

G U N D E R S O N D E T T M E R

LICENSING, STRATEGIC PARTNERING &
COMMERCIAL TRANSACTIONS GROUP

Coding with Generative AI (GAI):

Open Source Compliance and Practical Risk Management

May 31, 2023

Meet the Presenters



Aaron Fiske

Partner @ Gunderson Dettmer

*Licensing, Strategic Partnering &
Commercial Transactions*



Phil Odence

GM Black Duck Audits @ Synopsys

Gunderson Webinars

GAI Presentation Series



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 that patents can protect, such as data preparation, training processes, and functional applications of AI | [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., European Union and the United Kingdom, and practical steps companies can take to address legal requirements and mitigate privacy and security risks | [LINK](#)



Agenda

1 | **Potential Benefits of Using GAI Tools**

— MCLE Code —

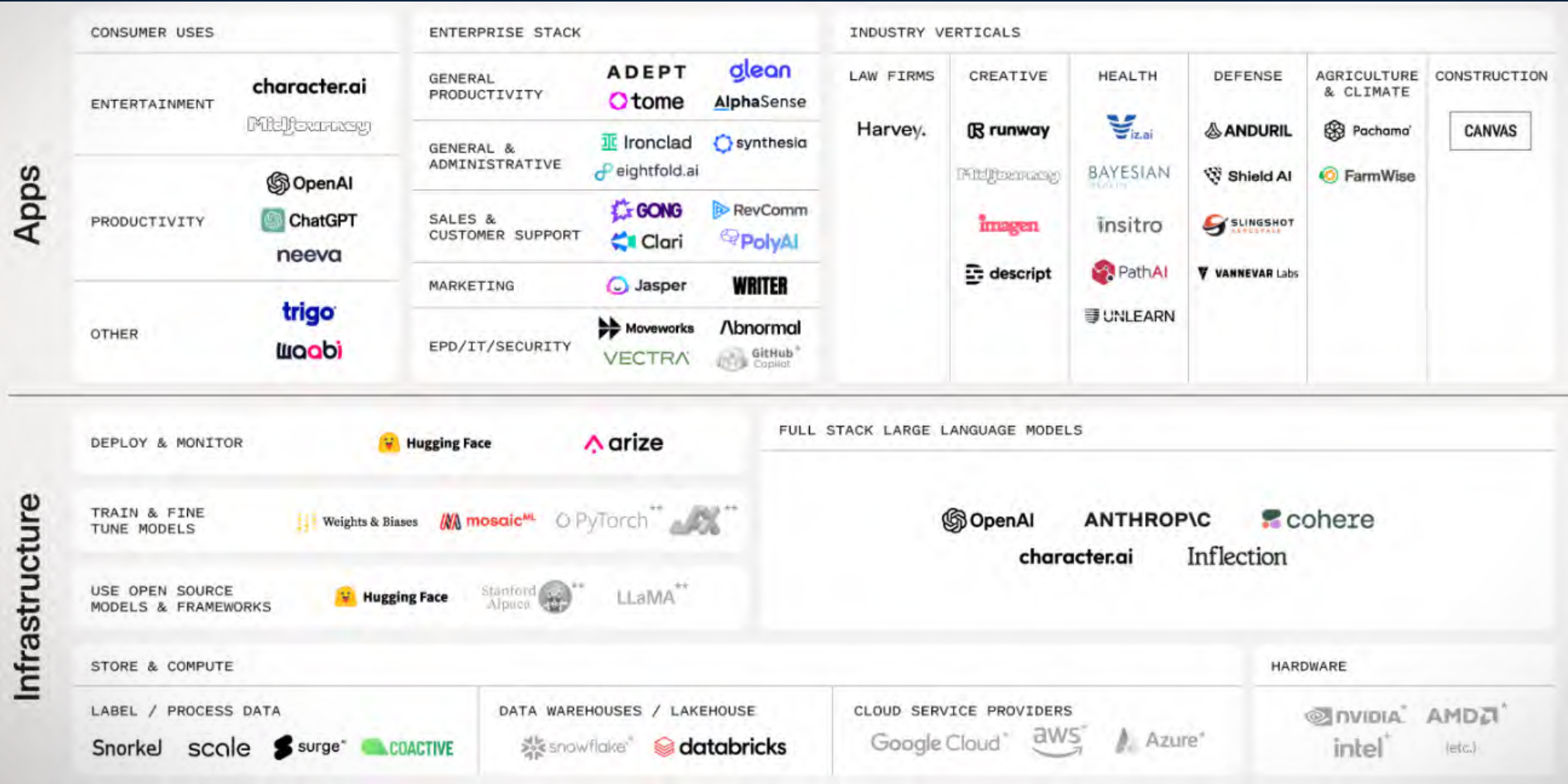
2 | Risks Associated With Using GAI Tools

3 | Practical Ideas for Managing Risks

— MCLE Code —

Potential Benefits of Using GAI Tools

GAI Landscape



Source:
Sequoia, AI 50

Potential Benefits of Using GAI Tools

Generating Content



Imagine you are a marketing executive at a private company, can you generate for me

Imagine you are a primary school teacher, can you suggest ...

Imagine you are the director of finance in a public company, can you state ...



As a small business owner, I'm interested in building a website or application for my business. Can you suggest some actionable steps for each option, including the use of freelancers and hiring a company? Additionally, what are some resources I can use to get started with my research and find the best solution for my needs?

I want to make sure that I choose the best approach for my business, whether that means building the website or application myself, hiring freelancers, or working with a company. I'm also interested in any resources that can help me get started with my research, such as websites or tools for comparing options and finding the best fit for my business.

AI image of a person in a white suit walking through time surrounded by people in gray suits in Times Square. The image was created on Midjourney

You can copy the text below to get a similar result.

A person wearing a full white suit walking through the time in the middle of Times Square. Many other people are also walking in Times Square, however all other individuals are wearing a full grey suit. Make the picture to be detailed and ultra-realistic 8K--or 4K--4 5--5 600



Source: UAE, 100 Practical Applications and Use Cases of Generative AI (ChatGPT and Midjourney prompt engineering examples)

Potential Benefits of Using GAI Tools

Generating Code: CoPilot



Source: Nira, [Blog](#)

CoPilot

- Increased productivity and efficiency
- Reduced development time
- Improved code quality and consistency (e.g., maintenance, testing, and debugging)
- Vast training dataset of publicly-available code

Potential Benefits of Using GAI Tools

Generating Code: CoPilot Examples

JS index.js

```
1 //dictionary of 2 letter ISO country codes and their corresponding country
```

Source: Github, Blog

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. */

- Code recommendations that go beyond mere autocompletion (e.g., creating dictionaries of lookup data)
- Code translation and analysis (e.g., summarization, interpretation, explanation)

Potential Benefits of Using GAI Tools

Corporate Open Source Adoption: Driven By Productivity and Complexity

Open Source

2000s

- Grassroots developer usage
- Companies clueless
- Cisco got sued

2010s

- Supply/usage explode
- Companies bewildered, banning and playing catch up
- Lawyers scrambling to understand risks
- Benefits demonstrated; employees insist
- Legal to vulnerabilities, big breaches
- Processes and tools evolve
- Importance of policy, process, education

2020s

- Fully mainstream
- SBOMs

Potential Benefits of Using GAI Tools

Corporate Open Source Adoption: Driven By Productivity and Complexity

Open Source
(decades)

***"THOSE WHO CANNOT REMEMBER THE
PAST ARE CONDEMNED TO REPEAT IT."***

— George Santayana

GAI
(years, months?)

- Grassroots... and deep roots
- Bewilderment
- Scrambling
- Banning
- [...]
- Policy/Process
Tools/Education
- Mainstream
- Breaches

MCLE Codes

- 2976
-





Agenda

- 1 | Potential Benefits of Using GAI Tools
— MCLE Code —
- 2 | **Risks Associated With Using GAI Tools**
- 3 | Practical Ideas for Managing Risks
— MCLE Code —

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

Risks Associated With Using GAI Tools

Direct Legal Liability: Actual Infringement



Source:
*Doe v. GitHub, Order Granting in Part
and Denying In Part Motions to
Dismiss* (May, 11, 2023)

- Unclear rights to underlying training data and right to use generated outputs.
 - ✦ Is the use of proprietary data for training AI defensible "fair use"? **Getty v. Stability AI**
 - ✦ Is there a breach of software licensing terms? **Doe v. GitHub**
 - ✦ Is there a violation of DMCA §1202(b)? **Doe v. GitHub**
- **Other open-ended issues:** violation of privacy laws, contractual interference, contract breach, tort liability, etc.

attribution, and copyright notices. Taking the facts of the complaint as true and construing all inferences in the Plaintiffs' favor, the Court can reasonably infer that, should Plaintiffs' code be reproduced as output, it will be reproduced in a manner that violates the open-source licenses under which Plaintiffs published their code.⁹

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**

CC0

Permissive License

MIT, BSD

**Copyleft / Viral
License**

AGPL, GPL

Limited License

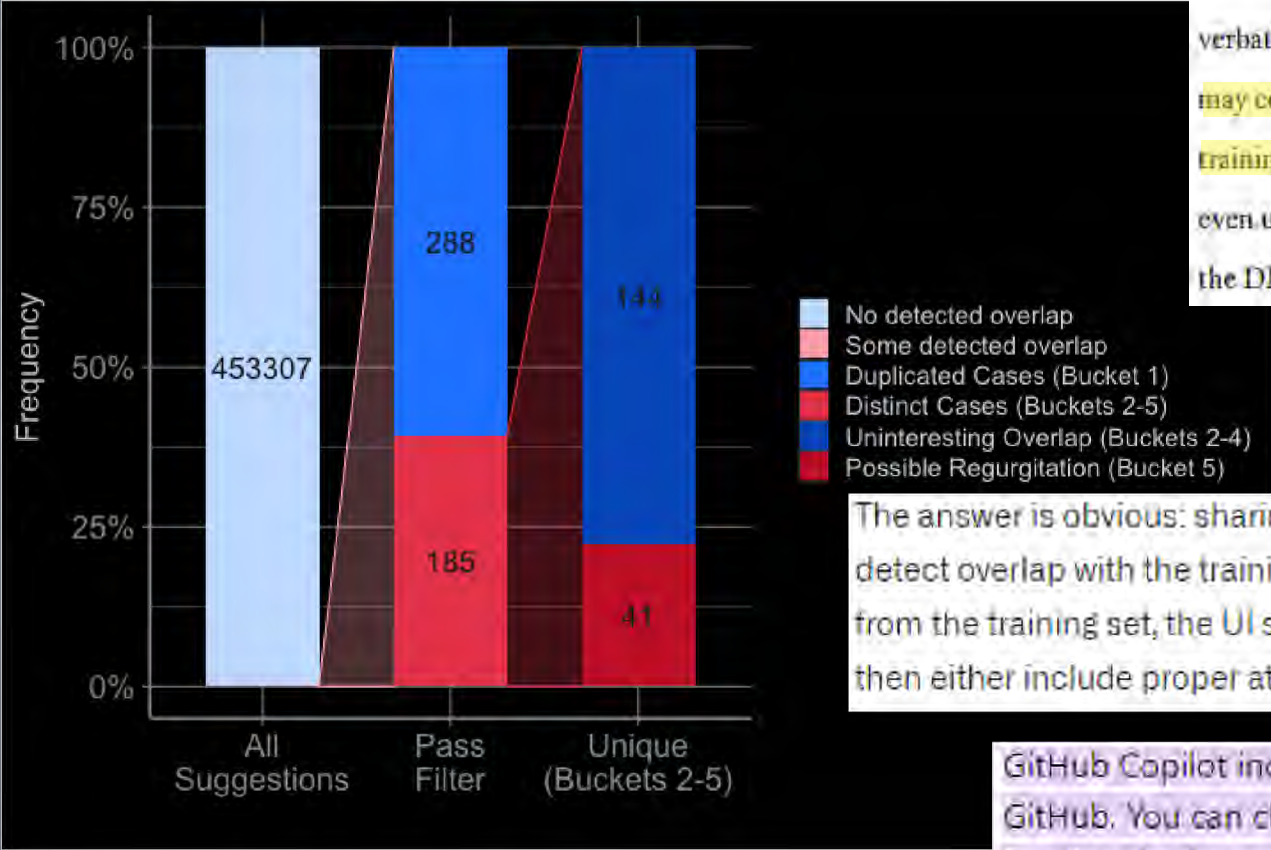
*custom proprietary
terms*

**Unspecified License
(or No License)**

*(likely subject to custom
proprietary terms)*

Risks Associated With Using GAI Tools

Direct Legal Liability: CoPilot Examples



Source: GitHub,
CoPilot Configuration Settings

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.

Risks Associated With Using GAI Tools

Direct Legal Liability: CoPilot Examples

CoPilot Output

```
def create_window():  
    import tkinter as tk  
    from tkinter import ttk  
  
    window = tk.TK()  
    window.title("Welcome to LikeGeeks app")  
    window.geometry('150x100')  
  
    lbl = ttk.Label(window, text="Hello")  
    lbl.grid(row=0, column=0)  
  
    def clicked():  
        lbl.configure(text="Button was clicked!")  
  
    btn = ttk.Button(window, text="Click Me", command=clicked)
```

Training Input

```
from tkinter import *  
  
window = Tk()  
window.title("Welcome to LikeGeeks app")  
window.geometry('150x100')  
  
lbl = Label(window, text="Hello")  
lbl.grid(row=0, column=0)  
  
def clicked():  
    lbl.configure(text="Button was clicked!")  
  
btn = Button(window, text="Click Me", command=clicked)
```

The entire contents of the Site are protected by international copyright and trademark laws. The owner of the copyrights and trademarks are All LikeGeeks, its affiliates or other third party licensors. YOU MAY NOT MODIFY, COPY, REPRODUCE, REPUBLISH, UPLOAD, POST, TRANSMIT, OR DISTRIBUTE, IN ANY MANNER, THE MATERIAL ON THE SITE, INCLUDING TEXT, GRAPHICS, CODE AND/OR SOFTWARE.

Source: LikeGeeks, [Tutorial](#) and [Terms of Service](#)

Risks Associated With Using GAI Tools

Direct Legal Liability: CoPilot Examples

CoPilot Output

```
def create_socket():
    # Create a socket object
    s = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
    # Get local machine name
    host = socket.gethostname()
    port = 8080
    # Bind to the port
    s.bind((host, port))
    # Now wait for client connection.
    s.listen(5)
    while True:
        # Establish connection with client.
        c, addr = s.accept()
        print('Got connection from', addr)
```

Training Input

```
s = socket.socket()
host = socket.gethostname()
port = 8080
s.bind((host, port))
s.listen(5)
while True:
    c, addr = s.accept()
    print('Got connection from', addr)
```

Create a socket object
Get local machine name
Reserve a port for your service
Bind to the port
Now wait for client connection.
Establish a connection with client.

The Network is protected by copyright as a collective work and/or compilation, pursuant to U.S. copyright laws, international covenants, and other copyright laws. Other than as expressly set forth in these Public Network Terms, you may not copy, modify, publish, transmit, upload, participate in the transfer or sale of, reproduce (except as provided in this Agreement), create derivative works based on, distribute, perform, display, or in any way exploit any of the Network Content, software, materials, or Services in whole or in part. You may download or copy the public Network Content, and other items displayed on the public Network for download or personal use provided that you maintain all copyright and other notices contained in such Public Content.

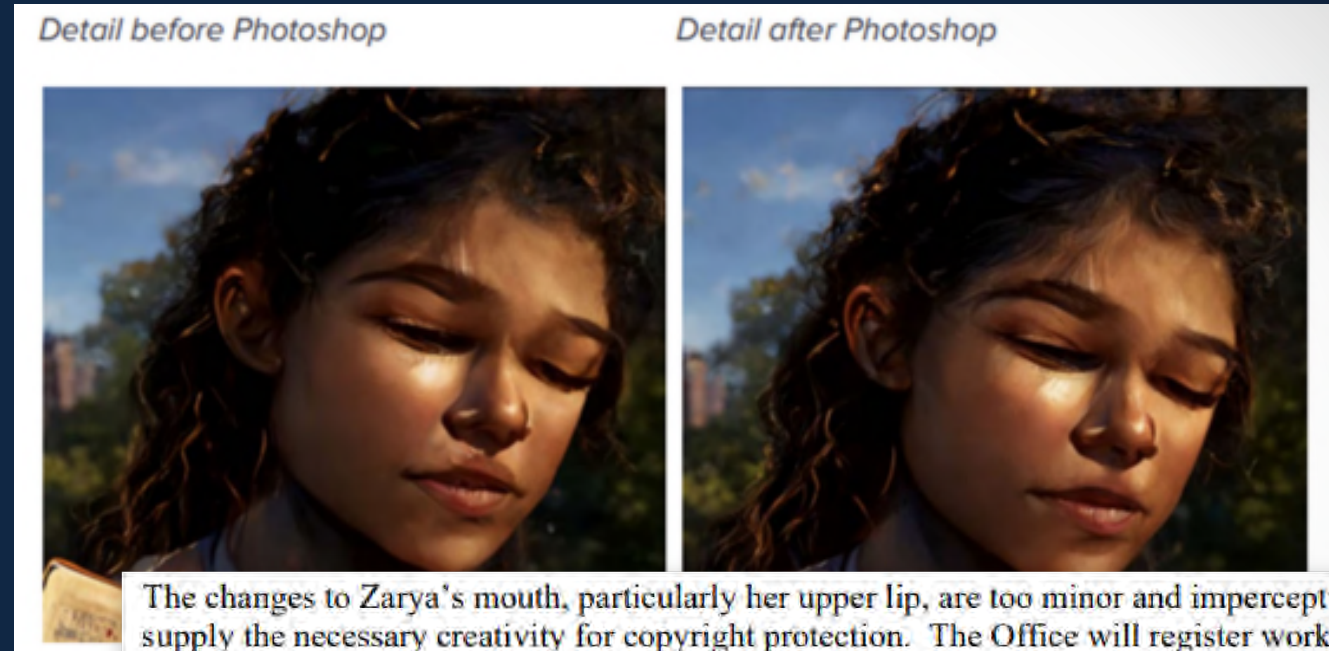
Source: Stack Overflow, Thread and Terms of Service

Risks Associated With Using GAI Tools

Direct Legal Liability: Protectability



Source: USCO Letter (Feb. 21, 2023) regarding "Zarya of the Dawn" by Kris Kashtanova (Midjourney generated images)



The changes to Zarya's mouth, particularly her upper lip, are too minor and imperceptible to supply the necessary creativity for copyright protection. The Office will register works that contain otherwise unprotectable material that has been edited, modified, or otherwise revised by a human author, but **only if the new work contains a "sufficient amount of original authorship"** to itself qualify for copyright protection. COMPENDIUM (THIRD) § 313.6(D). Ms. Kashtanova's

Rather than a tool that Ms. Kashtanova controlled and guided to reach her desired image, **Midjourney generates images in an unpredictable way**. Accordingly, Midjourney users are not the "authors" for copyright purposes of the images the technology generates. As the Supreme Court has explained, the "author" of a copyrighted work is the one "who has actually formed the picture;" the one who acts as "the inventive or master mind." *Burrow-Giles*, 111 U.S. at 61. A

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

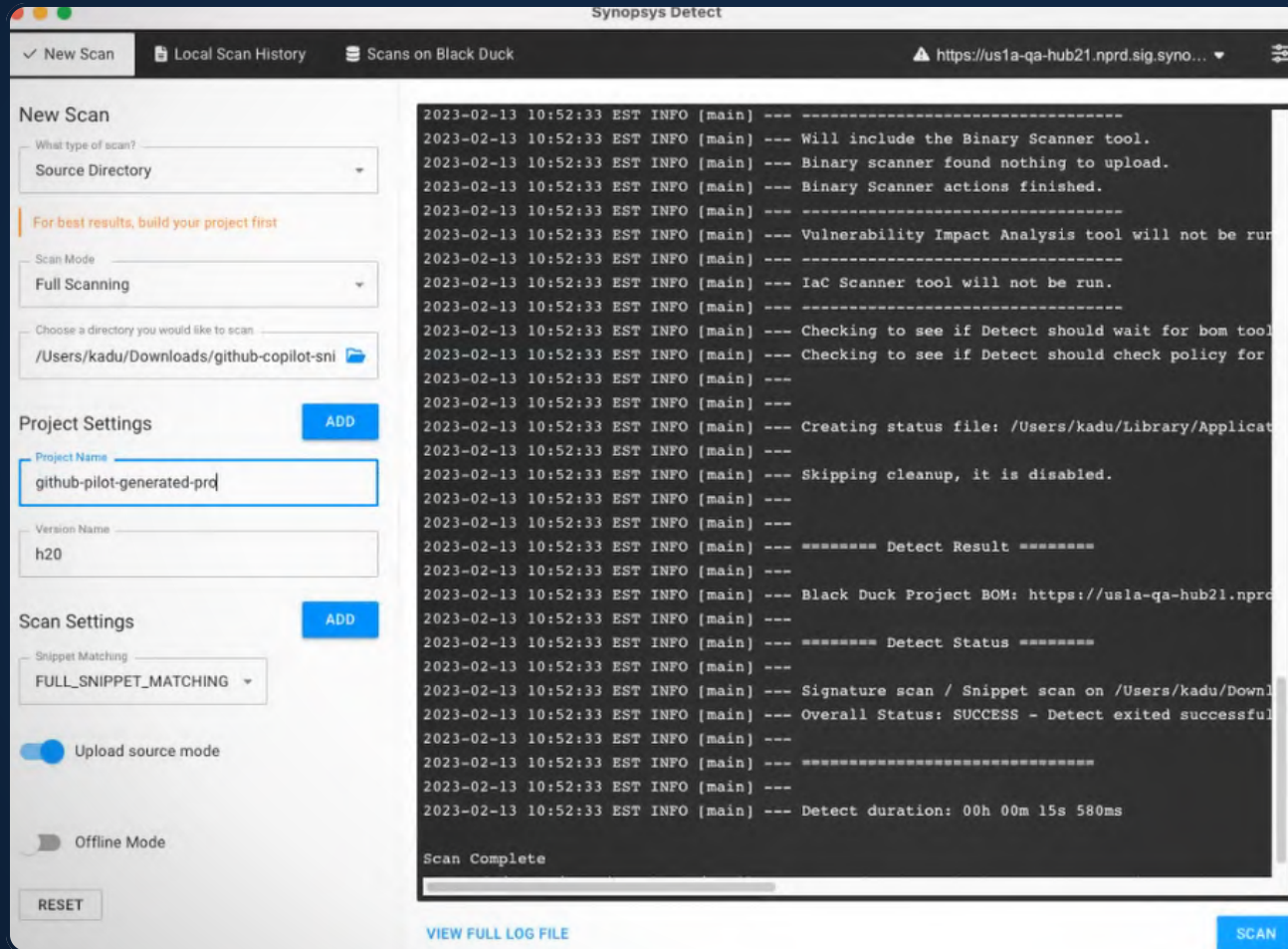


Vendor Risk

- Vendor dependency
- Information security

Risks Associated With Using GAI Tools

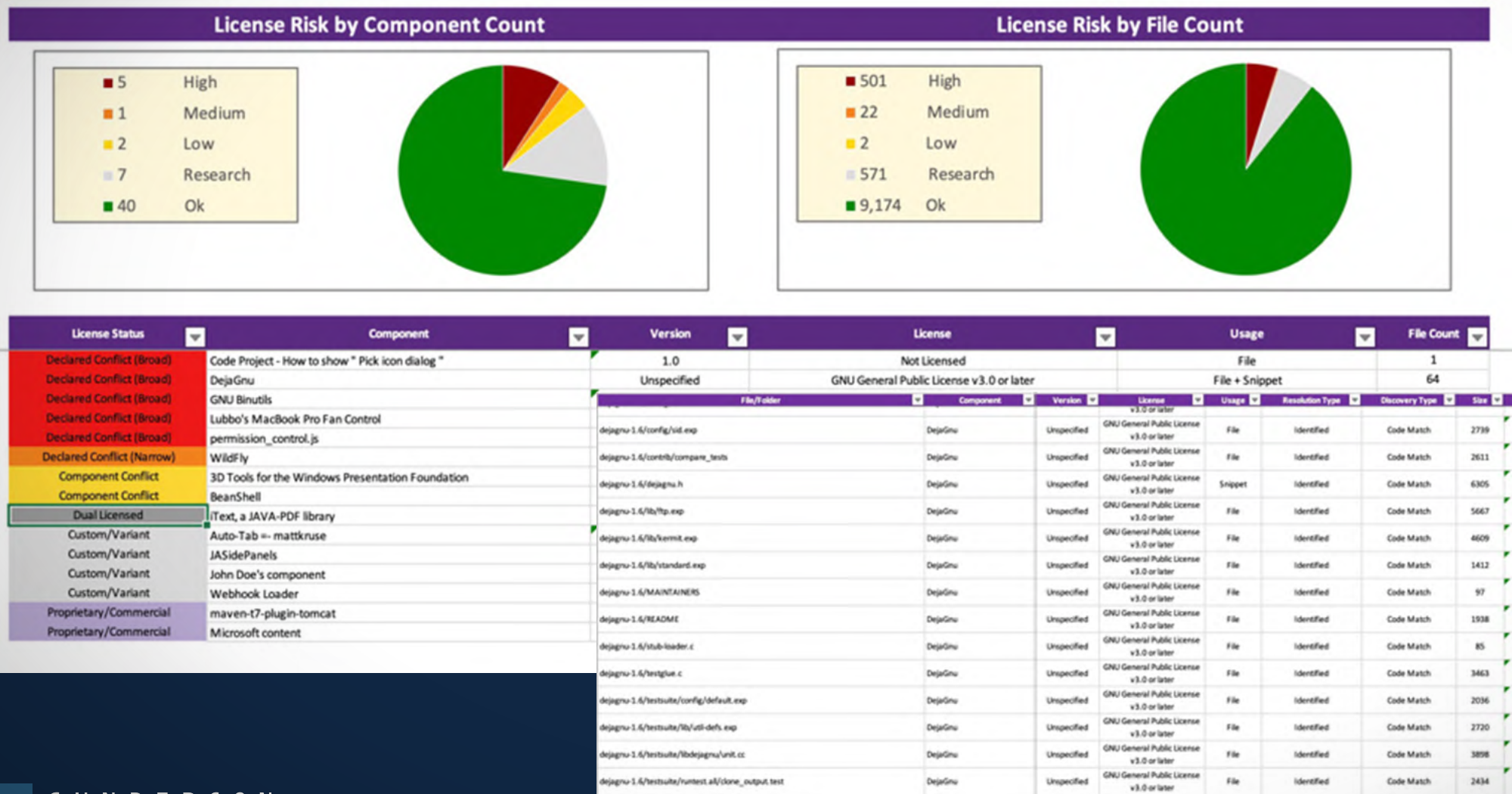
Secondary Transactional Risk: M&A and Financing Concerns



- **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

Secondary Transactional Risk: Technical Diligence



Risks Associated With Using GAI Tools

Secondary Transactional Risk: Technical Diligence

Snippet View

This adjustment will apply to all versions of this project - excluding archived versions

Comments Ignore Match Confirm Match Alternative Matches

Scanned File

github-copilot-generated.c

Scanned File Path: file:///Users/kadu/Downloads/github-copilot-snippet-scan/github-copilot-generated.c

File Size: 3.83 KB

(Line 88 to 99)

86 }
87
88 /* allocate a sparse matrix (triplet form or compressed-column form) */
89 cs *cs_spalloc (int m, int n, int nzmax, int values, int triplet)
90 {
91 cs *A = (cs *) cs_calloc (1, sizeof (cs)) ; /* allocate the cs struct */
92 if (!A) return (NULL) ; /* out of memory */
93 A->m = m ; /* define dimensions and nzmax */
94 A->n = n ;
95 A->nzmax = nzmax = CS_MAX (nzmax, 1) ;
96 A->nz = triplet ? 0 : -1 ; /* allocate triplet or comp.col */
97 A->p = (int *) cs_malloc (triplet ? nzmax : n+1, sizeof (int)) ;
98 A->i = (int *) cs_malloc (nzmax, sizeof (int)) ;
99 A->x = values ? (double *) cs_malloc (nzmax, sizeof (double)) : NULL ;
100 return ((!A->p || !A->i || (values && !A->x)) ? cs_done (A, NULL, NULL, 0)
101 : A) ;
102 }
103
104
105

Matched Component

acado last_1_1_0_beta

License: GNU Lesser General Public License v3.0 only

Release Date: Oct 6, 2013

Matched File Path: /acado-e2ba77daa5171e589c105c794cbf23c4c8385dea/external_packages/csparse/cs_util.c

Snippet Match: 17%
(Line 2 to 13)

Needs confirmation

1 #include "cs.h"
2 /* allocate a sparse matrix (triplet form or compressed-column form) */
3 cs *cs_spalloc (int m, int n, int nzmax, int values, int triplet)
4 {
5 cs *A = (cs *) cs_calloc (1, sizeof (cs)) ; /* allocate the cs struct */
6 if (!A) return (NULL) ; /* out of memory */
7 A->m = m ; /* define dimensions and nzmax */
8 A->n = n ;
9 A->nzmax = nzmax = CS_MAX (nzmax, 1) ;
10 A->nz = triplet ? 0 : -1 ; /* allocate triplet or comp.col */
11 A->p = (int *) cs_malloc (triplet ? nzmax : n+1, sizeof (int)) ;
12 A->i = (int *) cs_malloc (nzmax, sizeof (int)) ;
13 A->x = values ? (double *) cs_malloc (nzmax, sizeof (double)) : NULL ;
14 return ((!A->p || !A->i || (values && !A->x)) ? cs_spfree (A) : A) ;
15 }
16
17 /* change the max # of entries sparse matrix */
18 int cs_sprealloc (cs *A, int nzmax)
19 {
20 int ok, oki, okj = 1, okx = 1 ;

Close

Risks Associated With Using GAI Tools

Secondary Transactional Risk: M&A Deal Outcomes

- 1 | Prepare for deep diligence
- 2 | Consider risk shifting:
 - Software code remediation (as a signing or closing condition)
 - Enhanced Seller liability (expanded reps and warranties, special indemnities)
- 3 | Consider deal risk:
 - Time kills deals
 - Buyer risk aversion

ARTICLE VI PRE-CLOSING COVENANTS

Section 6.13 Remediation Plan. Section 6.13 sets forth the source code remediation plan agreed to by Parent and Company on or prior to the Agreement Date (the “Remediation Plan”). Prior to the Closing, Company shall take all steps necessary to comply with the Remediation Plan (and the Parties acknowledge that such pre-Closing efforts will focus on the subset of the Remediation Plan set forth on Schedule 8.01(o)). Prior to the Closing, Parent may, at its option, undertake or commission a scan of the source code described in the Remediation Plan.

ARTICLE VIII CONDITIONS TO CLOSING

Section 8.01 Conditions to Obligations of Parent. The obligation of Parent to consummate the Transactions shall be subject to the satisfaction or Parent’s waiver of, at or before the Closing, each of the following conditions:

...

(o) Code Remediation Steps. The code remediation action items set forth on Schedule 8.01(o) shall have been completed.

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

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

Agenda

- 1 | Potential Benefits of Using GAI Tools
— MCLE Code —
- 2 | Risks Associated With Using GAI Tools
- 3 | **Practical Ideas for Managing Risks**
— MCLE Code —

Practical Ideas for Managing Risks

Preparing for Technical Diligence

Get ahead of potential remediation efforts

1 | **Develop an internal compliance strategy**

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**

Certain GAI tools, including [CoPilot](#), include a filtering option that will block CoPilot from making 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

- 2976
- 1437



We want your feedback!

Please email us at insights@gunder.com and
phil.odence@synopsys.com

