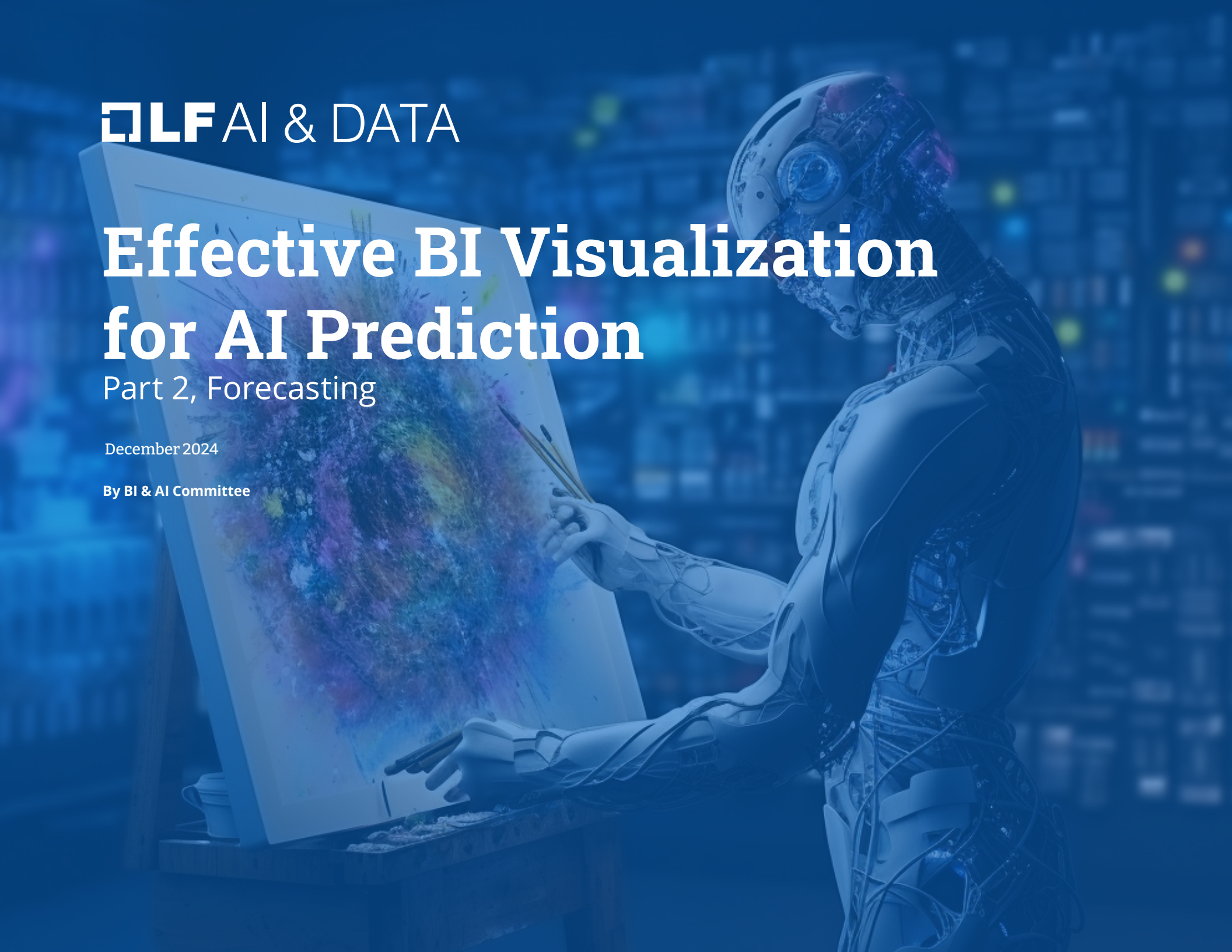
LF AI & DATA

Effective BI Visualization for AI Prediction

Part 2, Forecasting

December 2024

By BI & AI Committee

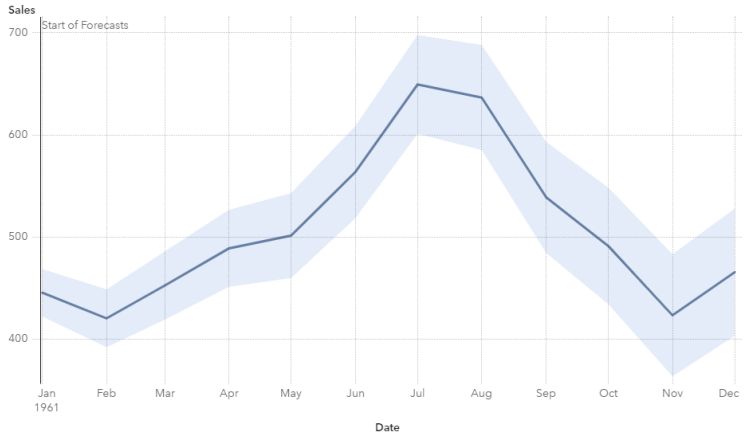


When we last left Sally Sue, she mastered visualizing continuous predictions. As she progressed in her data science career, she quickly learned that simply adding the element of time into a predictive model for continuous data creates a whole new field of study: forecasting. Forecasting is the process of predicting the future. In the business world, it is not uncommon for forecasts to be based on simple formulas, basic statistics, or even group consensus. Sally's been tasked with replacing those simple methods with a better AI model. No matter how it's done, the forecast ultimately comes to a value at a single point in time in the future, and Sally has to convey forecasts from a complicated AI model in a compelling way.

Forecasting is a very visual discipline: trends, patterns, seasons, and cycles over time can be measured statistically, but they must be reviewed visually to gain the greatest amount of insight. If the visuals are too complex, the numbers might get glossed over and ultimately end up unused. To make sure they're actually used by decision-makers, the following are six best practices Sally Sue uses for creating a simple yet compelling forecast dashboard. So you can play along at home, we'll be using the classic Box & Jenkins international airline data as our example with the forecasted data renamed to "Sales" for a more business context.

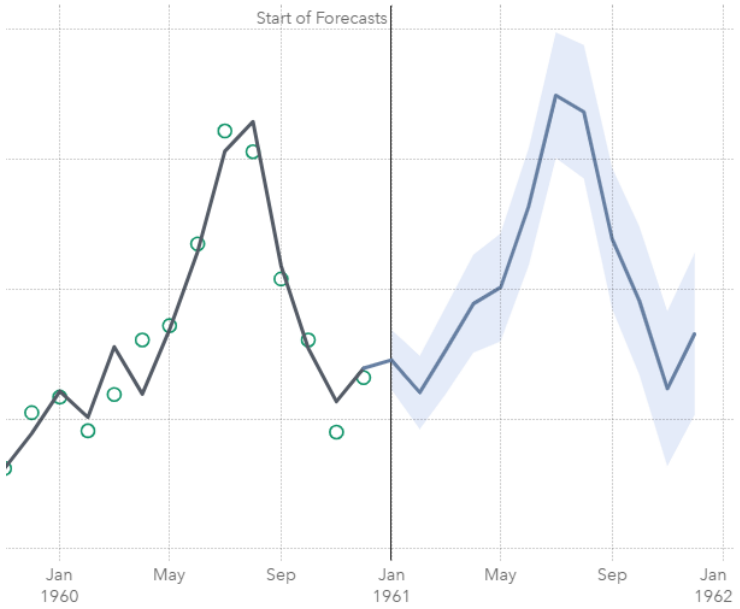
1. Always display confidence bands

Although every forecast is a single value at a point in time, there is always a level of uncertainty with each one. Whenever Sally was presented with forecasts, all too often she found that they were just a single number with no additional information. In reality, forecasts change quite often: even the most robust models are not immune to unexpected events. When Sally displays forecasts, she always includes confidence or prediction bands. These give decision-makers the ability to understand where the true value of the forecast may lie and give them a rough idea of how to prepare better for the future.



2. Separate fit from the forecast

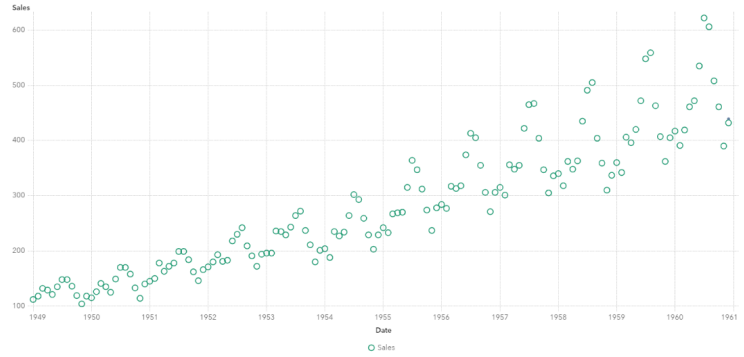
There are two parts to every AI time-series model: a fit and a forecast. The fit is how well the model predicts past data. This is important to display in order to help end-users detect patterns and instill confidence in your model. Both the fit and the forecast should be two distinct, separately colored lines. Since not all end-users understand what the fit is compared to the forecast, consider giving them the ability to turn off the fit and focus only on the forecast.



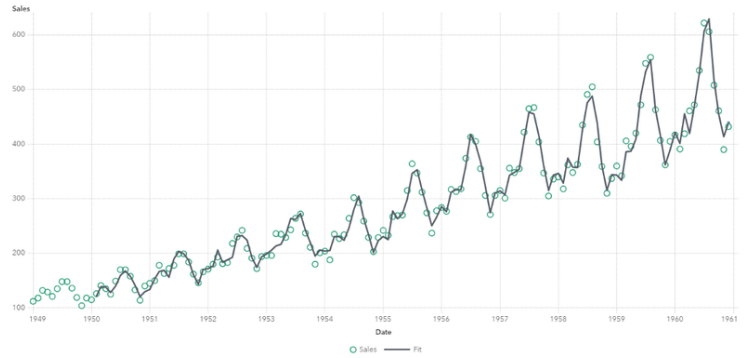
Model fit is how well your model predicts past data. Fit is important to display in order to help end-users detect patterns and instill confidence in your model.

3. Display actual or observed values as points

Actual or observed values should be displayed as points, ideally unfilled circles. Displaying both the fitted model and the actual values as lines simultaneously makes it difficult to perceive the two and causes the actual values to hide behind the fitted model or vis versa.

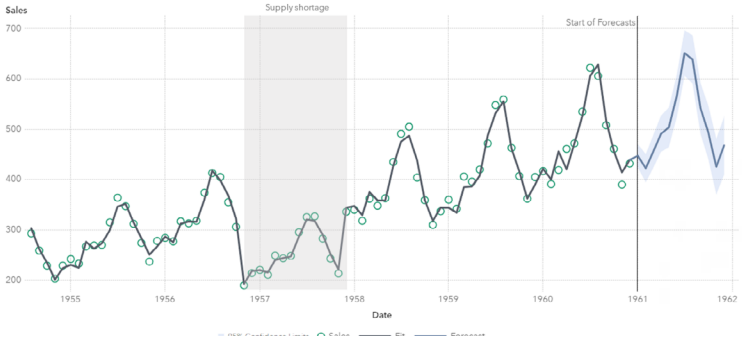


With the fitted model overlayed on top of the actual values, users get a clear sense of how well the model fits the data and allows them to perceive statistically identified trends, cycles, and seasons.



4. Selectively annotate outliers or significant events

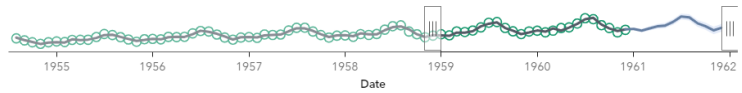
Outliers and significant events such as spikes, level shifts, or temporary changes should be annotated where appropriate. For example, an unexpected supply chain disruption that created a period of low inventory can be readily seen on the graph, but without annotation or background information, the end user may not understand why. Adding a simple annotation to the area via a tooltip, vertical line, or shaded region can help provide background information to the graph. Whenever Sally has people regularly asking her about a particular point in her graph, she considers annotating it; however, she annotates with care: too many annotations on a graph can make it unreadable.



5. Provide zoom controls

Allow the user to easily home in on any part of the forecast with zoom controls. These can be done via slide-based controls:

or through a window-based overview:



Optionally, consider adding a single button that allows the user zoom directly to the forecast in one click.

6. Always include a table

While forecasting is a visual discipline, not all of Sally's end-users are visual learners. She always has a table available for users to see the actuals and forecasts together. This table is minimal and includes at least the following information:

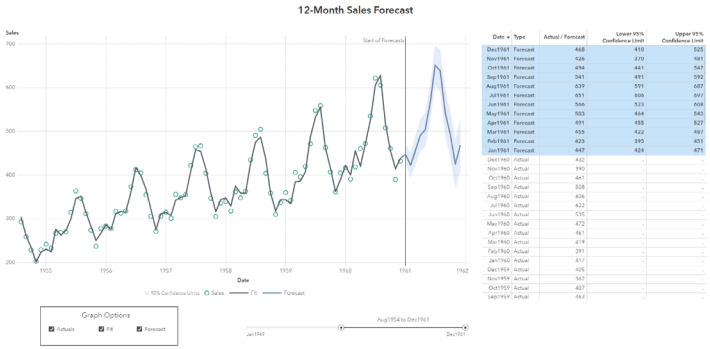
- Date
- Type: Actual or Forecast
- Forecast or Actual value
- Lower confidence or prediction limit
- Upper confidence or prediction limit

This table interacts with zoom controls, and users can download this table so they can quickly share it with decision-makers. It is also color-coded so that forecasts easily stand out from the rest of the data.

Date	Type	Actual / Forecast	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Dec1961	Forecast	466	403	528
Nov1961	Forecast	423	364	483
Oct1961	Forecast	491	434	548
Sep1961	Forecast	539	485	593
Aug1961	Forecast	637	585	688
Jul1961	Forecast	649	601	698
Jun1961	Forecast	564	519	609
May1961	Forecast	501	460	543
Apr1961	Forecast	489	451	527
Mar1961	Forecast	452	419	486
Feb1961	Forecast	421	392	449
Jan1961	Forecast	446	422	469
Dec1960	Actual	432		
Nov1960	Actual	390		
Oct1960	Actual	461		
Sep1960	Actual	508		
Aug1960	Actual	606		
Jul1960	Actual	622		
Jun1960	Actual	535		
May1960	Actual	472		
Apr1960	Actual	461		
Mar1960	Actual	419		
Feb1960	Actual	391		

Putting it all together

Using the best practices above, Sally puts everything together neatly onto a single page.



A Note on Underlying Factors

If your forecasts include significant underlying factors, also known as independent variables, add them to your dashboard on a separate page. Since independent variables are also often forecasted as well, you should follow these same best practices for displaying them with one exception: display the actual values as a line instead of the model fit. Your forecast is affected by the true underlying series rather than its fitted model. In this case, the actual historical values are very important.

Take advantage of pagination in your tool if you have multiple underlying factors so users can quickly flip between them. Order them from most impactful to least impactful, and consider adding a separate what if analysis page that allows users to create scenarios, run goal seeking, and interactively learn how each underlying factor affects the forecast.

Wrapping up Forecast Visualization

A simple, easy-to-read forecasting dashboard allows your end-user to consume the data without feeling flustered. All too often, forecasting dashboards experience information overload: hard-to-read graphs, excessive tables, lack of confidence intervals, and ambiguity about how well the model actually fits the data. By following these guidelines, Sally's business users and key stakeholders appreciate the simplicity and effectiveness of her dashboard, enabling them to make better decisions while still giving them the flexibility to export tables of detailed information for deep analysis.

Now that Sally Sue has mastered the art of forecasting visualization, she has a new challenge ahead of her that takes her far away from the world of structured data and continuous predictions: analyzing unstructured text. How will Sally handle this new challenge? Find out all about it in the next paper!

Forecasting is a very visual discipline: trends, patterns, seasons and cycles can be measured statistically, but they must be reviewed visually to gain the greatest amount of insight. If the visuals are too complex, the numbers might get glossed over and ultimately end up unused.

Authors

Cupid Chan

Cupid Chan, an industry veteran, holds the Senior Fellow position at CDHAI in Johns Hopkins and the University of Maryland. His career began in developing a top-tier BI platform and contributing to Linux Foundation ODPi. As Chair of LF AI & Data's BI & AI Committee, he drives the fusion of AI and BI, forging Cognitive Intelligence (CI), melding AI's speed with human-driven insights.

Deepak Karuppiah

Deepak Karuppiah, a Senior Architect in MicroStrategy's Augmented Analytics Group, merges AI and BI to craft technology stacks for AI-powered dashboards. Formerly, he specialized in secure data connectors for MicroStrategy. Proficient in machine vision and learning, he explores novel AI-BI intersections, evident through his role in the BI & AI committee. Beyond work, he volunteers technical skills to a local non-profit, showcasing dedication beyond his professional sphere.

Dalton Ruer

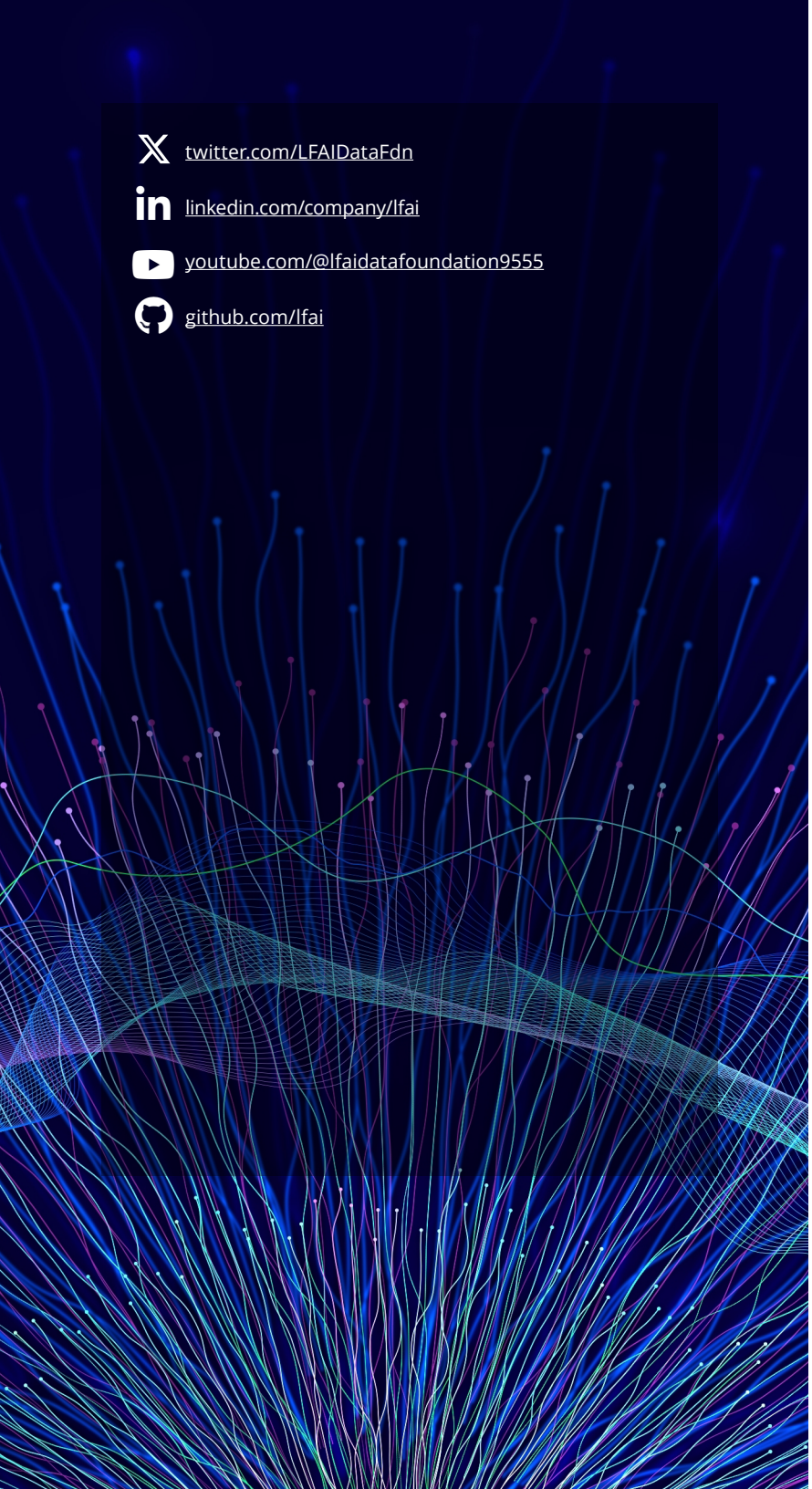
Dalton is a Senior Solution Architect at Qlik. He is a Data Scientist Storyteller, Analytics Evangelist and is an impassioned student of Generative AI. He is a seasoned author, speaker, blogger and YouTube video creator who is best known for dynamically sharing inconvenient truths and observations in a humorous manner. The passion which Dalton shares through all mediums, moves and motivates others to action.

Sachin Sinha

Sachin, Microsoft's Director of Technology Strategy, has honed his data engineering expertise post-graduation from the University of Maryland. Specializing in information management, he crafted systems aiding data-driven decision-making. His support extended to healthcare startups, fostering data-centric business models, and empowered public sector entities toward their missions via data-enabled decisions. Residing in Fairfax, VA, with his family, he passionately backs the Terps and Ravens while excelling in his professional pursuits.

Stu Sztukowski

Stu, a senior product manager at SAS, focuses on user-friendly AI through visual analytics. With a BS in Statistics from North Carolina State University and an MS in Advanced Analytics, he transitioned from data science, excelling in forecasting and BI. He champions accessible high-performance AI, aiming for universal comprehension. A mentor and versatile leader, he simplifies complex analytics for data scientists and business analysts, driven by a passion for public speaking and democratizing analytics.



twitter.com/LFAIDataFdn

linkedin.com/company/lfai

youtube.com/@lfaidatafoundation9555

github.com/lfai

LF AI & DATA

About the LF AI & Data Foundation

LF AI & Data is an umbrella foundation of the Linux Foundation that supports open source innovation in artificial intelligence (AI) and data. LF AI & Data was created to support open source AI and data, and to create a sustainable open source AI and data ecosystem that makes it easy to create AI and data products and services using open source technologies. We foster collaboration under a neutral environment with an open governance in support of the harmonization and acceleration of open source technical projects.Explore our current portfolio of projects and contact us to discuss hosting your open source AI or data project under our Foundation.

December 2023

To reference this work, please cite as follows:
https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=35160417#BI&AI-PastPublication