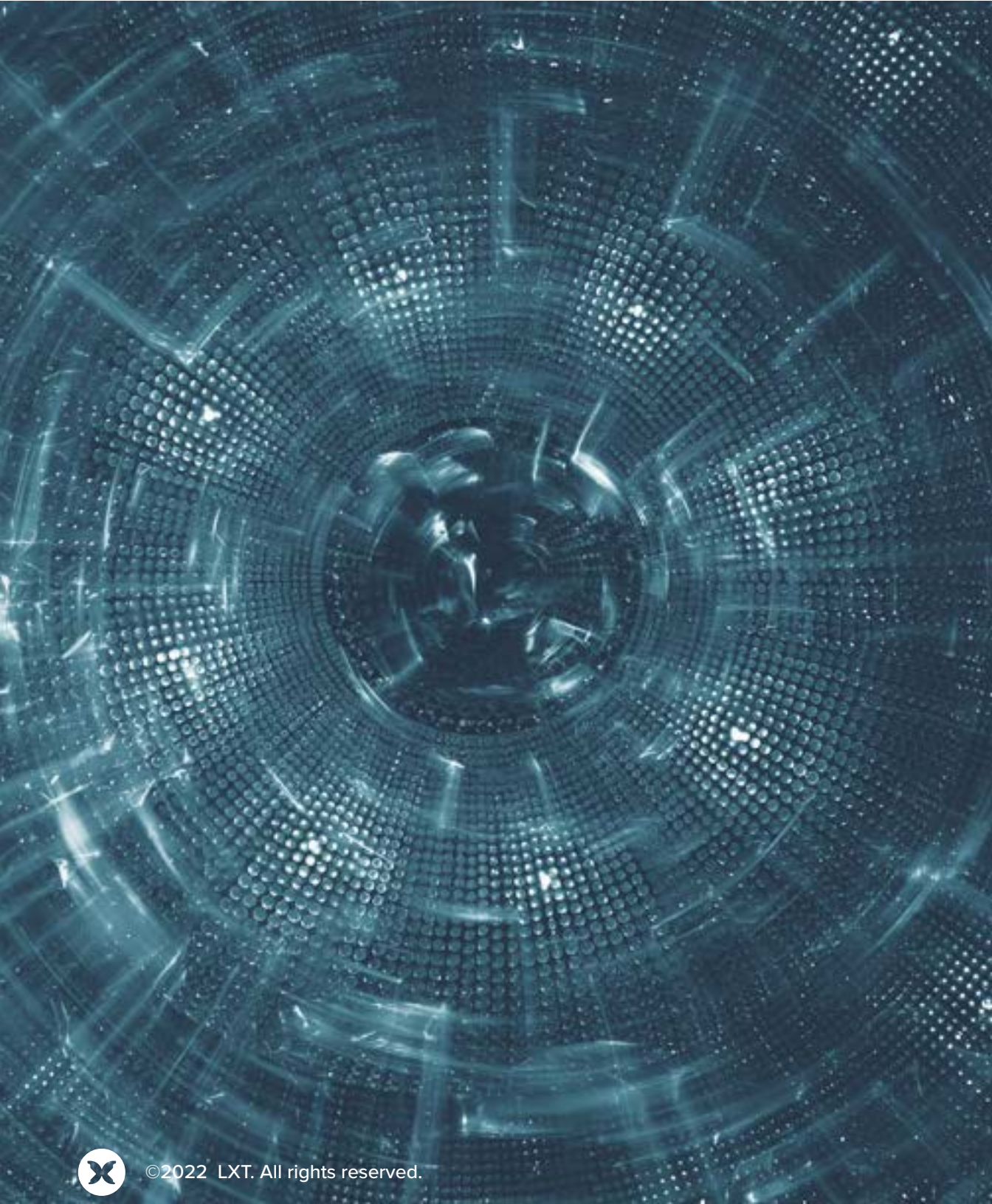
## Language support

Research has shown that 40% of people prefer self-service to human interaction, and providing users with human-like, in-language conversation experiences is key to meeting this need head-on. The beauty of CAI for multilingual support is its ability to detect a user's language via location (IP address), browser settings, or HTML attributes. Some CAI solutions use translation engines to respond to users in their preferred language, but others may require that responses be translated and fed to them in advance. Look for a partner that can provide multilingual support using data from native speakers for the best results.

## Integration

When we talk about integration, we have to think about it in two ways: the IT backend, and the user experience. From an infrastructure standpoint, make sure that things like related customer care platforms, backend systems, and knowledge bases all work seamlessly with the CAI solution to ensure that users can easily opt into self-service and not be forced to “hold for an agent,” or wait for a return call or email.





## **Cost savings**

The overhead costs of sourcing quality data from native speakers and managing the overall volume of data required to train algorithms on an ongoing basis can be significant. In attempting to build and deploy a CAI solution, you may find that those costs pile up quickly and can easily turn a straightforward project into a money pit.

## **Privacy**

Along with compliance, privacy is another hot-button topic when it comes to CAI. Get clear upfront about how (and how well) your solution stores and uses data and explore options around isolated single tenant environments and on-premises deployments for data.

## **Algorithmic bias**

Instances of algorithmic bias in AI have become an area of increased focus in recent years. Left to their own devices, algorithms can replicate or amplify human biases due to narrow data parameters and limited inputs—ultimately delivering negative impact to protected or underserved communities. Working with a partner that brings a diverse network of contributors to the table can help you build a more robust, inclusive experience.

## **Scalability**

Are you prepared for a significant spike (or plunge) in business? As we've seen recently, unforeseen events can quickly create repercussions that transcend geography and industry. Design your CAI solution with scalability in mind so it can easily adapt to shifts in demand in real time.

## **Standardization**

When first starting out with CAI, it may be tempting to be overly conservative, creating a dedicated chatbot for a specific function and later adding another one (or a voice assistant, or a virtual agent, etc.) in response to a new need. Instead, build out a single, all-inclusive platform that will support multiple CAI solutions that aid all users and offer options to maintain future requirements. This standardization strategy reduces maintenance costs and adds value by serving as a single source of data insight.

## **Market access**

Rolling out your CAI to new markets is key to increasing user adoption and satisfaction. With one CAI solution successfully deployed for a specific language or region, you'll be well positioned to expand into additional markets quickly. However, entering into new markets often opens the door to new risks and pitfalls, particularly if these are markets where your company doesn't have prior go-to-market experience. Working with a data partner with proven global reach and expertise in your target markets can help you avoid the hazards of new market entry, and offers a smooth experience for both you and your end users.



# Your data strategy

There's no way of getting around this one core fact: your CAI solution is only as good as its data. Investing in and developing a high-quality dataset that gets more and more precise over time is priceless.

Here are some expert tips for ways to approach your data strategy during product development:

**Design** – This five-part process will include the following: 1) Identifying the type of data you need; 2) Identifying a resource for getting your data; 3) Ensuring data quality; 4) Diversifying data sources and re-training data; and 5) Developing guidelines for how the algorithms will be trained.

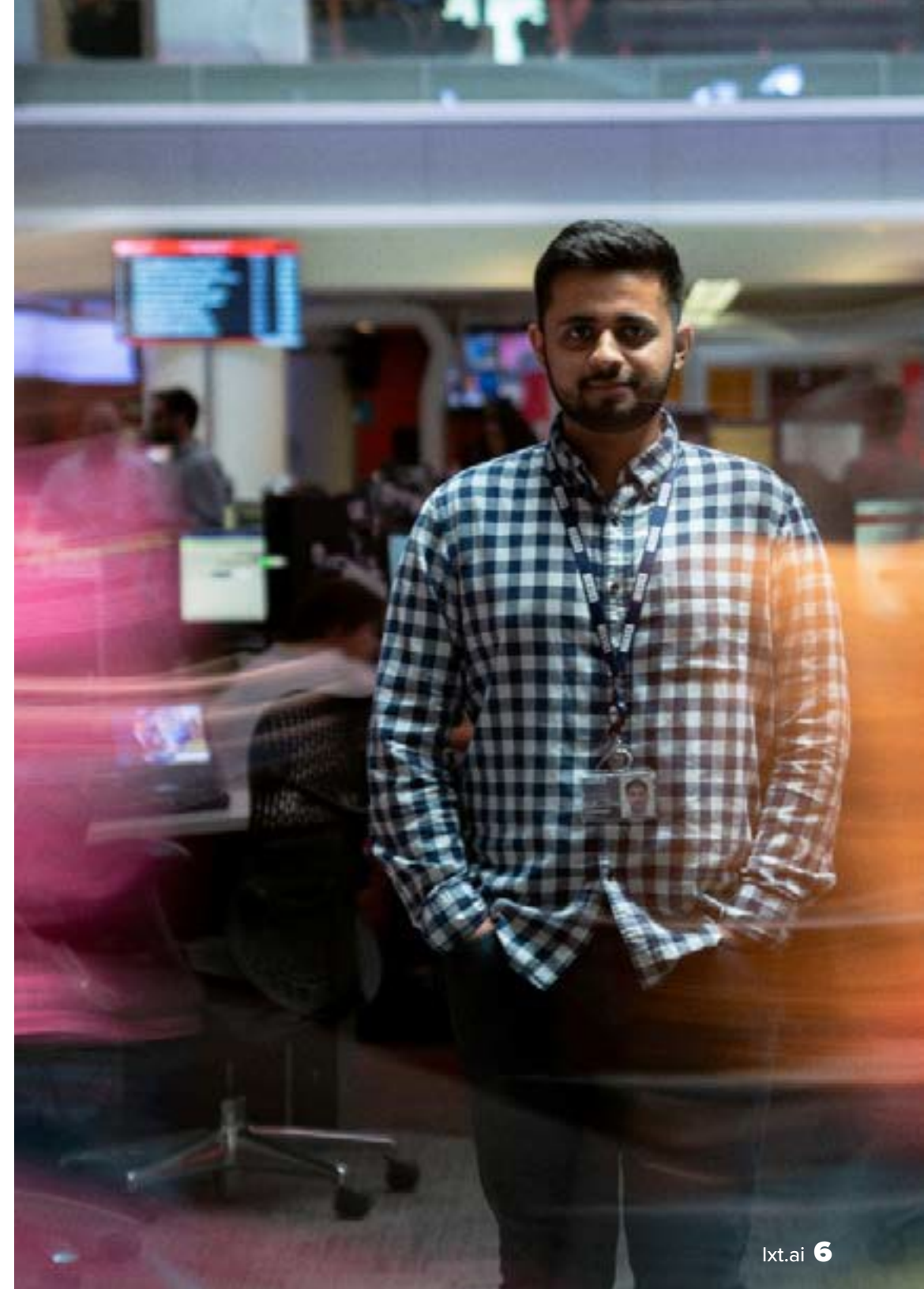
**Data collection** – If you don't have access to substantive data in-house or from a close partner, vet third-party providers carefully, examining their professional reputation, the sources of their data, and the scale and quality of their available data.

**Data organization** – Labeling and organizing your data can be done either in-house, or outsourced—and there's a case to be made for either method. Doing it in-house has the advantage of easy insight into user behaviors and preferences through proximity to the data; outsourcing reduces the risk of bias and can better support growth in data requirements and rollout in multiple languages.

**Training** – Establish your pipeline and seed it with enough data samples to enable the algorithm to analyze the dataset and start training itself. Some elements to consider when building out your training dataset include the languages and dialects the application may encounter, the ability to imitate human language, support for diversity and flagging potential bias, privacy and security, quality, and identifying human intent (more on this below).

**Application deployment** – Establish a target data accuracy rate (impact for users, business goals, etc.) that will serve as the benchmark for deployment. Once this rate has been reached or satisfied, the application can be rolled out either into a production environment, or directly released into the market.

**Live data collection for continual improvement** – After your application is deployed, collect and annotate actual live user data and funnel it back to the product development team to use it to re-train and refine the algorithms, improve the quality of the application, and make it more robust and accurate over time.





## More about training your ML algorithm

As we outlined previously, training your machine learning algorithm is an iterative process that works to improve the efficacy and value of your CAI application over time. The algorithm begins training itself by identifying the patterns and relationships in your data. As it works, it will develop a series of educated assumptions to help it make predictions. Using what it's learned, the algorithm will be able to make more accurate decisions and generate the most application-appropriate and valuable responses for you and your end users.

It's important to annotate and correct your algorithm's recognition hypotheses on a regular basis to keep your CAI solution continually improving and delivering the results you need. Depending on the nature of the data or your industry, you'll want to schedule annual, quarterly, monthly, or weekly updates (or even more frequently for certain rapidly changing markets).





# The risks of doing it yourself

As we've already discussed, there are a number of hot buttons when it comes to data, and if issues like data quality and collecting, transcribing, and annotating data to create compliant data sets aren't fully and carefully addressed by an experienced resource, the true cost of a failed CAI execution can be catastrophic.

Working with an experienced data training partner can take the worry out of developing a CAI solution. Having an expert resource by your side to help with strategy, data collection, solution development, and ongoing enhancements can help you avoid risk and make your Conversational rollout smoother and more seamless.





# LXT: Your partner for CAI training data

At LXT, we live and breathe data. We're obsessed.

Our one and only mission is to power the technologies of the future through data generation and enhancement across every language, culture, and modality. We're here to help you transform your business with Conversational AI.

## Our expertise

### Data generation & collection

Our network allows us to generate data anywhere in the world. We'll tailor what we collect to your requirements and rapidly bring it up to scale.

### Annotation & enhancement

Whether you need human insight into speech, emotion, images, or anything else, our training data services provide you with useful data to feed your models.

### Technology

Our secure platform delivers high-quality data by facilitating human insight to improve accuracy. It's intuitive design streamlines workflow to reduce costs and optimize turnaround times.

### Security

Our technology and solutions are GDPR, OWASP Top Ten, and HIPAA compliant. All solutions are protected with 256-bit SSL encryption, while maintaining flexibility and a fluid user interface.



# Delivering AI data to power innovation

We have the support, resources, and facilities to develop bespoke, targeted datasets that support all your CAI initiatives:

**Customization** – When it comes to data, one size never fits all. We work closely with you to understand your business and your users and design and implement a data strategy that helps you achieve your CAI goals.

**Agility** – Our proven best practices enable us to build out your custom strategy in weeks, rather than months.

**Language support** – We offer multilingual support, covering more than 750 language locales across more than 115 countries around the world.

**Secure facilities** – We operate five secure facilities around the world. Our facilities are ISO 27001 certified, PCI DSS compliant, and offer supervised annotation to safeguard customer data. Data can't leave your facility? No worries, we'll come to you. Our experts will train and work alongside your team, implementing solutions that integrate with existing infrastructure.



# LXT customer spotlight: Dubber

## An eye on growth

Dubber is a provider of cloud-based call recording and voice AI. Over 170 mobile networks and service providers worldwide partner with Dubber to provide scalable conversational recording and intelligence to their customers.

As international expansion is a key strategic goal for Dubber, ensuring that its products work effectively in 10 languages is a top priority. It was important to the company to have the most accurate and bias-free benchmarking data, collected and annotated by a reliable partner that could ensure on-time delivery for the upcoming Notes by Dubber launch at a leading industry conference.

## Benchmarking data

LXT worked hand in hand with the Dubber team to develop comprehensive speech collection and transcription guidelines that created a strong foundation for the project. The LXT team proceeded to screen project participants for language fluency and business experience, as the data collection task simulated a meeting environment which required participants to have a strong business background.

The Dubber team was given sample datasets for each language and dialect to establish data quality before full datasets were completed. Responding to feedback quickly, the LXT team delivered the high-quality data Dubber needed to complete its benchmarking across all languages and dialects in less than eight weeks.

## Impact

Within eight weeks, LXT delivered accurate and bias-free annotated benchmarking data that enabled Dubber to launch Notes by Dubber in multiple languages at a recent industry conference. As a result of the successful launch, the company is working with several of the world's leading telecommunications companies to secure distribution agreements for Notes by Dubber.



*LXT is a key partner in our product expansion efforts providing us with a high-quality data we need to launch Notes by Dubber into new markets. The benchmarking data LXT generates for us is critical in our ongoing success.*

Iain McCowan, Director of AI, Dubber

[Read more about The LXT Dubber project here.](#)

## Key benefits

- Initial project scoped for 10 languages and dialects
- Project completed in 8 weeks
- Strong results led to project expansion into more languages



# Ready to help your Conversational AI learn and solve problems better?

If a Conversational AI project is in your future, you know there can be many hurdles on the path to a successful launch—the biggest of which involves data. If you don't have the dedicated resources to build a strong data strategy, annotate your core datasets, and train algorithms, you could be setting yourself up for failure.

Ready to take charge of your Conversational AI success?  
Get in touch with an LXT data expert today!

## About LXT

LXT is an emerging leader in AI training data to power intelligent technology for global organizations. In partnership with an international network of contributors, LXT collects and annotates data across multiple modalities with the speed, scale and agility required by the enterprise. Our global expertise spans more than 115 countries and over 750 language locales. Founded in 2010, LXT is headquartered in Toronto, Canada with presence in the United States, UK, Egypt, Turkey, and Australia. The company serves customers in North America, Europe, Asia Pacific, and the Middle East.

Learn more at [lxt.ai](https://lxt.ai) or contact us: [info@lxt.ai](mailto:info@lxt.ai)

