GUNDERSON DETTMER

STRATEGIC TRANSACTIONS & LICENSING GROUP

5 Most Common Generative AI Use Cases:

Practical Guidance on How to Mitigate Privacy-Related Risk

April 17, 2024

Meet the Presenters

Cecilia Jeong
Senior Associate
Data Privacy
cjeong@gunder.com

Katie Gardner
Partner
Strategic Transactions & Licensing
kgardner@gunder.com

GD AI Resources and Webinars

GD AI Resources

GD has developed a diverse range of AI-focused resources, including **articles, client insights, events, podcasts and webinars** designed to enrich your understanding and engagement with this dynamic field.

For more information, please refer to the GD AI Resources page here.

Regulating AI in Employment: *How to Comply and Best Practices Webinar*

Labor and employment best practices to comply with current and anticipated regulations governing automated decision making technology | LINK

Generative AI Developments: *Latest Developments, Legal Risks and Best Practices*

Covers developments in the AI landscape, including potential risks associated with AI, the recent case law updates, and methods for mitigating risks | LINK

Patenting AI: *What does it mean, should we do it, and what does success look like?*

Examines various aspects of AI protection as patentable technology | LINK

Generative AI: *Navigating Privacy and Security Concerns in the U.S., EU and UK*

Overview of regulatory guidance and evolving legal requirements in the U.S., EU and the UK, and practical steps companies can take to mitigate privacy and security risks | LINK

The Latest in GenAI: *Updates on the Regulatory Landscape and Company Best Practices to Engage in Now*

Update on new AI developments, focusing on government regulation and enforcement, copyright development, legal diligence, licensing deals, financing, and acquisitions | LINK

Coding with Generative AI: *Open Source Compliance and Practical Risk Management*

Discussion of business and legal issues associated with using AI in software development, including insights into using AI-powered coding tools and strategies to manage risk | LINK

Bias and Hallucinations in AI Policy Development, Training and Risk Mitigation

Explores the challenges posed by bias and hallucinations in generative AI | LINK

GUNDERSON DETTMER

Agenda

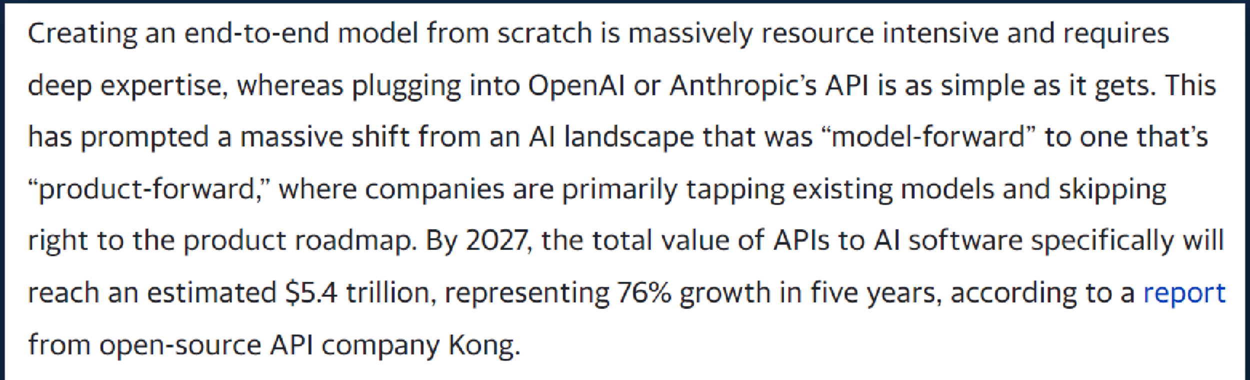
1 | **Use of Data to Train Proprietary AI Models**

2 | Building Products Using Third-Party APIs

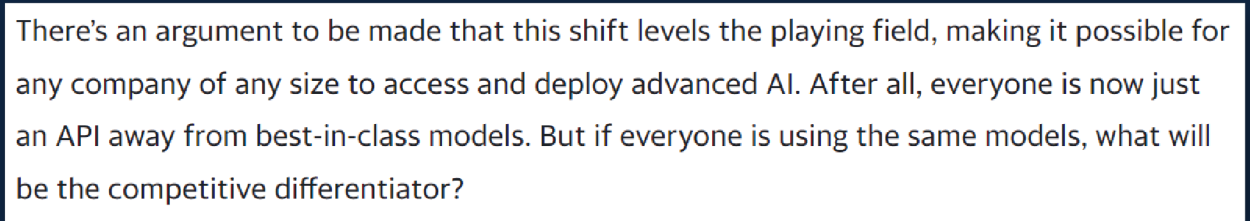
3 | Voice Assistants, Chatbots and Conversational AI

4 | Deepfakes & Image, Likeness and AI-Cloned Voice

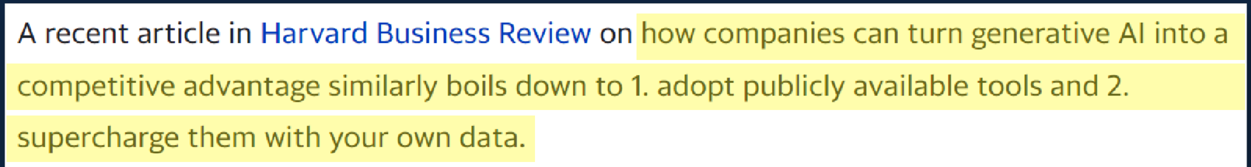
5 | Other High-Risk Areas for AI



Creating an end-to-end model from scratch is massively resource intensive and requires deep expertise, whereas plugging into OpenAI or Anthropic's API is as simple as it gets. This has prompted a massive shift from an AI landscape that was "model-forward" to one that's "product-forward," where companies are primarily tapping existing models and skipping right to the product roadmap. By 2027, the total value of APIs to AI software specifically will reach an estimated $5.4 trillion, representing 76% growth in five years, according to a report from open-source API company Kong.



There's an argument to be made that this shift levels the playing field, making it possible for any company of any size to access and deploy advanced AI. After all, everyone is now just an API away from best-in-class models. But if everyone is using the same models, what will be the competitive differentiator?



A recent article in Harvard Business Review on how companies can turn generative AI into a competitive advantage similarly boils down to 1. adopt publicly available tools and 2. supercharge them with your own data.

Source: Fortune, AI's big shift from 'model-forward' innovation to 'product-forward" (January 25, 2024)

Use of Data to Train Proprietary AI Models

Implications

Assess Data Sources

- **Customer Data:** B2B vs. B2C
- **Public Data:** open source vs. scraping
- **Data Types:** personal vs. de-identified

Evaluate Data Risks

- *JL v. Alphabet* (N.D. Cal 2023)
- *PM v. OpenAI* (N.D. Cal. 2023)
- *AT v. OpenAI and Microsoft* (N.D. Cal. 2023)

Align Use Cases

- **What Is The Use Case?**
  - ***Internal:*** analytics, optimization, benchmarking, etc.
  - ***External:*** product development, finetuning, enhancements, etc.
- **Data Treatment**
  - ***Training Data:*** used to train and finetune AI models
  - ***Test Data:*** used to evaluate and compare AI models
  - ***Validation Data:*** used to validate the final AI model

Use of Data to Train Proprietary AI Models

Best Practices

Use of Customer Data

- ✓ Assess permission under privacy policies and contractual agreements.
- ✓ Obtain specific authorizations (e.g., opt-in consent, license, etc.) as needed, especially for regulated personal or sensitive information.
- ✓ Verify compliance with applicable laws and regulations (e.g., U.S. state privacy laws, GDPR, etc.).

Use of Public Data

- ✓ Confirm that scraped data is "public" or "publically available."
- ✓ Verify the source of open source data and review any accompanying open source terms (e.g., restrictions on commercial use, attribution requirements, etc.)

Secure Testing & Development Environments

- ✓ Establish "clean rooms" or segregated spaces for development and testing.
- ✓ Control development and testing scenarios to segregate sensitive data, and ensure no impact to data security or operational integrity.

Anonymize, De-Identify, or Create Synthetic Data

- ✓ Suppress, de-identify, or anonymize data points in real datasets to reduce risk of exposure and enhance privacy compliance.
- ✓ Leverage artificially-generated synthetic data that does not contain personal or sensitive information to mimic real world datasets.

GUNDERSON DETTMER

Agenda

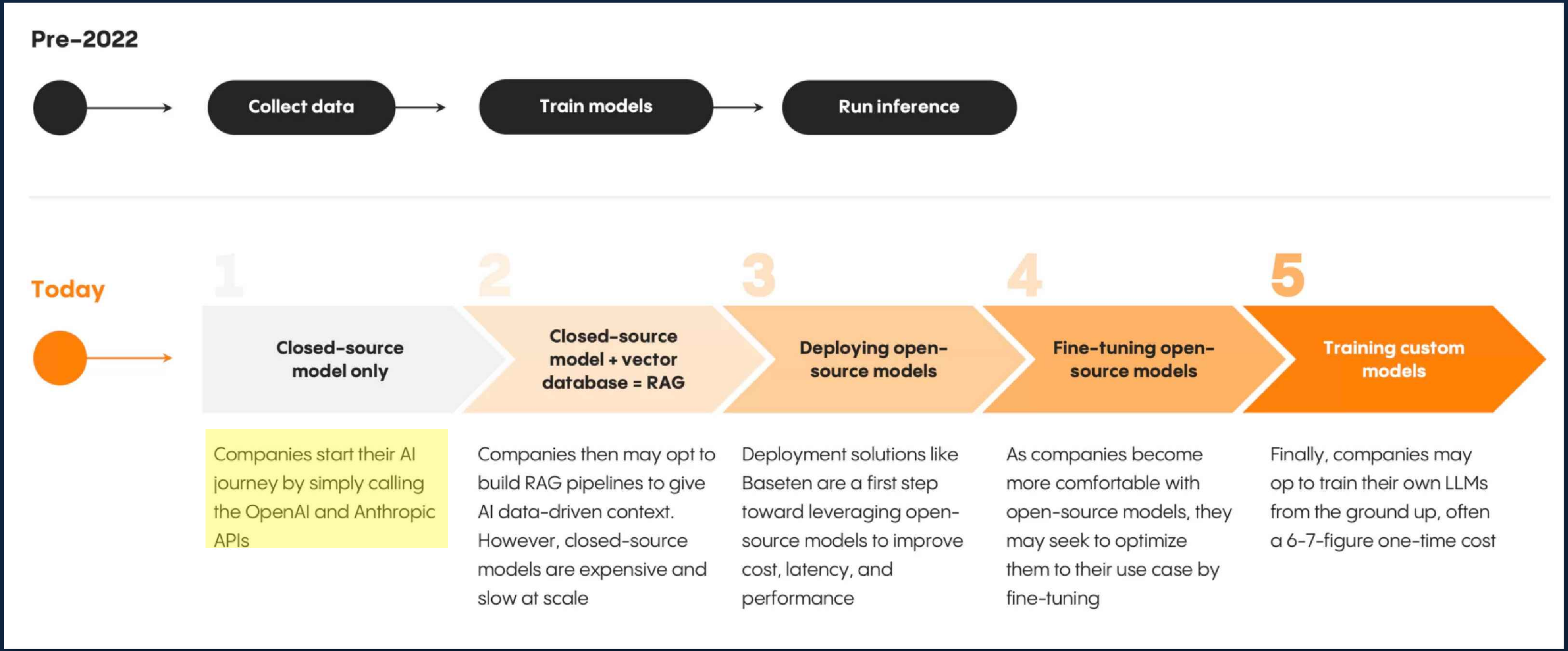
1 | Use of Data to Train Proprietary AI Models

2 | **Building Products Using Third-Party APIs**

3 | Voice Assistants, Chatbots and Conversational AI

4 | Deepfakes & Image, Likeness and AI-Cloned Voice

5 | Other High-Risk Areas for AI



Pre-2022

Collect data → Train models → Run inference

Today

1 Closed-source model only

2 Closed-source model + vector database = RAG

3 Deploying open-source models

4 Fine-tuning open-source models

5 Training custom models

Companies start their AI journey by simply calling the OpenAI and Anthropic APIs

Companies then may opt to build RAG pipelines to give AI data-driven context. However, closed-source models are expensive and slow at scale

Deployment solutions like Baseten are a first step toward leveraging open-source models to improve cost, latency, and performance

As companies become more comfortable with open-source models, they may seek to optimize them to their use case by fine-tuning

Finally, companies may op to train their own LLMs from the ground up, often a 6-7-figure one-time cost

Source: Menlo Ventures, The Modern AI Stack: Design Principles for the Future of Enterprise AI Architectures (January 18, 2024)

Building Products Using Third-Party APIs

Implications

Public API vs. Private Enterprise Instances

Public API

- ***Accessibility and Ease of Use:*** easy integration with minimal setup is an attractive option for companies seeking add AI functionalities quickly.
- ***Cost Effectiveness:*** tiered or "pay as you go" pricing models are advantageous for companies testing new features or operating with limited budgets.
- ***Limited Customization and Control:*** shared infrastructure does not allow companies to tailor use to specific business needs or meet all performance requirements.

Private Enterprise Instance

- ***Enhanced Customization and Integration:*** allows for customization and tailored integration with existing company infrastructure.
- ***Improved Performance and Reliability:*** better performance (e.g., lower latency, higher throughput) from dedicated resources, and more advantageous for critical applications.
- ***Higher Cost and Maintenance:*** likely requires dedicated implementation, maintenance, and support, thus driving up costs.

Privacy & Security Issues With Using APIs

Data Vulnerability

- ***Risks In Transit:*** use of third-party APIs exposes data to interception and unauthorized access in transit, so proper encryption is crucial.
- ***Storage Risks:*** transmitted data may be stored on third-party services or shared infrastructure, raising jurisdictional concerns and increasing the risk of unauthorized access or data breach.
- ***Limiting Use of Company Proprietary Data:*** third-party public APIs may not offer companies the ability to protect and limit use of data for enhancing provider's AI models.

Legal Compliance and Exposure

- ***Regulatory Compliance:*** companies need to ensure third-party APIs comply with relevant privacy standards and regulations.
- ***Third-Party Dependency:*** companies are inherently dependent on third-party provider's security measures, and need to ensure sufficient liability coverage for any breach or failure attributable to APIs.

Building Products Using Third-Party APIs

Best Practices

- **Due Diligence:** conduct thorough due diligence of API provider's privacy and security practices in accordance with industry/commercial standards (e.g., NIST Risk Management Framework).
  - **June 12, 2023:** Salesforce announces AI Cloud to bring a trust layer within its generative AI tools
- **Data Minimization:** only transmit necessary data, and consider using Privacy Enhancing Technologies (e.g., data suppression, de-identification, or anonymization) to protect sensitive information.
  - **March 1, 2023:** OpenAI updates licensing terms to state that it would not train models on API data
  - **June 12, 2023:** Meta updates privacy policy to clarify that user activity and information may be used to train its AI models
  - **June 20, 2023:** Github updates Copilot Terms stating user prompts are not used for training
  - **July 1, 2023:** Google updates privacy policy to clarify that it can use public data to train models
- **Contractual Guarantees:** ensure that provider terms (e.g., DPAs, SLAs, etc.) include strong privacy/security commitments, and uptime/support/maintenance guarantees.
  - **June 8, 2023:** Adobe announces that it will indemnify enterprise users of Firefly for copyright claims related to works created from the tool

Agenda

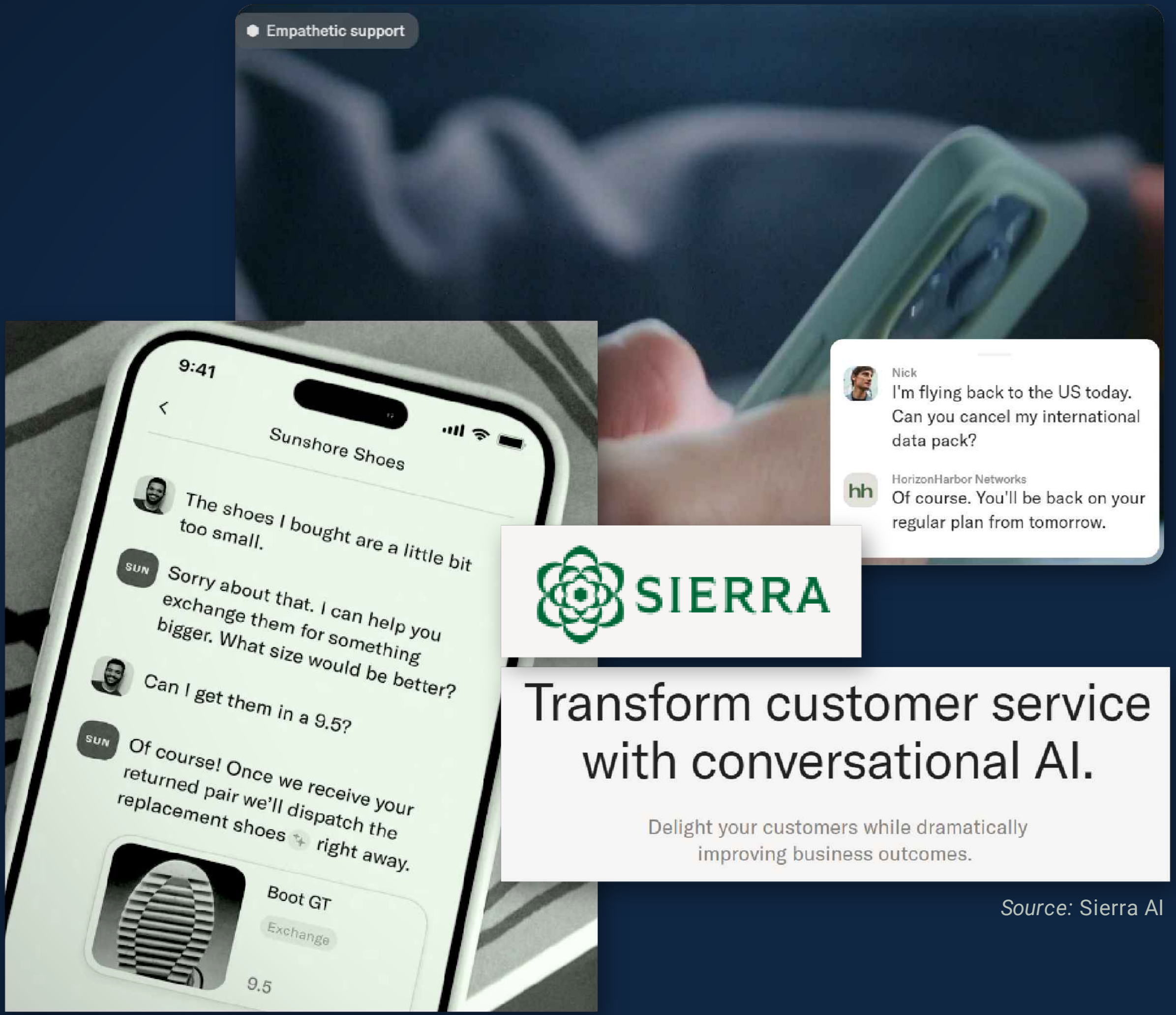
1 | Use of Data to Train Proprietary AI Models

2 | Building Products Using Third-Party APIs

3 | **Voice Assistants, Chatbots and Conversational AI**

4 | Deepfakes & Image, Likeness and AI-Cloned Voice

5 | Other High-Risk Areas for AI



Empathetic support

9:41

Sunshore Shoes

The shoes I bought are a little bit too small.

SUN Sorry about that. I can help you exchange them for something bigger. What size would be better?

Can I get them in a 9.5?

SUN Of course! Once we receive your returned pair we'll dispatch the replacement shoes right away.

Boot GT
Exchange
9.5

Nick
I'm flying back to the US today. Can you cancel my international data pack?

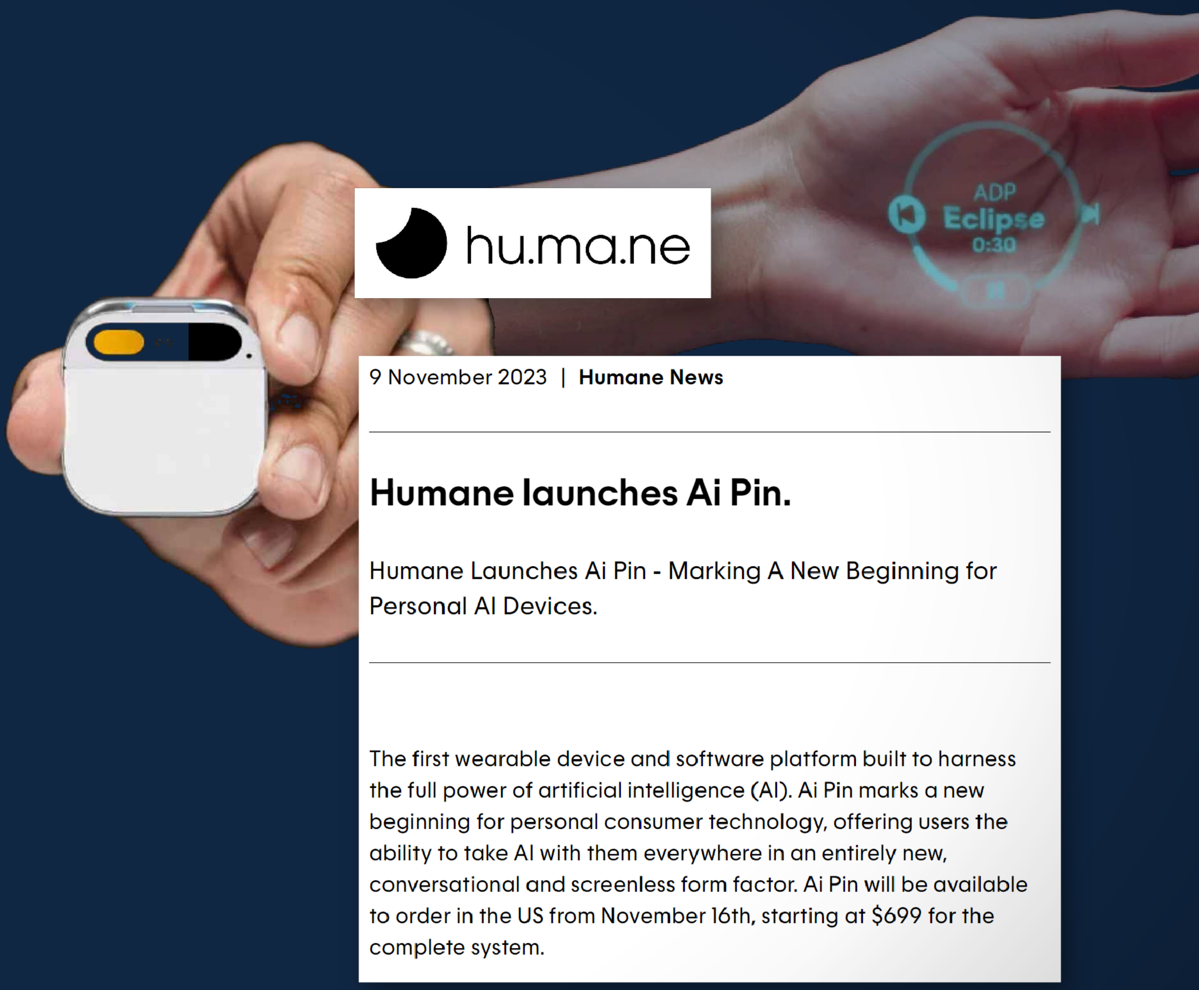
HorizonHarbor Networks
Of course. You'll be back on your regular plan from tomorrow.

SIERRA

Transform customer service with conversational AI.

Delight your customers while dramatically improving business outcomes.

Source: Sierra AI



hu.ma.ne

9 November 2023 | **Humane News**

Humane launches Ai Pin.

Humane Launches Ai Pin - Marking A New Beginning for Personal AI Devices.

The first wearable device and software platform built to harness the full power of artificial intelligence (AI). Ai Pin marks a new beginning for personal consumer technology, offering users the ability to take AI with them everywhere in an entirely new, conversational and screenless form factor. Ai Pin will be available to order in the US from November 16th, starting at $699 for the complete system.

Source: Humane, Announcement (November 9, 2024)

Voice Assistants, Chatbots and Conversational AI

Implications & Best Practices

Implications

Privacy Notices

- ***Privacy Policy:*** generally details use of data in both B2C and B2B contexts
- **[@CECILIA/KATIE - ANYTHING ELSE HERE?]**

Compliance With Applicable Laws

- ***California SB 1001 (Chatbot Law):*** [placeholder]
- ***Call Recordings & Wiretapping Laws:*** [placeholder]
- **[@CECILIA/KATIE - ANYTHING ELSE HERE?]**

Best Practices

Revise Privacy Notices

- Make sure privacy policy details the company's data collection, processing, and usage practices.

Compliance With Applicable Laws

- Disclose use of chatbots or automated call recording technologies, and obtain sufficient authorization or opt-in consent.
- Consider flowdown contractual requirements, including additional disclosures, if using a third-party service provider.
- Use Privacy Enhancing Technologies (PETs), such as synthetic data or suppression of sensitive data points.

Agenda

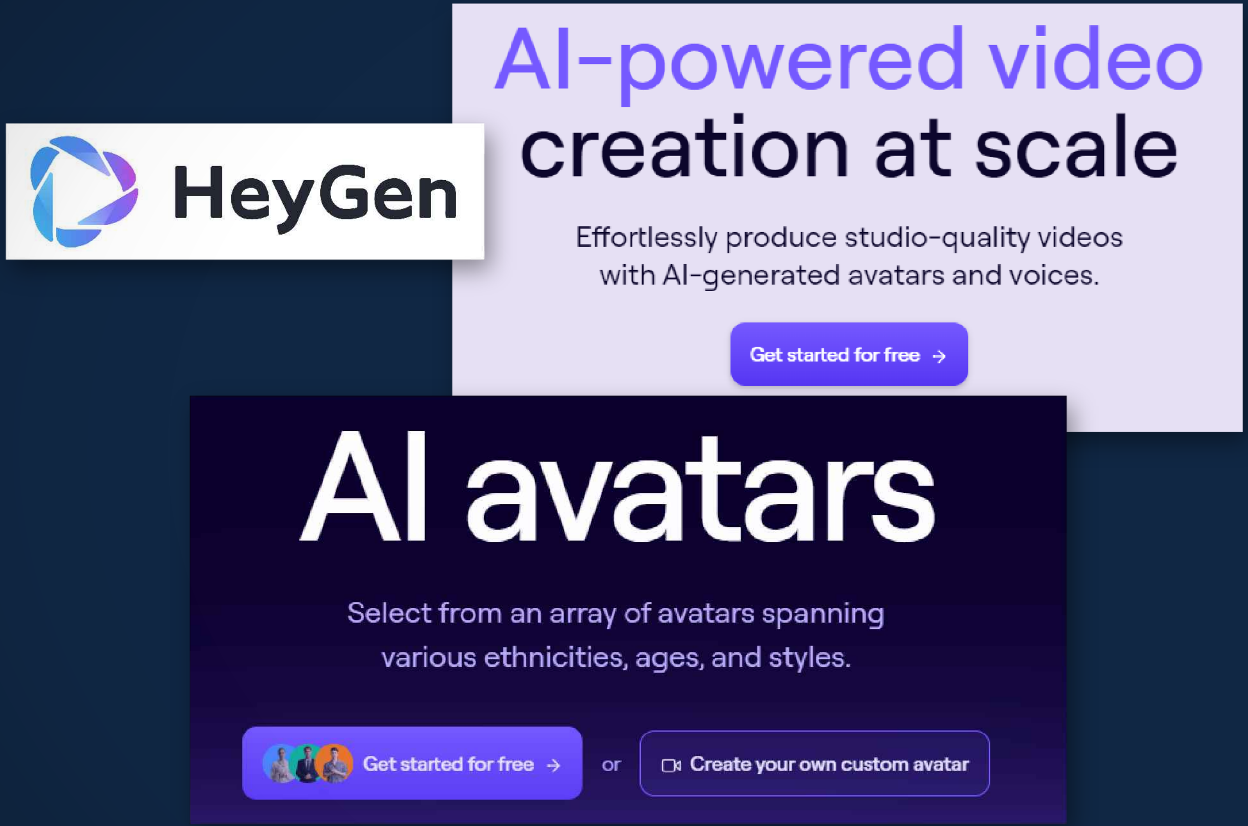
1 | Use of Data to Train Proprietary AI Models

2 | Building Products Using Third-Party APIs

3 | Voice Assistants, Chatbots and Conversational AI

4 | **Deepfakes & Image, Likeness and AI-Cloned Voice**

5 | Other High-Risk Areas for AI



HeyGen

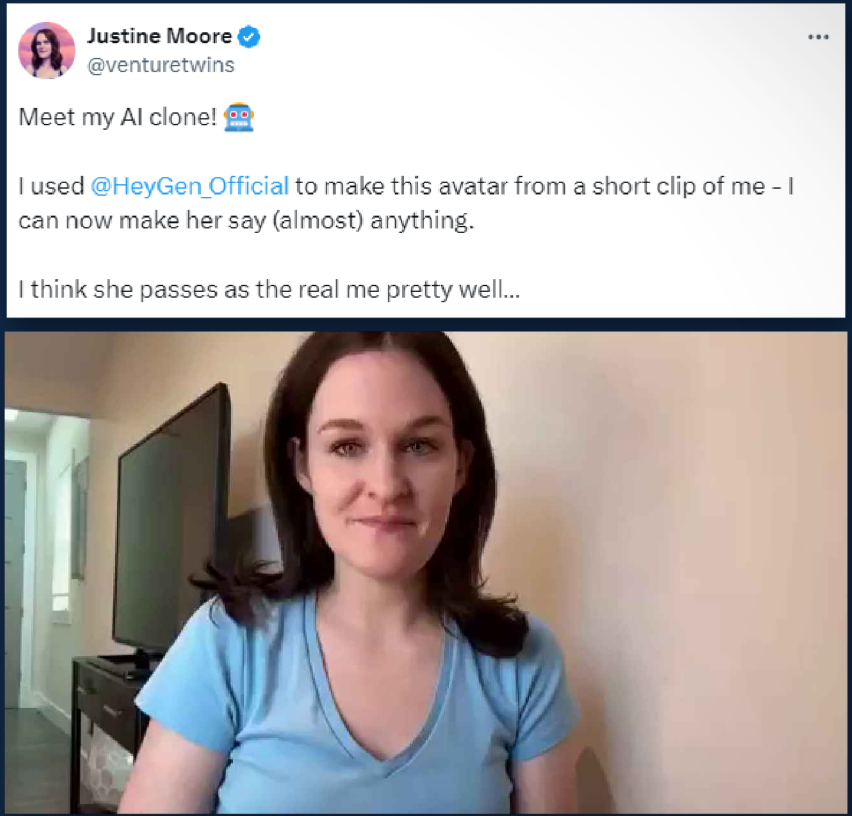
AI-powered video creation at scale

Effortlessly produce studio-quality videos with AI-generated avatars and voices.

AI avatars

Select from an array of avatars spanning various ethnicities, ages, and styles.

Source: HeyGen



Justine Moore
@venturetwins

Meet my AI clone! 🤖

I used @HeyGen_Official to make this avatar from a short clip of me - I can now make her say (almost) anything.

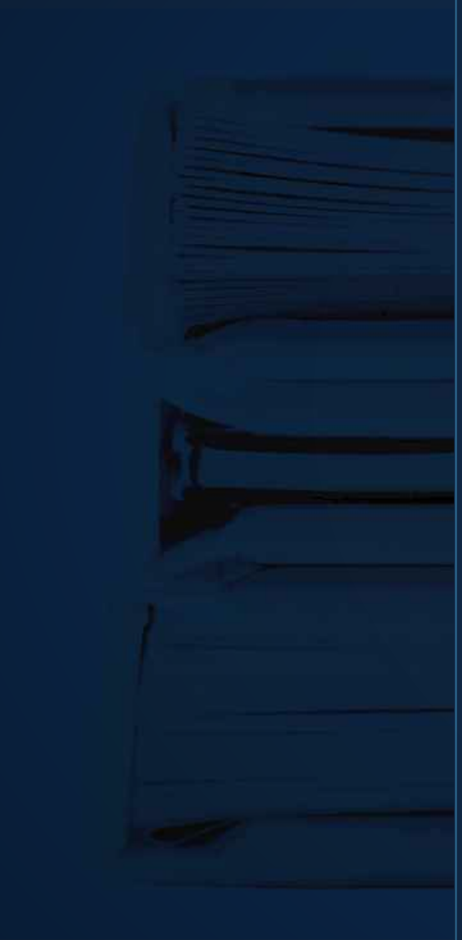
I think she passes as the real me pretty well...

Source: Justine Moore, X

GUNDERSON DETTMER

Deepfakes & Image, Likeness and AI-Cloned Voice

Implications



- **Industry Response**
  - **August 29, 2023:** Google announces SynthID, a tool for watermarking and identifying AI-generated images.
  - **November 8, 2023:** SAG-AFTRA and Hollywood producers reach tentative deal to end strike, hinging upon agreement to compensate actors/actresses for AI-generated content.
  - **November 8, 2023:** Meta requires advertisers to label and disclose political ads created using AI.
  - **November 14, 2023:** YouTube updates terms for creators to require disclosure and labeling of AI-generated content, and content moderation tools to allow users to request removal of deepfakes or AI-generated identifiable persons.
  - **February 6, 2024:** OpenAI announces that C2PA watermarks will appear in images generated through DALL-E. Other companies using the C2PA standard include Adobe, BC, Google, Intel, Microsoft, and Sony.
- **Regulatory Response**
  - **February 8, 2024:** FCC immediately outlaws scam robocalls featuring AI-created voices by expanding the TCPA.
  - **February 15, 2024:** FTC announces proposed rule that prohibits the impersonation of individuals, thus making it unlawful for AI providers to provide services that they know (or have reason to know) is being used to harm consumers through impersonation.
  - **As of April 2024,** there are **over 100 proposed bills in 39 state legislatures** containing provisions intended to regulate the potential for AI to produce election disinformation.
    - **April 3, 2024:** Arizona House passes HB 2394 which, in conjunction with SB 1359 (introduced in February), limits digital impersonation of a candidate or elected official through synthetic media. SB 1359 imposes criminal liability (including felony for repeat offenses), and HB 2394 creates a civil cause of action to seek injunction and monetary damages.
    - **March 1, 2024:** Florida introduces HB 919, which requires specific disclaimers for AI-generated products of certain size and/or length used in political advertisements. Failure to include require printed, audio, or visual disclaimers would be a misdemeanor punishable by up to 1 year of incarceration.
    - **March 21, 2024:** Wisconsin enacts AB 664, which requires political campaign-affiliated entities to add a disclaimer noting use of AI for any published content. Failure to comply can result in a fine of $1,000 per violation.

Deepfakes & Image, Likeness and AI-Cloned Voice

Addressing Risks & Best Practices

Specific, Affirmative & Informed Consent

It will become increasingly crucial to obtain **explicit, informed consent** for use of personal data (including image, voice, and likeness) for AI cloning or replication.

Expect highly-scrutinized negotiations to establish **clear compensation, guidelines, and individual protections** for use of AI to replicate image, likeness, and voice in media.

Authentication & Accountability

Industry pressure mounts to **implement standards authenticating provenance and origin of AI-generated content** (e.g., watermarking, embedded metadata, etc.).

Increase in state, federal, and international **legislation for the purpose of enhancing AI transparency and accountability** (e.g., audit, reporting, and self-certification requirements).

Rights of Publicity & Privacy

In addition to US federal and state regulations, expect increase in cases asserting **publicity and privacy rights**:

- **Privacy Rights:** common law tort claims such as:
  - unreasonable intrusion upon the seclusion
  - appropriation of a person's name or likeness
  - public disclosure of private facts
  - publicity placing a person in false light
- **Publicity Rights:** protects individuals against unauthorized commercial use of their likeness or identity. Enforcement in the global digital domain will require expanded international legal cooperation and harmonization.

Agenda

1 | Use of Data to Train Proprietary AI Models

2 | Building Products Using Third-Party APIs

3 | Voice Assistants, Chatbots and Conversational AI

4 | Deepfakes & Image, Likeness and AI-Cloned Voice

5 | **Other High-Risk Areas for AI**

Other High-Risk Areas for AI

U.S. Privacy & AI Regulations

- **Federal:**
  - American Privacy Rights Act (APRA)
  - National Executive Order to Ensure Safe Development and Use of AI
- **States:**
  - **More than 30 states** have enacted varying laws addressing AI (e.g. data privacy, discrimination, hiring practices, media manipulation), including:
    - Use of AI tools for profiling purposes or disclosing use of automated decision making tools (e.g., NY AEDT)
  - Application of **state privacy regulations** to certain data types used in or by AI tools, including:
    - Use of sensitive categories of data like children's data (e.g., Maryland, California), health data (e.g., Washington) and biometric data (e.g., Illinois BIPA)
    - Expanded application of existing personal privacy laws (e.g., California, Colorado)
    - Forthcoming personal privacy legislation (e.g., Washington, Texas, Oregon, Maryland, Montana)
- **Industry:**
  - Increase in stewardship by industry leaders (e.g., Responsible AI Coalition, Coalition for Content Provenance)
  - Companies to coalesce around industry standards (e.g., NIST, C2PA)
  - Creation of self-regulation systems and requirements (e.g., self-certification)

Other High-Risk Areas for AI

Global AI Regulations

Overview

31 countries have passed AI regulations and 13 more are debating AI laws with significant divergences on regulation approach.

- ○ **Israel/Japan/Australia**: revising existing laws to facilitate AI development
- ○ **UAE**: Focus on national strategy not regulation
- ○ **Russia**, **Iran**, **North Korea**, **Syria**, **Iraq**: Outright ban on services like ChatGPT

Europe

EU AI Act is world's first comprehensive framework for regulating AI systems.

- ○ AI is classified into different risk categories based on use cases, taking risk-based approach (similar to GDPR).
- ○ Will go into effect no sooner than 2025.
- ○ In February 2024, announced a new EU AI Office to promote the development and use of safe and trustworthy AI across Europe, which will be responsible for monitoring compliance.

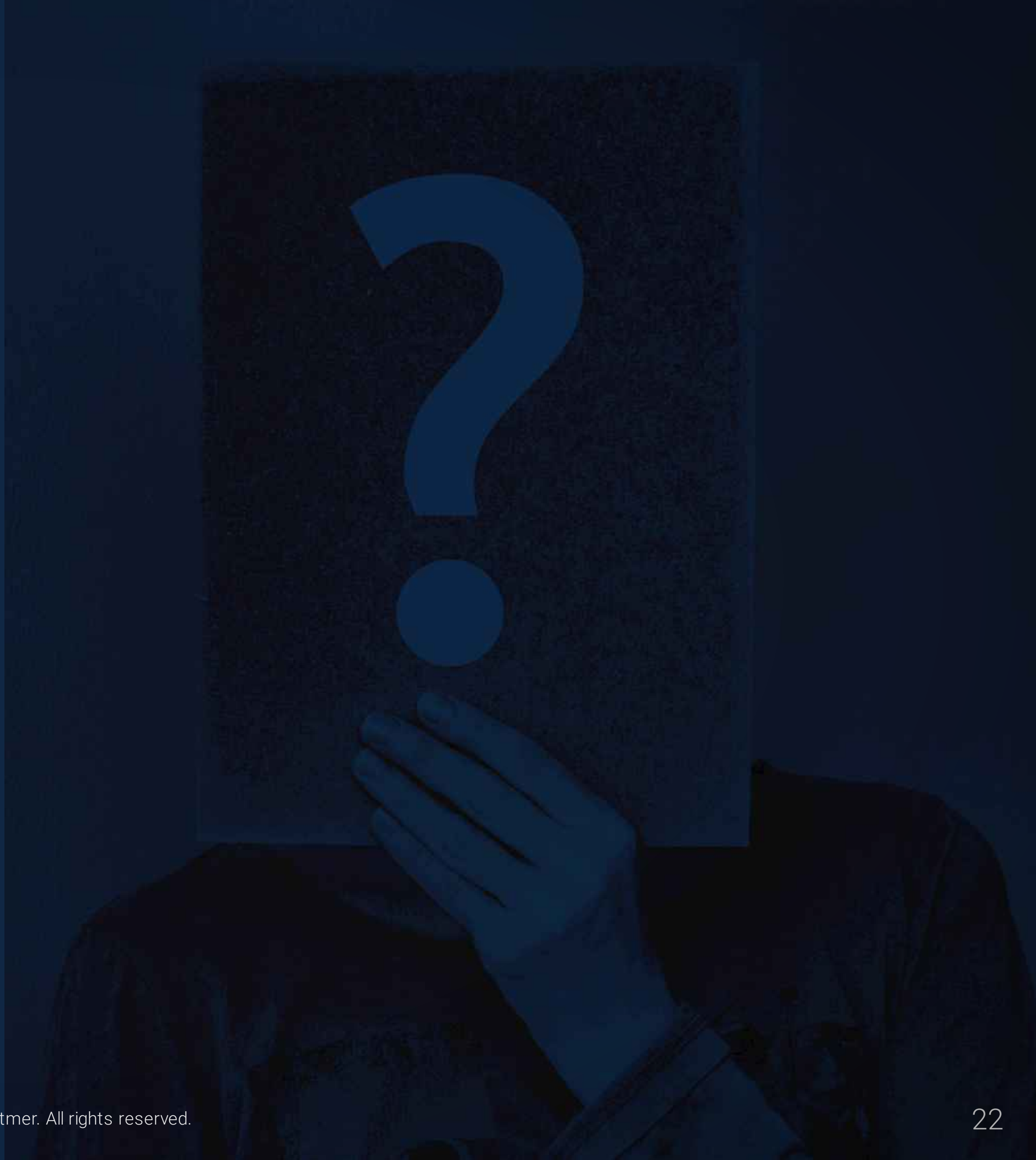
China

Takes a **vertical approach with regulations** targeting specific applications or sets of applications.

- ○ Some parts of regulations align with Western values (e.g. watermarking and accountability)
- ○ Other parts of regulations uniquely impact companies operating in China, including censorship, data sharing requirements and licensing requirements.

GUNDERSON DETTMER

Questions?



MCLE Codes

- ____
- ____

We want your feedback!

Please email us at *insights@gunder.com*

GUNDERSON DETTMER