Meet the Presenters

Strategic Transactions & Licensing Group

Aaron Fiske
Partner
afiske@gunder.com

Shu Hu
Associate
shu@gunder.com

GUNDERSON
DETTMER

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

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

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
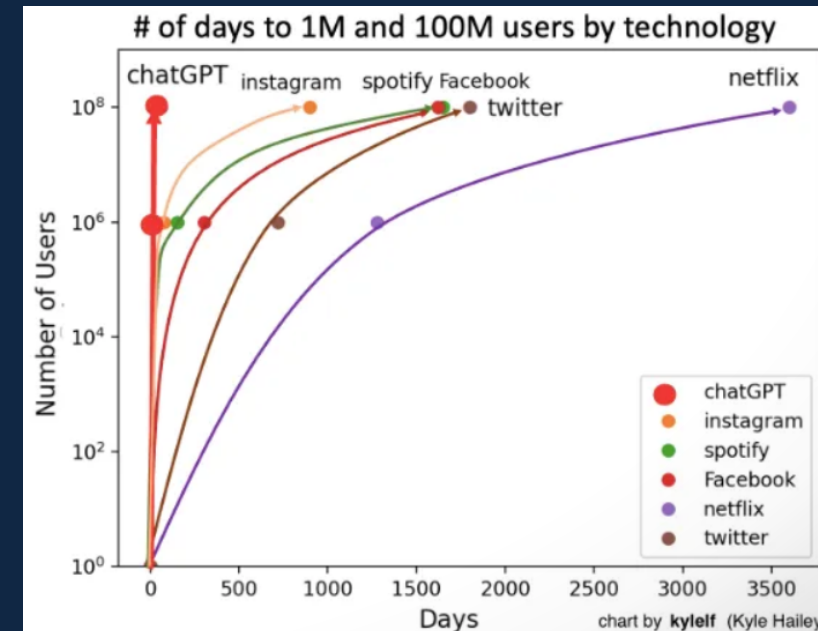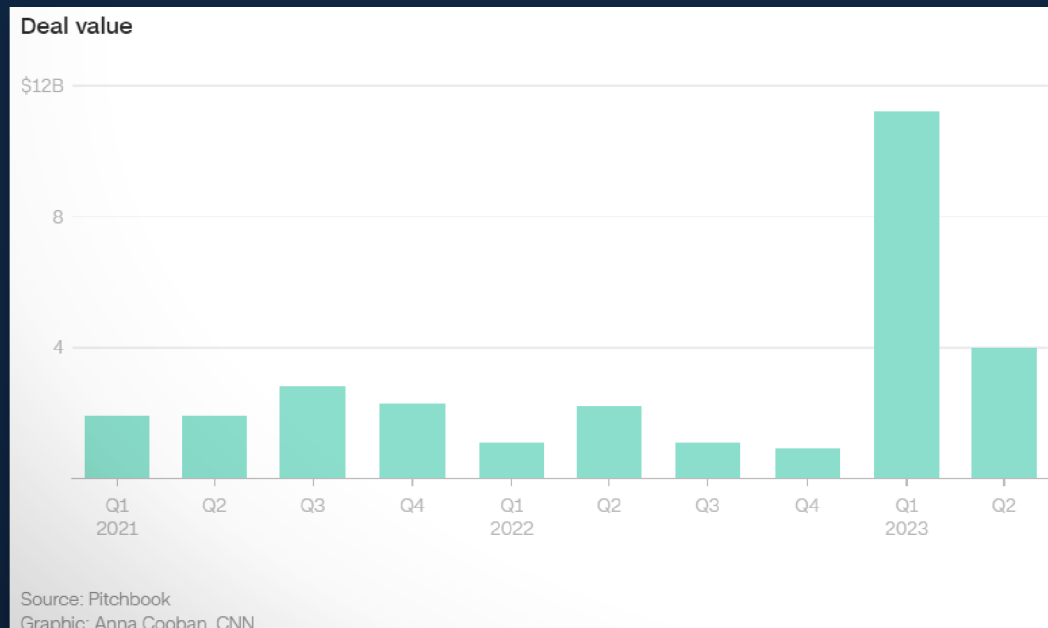
GUNDERSON DETTMER

Agenda

1 | **Overview of GAI Coding Tools**

2 | Risks Associated With Using GAI Tools

3 | Practical Ideas for Managing Risks

GUNDERSON DETTMER

Overview of GAI Coding Tools

Rapid Market Adoption and Investment

- **Copilot Adoption:** As of June 2023, a year into its launch, Copilot has been used by a million developers, and adopted by more than 20,000 org.

- **AI Use:** Based on predictions by Gartner, by 2025, 30% of enteprises will have implemented an AI-augmented development and testing strategy, up from 5% in 2021.

- **VC Funding: Over 25% of total investments in US-based startups this year** ($23B of $90B funding deployed in 2023) has gone to AI companies, compared to 2018-2022 where AI startups received 12% of funding capital.

- **Public Markets: +35%** increase into the top 5 AI-focused ETFs in 2023, compared to 2021-2022.

Deal value

Source: Pitchbook
Graphic: Anna Cooban, CNN

# of days to 1M and 100M users by technology

GUNDERSON DETTMER

Overview of GAI Coding Tools

Source: Nira, Blog

Components

- Training datasets

- Large language models (LLMs)

- Input/prompts

- Output/suggestions

Overview of GAI Coding Tools

Benefits

- Increased productivity and efficiency

- Reduced development time

- Improved code quality and consistency (e.g., maintenance, testing, and debugging)

- Improved developer satisfaction

Source: Github, Blog

```
JS index.js
1    //dictionary of 2 letter ISO country codes and their corresponding country
```

```
▸ EXPLAIN

function binarySearch(arr: any[], key: any) {
    let low = 0;
    let high = arr.length - 1;
    while (low <= high) {
        const mid = Math.floor((low + high) / 2);
        const guess = arr[mid];
        if (guess === key) return mid;
        if (guess > key) high = mid - 1;
        else low = mid + 1;
    }
    return -1;
}
```

Code does following (English)

▸ Advanced

🤖 Ask Copilot

```
RESULT

/* The code above does the following, explained in English:
1. Declare two variables, low and high, which will be used to
represent the start and end of the search.
2. Initialize low to 0 and high to the length of arr − 1.
3. While low is less than or equal to high, do the following:
  a. Declare a variable mid which will be used to store the
result of Math.floor((low + high) / 2).
  b. Declare a variable guess which will be used to store the
value at the index mid.
  c. If guess is equal to key, return mid.
  d. If guess is greater than key, set high to mid − 1.
  e. Otherwise, set low to mid + 1.
4. Return -1. */
```

GUNDERSON DETTMER

Overview of GAI Coding Tools

Comparison of Leading AI Coding Tools

	GitHub Copilot	Amazon CodeWhisperer	Google Duet AI
When was the tool first introduced or generally available?	Preview: October, 2021 GA: June, 2022	Preview: June, 2022 GA: April, 2023	Preview: May 2023
Scope/focus?	General code development	Tailored for AWS cloud development.	Only available to Google Cloud users.
Foundational AI model?	OpenAI Codex	Amazon's LLMs	Codey, a family of code models built on Google's PaLM-2 LLM
Where is the model's training data sourced from?	Source code from publicly-available source, including public repositories on GitHub.	Amazon code and open source code.	Google Cloud code and selected third-party code.

Agenda

1 | Overview of GAI Coding Tools

2 | **Risks Associated With Using GAI Tools**

3 | Practical Ideas for Managing Risks

Risks Associated With Using GAI Tools

Overview of Potential Risks and Exposure

Direct Legal Liability

- Actual infringement
- Protectability

Secondary Transactional Risk

- M&A and financing concerns
- Remediation

Vendor Risk

- Vendor dependency
- Information security

GUNDERSON DETTMER

Risks Associated With Using GAI Tools

Direct Legal Liability: Actual Infringement

Source:
Getty Images (original, left)
Stability Diffusion (generated, below)

Unclear rights to underlying training data and right to use generated outputs.

- ***Getty v. Stability AI***
  - Are the models and/or output infringing derivative works?
  - Is the use of proprietary data for training AI defensible "fair use"?

- ***Doe v. GitHub***
  - Is there a violation of DMCA §1202(b)?
  - Is there a breach of software licensing terms?
  - What about other claims, e.g. privacy violations, unjust enrichment, unfair competition, negligence?

Litigation Update

Copyright

Increased litigation in 2023 alleging infringement of LLM models trained on copyrighted works (i.e. images, books, writings, and other such content).

Class actions:

- *Authors Guild v. OpenAI* (SDNY 2023)
- *Chabon v. OpenAI* (N.D. Cal. 2023)
- *JL v. Alphabet* (N.D. Cal 2023)
- *Silverman v. OpenAI* (N.D. Cal. 2023)
- *Kadery v. Meta* (N.D. Cal. 2023)
- *Tremblay v. OpenAI* (N.D. Cal 2023)
- *Andersen v. Stability AI* (N.D. Cal. 2023)
- *Doe v. GitHub* (N.D. Cal. 2022)

Individual lawsuits:

- *Getty Images v. Stability AI* (D. Del. 2023)

Privacy

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

Discrimination

- *Mobley v. Workday* (N.D. Cal. 2023)
- *Huskey v. State Farm* (N.D. Ill. 2023)

Tort

- *Walter v. OpenAI* (GA Super. Ct. 2023)

Risks Associated With Using GAI Tools

Direct Legal Liability: Rights in Training Inputs, IP Restrictions

- Publicly available code is copyrighted and subject to license terms.

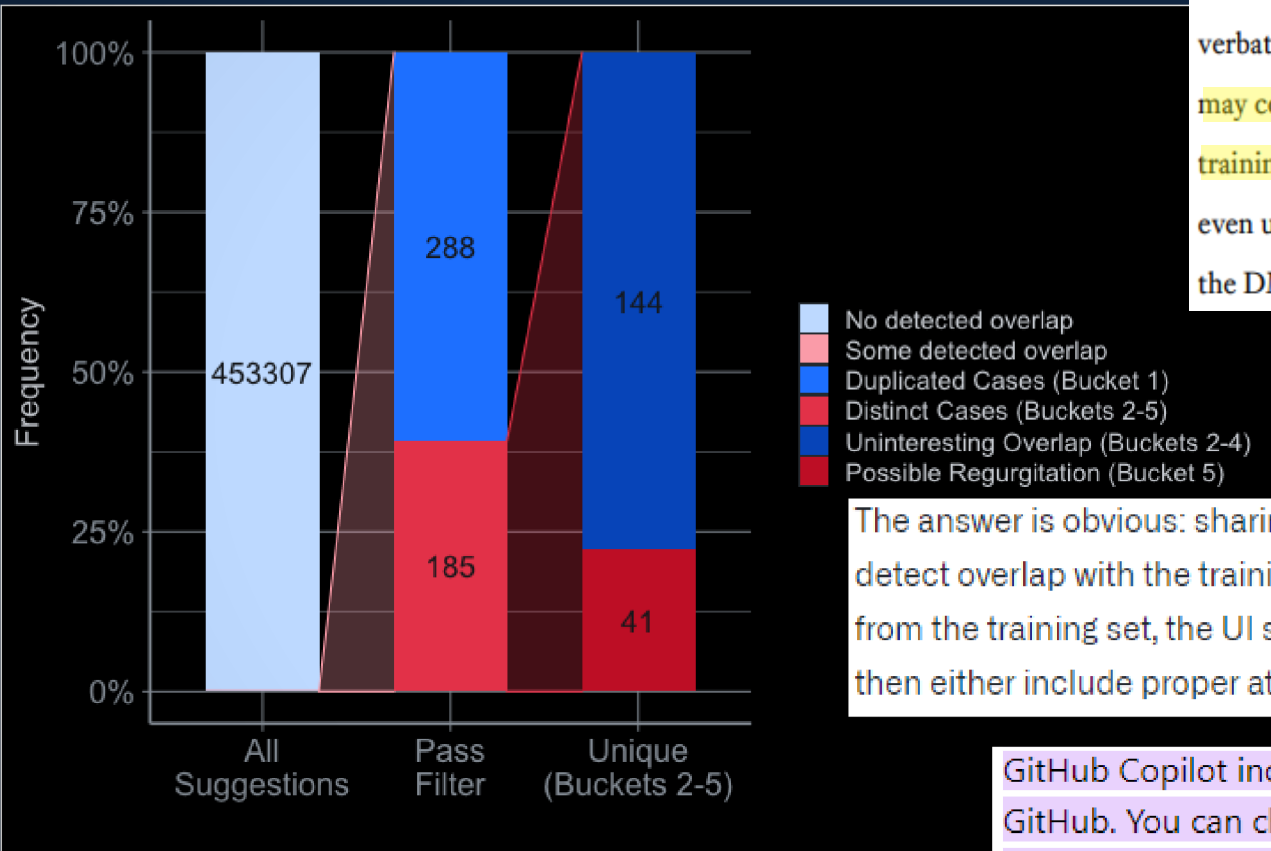| **Public Domain**	**Permissive License**	**Copyleft / Viral License**	**Limited License**
CC0	*MIT, BSD*	*AGPL, GPL*	*custom proprietary terms*

Unspecified License (or No License)

(likely subject to custom proprietary terms)

Risks Associated With Using GAI Tools

Direct Legal Liability: CoPilot Examples

90. GitHub concedes that in ordinary use, Copilot will reproduce passages of code verbatim: "Our latest internal research shows that about 1% of the time, a suggestion [Output] may contain some code snippets longer than ~150 characters that matches" code from the training data. This standard is more limited than is necessary for copyright infringement. But even using GitHub's own metric and the most conservative possible criteria, Copilot has violated the DMCA at least tens of thousands of times.

Source: Complaint, Doe v. GitHub

The answer is obvious: sharing the prefiltering solution we used in this analysis to detect overlap with the training set. When a suggestion contains snippets copied from the training set, the UI should simply tell you where it's quoted from. You can then either include proper attribution, or decide against using that code altogether.

Source: GitHub, CoPilot Research Recitation

GitHub Copilot includes a filter which detects code suggestions matching public code on GitHub. You can choose to enable or disable the filter. When the filter is enabled, GitHub Copilot checks code suggestions with their surrounding code of about 150 characters against public code on GitHub. If there is a match or near match, the suggestion will not be shown to you.

Source: GitHub, CoPilot Configuration Settings

Risks Associated With Using GAI Tools

Direct Legal Liability: CoPilot Examples

Risks Associated With Using GAI Coding Tools

Comparison of Risk Mitigation Features

	GitHub Copilot	Amazon CodeWhisperer	Google Duet AI
Where is the model's training data sourced from?	Source code from publicly-available source, including public repositories on GitHub.	Amazon code and open source code.	Google Cloud code and selected third-party code.
Are there any special features designed to reduce infringement risks?	**Duplicate Detection:** When the filter is enabled, GitHub Copilot checks code suggestions with their surrounding code of about 150 characters against public code on GitHub. If there is a match or near match (ignoring whitespace), the suggestion will not be shown to you.	**Reference Tracker:** The reference tracker can flag a suggestion that might be similar to open source training data with a repository URL and project license information. Users can opt to filter such suggestion out. All references are logged for users to review and decide whether to use it.	**Source Citation:** Duet AI provides source citations when suggestions directly quote at length from a source. Users also have the option to block those suggestions.
Other special features?	Code referencing (upcoming)	code scanning feature for security vulnerabilities Enterprise-customization (preview)	Safety and toxicity filtering Enterprise-customization

GUNDERSON DETTMER

Risks Associated With Using GAI Tools

Overview of Potential Risks and Exposure

Direct Legal Liability

- Actual infringement
- Protectability

Secondary Transactional Risk

- M&A and financing concerns
- Remediation

Vendor Risk

- Vendor dependency
- Information security

GUNDERSON DETTMER

Risks Associated With Using GAI Tools

Secondary Transactional Risk: M&A and Financing Concerns

(l)   [Generative AI. (i) The Company uses all Generative AI Tools (as defined below) in [material] compliance with the applicable license terms, consents, agreements and laws. (ii) The Company has not included and does not include any sensitive Personal Information, trade secrets or material confidential or proprietary information of the Company, or of any third Person under an obligation of confidentiality by the Company, in any prompts or inputs into any Generative AI Tools, except in cases where such Generative AI Tools do not use such information, prompts or services to train the machine learning or algorithm of such tools or improve the services related to such tools. (iii) The Company has not used Generative AI Tools to develop any material Company-Controlled Intellectual Property that the Company intended to maintain as proprietary in a manner that it believes would materially affect the Company's ownership or rights therein. (v) For purposes hereof, "**Generative AI Tools**" means generative artificial intelligence technology or similar tools capable of automatically producing various types of content (such as source code, text, images, audio, and synthetic data) based on user-supplied prompts.]

Source: NVCA Stock Purchase Agreement (revised 2023)

Generative AI Tools. Section ### of the Disclosure Schedule sets forth all third party Generative AI Tools used by the Company, together with the license terms applicable to each such Generative AI Tool, and the purposes for which the Company has used each such Generative AI Tool (including what outputs the Company has generated and how it uses the outputs). The Company has not used Generative AI Tools in its business, including in the development of any Company Products, to generate any Technology or Content which the Company intended to maintain as proprietary, or which is otherwise material to the Company, and has not included any Company Confidential Information or Company Source Code in any prompts or inputs into any Generative AI Tools. The Company uses all Generative AI Tools in compliance with the applicable license terms.

Source: Example Merger Agreement (2023)

- **GAI Reps and Warranties:** Market movement on requiring companies to disclose AI use and compliance

- **Technical Diligence:** Some code scans can pick up that copying has occurred from public sources

- **Remediation:** Consider timing, closing conditions and resource expenditure

- **Purchaser Risk Profile:** Purchaser willingness to accept certain risks (i.e. source code exposure) may vary

Risks Associated With Using GAI Tools

Overview of Potential Risks and Exposure

Direct Legal Liability
- Actual infringement
- Protectability

Secondary Transactional Risk
- M&A and financing concerns
- Remediation

Vendor Risk
- Vendor dependency
- Information security

GUNDERSON DETTMER

Risks Associated With Using GAI Tools

Vendor Risks

Vendor Risk and Dependency

- Availability of vendor services in uncertain legal environment

- Risks may vary between vendors (e.g., vendor's training sets and license compliance)

- Need for a negotiated agreement (more than vendor clickthrough terms)

- Consider alternatives (e.g., open source LLMs) – *many additional considerations*

- Incorporation of GAI technology within a product or service – *many additional considerations*

Information Security / Code Quality Concerns

- **Code-related concerns**, such as:
  - Code security
  - Code quality
  - Human supervision and review. For example, Stack Overflow continues to ban GPT-generated responses.

- **General data security and leakage issues.** For example, Samsung identified 3 instances of employees unintentionally leaking sensitive company info:
  - Pasting confidential source code into ChatGPT to check for errors
  - Requested code optimization
  - Uploading meeting recording to convert into presentation notes

MCLE Codes

- **1125**

- ____

GUNDERSON
DETTMER

Agenda

1 | Overview of GAI Coding Tools

2 | Risks Associated With Using GAI Tools

3 | **Practical Ideas for Managing Risks**

Practical Ideas for Managing Risk

Legal Due Diligence: Understand the terms, conditions and limitations

	GitHub Copilot	Amazon CodeWhisperer	Google Duet AI
Applicable terms	GitHub Copilot Product Specific Terms GitHub Customer Agreement – General Terms	AWS Service Terms	Google Cloud Platform Terms of Service Google Cloud Service Specific Terms
Who owns the inputs into the model?	User	User	User
What can the provider do with your inputs?	Prompts are transmitted only to generate Suggestions in real-time and are deleted once Suggestions are generated. Code snippets submitted through GitHub Copilot are deleted once the suggestion is provided.	User content is processed solely to provide and maintain the service, and is not stored or used for service improvement, nor for further training the model. Users' private code will never be reproduced in a code suggestion for another user.	Duet AI does not use user prompts (e.g. any input information or code that you submit to Duet AI) or its responses (e.g. answers or code completions) as data to train its model.
What data does the tool collect?	Copilot collects usage information, such as for completion acceptance rate, error rate, latency, and feature engagement, to provide the service and enable improvements.	AWS may collect and use client-side telemetry and usage metrics for service improvement purposes. User can opt out of this data collection.	Cloud Data Processing Addendum (Customers)
Who owns the outputs?	GitHub does not claim any ownership rights in Suggestions.	User. Output is "Your Content"	Google does not assert any ownership rights in the Generated Output.
Does the provider get any right to use your output?	No	No	No

Practical Ideas for Managing Risk

New Trend: Indemnities offered by AI coding tool providers and their limitations (*GitHub Copilot*)

4. Defense of Third Party Claims.

If your Agreement provides for the defense of third party claims, that provision will apply to your use of GitHub Copilot. Notwithstanding any other language in your Agreement, any GitHub defense obligations related to your use of GitHub Copilot do not apply if you have not set the Duplicate Detection filtering feature available in GitHub Copilot to its "Block" setting. You can learn how to enable the Duplicate Detection filter at gh.io/cfb-dd.

Source: Copilot Product Specific Terms (Sec. 4) and General Terms (Sec. 6)

6 Third party claims

6.1 The parties will defend each other against third party claims described in this section and will pay the amount of any resulting adverse final judgment or approved settlement, but only if the defending party is promptly notified in writing of the claim and has the right to control the defense and any settlement of it.

6.2 The party being defended must provide the defending party with all requested assistance, information, and authority. The defending party will then reimburse the other party for reasonable out-of-pocket expenses it incurs in providing such assistance.

6.3 This section describes the parties' sole remedies and entire liability for such claims, subject to the limitations set out in Section 7 ("Limitation of liability").

(a) **By GitHub.** GitHub will defend Customer against any third-party claim that a Product made available by GitHub for a fee and used within the scope of this Agreement (unmodified as provided by GitHub and not combined with anything else), misappropriated a trade secret or directly infringes a patent, copyright, trademark, or other proprietary right of a third party. If GitHub is unable to resolve a claim of misappropriation or infringement, it may, at its option, either (1) modify or replace the Product with a functional equivalent or (2) terminate Customer's license and refund any license fees, including amounts paid in advance for any usage period after the termination date. GitHub will not be liable for any claims or damages due to Customer's continued use of a Product after being notified to stop due to a third-party claim.

(b) **By Customer.** To the extent permitted by applicable law, Customer will defend GitHub and its Affiliates against any third-party claim that: (1) any Customer Content misappropriated a trade secret or directly infringes a patent, copyright, trademark, or other proprietary right of a third party; or (2) Customer's use of any Product, alone or in combination with anything else, violates the law or harms a third party.

Practical Ideas for Managing Risk

New Trend: Indemnities offered by AI coding tool providers and their limitations (*Microsoft CoPilot*)

"The **Copilot Copyright Commitment extends Microsoft's existing IP indemnification coverage to copyright claims relating to the use of our AI-powered Copilots**, including the output they generate, specifically for paid versions of Microsoft commercial Copilot services and Bing Chat Enterprise."

Covers:

- **Third-party IP claims** based on *direct* infringement of a patent, copyright, trademark, or other proprietary right of a third party, or misappropriation of a trade secret.

- Customer's **use of Copilot**, including the **output** generated by Copilot.

Conditions include:

- Customers must use the content filters and other safety systems built into the product;

- Customers must not attempt to generate infringing materials, including not providing input to a Copilot service that the customer does not have appropriate rights to use.

Exclusions include:

- (i) the customer's *input*, (ii) *modifications* of the model or output, (iii) *combination* of the model or output with anything else, or (iv) any use that breaches the terms, violates the law, or harms a third party.

Practical Ideas for Managing Risk

New Trend: Indemnities offered by AI coding tool providers and their limitations (*Amazon CodeWhisperer*)

50.10. Amazon CodeWhisperer Professional Defense of Claims and Indemnity. The following terms apply to Amazon CodeWhisperer Professional:

50.10.1. Subject to the limitations in this Section 50.10, AWS will defend you and your employees, officers, and directors against any third-party claim alleging that the output generated by Amazon CodeWhisperer Professional infringes or misappropriates that third party's intellectual property rights, and will pay the amount of any adverse final judgment or settlement.

50.10.2. AWS will have no obligations or liability under Section 50.10.1 with respect to any claim: (i) arising from Content that differs from output generated by Amazon CodeWhisperer Professional; (ii) arising from output generated in connection with inputs or other data provided by you that, alone or in combination, infringe or misappropriate another party's intellectual property rights; (iii) if you intentionally generate output that infringes or misappropriates another third party's intellectual property rights; (iv) if you have fine-tuned, refined, customized, or otherwise modified Amazon CodeWhisperer Professional and the alleged infringement or misappropriation would not have occurred but for this fine-tuning, refinement, customization, or modification; (v) if you have not enabled all filtering features made available for Amazon CodeWhisperer Professional; (vi) if you disregard an indication by the Service of the output's similarity to other data; (vii) if your use of Amazon CodeWhisperer Professional violates the Agreement; or (viii) arising after you receive notice to stop using the output or become aware of the output's potential infringement or misappropriation. The remedies in this Section 50.10 are the sole and exclusive remedies under the Agreement for any third-party claims alleging that the output generated by Amazon CodeWhisperer Professional infringes or misappropriates a third party's intellectual property rights. AWS's defense and payment obligations under this Section 50.10.1 will not be subject to any damages cap under the Agreement.

50.10.3. The obligations under this Section 50.10 will apply only if you: (a) give AWS prompt written notice of the claim; (b) permit AWS to control the defense of the claim; and (c) reasonably cooperate with AWS (at AWS's expense) in the defense and settlement of the claim. AWS may settle the claim as AWS deems appropriate, provided that AWS obtains your prior written consent (not to be unreasonably withheld) before entering into any settlement.

Source: AWS Service Terms (Sec. 50.10)

Practical Ideas for Managing Risk

New Trend: Indemnities offered by AI coding tool providers and their limitations (*Google Duet AI*)

Additional Google Indemnification Obligations.

(i) *Generated Output.* Google's indemnification obligations under the Agreement also apply to allegations that an unmodified Generated Output from a Generative AI Indemnified Service using a Google Pre-Trained Model infringes a third party's Intellectual Property Rights. This subsection (i) (Generated Output) does not apply if the allegation relates to a Generated Output where: (1) Customer creates or uses such Generated Output that it knew or should have known was likely infringing, (2) Customer disregards, disables, or circumvents source citations, filters, instructions, or other tools that will be made available by Google that are intended to help Customer use Generated Output responsibly, (3) Customer uses such Generated Output after receiving notice of an infringement claim from the rightsholder or its authorized agent, or (4) the allegation is based on a trademark-related right as a result of Customer's use of such Generated Output in trade or commerce. "Generative AI Indemnified Service" means the Services listed at https://cloud.google.com/terms/generative-ai-indemnified-services, where the use of such Service is paid for by Customer and not subject to credits or free tier usage.

(ii) *Training Data.* Google's existing indemnification obligation for the Services under the Agreement covers allegations that Google's use of training data to create any Google Pre-Trained Model utilized by a Generative AI Service infringes a third party's Intellectual Property Rights. This indemnity does not cover allegations related to a specific Generated Output, which may be covered by subsection (i) (Generated Output) above.

Source: Google Cloud Service Specific Terms (Sec. 17(j) of Service Terms, last updated October 12, 2023)

Practical Ideas for Managing Risks

Preparing for Technical Diligence

Get ahead of potential remediation efforts

1 | **Develop an internal compliance strategy**

Always have human review and avoid relying exclusively on GAI tools (e.g., humans as the ultimate authors); and arm humans with code analysis tools.

Use tools from reputable sources and for lower-risk activities where the outputs are replicable and easily replaced.

2 | **Turn on filtering and reference/citation tracking**

Certain GAI tools, including Copilot, CodeWhisperer, Duet AI, include filters and reference trackers that can block recommendations that match public code.

3 | **Implement a routine, ongoing code scan**

Consider using third-party code scanning services as needed (or in sandbox environments).

4 | **Review internal compliance strategies**

GAI landscape is rapidly changing, and compliance tools are a rapidly developing area. Develop methods for routine review of your organization's internal compliance strategies.

MCLE Codes

- **1125**

- **9598**

GUNDERSON DETTMER

We want your feedback!

Please email us at *insights@gunder.com*

GUNDERSON
DETTMER