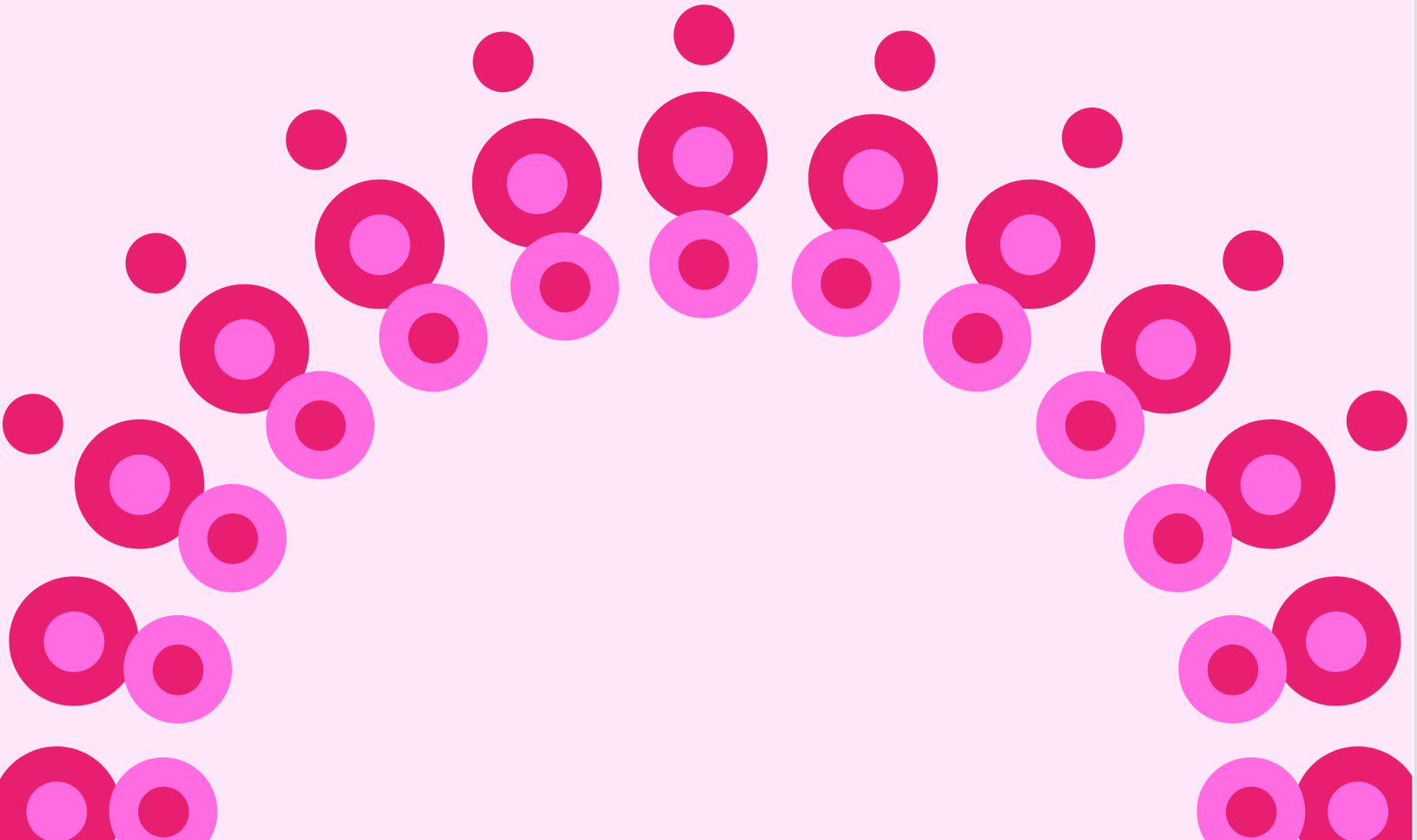
Arcee AI

Enterprise AI Agent Guide

Authors:
Nora He, Andrew Walko, Mary MacCarthy



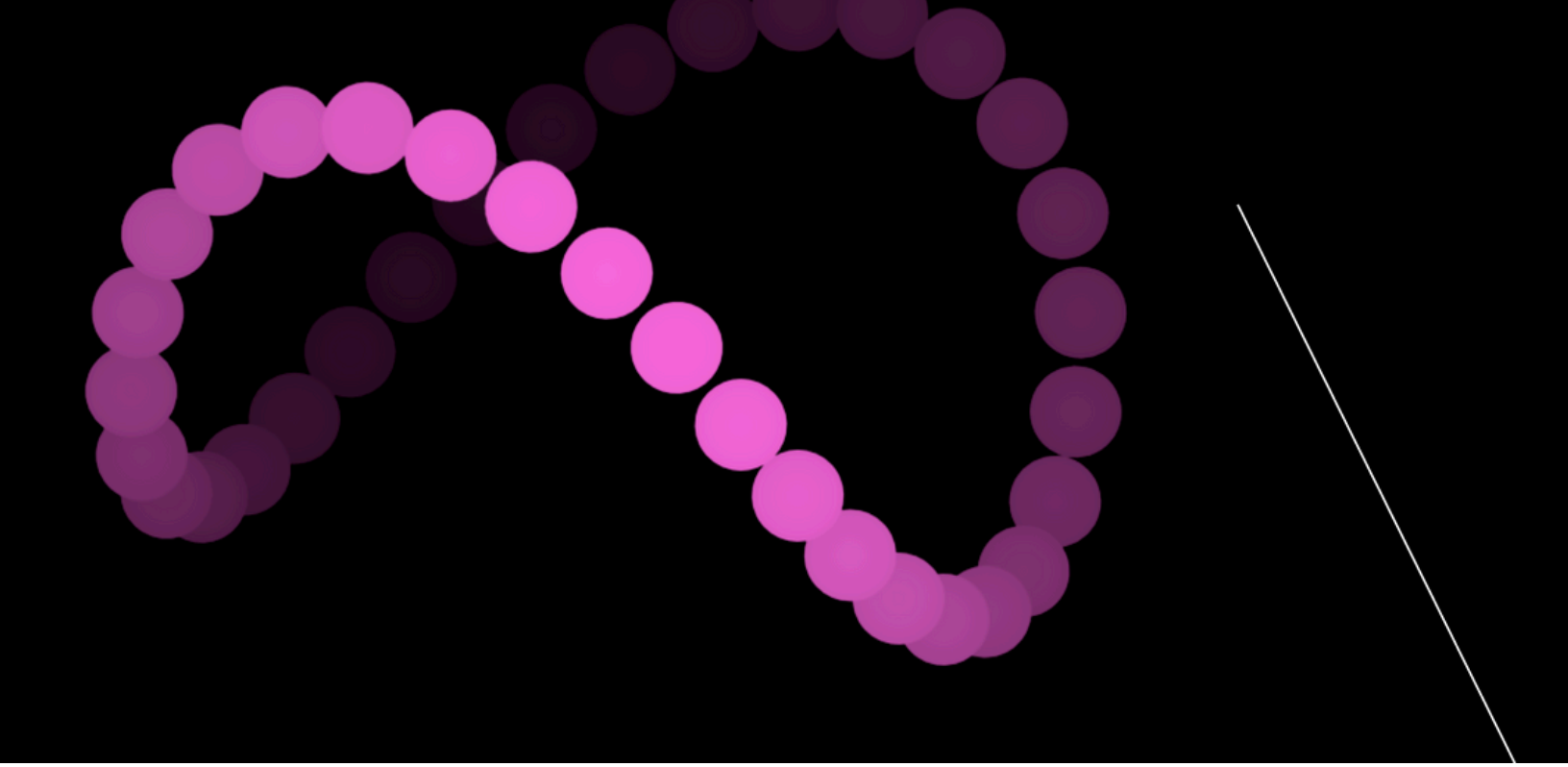
Arcee AI

Table of Contents

01

AI Agents: *The Fundamentals*

01



Introduction:
From Pipe Dream to the Present

AI agents are no longer just hype. They're very much real, and the results are out there. Organizations are transforming tasks that once took hours into work completed in minutes.

Think of them as your organization's digital workforce. They do more than just follow instructions. They can process complex information, execute multi-step workflows, and adapt to different situations and goals.

In this section, you'll discover how AI agents are transforming businesses. Here's what you'll learn:

- What AI agents are
- How AI agents are different from traditional automation and AI workflows
- Why leading companies are rapidly adopting AI agents.

Our aim is not to dive deep into the technical details. We do provide an overview of the essential technical concepts, but our main goal is to provide you with the **strategic view** that today's industry leaders need in order to truly understand the potential of AI agents.

By the end of this section, you'll understand what AI agents are, and why they might become your next most powerful ally in the workplace. These aren't just tools–they're potential game-changers for your organization's efficiency and productivity.

The Basics: What Are AI Agents?

The term "AI agent" has become a buzzword, with companies everywhere claiming to build or use them. But what exactly makes something an AI agent?

Let's cut through the hype.

To understand AI agents, let's start with something more familiar to most of us: large language models (LLMs), like GPT-4o (which powers ChatGPT). This gives us a clear reference point to highlight the key differences.

GPT-4o excels at generating content and offering recommendations, but it has a key limitation: it can't take action on its own. You must manually implement any insights or suggestions it provides.

AI agents, however, take capabilities to the next level. They can **actively execute tasks** by interfacing with your company's systems, accessing your data, and taking concrete actions–all with minimal human oversight. This moves beyond simple suggestions into actual task completion.

Consider this practical example: with GPT-4o, you might ask it to, "Summarize this sales call transcript for me." But with an AI agent, you can say, "I just had a call with Globomax Insurance; please send me all the deal info." The agent will then determine the best course of action–whether that's creating a summary report, updating your CRM, scheduling follow-up calls, or all of the above.

It's important to note that not everything marketed as an "AI agent" truly qualifies as one. Many products are simply LLMs or are, in fact, basic automated workflows without the intelligent, autonomous capabilities that define true AI agents.

So, let's explore what makes a real AI agent different.

What Makes an AI Agent an AI Agent: Six Key Characteristics

What distinguishes AI agents is their **true autonomy**. They operate independently, requiring minimal human intervention.

Here are six additional characteristics that define how AI agents work and give you a sense of their potential.

- AI agents have ***a specific goal.***

Each AI agent is built with a highly-specific objective. Whether it's managing customer service inquiries, overseeing shipping logistics, or monitoring system security, AI agents operate with a clear purpose that guides every decision and action.

- AI agents can ***plan on their own.***

Another key characteristic of AI agents is their ability to think and plan before taking any actions. They don't just follow a simple to-do list; they analyze situations, consider different approaches, and pick the best way to reach their goals.

This planning process is done by **models**:

Models: Usually categorized into either small language models (SLMs) or large language models (LLMs), models are like the left part of the human brain that processes language. They are responsible for understanding and taking input (such as "summarize this call transcript") and producing output (such as the actual call summary). During planning, instead of immediately outputting the answer or response, the model breaks down the task and determines if prerequisite actions need to be taken before providing an answer.

For multi-agent solutions or "agentic networks" where multiple agents work together to solve a complex task, an additional component is required–a **router**:

Router: AI agents can run on different models, each for a specific task. That's where the router comes in. It's exactly what it sounds like. The router figures out what needs to be done and picks the best model for the job, kind of like a conductor telling each instrument when and what to play in a symphony.

For example, when tasked with creating a marketing campaign for a new product, an AI agent coordinates the process by 1) routing campaign requirements to a large language model that generates basic campaign guidelines and 2) sending product specifications to a small language model good at marketing to develop the campaign proposal.

The agent breaks down complex tasks into steps, assigns them to appropriate resources, and manages the workflow efficiently.

- AI agents can ***interact with any environment.***

AI agents aren't restricted to their training data. They can interact with any digital environment through tools and integrations. This connectivity is achieved through APIs (Application Programming Interfaces), allowing agents to communicate with various systems and execute tasks.

For example, AI agents can use tools to:

- Retrieve customer data from CRM systems like Salesforce or HubSpot
- Make travel reservations
- Automate invoice processing workflows.

With API access, an AI agent can interface with virtually any tool in your company's technology stack.

- AI agents can ***remember.***

AI agents don't just act in the moment; they remember. They retain historical data, such as past interactions or decisions, to continually refine their approach. This lets them personalize user experiences, improve efficiency, and avoid repeating mistakes.

For example, a B2B sales AI agent can independently manage lead qualification by analyzing company profiles, financial status, and market position to make strategic decisions about lead prioritization. Unlike a simple algorithm that scores leads, this AI agent actively decides which prospects to pursue and automatically initiates appropriate follow-up actions like scheduling calls or sending sales materials.

- AI agents can ***make decisions independently.***

Once an AI agent has planned and gathered information, it takes action. AI agents can work independently to complete tasks within set limits, depending on the agent's goal. Here are some everyday examples of what they can do:

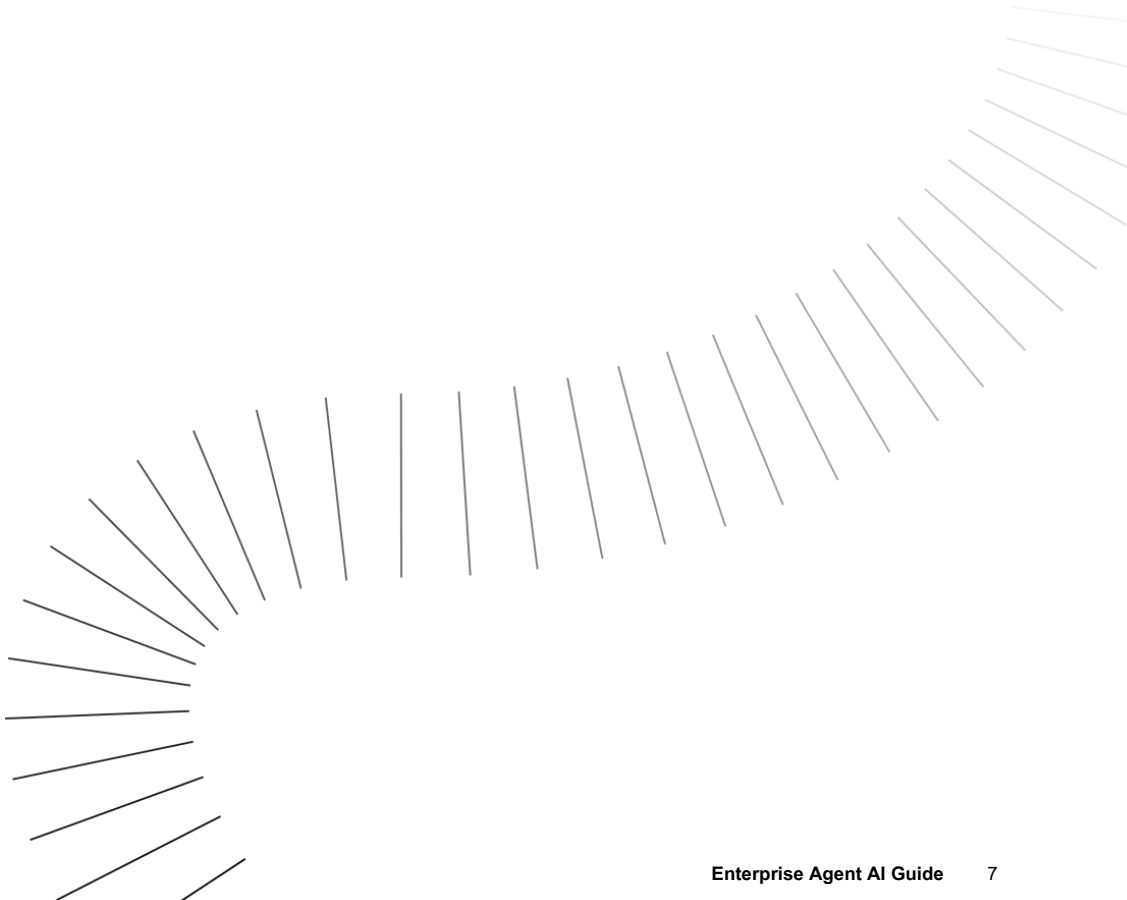
- Process routine expense reports
- Optimize and schedule meetings
- Manage customer refund requests.

This autonomy dramatically improves efficiency by eliminating the need for constant human oversight. However, the level of independence is customizable. For high-stakes decisions, human-in-the-loop checkpoints can also be necessary to integrate into the workflow.

- AI agents ***learn and improve.***

The best thing about AI agents is that they get smarter with use. With proper monitoring tools in place, they refine their processes and improve their performance over time by analyzing feedback, outcomes, and new data.

For example, a marketing AI agent may learn which campaign strategies produce the best results and adjust plans accordingly.



Why AI Agents Matter *Now*

Although the concept of AI agents has existed since the 1970s, we saw a pivotal shift in their accessibility and capabilities not until Microsoft launched Copilot Studio in November 2023 (Stallbaumer, 2023).

Copilot Studio is a platform that helps companies build AI agents without needing to code. For the first time, it seemed possible to take business information and turn it into intelligent, automated workflows without needing to code. With this big shift from Microsoft, the lure of fully autonomous systems in the workplace was no longer a dream.

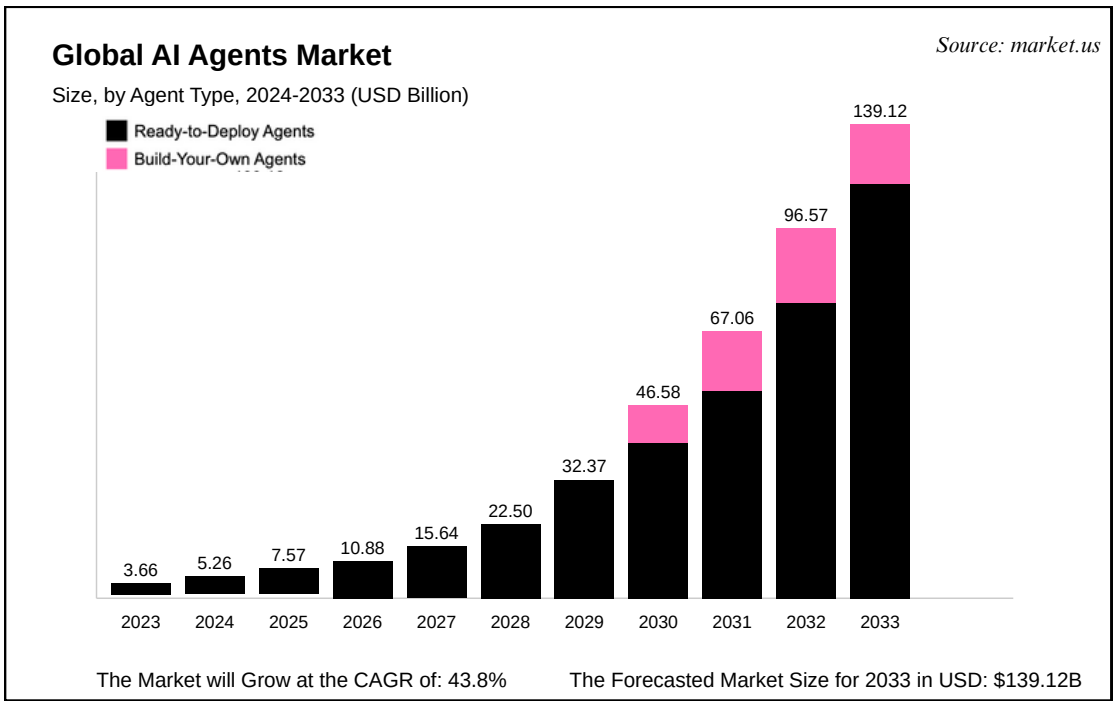
AI agents are expected to automate up to 50% of business tasks by 2027. And for many leaders, that kind of automation can bring massive benefits. According to Gartner, businesses can save up to 30% on operational expenses by using AI-driven automation (Rimol, 2021).

Let's examine why implementing AI agents should be a strategic priority for forward-thinking businesses.

Market Trends

The AI agent industry is growing fast. The market was valued at an estimated $3.66 billion in 2023, and experts predict it will grow to $139 billion in 2033 (Grandview Research, 2023).

A growing number of businesses now see AI agents as must-have tools, not just nice-to-have extras. In fact, 65% of Fortune 500 companies already use or test AI agents to make their work easier and more efficient (Nyhan, 2024).



Changing Customer Expectations

AI is projected to handle 80% of all customer interactions by 2030 (Nyoni, 2024), and while complex or emotionally sensitive customer issues will likely still require human intervention, this forecast reflects growing consumer comfort with automated AI interactions.

If you're in an industry like hospitality and finance, customer experience is critical to success. Customers will expect instant, personalized service available at any time of day. This shift in expectations could make AI agents particularly valuable, since they can provide consistent, high-quality service at scale.

Competitive Edge

The rapid adoption of AI agents isn't just a trend. It's becoming a critical factor in business performance. As more organizations adopt these tools for round-the-clock operations and automated decision-making, early adopters are gaining significant advantages. The performance gap between AI-enabled companies and those yet to implement these technologies continues to widen.

Here's how AI agents provide a decisive competitive advantage:

- **24/7 Operations**: AI agents operate around the clock, ensuring uninterrupted service delivery and rapid response times that human teams alone cannot match.
- **Enhanced Human Potential**: By managing routine tasks, AI agents free your team to focus on strategic initiatives, creative problem-solving, and building meaningful relationships with customers and partners.
- **Unlimited Scalability**: AI agents can seamlessly extend across your entire organization. Imagine deploying hundreds of AI workers simultaneously across departments, each handling specific tasks while maintaining consistent quality and performance.
- **Operational Efficiency**: Beyond the initial investment, AI agents significantly reduce operational costs while improving service quality and speed.

The combination of these advantages creates a powerful multiplier effect. Your organization becomes faster, more efficient, and more responsive than competitors. In today's fast-paced business environment, this enhanced capability isn't just an advantage but becoming a key differentiator for market leadership.

From Automation to AI Agents

We've seen how AI agents differ from generative AI. Now, let's go a step further and examine what truly distinguishes AI agents from traditional automation and AI workflows.

Understanding these differences will help you make informed decisions about your technology investments and choose solutions that best align with your business objectives.

	Traditional Automation	AI Workflows	AI Agents
Definition	Codified rules based on conditions, such as, “If this, then that”	Controlled automations which combine generative capabilities of models and codified conditions	Autonomous automations which make decisions based on context to solve the task
Capabilities	• Executes tasks based on triggers • Static, rigid workflows	• Combines static workflows with AI for smarter tasks • Enhances specific steps	• Plans and decides independently • Learns and adapts
Adaptability	**Minimal** Fails with new variables	**Moderate** Confined to tasks	**High** Dynamic and flexible
Example	Sends a follow-up email based on the sign-up trigger	Draft an email with guidance	Decides the best course of action is to run an email campaign
Strengths	• Easy setup • Cost-effective • Repetitive tasks	• Adds intelligence • Great for structured insights	• Handles complexity • Reduces human oversight
Limitations	• Can’t adapt to changes • Requires frequent updates	Stuck in task-specific boundaries	• Complex to implement (needs advanced resources) • Unreliable output
Best Use Cases	• Sending email confirmations • CRM updates	• Drafting email templates • Sorting data	• Automated code review • Intelligent customer support escalation

Traditional Automation

Traditional automation relies on predefined rules to handle routine tasks. While effective for static workflows, it struggles to adapt when new variables or complexities arise.

Example: A basic rule-based system sends a pre-determined welcome email sequence when a customer signs up for an app.

AI Workflows

AI workflows improve automation by using AI models to do specific jobs at specific steps. However, these workflows still operate within predetermined parameters.

Example:
- An AI-note-taking tool like Tactiq automatically transcribes the transcript and generates action items
- The action items are loaded into Linear with deadlines based on the transcript.

The AI workflow is still predetermined based on the type of task, the sequence of actions, and the tools you've chosen (Tactiq, Linear). This gives a balance between deterministic actions from codified steps and generative actions from model steps.

AI Agents

AI agents take AI workflows a step further. AI agents are software programs that use artificial intelligence to interact with their environment, collect data, and perform tasks.

For example, in customer service:

- Traditional Automation: Automatically sends a follow-up email based on a trigger
- AI Workflow: Uses limited AI model in specific steps of workflow; follows the instructions
- AI Agent: Independently determines the best follow-up action, such as sending an email, scheduling a call, or escalating the issue to a human agent.

Each type of automation has its place in a modern business, serving different needs and complexities.

Traditional automation excels at straightforward, repetitive tasks–like sending order confirmations or processing routine data entries.

AI workflows excel at tasks that need intelligence but follow consistent patterns. They can handle work that has some complexity, such as drafting customized emails or categorizing incoming requests based on content.

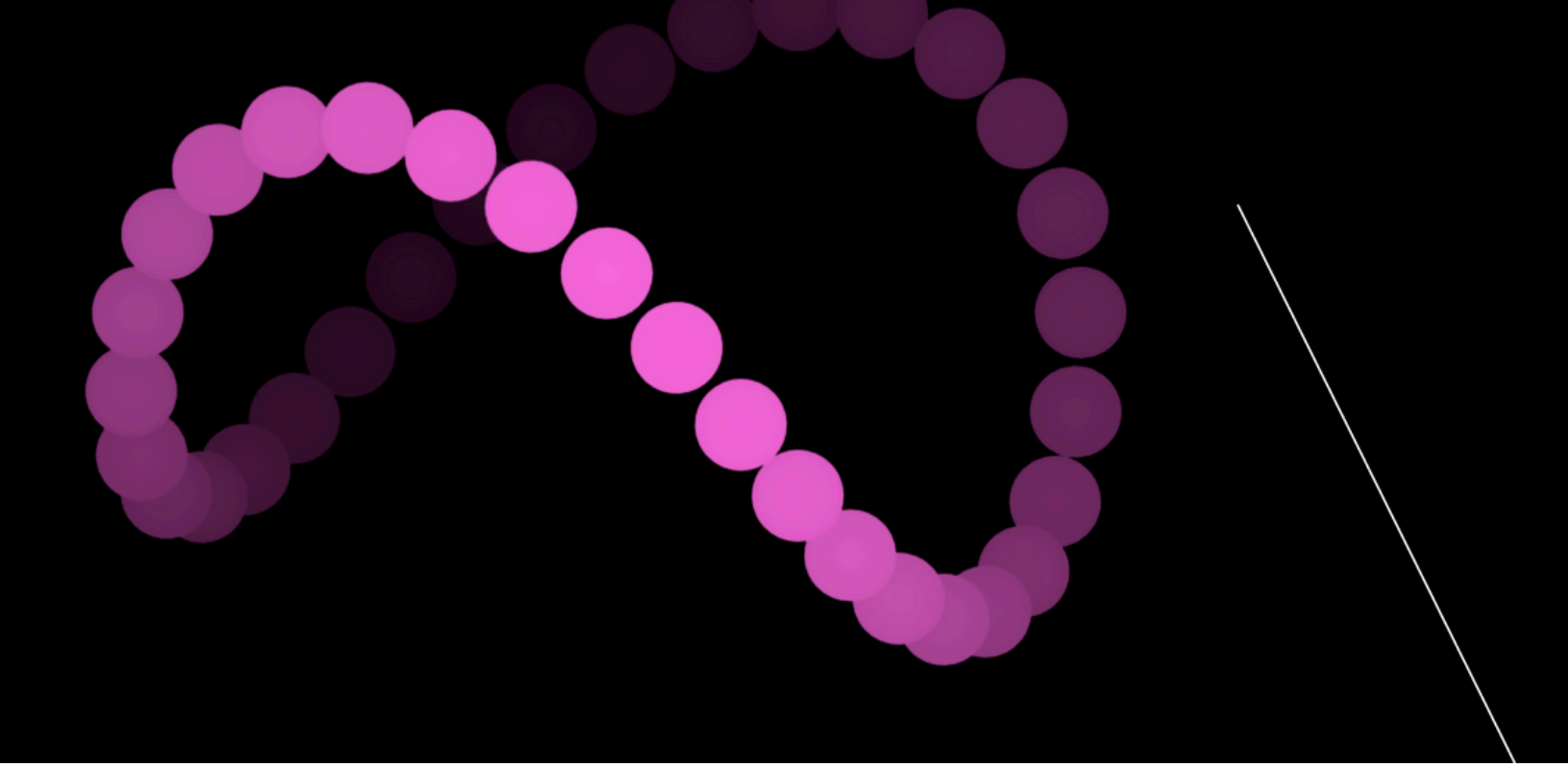
AI agents can handle all of these capabilities and are also well-suited for multi-step tasks that require judgment and adaptability. They can assess situations, make decisions, and adjust their approach, whether that's managing customer inquiries, optimizing schedules, or coordinating multi-step processes.

In the following sections, we'll explore practical applications and implementation strategies to help you leverage AI agents effectively in your organization.

02

AI Agents:
Preventing Headaches and Solving Problems

02



AI Agents: Preventing Headaches & Solving Problems

In today's business world, work happens at a frenetic pace, and at what seems to be an ever-growing level of complexity.

Leaders have to keep up not just with AI developments that move forward at lightning speed, but also with many other constant developments in their industries.

Recent data shows that 83% of business leaders face greater stress when making decisions, managing a daily decision load that has grown tenfold in just three years (Purdy and Williams, 2023).

This mounting pressure affects more than just leadership performance; it impacts the bottom line.

Studies indicate that decision fatigue and stress-impaired choices can reduce profit margins by 3% (Purdy and Williams, 2023).

Fear not, however. This section isn't meant to pile more stress on you! Our mission is to help alleviate some of the pressure by showing you how AI agents can help with smarter decision-making–for individual leaders and departments, as well as across entire organizations.

This section gives you real-world examples of how **AI agents are *already* improving how businesses operate**, from saving time to getting better ROI to getting and keeping a competitive edge.

AI Agents are All About Optimization

Let's start with a look at how AI agents can help optimize your operations–saving time and resources, which has an immediate and powerful impact on your ROI and boosts your competitive edge.

Save Time

Time optimization is a critical business priority with far-reaching impacts across organizations.

Office interruptions cost U.S. businesses approximately $650 billion annually in lost productivity (Bebo, 2024).

And here's the truth behind the big productivity loss: employees spend 51% of their average workday on low-value tasks such as email management, scheduling meetings, and generating routine reports (Homebase Team, 2024).

AI agents offer a solution by automating these time-consuming, low-value tasks. One recruitment firm tackled this challenge by implementing an AI agent with Arcee AI to evaluate candidates against company-specific criteria and internal database metrics. Expected to reduce resume screening time by 30-50%, **saving over 100 hours per person monthly**.

Save Resources

Beyond time savings, effective resource allocation is crucial for driving business growth, yet it remains a significant challenge for many organizations.

The impact is clear: companies that actively reallocate resources achieve 10% higher total returns to shareholders (Atsmon, 2016). However, despite 83% of executives recognizing resource allocation as critical for growth, most organizations only reallocate about 1% of their capital annually due to organizational barriers (Atsmon, 2016).

A major obstacle to the effective allocation of resources is fragmented data across departments, a reality that complicates decision-making.

AI agents solve this by:

- Analyzing data across departments in real-time
- Tracking resource utilization continuously
- Identifying optimization opportunities automatically.

For example, a real estate company implemented AI agents with Arcee AI to evaluate commercial properties. The system expects to **reduce property evaluation time by 35%** while maintaining high assessment quality through simultaneous analysis of multiple data points.

Better ROI

Leaders must demonstrate the value of their technology investments, with 94% prioritizing digital transformation initiatives, including AI implementations, in their strategic planning (Mittal, 2022).

AI agents are delivering compelling returns across industries. In customer service alone, organizations have reduced costs by 30% while improving response times. Looking ahead, financial institutions implementing AI automation are projected to increase profits by 38% by 2035 through enhanced capabilities like fraud detection.

Consider JP Morgan Chase's implementation of AI agents for email marketing automation. The system, which manages everything from copywriting to audience segmentation and timing optimization, achieved a 450% increase in click-through rates (Nawaz, 2024).

To put that result in perspective, such a dramatic spike in click-through rates can **boost campaign revenue by approximately 8%**, even for just one campaign!

Competitive Advantages

Business leaders face constant pressure to keep up, particularly as new competitors emerge with cutting-edge technology. AI agents can help organizations gain and maintain a competitive edge. In fact, 90% of businesses believe that AI agents are a competitive advantage for the next five years (Andre, 2025).

The pharmaceutical industry exemplifies this competitive challenge. Johnson and Johnson (J&J) leveraged AI agents to transform their drug research and discovery process. Traditionally, developing a new drug takes 10–15 years, with scientists spending significant time analyzing vast datasets to identify effective compounds. J&J uses AI agents to analyze extensive patient records and laboratory data to understand disease patterns and potential treatments. By simulating and predicting compound interactions with diseases, these AI agents help J&J accelerate drug development and stay ahead of competitors in bringing new treatments to market..

Human Agents & AI Agents: Working Together

In the financial industry, managing risk is essential to protecting organizations from financial losses.

Moody's, one of the world's top three credit rating agencies, uses AI agents to enhance its risk evaluation process. By assessing the creditworthiness of companies, large institutions, and government entities, Moody's helps stakeholders make informed financial decisions.

Challenge

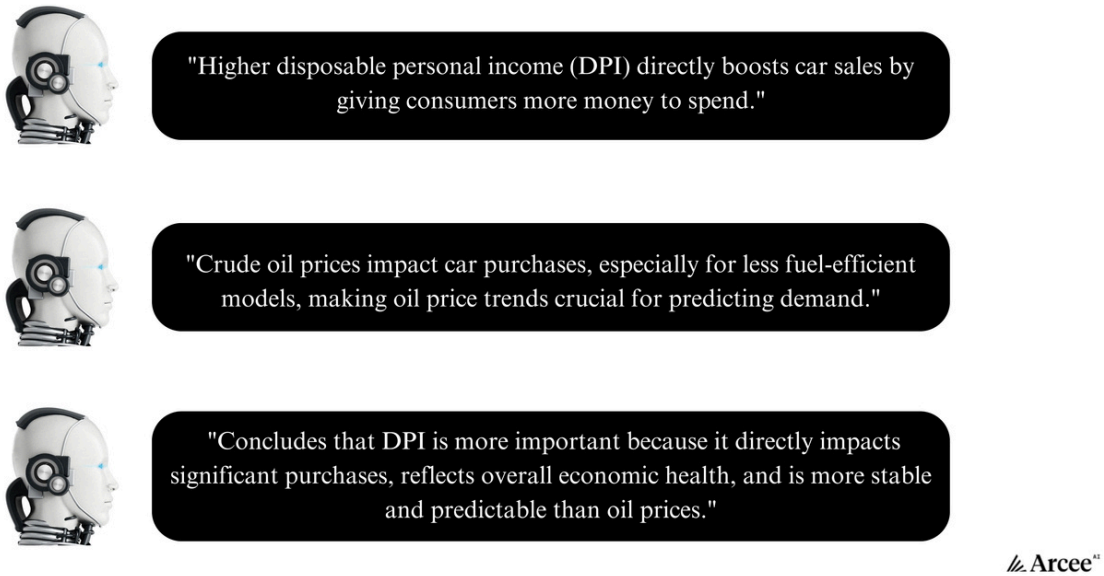
Risk assessment is complex due to intricate business structures and vast amounts of data spread across multiple departments. Additionally, regulatory requirements vary depending on the organization's location and that of its stakeholders and clients. Evaluating credit risk requires collaboration among specialists, teams, and financial experts to determine an institution's creditworthiness.

Solution

Moody's built a multi-agent system that mimics how their team workflows are structured, but with a few key differences. A multi-agent system involves a group of specific agents working together to achieve one goal (Lehavi et al., 2024):

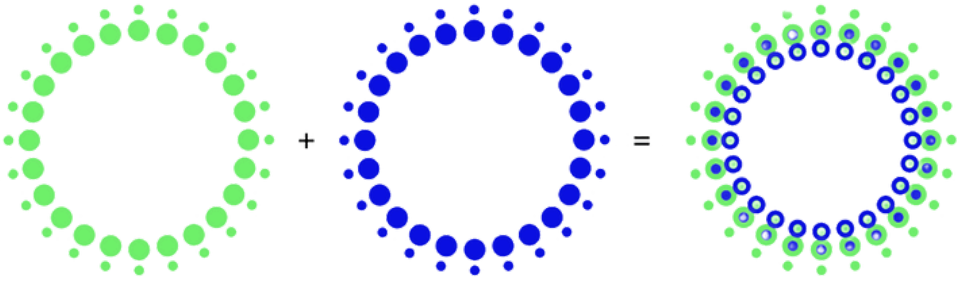
Agent Types	Responsibilities
Agent Orchestrator	• Coordinates workflow between all AI agents • Manages information flow and task distribution • Ensures seamless integration of analysis results
Financial Risk Analyst Agent	• Conducts comprehensive financial analysis • Performs trend analysis and cash flow assessment • Executes DuPont analysis for deeper insights
Macroeconomic Analyst Agent	• Evaluates broader economic conditions • Monitors GDP growth, inflation, and interest rates • Synthesizes industry trends and regulatory impacts
Knowledge Base Management Agent	• Maintains a comprehensive knowledge repository • Eliminates redundant information • Ensures narrative consistency across analyses
Credit Risk Analyst Agent	• Reviews credit history and behavior • Examines financial statements and income records • Evaluates the ability to meet financial obligations
Quality Assurance Agent	• Verifies the accuracy of all analyses • Removes potential misinformation • Ensures timely completion of all agent tasks

The AI agents team up to analyze risks from different angles. For example, this might be how one of the “conversations” works between the agents:



Source: Moody’s

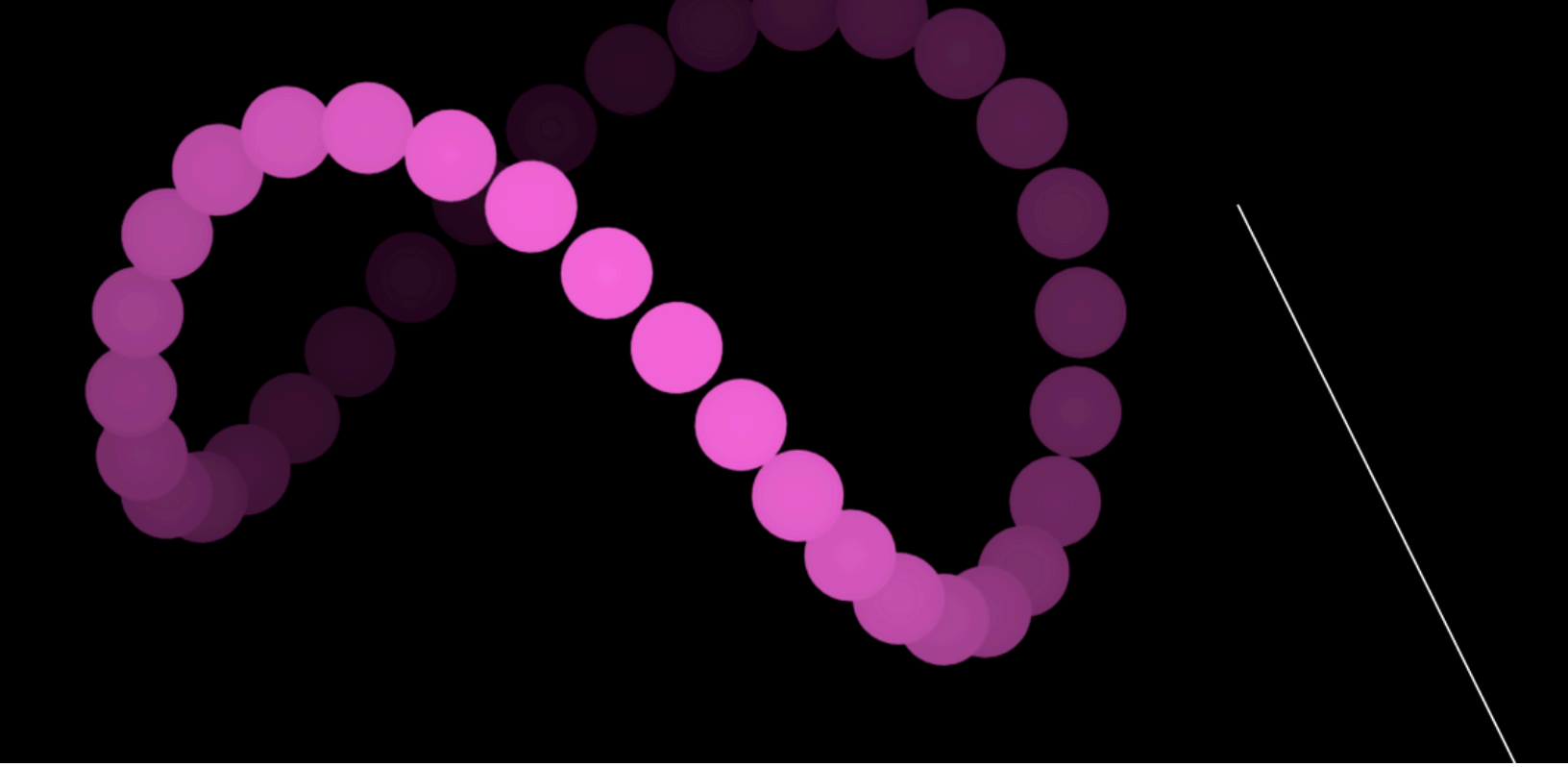
The AI agents work together to check credit risks more quickly and accurately than traditional methods, while maintaining the same high standards for evaluating companies. Note that **human agents are still part of the process**! It’s collaborative work between humans and AI agents. The humans actually work as the “master orchestrator” and make the final calls.



03

How to *Get Started* With AI Agents

03



How to Get Started With AI Agents

It's time to transform theory into action. We've explored AI agents' fundamentals and real-world applications. Now, let's put that knowledge to work.

This section walks you through **each step of the planning process**: evaluating potential use cases, mapping out specifications, checking technical foundations, and planning operation and implementation steps. We understand that every organization is unique and has its own distinct goals, structures, and resources. Rather than a rigid manual, consider this a practical guide to navigating what might otherwise feel like an overwhelming process.

Keep in mind that AI agents are just one tool in your broader toolkit. The goal isn't to chase trends; it's to create meaningful improvements in your operations, empower your employees, and strengthen your bottom line.

Arcee AI

AI Agent: Sample Checklist

We've developed a pre-implementation checklist that breaks down the essential phases and tasks you'll need to consider. There's no one-size-fits-all approach. Projects can be as small or as big as you want. Later, we'll discuss each phase in detail.

AI Agents: Sample Checklist

1	**Opportunity Assessment** - Identify current operational inefficiencies or areas that could be enhanced - Understand what AI agents can and cannot do - Map specific problems that AI agents could potentially solve - Prioritize problems based on complexity vs. potential impact	☐
2	**Process Documentation & Requirements** - Document current process steps for prioritized tasks in detail. - List required inputs/outputs for each task. - Identify what success looks like for each task. - Map critical decision points and requirements	☐
3	**Technical Readiness** - Verify that data is organized and complies with privacy standards - Select the Infra/base that your agents will run on - Identify what data/system/tools(s) the AI agent needs access to - Choose the right models for specific processes - Evaluate the Agentic Framework - Review the capability of orchestration - Select the best monitoring tools for visibility - Outline expected usage patterns to plan for scalability	☐
4	**Operation Planning** - Select a specific pilot task with a clear scope. - Assign team roles and responsibilities.	☐
5	**Budget & Resource Planning** - Calculate initial setup costs. - Estimate ongoing operational costs. - Plan for potential scaling costs. - Account for training and support costs.	☐

Step-by-Step Guide to Starting an AI Agent Project

Starting an AI agent project can feel overwhelming. Whether you're automating redundant tasks or planning an organization-wide AI transformation, this section provides a structured approach you can customize based on your goals, resources, and timeline.

Step 1: Assess opportunities.

The first step in implementing AI agents is conducting a **thorough assessment of your existing operations**.

Start by reflecting on your organization's daily operations, looking for areas with frequent complaints or inefficiencies. These pain points often emerge in customer service departments, where representatives handle the same queries repeatedly, or in financial operations, where invoice entry consumes valuable hours.

Once you have identified processes that could be improved, evaluate whether an AI agent is the right solution. AI agents excel at automating repetitive tasks, analyzing large datasets, and streamlining decision-making processes. However, they can't handle original creativity, ethical decision-making, or complex problem-solving that requires nuances and deep contextual understanding. Most AI agents also can't guarantee 100% reliability in every scenario.

The anticipated impact on your financial performance, the intricacy of implementation, and your team's preparedness to embrace new technologies will also influence your path to success.

Tip: Don't rush the assessment phase.

Most leaders rush to implement AI agents for obvious use cases, but the biggest opportunities often lie in unexpected areas like back-office operations or internal communications.

Step 2: Discuss documentation and requirements.

Now that you've identified potential opportunities, it's time to document the details of each process.

Start by selecting your top processes for improvement, then break down their current steps in detail. While you may already have process documentation, this is the time to create or update it. Thorough documentation helps establish a baseline for measuring improvements and determines which processes are best suited for AI agents.

Next, clearly define the inputs and outputs for each task. What information or resources are needed to start the process? What should the end result look like? Understanding these requirements helps identify what data and systems the AI agent will need to access and what specific outcomes it should achieve.

It's also **crucial to establish what success looks like for each task**. Define specific, measurable metrics that will indicate whether the AI agent is performing effectively. These could include processing time, accuracy rates, cost savings, revenue increases, employee or customer satisfaction scores, and more.

Finally, map out all critical decision points in the process. What rules or criteria are used to make decisions? What exceptions need to be handled? Understanding both standard operations and how exceptions are handled helps ensure that the AI agent can replicate human decision-making effectively and handle unexpected scenarios appropriately.

Step 3: Be technically ready.

Successfully implementing AI agents requires both a robust technical foundation and comprehensive security measures. Let's examine the essential things you'll need to be technically prepared.

Tools and Technologies

To make a useful and reliable agentic AI network, there are six essential components:

- **Models:** SLMs and LLMs that power the natural language processing capabilities, handling everything from fundamental analysis to complex function calling.
- **Agentic Framework:** The fundamental architecture determines how agents connect and interact.
- **Router:** The intelligent routing system ensures each request is directed to the most appropriate endpoint.
- **Tools and Integrations:** The vital APIs and interfaces that allow your AI agents to communicate with external systems and databases.
- **Monitoring and Control Systems:** Your command center for overseeing performance, maintaining security, and ensuring everything runs smoothly.
- **Scalable Hardware:** The infrastructure that your agents run on, which should scale up and down as the agents do more or less work.

The success of your AI agent implementation depends on how well these components are integrated and managed. When starting your AI agent journey, you'll likely navigate between two paths: developing in-house or seeking external solutions.

To successfully build an AI agent in-house, you need to connect many different components, such as models by providers like OpenAI and Anthropic, orchestration frameworks like CrewAI, monitoring tools like LangSmith, and many more. However, combining components from different vendors often leads to hidden complexities and challenges.

As your AI initiatives scale, this fragmented approach not only increases technical complexity and costs but also creates two major risks. First, distributing sensitive data across multiple external platforms introduces security vulnerabilities. Second, as your system grows, the connections between different vendor solutions become more fragile, reducing overall reliability.

For organizations without extensive AI capabilities, end-to-end platforms like **Arcee Orchestra** can simplify the process by providing all essential components in a single, integrated system.

Tip: Look for vendors that offer comprehensive solutions, rather than assembling components from multiple providers.

Arcee Orchestra is a turnkey, end-to-end platform with all components needed to build, test, deploy, and scale AI agents and agentic networks, ensuring enterprise-grade scalability and security.

Security

Security for AI agent systems requires special attention. **Traditional IT security measures aren't sufficient** because AI agents handle sensitive information across multiple interconnected systems. This requires comprehensive protection throughout the entire process.

Start by ensuring proper configuration and access permissions across all tools. Your technical team should evaluate which tools are essential for your specific use cases and verify their compatibility with your chosen AI agent framework.

For example, if you're using AI agents to process customer support tickets, they might need to access your CRM system, internal knowledge base, and communication platforms. This requires secure authentication and encrypted data transfer between these interconnected systems

Your security strategy must encompass data storage locations, infrastructure deployment options, authentication mechanisms, access controls, and continuous monitoring systems. Whether you choose to deploy on-premises, in your own Virtual Private Cloud (VPC), or in a managed cloud environment, you should maintain a security level that matches your organization's requirements.

Teams

The success of AI agent implementation also depends on your team's readiness and capabilities. Before diving into implementation, assess your current team's experience with AI workflows and development. Teams new to AI often benefit from starting with smaller projects to build expertise gradually.

Understanding your expected usage patterns is also crucial for technical readiness. Consider how your organization will interact with AI agents, the volume of prompts they will handle, and how demand may scale over time. These factors will influence how you design and deploy your AI agents to ensure optimal performance and scalability in the long run.

***Step 4*: Start planning the operation.**

After establishing your technical foundation and preparing your team, it's time to plan your implementation. Starting with a small, well-defined pilot project allows you to test the waters, learn from experience, and demonstrate value before scaling up. Let's break down the key considerations.

Choosing Your Test Pilot Project

When selecting your first AI agent implementation project, **look for a process that balances impact with manageable risk**. The ideal pilot should be important enough to demonstrate real value but not so critical that initial challenges would severely impact your operations.

Strong candidates for pilot projects include processes like customer support ticket routing, market research data analysis, sales content generation, code review, and data entry and validation. These processes typically can demonstrate immediate value without disrupting core operations

Assembling Your Implementation Team

The success of your pilot project depends heavily on having the right team in place. Your implementation team could include a project manager to coordinate all aspects of the implementation, security experts to ensure data protection, engineers to handle technical integration, and end users to provide practical insights and feedback. You might also want to add a change management specialist to facilitate smooth adoption.

The journey to implementing AI agents is exciting and challenging, bringing significant rewards when done right. Success comes from starting small with well-defined pilots while maintaining a clear vision of your strategic goals.

Tip: Already have automation engineers or RPA experts? Get them involved.

Their experience with process flows and system integration will be invaluable for your AI agent implementation.

***Step 5*: Calculate the costs.**

Implementing AI agents involves various costs that need careful planning. **Think of these costs as an investment in automating operations that will be able to work continuously with consistent performance.**

Consider these essential cost components:

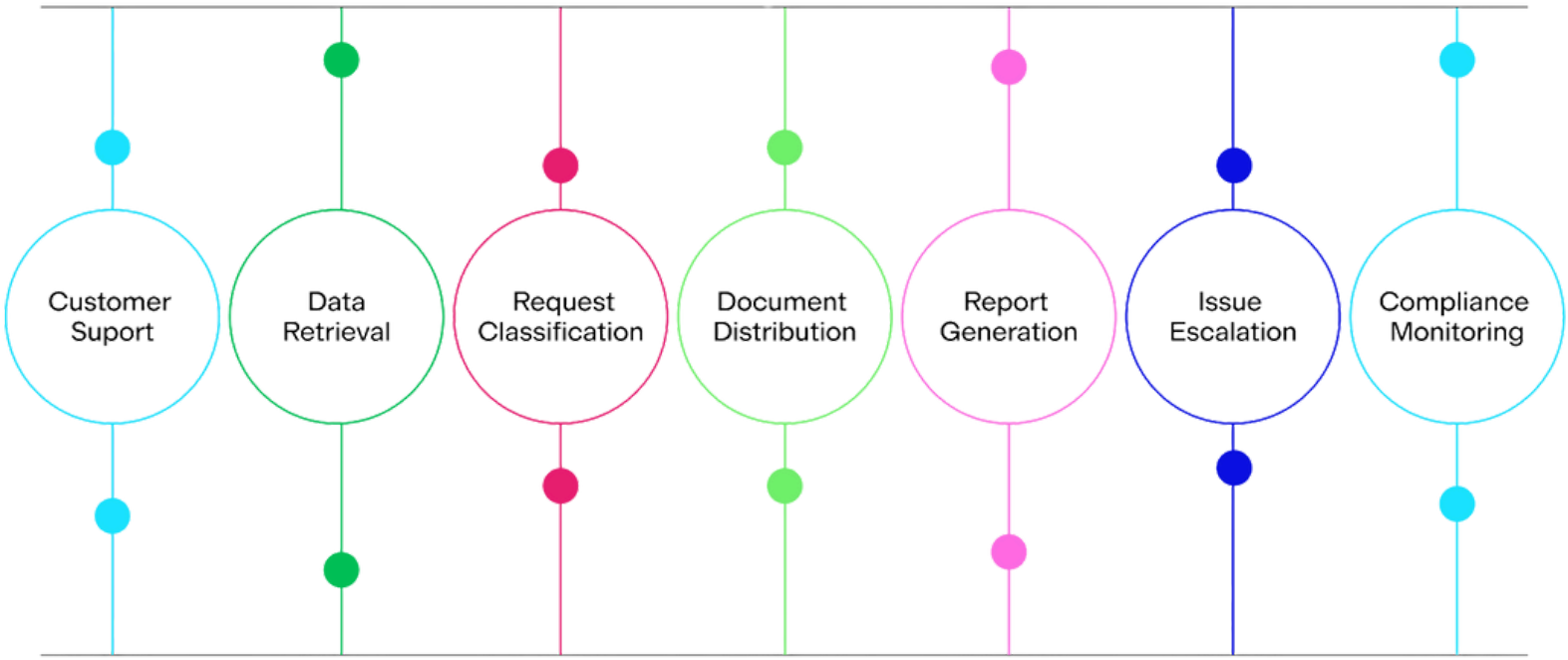
- Initial Setup: Infrastructure setup, platform licensing, development, security
- Operation: Model token counts, API usage, maintenance, cloud computing, technical support
- Human Resources & Training: Staff training, technical hires
- Scaling: Additional agents, system upgrades, new integrations.

Smart cost management starts with optimizing each component of your AI agent solution. For example, using small language models (SLMs) significantly reduces costs while maintaining performance. SLMs are more compact and can often achieve similar results while cutting compute spend by up to 50% compared to larger models.

Additionally, keep in mind that delaying AI implementation can end up costing you more than the implementation itself. As competitors put AI Agents to work to improve efficiency and customer experience, the competitive gap widens. The right investment today can position you for long-term success in an AI-driven market.

Up next: we introduce you to our own agentic AI platform, Arcee Orchestra.

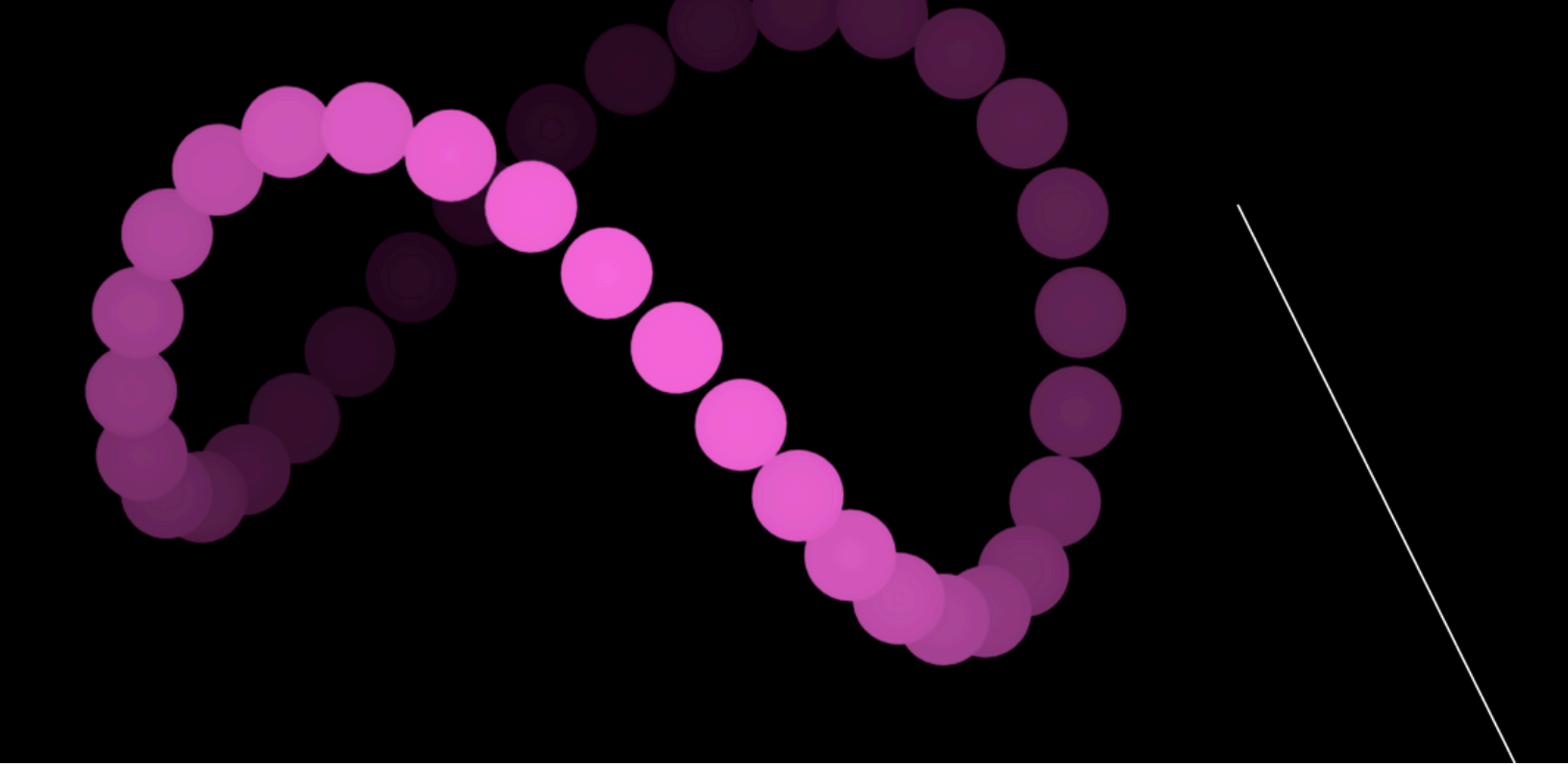
What can AI Agents do?
Just about ***anything.***



04

Introduction to Our Agentic AI Platform: *Arcee Orchestra*

04



What Makes *Arcee Orchestra* Special: The Models that Power the AI Agents

Here at Arcee AI, we built an end-to-end agentic AI platform called Arcee Orchestra.

When we first show it to businesses, the most common question we hear is, “Why would I need agentic AI that uses *your* models? I’m accustomed to using Claude [or insert any other popular LLM] and I don’t think a smaller model will do as well.”

The thing is, just like you, we also think the LLMs on the market are pretty incredible. These models are great to chat with because they are designed to interact with you, a human being. Yet, as we know, the whole point of agentic AI is to get computers and machines doing work with minimal involvement from human beings.

This is where Arcee AI’s small language models (SLMs) stand out.

While other models are optimized for back-and-forth discussion, our SLMs are built to excel within automated systems.

They’re trained to follow instructions precisely and have a deep understanding of API data. Here’s just one example: our general purpose model, Virtuoso-Large, is just 72 billion parameters–but it *outperforms* the largest LLMs (estimated to be over 1.3 trillion parameters) on instruction following benchmarks like *IFEval*.

In other words, the SLMs that power our agentic AI platform, Arcee Orchestra, are masters at getting stuff done–and they’re the ideal technology to power agentic AI.

Small Language Models (SLMs): The Logical Fit for Agentic AI

Historically, in the machine learning/AI industry, there has been a perception that the bigger a model is, the more capable it is. This led to companies generally seeking to use the largest models they could find.

But in today's world, where models have breached trillions of parameters, using the largest models often leads to out-of-control costs and scaling issues. Also, large models are very difficult to train for specialized tasks. These challenges become even more insurmountable when companies try to incorporate them into an agentic AI solution–especially for tasks that require highly-specialized knowledge.

Here at Arcee AI, we're the industry leaders in small language models (SLMs). We've pioneered some of the top techniques (model merging, distillation, Spectrum) for training small models that compete with, and often outperform, their large counterparts.

With the rise of agentic AI, we knew that we were perfectly positioned to provide the ideal models to power this technology. That's what inspired us to build Arcee Orchestra, a platform that combines our SLMs with intelligent model routing and orchestration–enabling you to use **the right model for the right tasks**, which reduces cost and latency, while maintaining or even improving accuracy.

When it comes to language models, bigger is not always necessary.

For most businesses, your teams sometimes *do* need larger models. But they certainly do *not* need the biggest, most expensive LLM for routine work like summarizing meeting transcripts or generating follow-up emails. By offering a variety of model sizes, and intelligently routing your tasks to the right model, Arcee Orchestra dramatically reduces your cost without sacrificing performance.

We also realized that **for any agentic system to be valuable, it needed to be able to integrate with existing systems**. Arcee Orchestra includes more than 200 pre-built integrations, many of them to popular applications like Salesforce, Slack, Dropbox, MS Office, and GSuite. This means that *you* don't have to be the integrator (which is often one of the most costly barriers to successful agentic AI deployments). You can use the integrations directly out-of-the-box without any custom code.

The fully integrated nature of Arcee Orchestra makes it incredibly easy not only to get started, but also to maintain and evolve your workflows. One development challenge that many customers don't consider until it's too late is that when models change, the optimal way to integrate and interact with them also changes. An AI agent that performs perfectly today may not work as well tomorrow, if a new model comes out and is integrated into the agent. With Arcee Orchestra, this complexity is abstracted away, and you get to utilize SOTA models without having to worry about a new model not working in your solution.

For Agentic AI, You Want Models *Specialized* in Automated Systems

Arcee Orchestra comes with four SLMs out-of-the-box: three general purpose models (72B parameters, 32B parameters, 14B parameters) and a coding model (32B parameters).

What all of the Orchestra models have in common: we put them through a highly-specific training process to make them excel at automated workflows and agentic systems.

Let's take a closer look at some of the challenges that arise in the context of agentic AI.

When you have different providers for models and frameworks/platforms, there's always an inherent risk that performance will suffer–since the two components were not built to work together. Also, different models require different prompting strategies, which means that–in order to get various models and platforms working well together–you need expertise in all of them. This is expensive in terms of labor and time-to-value.

Large third-party models are trained to excel at satisfying the (human) end user, by being conversational and having the ability to answer any query, such as "Create a song from the perspective of a giraffe." These types of capabilities are great for consumers… But they increase the size of the model, and add nothing to what companies actually need the model to do for their business tasks.

The Arcee AI SLMs that power Orchestra provide the requisite expertise to enable the platform to interface smoothly with a diversity of models and platforms, and to be able to execute business tasks. Our SLMs are carefully fine-tuned to be highly capable at instruction-following and function calling. Additionally, they understand the nuances of API inputs and outputs, which drastically improves the accuracy of AI agents.

SLMs for Agentic AI: *Affordable, Efficient, & Secure*

The models in Arcee Orchestra have been specially trained for agentic AI, which is what makes them so good at it. And with Orchestra, you get access to multiple models, with intelligent routing so the right model is used for each task. But... that's still not all. **Our SLMs also solve one of the other major challenges of working with third-party, closed-source LLMs: security and compliance.**

Thanks to their relatively compact size, SLMs can be run in your own environment. One of the largest blockers we hear from customers is that they can't utilize models hosted in another provider's environment due to regulatory and compliance requirements.

Since SLMs can run efficiently on smaller servers, Arcee Orchestra–including all Arcee AI models–can be deployed in your own environment, utilizing your security controls. Your requests and your data never leave your security perimeter. You get to decide how the models are used, what data they have access to, and who can access them. As you can imagine, this is music to the ears of your CISO and other security experts and data governance experts.

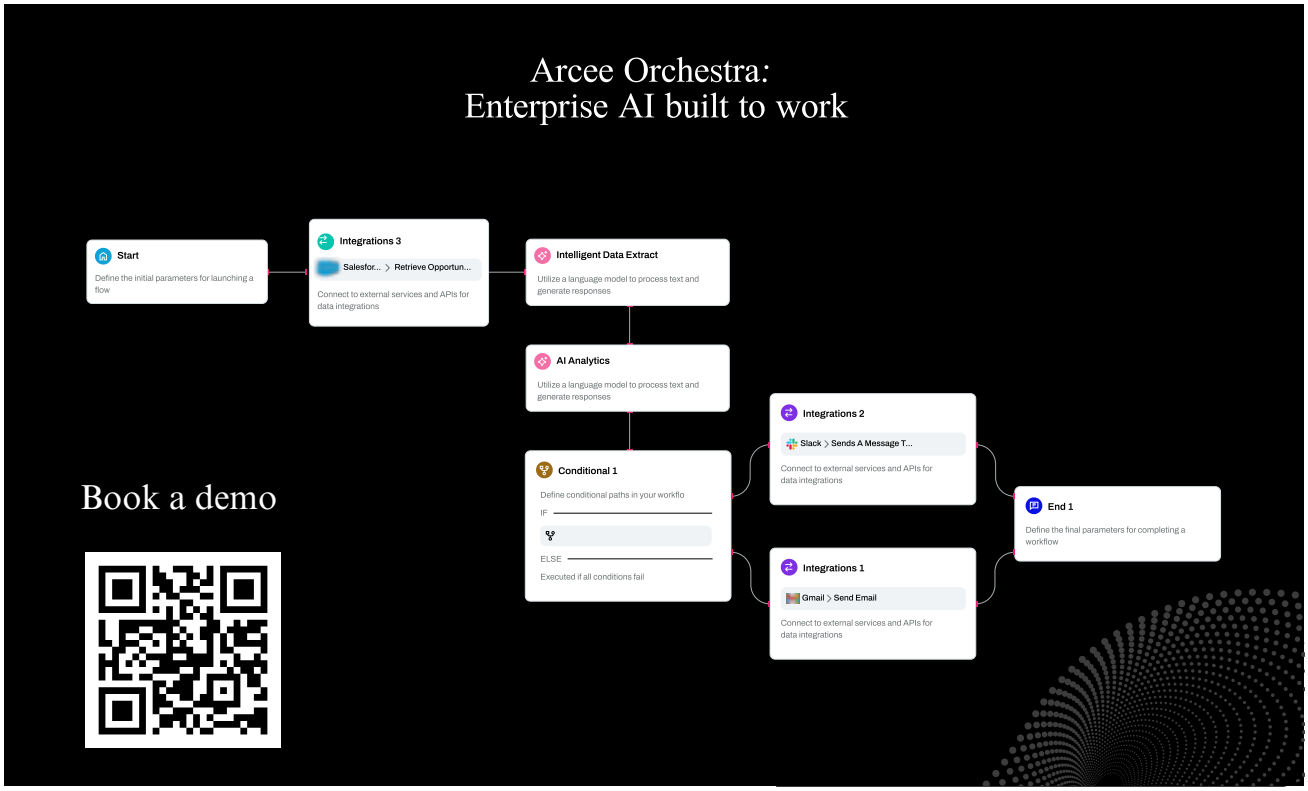
Agentic AI & Domain-Specific Models

We would be remiss if we didn't also mention that Orchestra can include one or more **SLMs that have been fine-tuned on data of your choice**: your company's proprietary data, industry data, data specific to a certain use case, etc.

Interestingly, we've found that in most cases where a company thinks they need a fine-tuned model, they can in fact solve their task with a well-built workflow–which ultimately is the most cost-effective solution. There are certainly some cases where we do recommend fine-tuning our SLMs for your needs; examples of this are when models are required to have a deeper understanding of a specific domain, or when they need to change how they perform a specific task. And needless to say, this process is much more affordable and effective compared to what is generally an exorbitant cost for fine-tuning an LLM.

Arcee Orchestra:
The Easiest Tool for World-Class AI Agents

We hope that this guide has given you a sense not only of what agentic AI is, but also of how it's currently being used, and how you can help your organization to get started with it. We're confident that our turnkey agentic AI platform, Arcee Orchestra, is the best solution for agentic AI that can truly make an impact from day one. We'd love to talk to you about evaluating where your organization stands with its AI strategy, and how to get started on your journey towards agentic AI. **Get in touch** today and our team will book a time!



References

Andre, D. (2025, January 2). *AI agents statistics for 2025: Market impact & industry insights.* All About AI. https://www.allaboutai.com/ai-agents/statistics/

Atsmon, Y. (2016, August). *Strategy and corporate finance special collection: Resource allocation.* McKinsey & Company. https://www.mckinsey.com/~/media/mckinsey/business%20functions/strategy%20and%20corporate%20finance/our%20insights/strategy%20and%20corporate%20finance%20special%20collection/final%20pdfs/mckinsey-special-collections_resourceallocation.pdf

Bebo, K. (2024, July 17). *The true cost of workplace distractions.* Timely. https://www.timely.com/blog/the-true-cost-of-workplace-distractions

Chopra, G. (2024, March 17). *Evolution of AI agents: The beginning (Part 1).* EightGen AI. https://eightgen.ai/evolution-of-ai-agents-the-beginning-part-1/

Forbes. (2024, September 26). *AI agents will be the key to achieving ROI from AI.* Moor Insights & Strategy. https://www.forbes.com/sites/moorinsights/2024/09/26/ai-agents-will-be-the-key-to-achieving-roi-from-ai/

Genius. (2024, September). *Average time to hire by industry (2025 statistics).* Genius. https://joingenius.com/statistics/average-time-to-hire/

Glassdoor. (2015). *50+ HR recruiting stats that make you think.* Glassdoor for Employers. https://www.glassdoor.com/employers/blog/50-hr-recruiting-stats-make-think/

Grand View Research. (2023). *AI agents market report.* https://www.grandviewresearch.com/industry-analysis/ai-agents-market-report

Harvard Business Review. (2023, October). *How AI can help leaders make better decisions under pressure.* https://hbr.org/2023/10/how-ai-can-help-leaders-make-better-decisions-under-pressure

Homebase. (2024, February). *Time management statistics.* https://www.joinhomebase.com/blog/time-management-statistics#:~:text=The

Hubstaff. (2025, January 17). *Time management statistics.* https://hubstaff.com/blog/time-management-statistics/

Jobanputra, K. (2024, August 22). *Customer service: How AI is transforming interactions.* Forbes Business Council. https://www.forbes.com/councils/forbesbusinesscouncil/2024/08/22/customer-service-how-ai-is-transforming-interactions

Johnson & Johnson. (2024, October 10). *6 ways Johnson & Johnson is using AI to help advance healthcare.* https://www.jnj.com/innovation/artificial-intelligence-in-healthcare

References (cont'd)

McKinsey & Company. (2025, January). *Superagency in the workplace: Empowering people to unlock AI's full potential at work.* https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work

Microsoft. (2023, November 15). *Introducing Microsoft Copilot Studio and new features in Copilot for Microsoft 365.* Microsoft 365 Blog. https://www.microsoft.com/en-us/microsoft-365/blog/2023/11/15/introducing-microsoft-copilot-studio-and-new-features-in-copilot-for-microsoft-365/

Mittal, N., et al. (2022, October). *State of AI: Fifth edition.* Deloitte AI Institute. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-ai-institute-state-of-ai-fifth-edition.pdf

Levahi, et al. (2024). *The rise of the digital colleague.* https://www.moodys.com/web/en/us/insights/resources/the-rise-of-the-digital-colleague.pdf

Nawaz, S. (2024, April). *Top 24 use cases of AI agents in business.* Ampcome. https://www.ampcome.com/post/24-use-cases-of-ai-agents-in-business

Nyoni, E. (2024, March 20). *Customer self-service: Why it matters and how to improve it.* Helpjuice. https://helpjuice.com/blog/customer-self-service

Purdy, M., & Williams, A. M. (2023, October). *How AI can help leaders make better decisions under pressure.* Harvard Business Review. https://hbr.org/2023/10/how-ai-can-help-leaders-make-better-decisions-under-pressure

Recruiting from Scratch. (2024, November). *How hiring takes time.* https://www.recruitingfromscratch.com/blog/time-spent-hiring

Rimol, M. (2021, April 28). *Gartner forecasts worldwide hyperautomation-enabling software market to reach nearly $600 billion by 2022.* Gartner. https://www.gartner.com/en/newsroom/press-releases/2021-04-28-gartner-forecasts-worldwide-hyperautomation-enabling-software-market-to-reach-nearly-600-billion-by-2022

Rosales, I. G. (2024, September). *Average time to hire by industry (2025 statistics).* Genius. https://joingenius.com/statistics/average-time-to-hire/

Security Magazine. (2023, July 24). *80% of data experts believe AI increases data security challenges.* https://www.securitymagazine.com/articles/100631-80-of-data-experts-believe-ai-increases-data-security-challenges

References (cont'd)

SellersCommerce. (2025, January). *AI agents statistics: Usage and market insights (2025).* SellersCommerce Blog. https://www.sellerscommerce.com/blog/ai-agents-statistics

Sobot. (2025, February). *Steps to reduce costs with AI customer service.* Sobot. https://www.sobot.io/article/how-ai-improves-customer-service/

Stallbaumer, C. (2023, November 15). *Introducing Microsoft Copilot Studio and new features in Copilot for Microsoft 365.* Microsoft 365 Blog. https://www.microsoft.com/en-us/microsoft-365/blog/2023/11/15/introducing-microsoft-copilot-studio-and-new-features-in-copilot-for-microsoft-365/

Toyota Global. (n.d.). *Toyota production system.* Toyota. https://global.toyota/en/company/vision-and-philosophy/production-system