



AHEAD

Adopting Generative AI: *From Idea to Execution*

Executive Summary

Adopting Generative AI requires strategic planning and execution for successful integration. This whitepaper offers a comprehensive roadmap for leveraging GenAI, advocating for a measured approach to minimize risks and maximize business impact, thereby empowering organizations to harness the potential of generative artificial intelligence effectively.

Key points include the historical context of AI, the impact of the transformer model introduced in 2017, and the rapid improvements in AI capabilities. The document also outlines essential steps for businesses: assessing organizational needs, fostering innovation, prioritizing initiatives, and choosing between in-house development and existing solutions.

Introduction

In the ever-evolving landscape of technology, generative artificial intelligence (GenAI) stands out as a revolutionary force. As businesses strive to stay ahead in an increasingly competitive market, generative AI presents a unique opportunity to innovate, streamline operations, and enhance customer experiences.

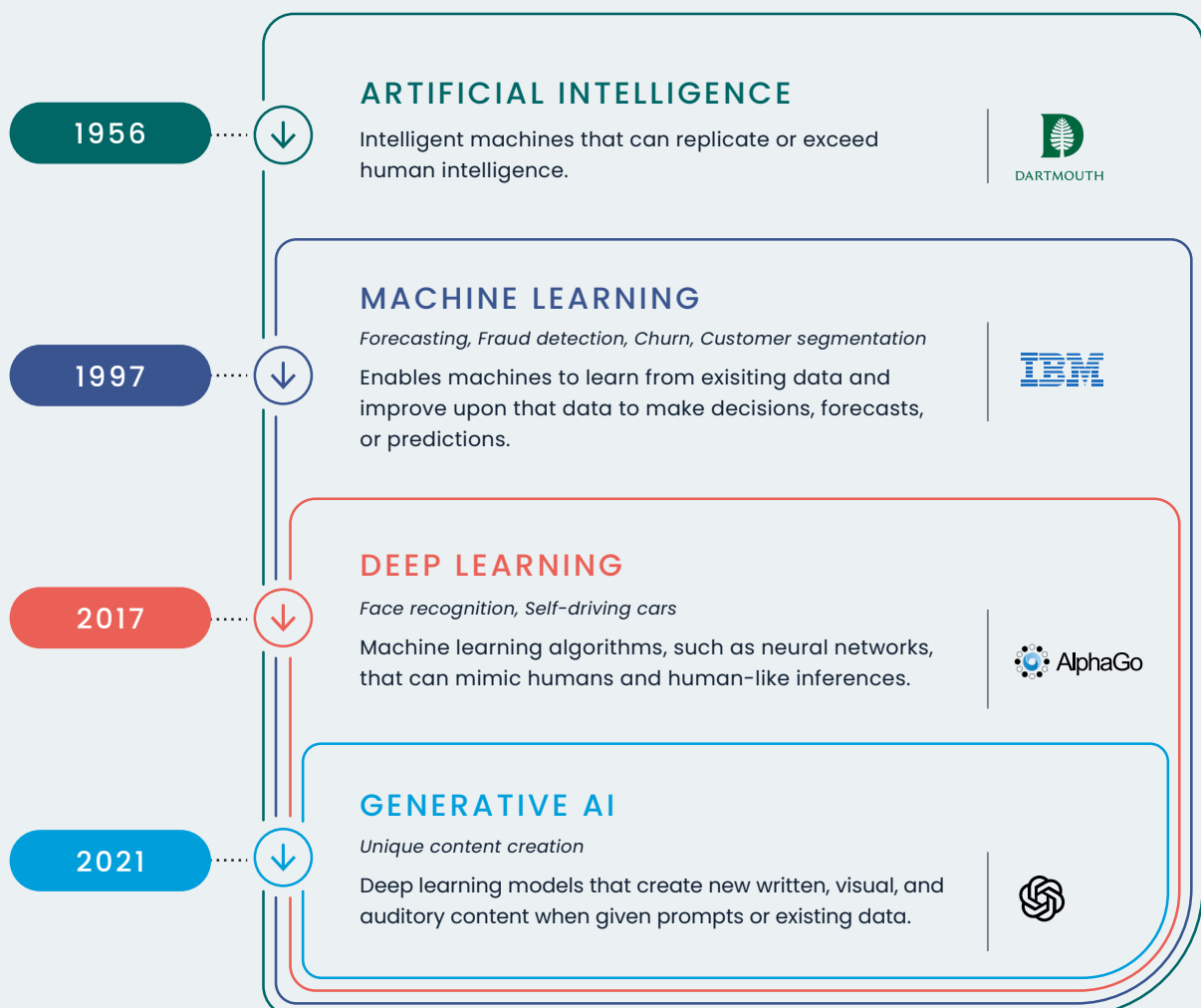


A Brief History of AI Innovation

While the current buzz around GenAI is certainly based on new and groundbreaking applications of mathematics and data science, the overarching concept of artificial intelligence (AI) dates as far back as 1956. Continuing innovation in the study of AI has produced several defining moments and created numerous sub-specialties in the general AI field. In the late 1990's, machine learning (ML) became the first of these AI subsets, allowing algorithms to analyze data, make predictions, and continuously learn over time with the help of human intervention. Fast forwarding two decades to the late 2010's, deep learning (DL) emerged as a subset of ML. Using artificial neural networks designed to mimic the function of the human brain, deep learning allowed computers to analyze more complex data sets with limited intervention from humans. Traditional machine learning and deep learning methods are still wildly powerful and applicable to many of today's business challenges. Capabilities in natural language processing (NLP), object detection, and speech-to-text are all examples of this.

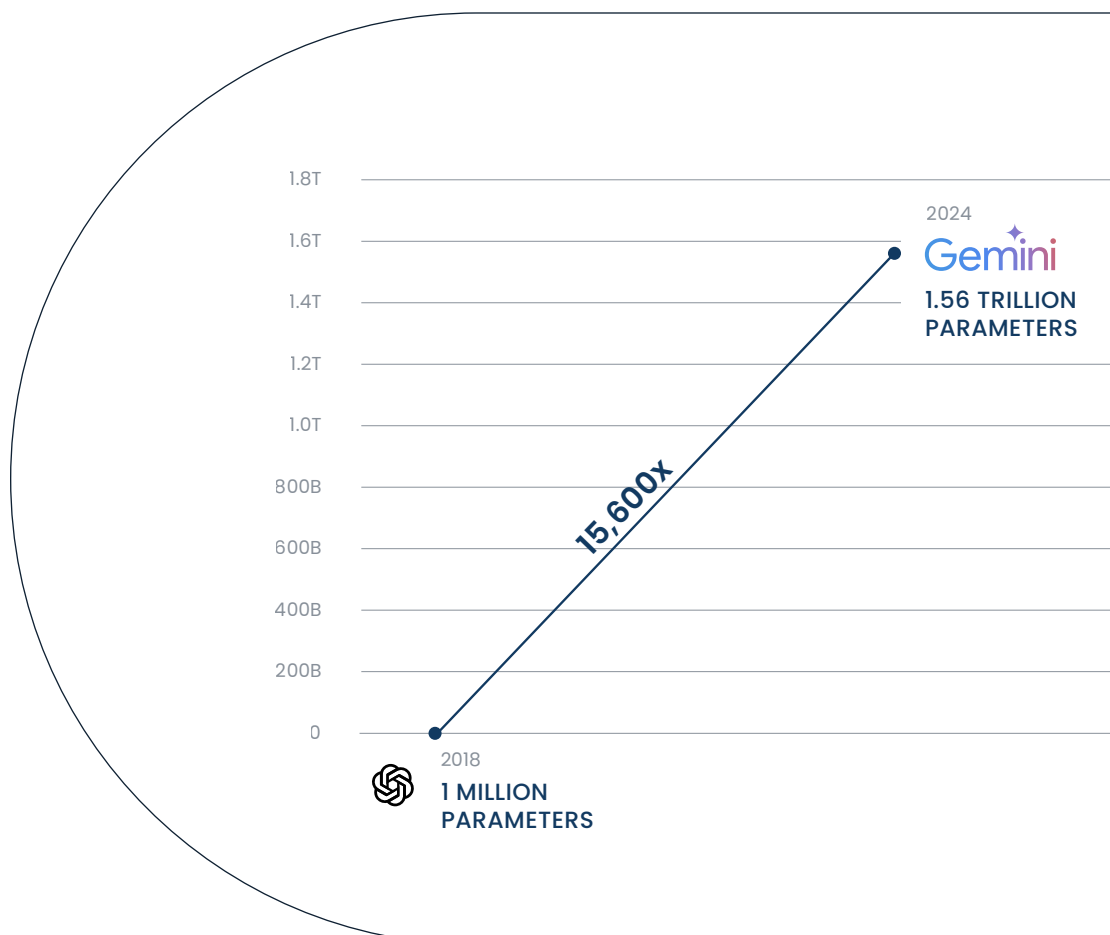
To this point, while interacting with human language was possible, it was far from perfect and relied heavily on pattern matching with little to no understanding of context. 2017 brought the early underpinnings of GenAI, led by the publishing of Google's research paper: "[Attention is All You Need.](#)" The attention mechanism thrust deep learning into the modern age, setting the framework for the transformer model, which is responsible for powering all of the large language models (LLMs) we hear about today. In the purest of definitions, GenAI is able to create net-new, never before existing content using the contextually aware capabilities of LLMs. On November 30, 2022, OpenAI captivated the world with the release of ChatGPT, kicking off the frenzied interest and innovation we are seeing today.

AI Evolution and Subsets



The Hype Cycle & Adoption

US markets alone have created an AI market cap increase of over 7.8 trillion dollars and spawned over 1,500 new AI startup companies in 2024. It is no wonder that businesses are feverishly seeking to embrace this new technology and uncover new ways to leverage it within their organizations. With all the major technology players racing to introduce AI enabled platforms, sell AI tools, and win the market, navigating the AI minefield to find success can be daunting. With [Harvard Business Review](#) estimating that as high as 80% of AI projects fail, it is paramount to have a measurable approach and methodology to embrace AI in a manner that reduces risk and cost while limiting efforts solely to those that have a higher likelihood of producing a positive business impact. Compounding the problem statement further, the speed of innovation in the GenAI space is showing an expanding field of foundational models and capabilities like nothing else in the history of technology. To put this in perspective, consider the fact that in 2018, OpenAI's GPT-1 had 100 million parameters, but by 2024, Google's Gemini Ultra was released at 1.56 trillion parameters. This represents a 15,600x improvement in only six short years.



Increasing the Success Rate of GenAI Initiatives

As we delve further into the practical implementation of GenAI in business contexts, it is essential to address the fundamental steps and considerations for successful integration. The subsequent sections will provide guidance on evaluating your organizational needs, fostering creative ideation, prioritizing initiatives, and strategically deciding between developing in-house solutions or purchasing existing ones. Additionally, we will explore the distinctions between public and private deployments, aiming to empower your organization to effectively harness the transformative potential of Generative AI.

“ *A problem well-stated is a problem half-solved*
— *Charles Kettering*

Evaluating Business Needs

It should go without saying that solving any problem is impossible without a problem definition. While a seemingly obvious step, companies are frequently guilty of skipping the basics, rushing to invest in expensive AI technologies, getting in line to gain access to big iron GPU-enabled compute, and generally allocating budget to “AI initiatives” without an understanding of which tools and technologies are the right fit to solve the problems they are facing. Solving problems with generative AI must start at the business level, but with calculated involvement from technology teams.

Developing the Ideation Process

The ideation process is the first step toward success, so long as it is inclusive of all key stakeholders, such as business leaders, IT leaders, and individuals who have in-depth understanding of the organization’s data and data availability. Naming is not necessarily important, but for simplicity, we may call this team the AI innovation Board (AIB). Once this team is established, step one is to document ideas that clearly address the business problems that the organization is hoping to solve. It’s not uncommon for hundreds of ideas to be presented, but in most cases, only a few will end up with the right parameters to be a target for innovation.



Prioritizing & Down-Selecting

Not every business problem has an AI solution, and AI is not the solution to every business problem. Don't expect what you can't inspect. An organization should be able to quickly pare off ideas by methodically evaluating each idea for its ability to have meaningful measurement. Any problem that cannot have a key performance indicator applied to it is not an idea worth pursuing. Ranking the remaining ideas by perceived measurable business value is the next step. Stack rank all ideas and assign a number that can be used for score carding. What follows is certainly one of the more challenging steps, but also the most important. Feasibility needs to be scored, taking into account the current state of the organizations' data, governance, data quality, and data availability. Once again, stack rank and score accordingly.

Deciding to Build or Buy

While it is easy to get excited about building GenAI software from scratch, it is a significant undertaking. One of the simplest missteps in evaluating AI strategy is a failure to [identify which projects should simply be a software or capability purchase.](#) The easiest way to overcome this conundrum is by thinking through your business problem statement and trying to determine if this is a problem that impacts other organizations in a similar way. If the answer is yes, you are looking to procure a solution from a vendor. Universal problems will almost always be solved by companies who are dedicated to solving that problem and better-equipped to provide full-lifecycle support. However, it's not always necessary to look to new players for solutions to universal challenges. An evaluation of your organization's current investments may uncover AI features that already exist, but have yet to be leveraged.



Choosing Between Public & Private Models

If the business has identified problems GenAI can solve that require customization, or interact with critical or sensitive data, it's time to evaluate where to start. It is important to understand that an organization's starting point does not need to be its ending point. Considering that most AI initiatives fail, it is imperative to experiment in a way that is low-risk and embrace a fail-fast mentality. Publicly available large language models and cloud APIs represent a simple place to prove value and determine whether or not an idea will work. While this may need to be proven with dummy or non-sensitive data, it is possible nonetheless. It is also important to note that "publicly available" does not always mean that the models must be run in a public environment – many LLMs can be run in a private environment.

Assessing Model Requirements

Success within a generic large language model does not necessarily mean you've found the right tool for long term production. For example, while an organization could certainly prove out that legal document summarization is possible and adds value using GPT-4, the cost and token utilization of such a model for this purpose becomes irresponsible when considering that you are using a model that also knows the history of art and the works of Shakespeare in totality. Upon completing a proof of value, organizations are finding that a "Composition of Experts" approach—where we use a collection of smaller, more lightweight purpose-driven models—is more efficient, less costly, and easier to maintain over time. This methodology produces AI applications similar to that of a microservices application. As the model capabilities mature and change, it is much simpler to upgrade a portion of an AI application than the entire thing.



Final Thoughts

AI and GenAI are propelling organizations to new levels of efficiency and capability, but only after careful consideration of their specific use cases. Thus, a methodical approach to identifying valid opportunities and experimenting with those possibilities is key to long-term success. While several clichés are appropriate, the one that rings truest is: “slow is smooth, and smooth is fast.”

Contributing Author:

Mike Menke, Field Chief Technology Officer

AHEAD

Combining cloud-native capabilities in software and data engineering with an unparalleled track record of modernizing infrastructure, we're uniquely positioned to help accelerate the promise of digital transformation.

Visit us at ahead.com.

National Hubs

CHICAGO

444 W. Lake Street
Suite 3000
Chicago, IL 60606

NEW YORK

500 Fifth Avenue
Suite 1500
New York, NY 10110

ATLANTA

1117 Perimeter Center
W406
Atlanta, GA 30338

SAN FRANCISCO

2000 Crow Canyon Place
Suite 250
San Ramon, CA 94583