

# Making Bots Better by Leveraging Transformer-Based Solutions



 **opusresearch**

March 2023

Dan Miller | Opus Research

Opus Research, Inc.  
893 Hague Ave.  
Saint Paul, MN 55104

[www.opusresearch.net](http://www.opusresearch.net)

Published March 2023 © Opus Research, Inc. All rights reserved.



“Transformers” are no longer just toys that change from a vehicle, device, or animal, to a robot action figure and back again. Instead, in the world of conversational AI, transformers are a type of artificial neural network architecture that evolve conversational AI solutions far beyond what was previously possible. These capabilities span from improved understanding, free-flowing context, and generation of personalized content. Rather than enduring the development slog associated with previous chat technologies, transformers empower practitioners to build advanced conversational flows in a fraction of the time, particularly when accompanied by a heavily optimized conversational AI platform.

ChatGPT, an application of transformer – the “T” in GPT – technology that achieved rockstar status, empowers any user to search for answers via natural conversation, compose drafts of essays, and even aid software developers in writing code. Transformers enable applications like ChatGPT by providing an efficient pre-training phase on massive corpora spanning many topics and phrasings. These pre-trained – the “P” in GPT – models understand semantics and relationships between words, phrases, and sentences. In addition to understanding users, as much of previous technology was limited to, the unprecedented language skill of transformers expands conversational AI to generative – the “G” in GPT – functionality. While the ability to generate content catalyzes public interest and pressure on enterprise conversational AI usage, brands are paralyzed by the concept of employing technology lacking mechanisms to enforce compliance and their image.

### The Challenges of ChatGPT and Transformer-based Solutions

Accelerated expectations from ChatGPT users pose a large challenge for enterprise executives in charge of customer experience and contact centers, but the capabilities of ChatGPT, and generative solutions in general, yield nothing but unacceptable solutions to these enterprise leaders. In particular, these solutions are known to produce factually incorrect responses frequently and even offensive responses on occasion. While this is no secret, with companies like OpenAI (ChatGPT's creator) openly publishing these limitations, most companies building transformer-based conversational AI do not currently offer solutions to these problems.

In addition to deal-breaking properties like compliance and brand image violation, transformer-based solutions present other substantial challenges. These challenges are rooted in fundamental properties of massive black box, pre-trained models:

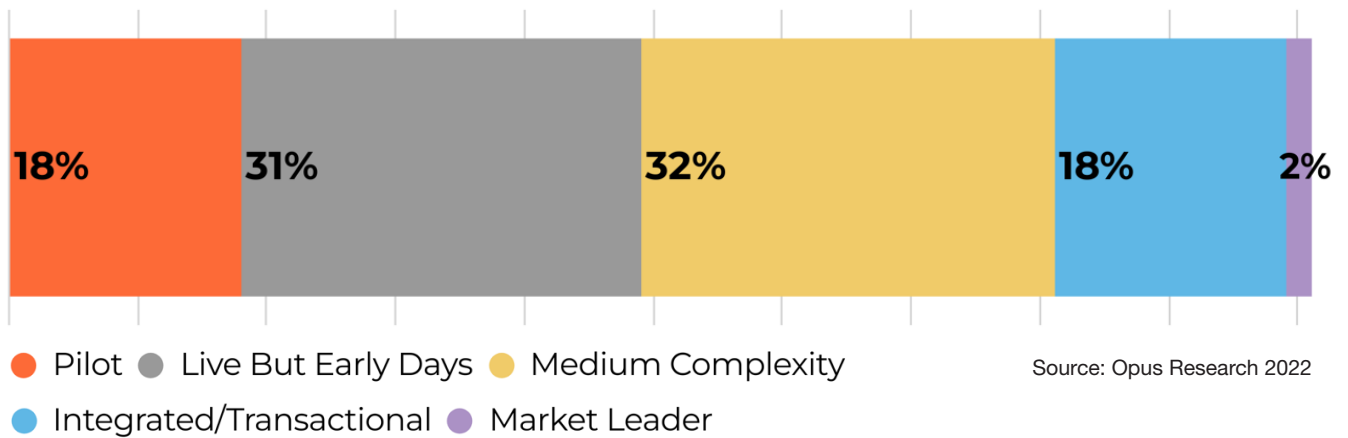
- It's difficult to inspect or explain the behavior of the pre-trained model. When the AI misunderstands the user, brands need to lean on machine learning experts to monitor usage and suggest paths to improvement.
- The improvement process is a dark art at best and intractable at worst. In its simplest form, this improvement process is the task of adding/removing data from the model's training corpora so that it handles a topic better while also avoiding performance degradation on other topics. In practice, this frequently devolves into a whack-a-mole situation where developers fix one issue, but create another, and then repeat the process until it's better "on average". When working with models trained on tens of billions of words, like ChatGPT, a manual brute force approach to improving training data is simply intractable.
- The resource requirements for training and deployment can be prohibitive for companies that do not specialize in or have a large budget for AI R&D. For example, the model preceding ChatGPT's GPT-3.5 foundation, GPT3, cost around \$4.6M to train. With the size of ChatGPT and the number of active users, it has been estimated that it costs around \$100K to operate daily.



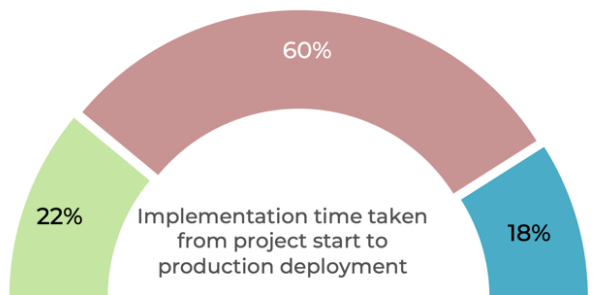
### The Importance of Adopting Transformers and a Path Forward

Although there are several challenges to adopt transformer-based conversational AI solutions, enterprises know they need to move past the disappointing chatbots of the past. In a recent survey, Opus Research asked business decision-makers about the maturity level of their intelligent assistant bot offerings (Figure 1 below). Only 2% see themselves as “market leaders with sophisticated voice or chatbots” and another 18% see their solutions as “mature,” with the remainder either considering a solution, early in production, or handling conversations of limited complexity. Additionally, many of these brands see their expenses (10+ employees to support) and timelines (weeks-to-months) of chatbot implementations as major challenges, prompting respondents to indicate that ease and speed of implementation as top considerations in selection of solution providers (Figure 2).

Figure 1: Current State of Maturity Levels for Intelligent Assistants

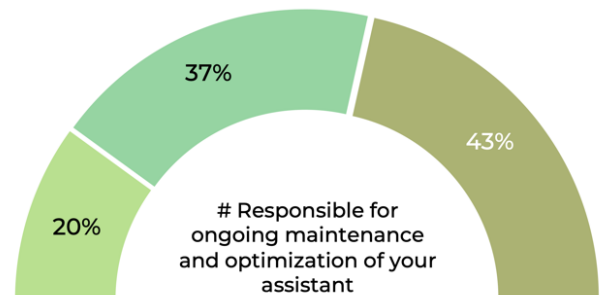


#### Time to Implement Solution



■ <2 weeks ■ 2-4 weeks ■ 4-8 weeks ■ >8 weeks

#### Number of People to Maintain

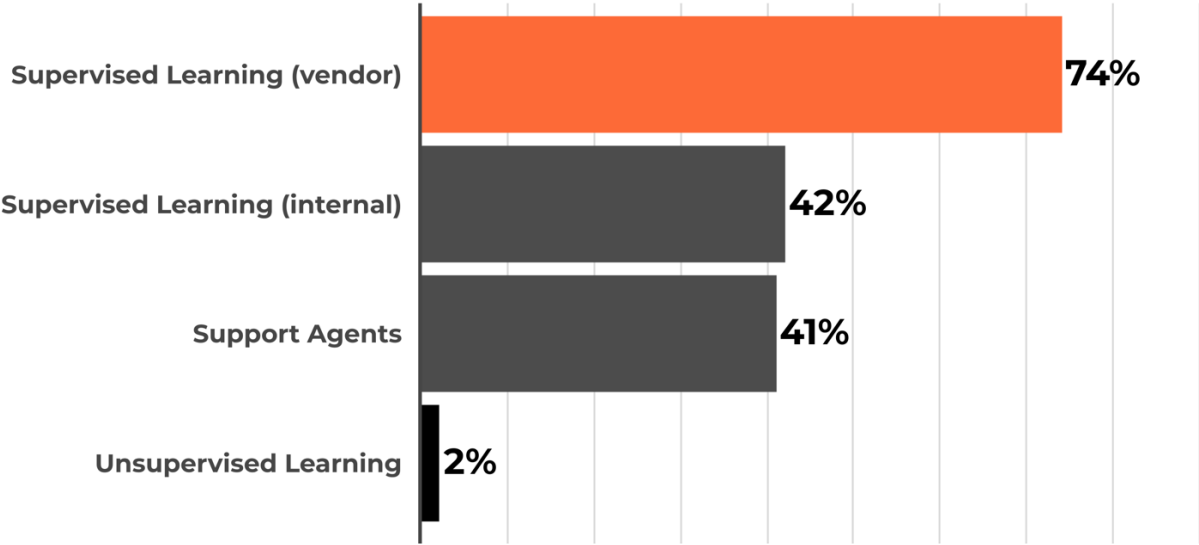


■ Less than 10 ■ More than 10 ■ NA, we rely on our vendor



Beyond the perception and staffing of their experiences, a small 2% of respondents are using unsupervised assistants (e.g., transformers), highlighting the abundance of previous-generation technology in the field – an easy explanation for why it takes so long to build experiences and why brands do not see themselves as leaders in the space.

Figure 3: Methods Used to Train Intelligent Assistants



Source: Opus Research 2022

While Opus Research did not survey reluctance to adopt state-of-the-art unsupervised technology by respondents, challenges of brand image, compliance, and expertise requirements to wield such technology were likely causes. Addressing these challenges, yielding reduced expenses and increased capabilities of chatbots for brands creates a new landscape of conversational AI vendors to succeed those stuck in the last generation of technology.

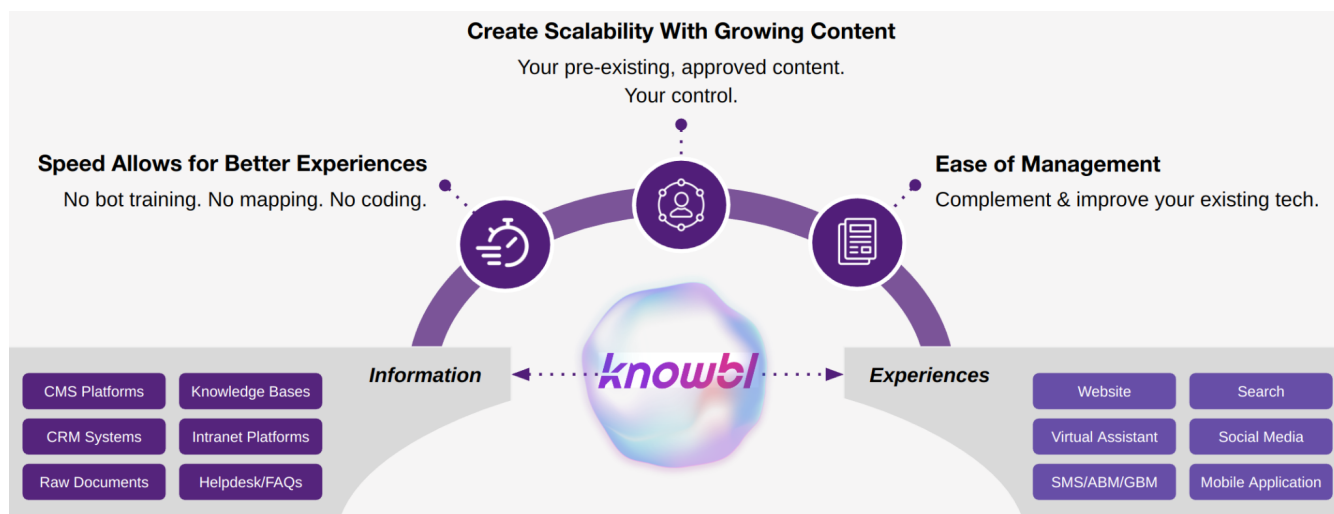
**Knowbl Capabilities & Use Case**

A new category of firms, such as Knowbl, has emerged to address these challenges and bring transformer technology to production on brand's digital properties. Specifically, Knowbl:

- Avoids the compliance and brand image risk associated with generative transformer models by constraining responses to content approved by the brand. Their model selects the best brand content to respond with rather than attempting to generate novel responses that may not be acceptable to the brand.
- Leverages pre-trained models to provide brands the building blocks they need to construct robust conversational experiences quickly rather than sticking them with an effectively immutable, black box model.
- Provides explainability tooling to enable non-technical users to quickly diagnose and fix issues in the AI's ability to handle their content or data.
- Hides the complexities and expenses of deploying and managing large language models.



Figure 4: Knowbl's Solution



Due to customer and operating expense pressures, the decision maker's appetite for change has already shifted. A financial services credit card company business executive sought Knowbl to help her control how her organization improves her rewards chatbot, because "I can't get access to our scarce technical resources", who are distributed and only work on activities contained to their specific business unit, versus hers, who owns the relationship with the paying customer. This governance model hasn't allowed her team to rapidly care for the way customers are being treated, causing the business loss of satisfaction and subscribers, affecting P&L.

Knowbl suggested a better way for her and the organization, giving the business executive the tools to ingest her brand's knowledge from five out of their 200 branded properties, to test Knowbl's ability to learn, contextualize, and answer her test cases, immediately.

The business case to deploy became palpable when comparing the current time it took her customers to search the website or engage her agents to find answers. The use of the concierge resulted in immediate satisfaction to clients, reducing the agent search time to seconds vs minutes for agents handling those basic inquiries missed by user searches. Beyond the drastic experience improvements, the business case became even more obvious when considering the time to market and development costs. Her non-technical skills were all she needed to launch the experience, test it accordingly, and leave her with confidence the concierge would handle all the ways her customers had been engaging the knowledge previously. This process also amplified her confidence she could actually control experience misdirections in real time using Knowbl's few-shot and automated context management, limiting her need and reliance on the internal and limited developer and data science resources. Her experience team gained speed of experience improvement, making it hours vs days or months.

### Conclusions

With the excitement and increased expectations from ChatGPT's release, brands are looking for solutions to bring ChatGPT-like experiences to their digital properties. However, these technologies present a new set of production challenges to their data science and compliance teams. Tech giants like Microsoft and Google compete to bring the smartest large language models to the public, but brands need solutions that address production challenges of these models. Emerging companies, like Knowbl, aiming to bridge these gaps are now at the center of attention for many of these brands.



## About Opus Research

Opus Research is a diversified advisory and analysis firm providing critical insight on software and services that support multimodal customer care. Opus Research is focused on “Conversational Commerce,” the merging of intelligent assistant technologies, conversational intelligence, intelligent authentication, enterprise collaboration and digital commerce.

---

**For sales inquires please e-mail [info@opusresearch.net](mailto:info@opusresearch.net) or call +1(415) 904-7666**

This report shall be used solely for internal information purposes. Reproduction of this report without prior written permission is forbidden. Access to this report is limited to the license terms agreed to originally and any changes must be agreed upon in writing. The information contained herein has been obtained from sources believe to be reliable. However, Opus Research, Inc. accepts no responsibility whatsoever for the content or legality of the report. Opus Research, Inc. disclaims all warranties as to the accuracy, completeness or adequacy of such information. Further, Opus Research, Inc. shall have no liability for errors, omissions or inadequacies in the information contained herein or interpretations thereof. The opinions expressed herein may not necessarily coincide with the opinions and viewpoints of Opus Research, Inc. and are subject to change without notice.

Published March 2023 © Opus Research, Inc. All rights reserved.