



ACCELERATING TIME TO AGENTIC AI VALUE

A VALOIR REPORT
February 2025

Agentic artificial intelligence (AI) promises to deliver exponential benefits from AI by automating complex tasks and interactions without human intervention. However, creating agentic AI that can handle complex tasks with acceptable performance has been a challenge. Valoir found using a platform optimized for agentic AI development like Salesforce Agentforce enables organizations to deliver autonomous AI agents an average of 16 times faster than other approaches while increasing accuracy by 75 percent.

The promise of agentic artificial intelligence (AI) is to exponentially increase the benefits from AI with autonomous agents that can handle complex tasks and interactions without human intervention. However, many generative AI projects never moved beyond the pilot stage. Even organizations with the resources to build models, integrate data, engineer prompts, and build guardrails and data security found they still couldn't achieve an acceptable level of performance and accuracy. These early failures led many organizations to shift from AI fear of missing out (FOMO to AI FOMU (fear of messing up), putting generative AI projects on hold until the technology – and vendor approaches to delivering accuracy – improved.

At Dreamforce in September 2024, Salesforce announced Agentforce to help customers build and deploy autonomous AI agents more quickly and with greater accuracy and performance than DIY approaches, and many other vendors followed with similar announcements. To understand the potential for accelerating time to value with a platform approach, Valoir analyzed the experiences of Salesforce customers.

Salesforce Agentforce

Salesforce Agentforce is designed to leverage the Salesforce Platform to deliver agentic artificial intelligence (AI) automation. Components include:

- Agentforce Service Agent is a customer-facing autonomous AI agent that helps deliver self-service resolution to customers with pre-built topics and actions for key service use cases.
- Agent Builder enables users to customize, set up, test, and activate agents with an Agent Wizard and low-code tools.
- The Atlas Reasoning Engine autonomously analyzes data, makes decisions, and completes complex multi-step tasks intelligently.

- Testing Center enables teams to test Agentforce using synthetically generated data at scale.

Agentforce is a new layer that is integrated in the Salesforce platform so customers can leverage other platform assets including:

- Customer 360 applications (Sales, Service, Marketing, and Commerce Cloud) can be leveraged by Agentforce to extend and augment the work of existing Salesforce users.
- Data Cloud connects, unifies, and harmonizes customer data and metadata for grounding precise and context-aware responses. Data Cloud retrieval augmented generation (RAG) capabilities enable users to upload unstructured data for further grounding. Tableau helps Agentforce see and understand data.
- The Einstein Trust Layer protects customer data through security features and guardrails including zero data retention, toxicity detection, secure data retrieval, and dynamic grounding. The Trust Layer Audit Trail enables users to track AI agent actions and outputs for compliance.
- The MuleSoft API management platform enables developers and admins to bring in third-party data to enable Agentforce to take action across any system.
- Salesforce Flow enables Agentforce to invoke any process flow created in the low-code process automation tool Flow Builder, or enable users to create new flows to support specific Agentforce jobs.
- Slack enables Salesforce users to deploy Agentforce within its conversational interface for employee-facing autonomous agents.

To assess Agentforce's comparative time to value Valoir conducted in-depth interviews with more than 20 Salesforce customers representing a range of industries, company sizes, and geographies (predominantly in North America). All had experimented with some form of generative AI to build their own agentic apps before Agentforce was released with varying degrees of success, and were piloting Agentforce or using it in production. For our analysis, we looked at seven phases of agentic development:

- Model setup
- Data and application integration
- Prompt engineering
- AI guardrails and security
- User interface and workflow/application development
- Tuning

Obviously, the complexity of agentic tasks and volume, sources, and hygiene of data varied by customer, as did the size and level of data

science and technical expertise of their teams. Valoir normalized the data by calculating the number of person-hours each team spent on each phase of agentic AI development with Agentforce versus the amount of time spent on the same phases of their DIY agentic projects to quantify the average time savings for each phase and the overall average acceleration of time to value with Agentforce.

Agentforce versus DIY

MODEL SETUP

Valoir found that Salesforce customers were able to leverage Agentforce's pre-tuned large language model (LLM) and integrated retrieval augmented generation (RAG) capabilities rather than investing in building their own LLM or RAG databases or other pre-built models. RAG enables more accurate and precise answers by enabling an LLM to reference a specified set of documents to inform its answers so it can use domain-specific or current information.

Agentforce's pre-tuned LLM and RAG capabilities enabled them to largely eliminate the time DIY projects devoted to initial model evaluation, prototyping, and tuning. Customers said:

We have a beta of Agentforce that someone developed within weeks. If we were trying to build all this in our LLM it would have taken months to do.

We spent the last year building our version of the Salesforce platform for my research. Four guys working on one team, five on another. To do RAG, architecture design, custom tables and schemas, figure out how we index and what vector database took six months and we weren't close to testing.

The time for development of a custom LLM could be years. However, in reality, most organizations taking a DIY approach are using pre-built models, which typically required three to 12 months to set up. In contrast, Agentforce's models are pre-integrated and pre-tuned, requiring little to no set up time.



Agentforce AI model setup is 7.5x faster than a DIY approach using pre-built models.

DATA INTEGRATION

Valoir found that the investments Salesforce had made in Data Cloud and Data Cloud Connectors and Service Cloud Knowledge, as well as the fact that customers' key customer relationship management (CRM) data

was already in Salesforce, significantly accelerated their ability to get value from agentic AI with Agentforce. Customers said:

We're getting value [from AI] in days or weeks versus months to years, and the difference is because we have our data in Salesforce.

Our data lives in Salesforce and for ChatGPT to grab all of that I wouldn't even know where to start. If you know how to build a flow it can grab the data in Salesforce where you have the data and just use those prompts. To pull that data out of Salesforce to put in ChatGPT would be way more time and costly than we could do. That's why we started down the AI path with Salesforce because all the data is there.

There's a big barrier to connecting data – you're not going to get a SMB or midsized company to figure out how to pull that data in. With Salesforce, I haven't had to figure out where to get data from.

All our knowledge is in Salesforce. With the help of one of our SEs we built our RAG in Data Cloud in one of our sandboxes and it only took us a few hours.

Realistically, organizations not using Agentforce would use some other data integration tool or customer data platform (CDP) to integrate and prepare their data for AI. Data Cloud and its pre-built RAG capabilities in Agentforce are the two accelerators for data integration compared with a DIY approach.

Using open source or other alternatives, organizations spent at least a month selecting a RAG approach; integrating document ingestion, retrieval, and storage tools; and integrating the RAG with their generative model and an additional two to three months training the retriever and model with domain specific data. With Agentforce they were able to complete needed data integration within weeks.



Agentforce data integration is 3.5x faster than a DIY approach.

PROMPT ENGINEERING

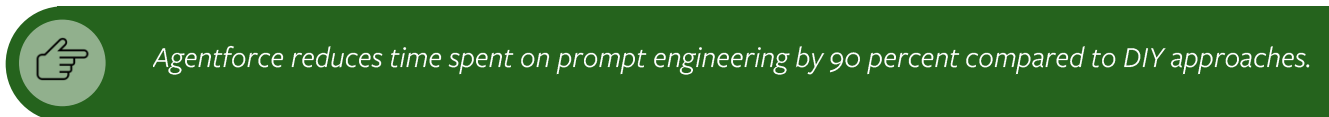
Valoir found that the investments Salesforce has made in prompt engineering with Prompt Builder, now part of Agentforce, enables organizations to easily build and reuse prompt templates that can then be leveraged by Agentforce agents to complete specific actions. Customers said:

We were trying to use a prompt in Excel that replaced names with customers, but it was a huge manual effort, and also not a

streamlined process. We spent a year working on it. With Agentforce, it took a month to operationalize.

They've created the structure to plug in your data, plug in your model, and enable SMEs to do it without prompt engineering expertise. Companies can use lesser paid resources without advanced degrees in data science and programming to deploy AI in a responsible way with Salesforce's reputation behind it.

Additionally, the pre-built prompts associated with the out-of-the-box tasks and skills in Agentforce further reduce the burden on teams to build at the prompt level.



AI GUARDRAILS AND SECURITY

Valoir found that the guardrails and security in the Salesforce Trust Layer were a key factor in enabling organizations to move beyond generative AI pilots to agentic AI. Customers said:

The team was trying to train an open AI model by feeding it all our compliance data and clinical trial data. We were concerned about putting personal data into ChatGPT, and I don't think we could build the Trust Layer.

We're having Agentforce explain a specific [person's financial details] in an authenticated experience. We can't be violating [privacy laws]. Any potential leak of data would be disastrous. That's when you really run into issues with exporting confidential data - you just can't do it with [Microsoft] Copilot. You could build it, but it would be a 20-30x effort.

Before when a service appointment was completed Zapier would take the service notes and appointment record and send it to ChatGPT and chat would summarize. It worked ok but we decided not to keep going forward with the announcement of Einstein Trust Layer. We couldn't trust anyone not putting it out there in the ChatGPT database.

None of the Salesforce customers Valoir interviewed for this analysis had successfully built the equivalent of a Trust Layer in their DIY projects. Teams with significant expertise that tried to do so worked on projects for more than 12 months.



Development teams with significant development and data science expertise would need more than 12 months to develop the equivalent of a Trust Layer.

USER INTERFACE AND APPLICATION DEVELOPMENT

Without Agentforce and the Salesforce platform, Valoir found that companies seeking to deliver agentic AI would increase their development burden in two main areas: the development of a user interface or integration into an existing conversational interface (such as a bot) and the development of workflows between agents and the LLM. Agent Builder's drag and drop interface and conversational instruction dialogues automated much of that process. Customers said:

The agentic piece is easy. To create a new agent you walk through a few screens of what topics you want to include, where to get its knowledge from in Data Cloud, and then you're testing and tuning the topic to make sure you're getting the right quality of response. You're not coding and building stuff – it's really quite simple. We had a functioning prototype in fewer than three weeks and now we're doing a 6-week sprint to production with a few people.

Copilot didn't take action – we would have to set up flows to trigger, and connect it to mail and a conversational UI, and each org would have to figure it out themselves.

With Microsoft it would be five times longer because you cut out the UI layer. It's putting the right data in the prompt and passing the right things to the LLM - because we're all using the same models.

Organizations already using Salesforce Flow and Einstein had the most rapid road to agentic AI development. Customers estimated it would take them, on average, at least six times longer to develop a customer-facing UI and associated workflows without Agentforce.



Agentforce accelerates user interface and agentic application development by 6x.

TUNING

Valoir found that tuning agentic AI to deliver an acceptable level of accurate results without hallucinations or toxicity was the biggest hurdle for DIY agentic AI projects. In fact, the majority of DIY projects Valoir analyzed never reached acceptable levels of accuracy despite having been tuned for significant periods of time. In contrast, the combination of the Atlas Reasoning Engine, Testing Center, and the ability to tune agents with conversational language instructions enabled

business users to rapidly tune and test models to ensure they could deliver accurate results. Customers said:

We took the open AI API from Microsoft and built a connector from Einstein to Copilot so when it detected certain topics it would route to a RAG built on Azure and then ChatGPT's LLM. It didn't work. It hallucinated way too much because it had no logic or reasoning layer for complex tasks. To tune all the scenarios and eliminate hallucinations it would have been years of work for our 4000 SKUs.

Topics are how you guide the agent and instructions are on specific tuning. If the accuracy isn't there you test it and say, "where do I need to add more instructions or be clearer?"

Now tuning is just adding instructions to the AI agent. Someone with a few years in industry and that understands business language can manage Agentforce.

Clearly, the time needed to tune agentic AI varies broadly based on the complexity of tasks to be performed. However, Valoir found that tuning for accuracy was a key area where Agentforce delivered, with customers finding increased accuracy of 75 percent over DIY with minor tuning:

With DIY we were able to get to 60 percent accurate. With what I'm seeing with Agentforce we've been able to push it up to 85 percent.

If I were to build it in a custom GPT it would be less accurate because you don't have business process and context around it, and it will guess at the best plausible answer. When I asked ours, I had a 50-50 chance of it getting it right. Now we're at 95 percent.

Customers estimated they would need years to achieve the same level of accuracy that they achieved in just weeks with Agentforce.



Agentforce can be tuned to deliver accurate results 12 to 24 times faster than DIY Agentic AI.

Data accuracy

A key factor in time to value was time to accuracy; that is, the time needed to build and train AI agents to deliver acceptable levels of correct responses. Obviously it is a lot faster to build and train an agent to respond to simple queries like password change requests than it is to train an agent to successfully navigate multi-level decision trees and responses depending on data from multiple sources. When it came to

the Salesforce customers interviewed for our research, Valoir found that:

- Approximately two-thirds of customers had reached the point with their DIY project where they were able to test for accuracy using multiple sets of simulated queries. They found that 52 percent of the answers or resolutions their agents delivered, on average, had what they deemed an acceptable level of accuracy. The other customers who had abandoned their DIY efforts at some point because of low accuracy, security concerns, or resource constraints – or because Agentforce became available and it was deemed a better option than DIY – didn't believe their DIY efforts, had they continued, would have delivered more than 50 percent accuracy.
- Sixty-two percent of customers had, by late January or early February, one or multiple Agentforce agents in pilot or production long enough to test the accuracy of agent responses, and found that 91 percent of answers or resolutions their agents delivered, on average, had an acceptable level of accuracy – representing a 75 percent increase in accuracy over DIY.

SELECTED CUSTOMER DIY-AGENTFORCE AGENT EXAMPLES

| Agent complexity | Description | DIY accuracy | Agentforce accuracy | % change |
|------------------|---|--------------|---------------------|----------|
| Simple | Employee-facing HR leave request agent | 50% | 95% | 90% |
| Moderate | Customer self-service agent for complex product catalog with 1000+ SKUs | 60% | 85% | 42% |
| Complex | Customer self-service agent with authentication and personalization based on PII | 40% | 80% | 100% |
| Complex | Complex sales coaching using conversation insights for improvement suggestions and role playing | 40% | 90% | 77% |

Beyond simple use cases, we found that the greatest change in accuracy in Agentforce agents versus DIY agents was often in more complex agents that were dependent on multiple data sources for grounding and often relied on more complex reasoning and workflows to come to the accurate answer based on the context of the customer query.



Moving from DIY to Agentforce can increase agent accuracy by 75 percent.

Ongoing support and maintenance

As Agentforce is relatively new, customers are still discovering how much tuning and support will be needed on an ongoing basis. However, customers recognize they are achieving agentic AI results that would likely be too resource-intensive to support without the Agentforce platform, saying:

We don't have the skill, time, or money to build guardrails and then audit them on an ongoing basis to make sure that the model doesn't hallucinate.

A big part of AI is the ongoing maintenance. You have to check it all the time, look at the logs, and it ends up being a product to manage. Either way you're doing some work, but do you want to manage and tune an LLM on an ongoing basis? And build all the tools to monitor it, because they don't exist? They exist with Agentforce.

Colossally difficult to maintain – it's a much different thing to maintain the training you're doing against AI versus these managed packages.

Like cloud computing, a platform approach to agentic AI like Agentforce enables organizations to focus their technology resources on fine tuning and differentiation rather than day-to-day application support.

Data summary

| Agentic AI phase | DIY months (average) | Agentforce months (average) |
|---------------------------------|----------------------|-----------------------------|
| Model setup | 12.0 | 1.0 |
| Data integration | 3.5 | 0.3 |
| Prompt engineering | 12.0 | 1.0 |
| Guardrails | 18.0 | 0.0 |
| UI and app/workflow development | 6.0 | 1.0 |
| Tuning | 24.0 | 1.6 |
| Total | 75.5 | 4.8 |

Valoir conducted in-depth interviews with more than 20 Salesforce Agentforce pilot customers in various industries to determine the average number of months they spent working on each phase of their agentic AI projects both on a DIY basis and with Agentforce. We found the average months spent on DIY projects was 75.5 and the average time needed to bring an Agentforce project to productive accuracy was 4.8 months – making it 16 times faster.

Looking ahead

Agentic AI has the potential to deliver significant benefits by automating complex tasks and interactions without human involvement. However, early experiences have led technology teams from FOMO to fear of messing up (FOMU), and creating agentic AI systems from scratch has been prohibitively expensive and often unsuccessful. Even with extensive testing and tuning, achieving the reliability needed for agentic AI has remained elusive.

In contrast, a platform approach that enables individual organizations to leverage vendor investments in model setup and tuning, data integration, prompt engineering, guardrails and security, UI and workflow development, and low-code tuning and testing – like Agentforce – will drive broader adoption of agentic AI by lowering technical and financial barriers and making the benefits of agentic AI accessible to all organizations.

Valoir is a technology analyst firm providing research and advisory services with a focus on the value technology delivers. With deep expertise in CRM, HCM, customer and employee experience, and enterprise applications, Valoir helps clients understand and maximize the value of technology. For more information, contact Valoir at www.Valoir.com or 1-617-515-3699.