

Navigating the Boundaries: Safeguarding Generative AI in a Rapidly Evolving Landscape

Before We Get Started...



Recording

A link to the recording will be made available.



Questions

Type in the question box and we will answer in real time or during the Q&A.



Social

Follow us on LinkedIn and go to blackhillsip.com to learn more about our webinars and other resources.

Today's Presenters...



Jim Hallenbeck

*President & CEO, Black Hills IP
Registered Patent Attorney*



Manjeet Rege, PhD.

*Director of Center of Applied
Artificial Intelligence, Professor at
the University of St. Thomas,
Advisor to
Black Hills IP*



Thomas Marlow

*President, BHIP Renewals,
CTO, Black Hills IP,
Former Chief Patent Counsel,
Fairchild Semiconductor
Registered Patent Attorney*

Understanding Generative AI

- Generative AI refers to a class of algorithms that can create or generate new content, such as images, text, music, etc.
- It utilizes machine learning techniques, particularly deep learning, to learn patterns and generate new, original content.
- Generative AI has gained significant attention due to its ability to create realistic and sophisticated outputs.

Generative AI and Industry buzz

Microsoft pledges to watermark images and videos

Kyle Wiggers @kyle_l_wiggers / 10:00 AM CDT • May 23, 2023

ChatGPT vs Bing Chat vs Google Bard: Which is the best AI chatbot?

AI chatbots are more popular than ever, but how do you figure out which one is the best?



Written by Maria Diaz, Staff Writer on May 30, 2023

Meta announces generative AI features for advertisers

Ivan Mehta @indianidle / 4:00 PM CDT • May 11, 2023

FORBES > INNOVATION > ENTERPRISE TECH

The Future Of Generative AI Beyond ChatGPT

Bernard Marr Contributor

Follow

0

May 31, 2023, 03:42am EDT

DID ANYONE ASK FOR THIS? —

Google Search starts rolling out ChatGPT-style generative AI results

If you opt-in to generative AI, big, colorful boxes will appear in search.

RON AMADEO - 5/25/2023, 12:39 PM

Potential Benefits of Generative AI

- **Creative Content Generation:** Generative AI can augment human creativity by generating new and original content, such as artwork, music, and literature.
- **Design and Prototyping:** It enables rapid design iterations and prototyping by generating diverse design options.
- **Personalization:** Generative AI can personalize user experiences by creating tailored recommendations, products, or services.
- **Automation:** It can automate tasks like content creation, data synthesis, and even programming.

Concerns raised about AI

OpenAI CEO in "historic" move calls for regulation before Congress



Ashley Gold, author of [Axios Pro: Tech Policy](#)



Credit: Alamy

IDEAS MADE TO MATTER | ARTIFICIAL INTELLIGENCE

Why neural net pioneer Geoffrey Hinton is sounding the alarm on AI

Ethical and Societal Concerns

- **Bias and Discrimination:** Generative AI algorithms can inadvertently perpetuate biases present in the training data, leading to discriminatory outcomes.
- **Deepfakes and Misinformation:** The technology can be misused to create convincing deepfakes or spread misinformation, posing risks to individuals and societies.
- **Intellectual Property Infringement:** Generative AI raises challenges regarding the ownership and copyright of generated content.
- **Security and Privacy:** Malicious use of generative AI can compromise security and privacy, such as identity theft or manipulation of personal data.

Deep Fake Video



Deep Fake Video



Detecting Deep Fakes



Detecting Deep Fakes



Deep Fake Audio



Who won the superbowl?

- ✓ Searching for: **superbowl winner**
- ✓ Generating answers for you...

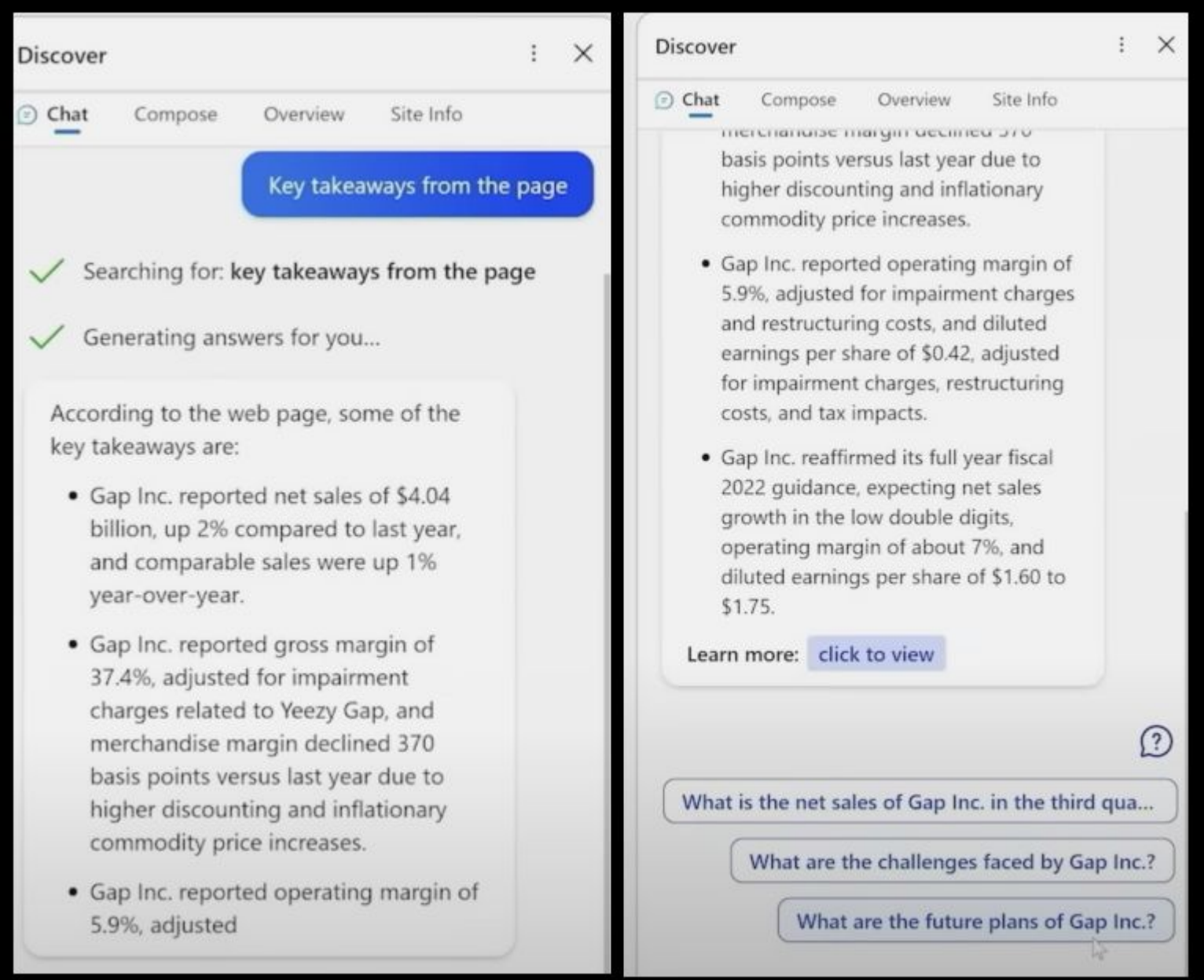
The Super Bowl is the annual American football game that determines the champion of the National Football League (NFL) ¹. The most recent Super Bowl was **Super Bowl LVI**, which was held on **February 6, 2023** at **SoFi Stadium** in **Inglewood, California** ². The winner of that game was the **Philadelphia Eagles**, who defeated the **Kansas City Chiefs** by **31-24** ². It was the second Super Bowl title for the Eagles, who previously won Super Bowl LII in 2018 ¹.

The previous Super Bowl was **Super Bowl LV**, which was held on **February 7, 2021** at **Raymond James Stadium** in **Tampa, Florida** ³. The winner of that game was the **Tampa Bay Buccaneers**, who defeated the **Kansas City Chiefs** by **31-9** ³. It was the second Super Bowl title for the Buccaneers, who previously won Super Bowl XXXVII in 2003 ¹. The Buccaneers also became the first

Learn more: [1. en.wikipedia.org](https://en.wikipedia.org) [2. sportingnews.com](https://sportingnews.com) [3. cbssports.com](https://cbssports.com)

This was asked before the Super Bowl. Also, it got the date and the location wrong.

Chatbot inaccuracies



Bing AI hallucinates while summarizing financial information

Copyright Issues



Delos Prime May 30, 2023

Japan Goes All In: Copyright Doesn't Apply To AI Training

In a surprising move, Japan's government recently reaffirmed that it will not enforce copyrights on data used in AI training. The policy allows AI to use any data "regardless of whether it is for non-profit or commercial purposes, whether it is an act other than reproduction, or whether it is content obtained from illegal sites or otherwise." Keiko Nagaoka, Japanese Minister of Education, Culture, Sports, Science, and Technology, **confirmed** the bold stance to local meeting, saying that Japan's laws won't protect copyrighted materials used in AI datasets.

Safeguarding Generative AI

- **Data Ethics and Bias Mitigation:** Implementing rigorous data collection and preprocessing methods to minimize biases and ensure fair representation.
- **Explainability and Interpretability:** Developing techniques to make generative AI algorithms more transparent and interpretable.
- **Robustness and Security:** Enhancing the resilience of generative AI systems against adversarial attacks and ensuring data security.
- **User Consent and Control:** Empowering users with control over their data and generated content and obtaining informed consent.
- **Continuous Monitoring and Auditing:** Regularly monitoring and auditing generative AI systems to identify and mitigate potential risks.

Legal and Regulatory Landscape

- **Intellectual Property Laws:** Assessing the legal implications of generative AI in terms of copyright, ownership, and licensing.
- **Privacy and Data Protection:** Adhering to data protection regulations and ensuring the privacy of individuals' data.
- **Liability and Accountability:** Defining legal frameworks to address liability and accountability in cases of harm caused by generative AI systems.
- **International Cooperation:** Encouraging global collaboration and coordination to develop consistent legal standards and guidelines.

Best practices for AI developers

- **Responsible Data Collection:** Collecting diverse and representative data while ensuring privacy protection and obtaining consent.
- **Robust Testing and Validation:** Conducting extensive testing and validation to detect and mitigate biases, errors, and unintended consequences.
- **Transparent Documentation:** Documenting the development process, data sources, and model architectures to ensure transparency and facilitate audits.
- **Regular Model Updates:** Implementing a system for regular updates and improvements to keep pace with evolving challenges and address emerging risks.

Mitigating Bias in Generative AI

- **Diverse and Representative Training Data:** Ensuring that training data is diverse, representative, and free from biases to minimize the perpetuation of discriminatory outcomes.
- **Bias Detection and Correction:** Implementing mechanisms to detect and address biases in generative AI algorithms during the training process.
- **Regular Auditing and Evaluation:** Conducting regular audits and evaluations of generative AI systems to identify and mitigate bias-related issues.
- **User Feedback and Iterative Improvement:** Encouraging user feedback to identify biases and making iterative improvements to enhance fairness and inclusivity.

Responsible Deployment of Generative AI

- **Ethical Considerations:** Incorporating ethical frameworks and principles into the development and deployment of generative AI systems.
- **Human-in-the-Loop Approach:** Integrating human oversight and control in generative AI processes to ensure responsible decision-making.
- **User Education and Awareness:** Promoting education and awareness about generative AI technology, its capabilities, and potential risks to users.
- **Impact Assessment:** Conducting comprehensive impact assessments to evaluate the potential social, economic, and ethical consequences of generative AI deployment.

Ensuring Transparency and Accountability

- **Interpretable Models:** Developing techniques to make generative AI models more explainable and interpretable.
- **Algorithmic Auditing:** Implementing mechanisms to audit and assess the decision-making processes of generative AI algorithms.
- **Open-Source Collaboration:** Encouraging open-source initiatives and collaboration to foster transparency and accountability.
- **Ethical Guidelines and Frameworks:** Establishing industry-wide ethical guidelines and frameworks for the development and deployment of generative AI systems.

Collaboration and Industry Standards

- **Partnerships and Knowledge Sharing:** Encouraging collaboration among industry stakeholders, researchers, and policymakers to exchange insights and best practices.
- **Industry Standards:** Developing and promoting industry-wide standards to ensure the responsible and ethical development and use of generative AI.
- **Regulatory Engagement:** Engaging with policymakers and regulatory bodies to provide input and expertise in the formulation of policies and regulations.
- **Continuous Learning and Improvement:** Creating platforms and forums for continuous learning, feedback, and improvement in the field of generative AI.

Data Collection

- Clearview AI
- OpenAI / ChatGPT
 - Age Verification

Data Management

- Data Maps
 - Data
 - Purpose
 - How Collected
 - Where Stored
 - Processing/Movement

Tools



Duties as Lawyers

- Mata v. Avianca, Inc. (Schwartz and LoDuca)
Varghese v. China South Airlines Ltd,
925 F.3d 1339 (11th Cir. 2019)

“Its legal analysis is gibberish,” he wrote, adding, “The summary of the case’s procedural history is difficult to follow and borders on nonsensical.”

- Tech Competence

Quality

- Do not rely solely on generative AI for decision-making purposes. Use the results to inform your research while making decisions based on additional factors and verified evidence.
- Responses often paraphrase other sources.
 - concerns regarding plagiarism and intellectual property rights depending on your use case
- Responses may be inaccurate or biased.
 - Always validate content against trusted sources.

Confidentiality and Export Control



OpenAI APIs

Starting on March 1, 2023, we are making two changes to our data usage and retention policies:

1. OpenAI will not use data submitted by customers via our API to train or improve our models, unless you explicitly decide to share your data with us for this purpose. You can opt-in to share data.
2. Any data sent through the API will be retained for abuse and misuse monitoring purposes for a maximum of 30 days, after which it will be deleted (unless otherwise required by law).

ChatGPT

For non-API consumer products like ChatGPT and DALL-E, we may use content such as prompts, responses, uploaded images, and generated images to improve our services.

Can I permanently delete all my chat history?

You can clear specific chat conversations one at a time from your chat history. After 30 days the cleared conversation will be permanently deleted or de-identified if you haven't opted out of having your content used to improve our services.

You can request to opt out of having your content used to improve our services at any time by filling out this [form](#)

Google Bard

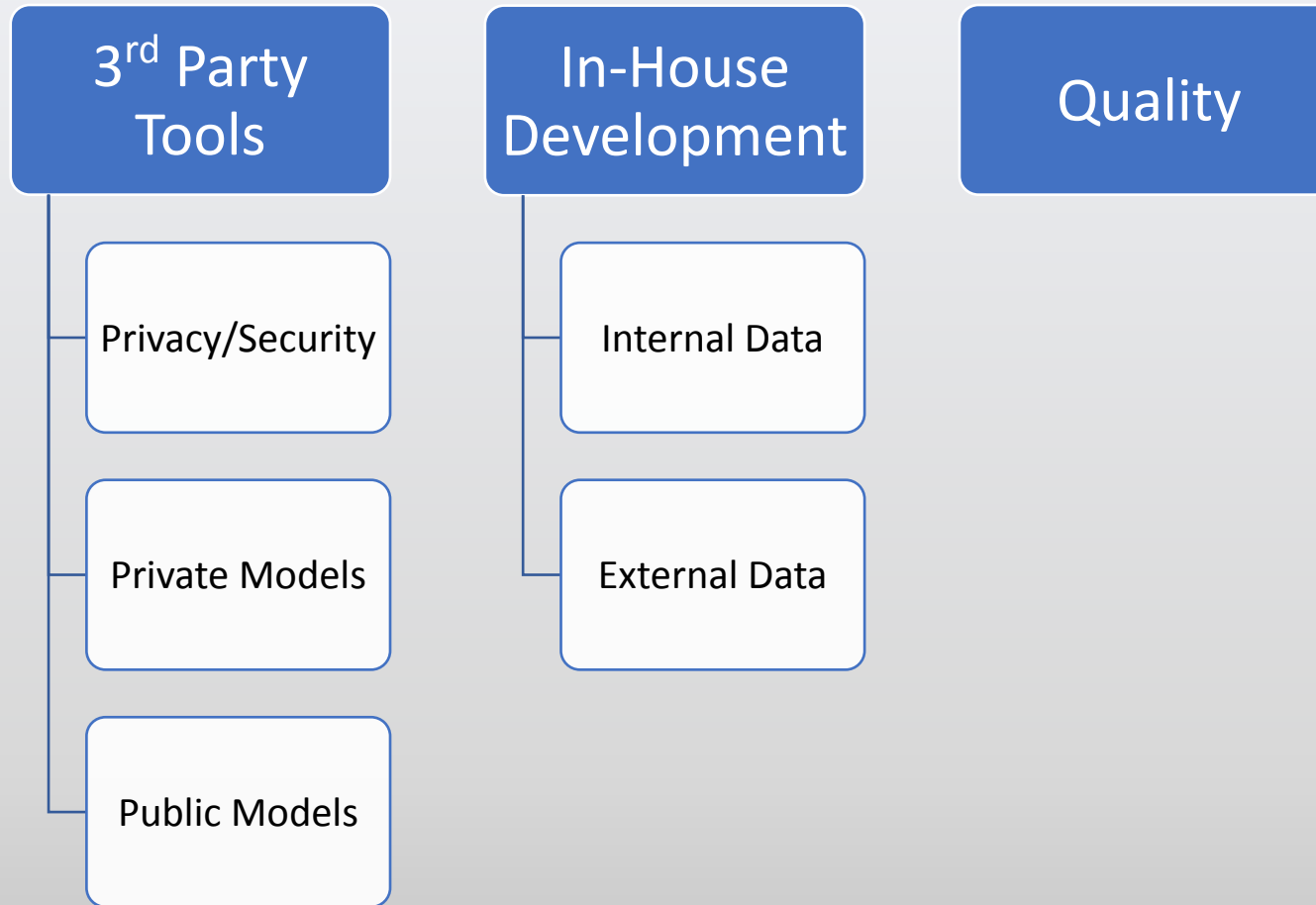
Google collects your Bard conversations, related product usage information, info about your location, and your feedback. Google uses this data, consistent with our [Privacy Policy](#), to provide, improve, and develop Google products and services and machine learning technologies, including Google's enterprise products such as Google Cloud.

By default, Google stores your [Bard activity](#) with your Google Account for up to 18 months, which you can change to 3 or 36 months at myactivity.google.com/product/bard. Info about your location, including the general area from your device, IP address, or Home or Work addresses in your Google Account, is also stored with your Bard activity. Learn more at g.co/privacypolicy/location.

If you want to use Bard without saving your conversations to your Google Account, you can [pause saving your Bard activity](#). You can review your prompts or delete your Bard conversations from your Bard activity at myactivity.google.com/product/bard. Bard conversations that have been reviewed or annotated by human reviewers are not deleted when you delete your Bard activity because they are kept separately and are not connected to your Google Account. Instead, they are retained for up to three years.

- Don't include confidential or sensitive information in your Bard conversations.

Considerations



Take-Away

- Track your data
- Understand how vendors use your data
- Evolving Landscape