ROYAL SIGNALS INSTITUTION

promoting professional dialogue, recognising technical achievement

Cloud Security in the Modern Age

27 February 2025

Capt Heath Groves

```
[~]$ whoami
```

ROYAL SIGNALS INSTITUTION

- **Who:** SO2 digital transformation, have been in the unit since Jan 7th, 2019
- **Tech skills:** Microsoft Certified Master (SharePoint), FBCS, CITP, 20-year MCT and O365 Expert Admin
- **Current Employer**: Sundown Solutions Ltd
- **Why I joined SGIS:** A had a feeling that my skillset might be in demand, and this proved correct. Joining a unit of likeminded individuals with a common shared experience is rare, certainly at the top end of the tech industry.
- **Where I came from:** 8.5 years in the REME (joined at 16) Then worked for MS on the O365 Product group as an SRE. Owned and operated Sundown Solutions since 2009
- **Where I live:** Hyndburn, Lancashire

Research – The Problem…

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

BTBDO - Dinner One

Monday, 6 June 2022 13:40

Diner number 1 - 30th June
Venue: UJC (Booked) Reserve Bar Lounge
Seats: 25 (+5 Sundown/speakers)

Speakers:
Heath Groves
Lucas Neill
Guy Batchelor

Order of March:

Timings

- 18.00 - Welcome drinks & Networking
- 18.45 - BEC to introduce and sit down
- 19.00 - First course
- 19.15 - Speaker Slot 1 - Heath Groves
  TOPIC: IT Sector Horizon Scanning, It's not just random guess work...
- 19.40 - Second course
- 20.15 - Speaker Slot 2 - Lucas Neill
  TOPIC: The perils & pitfalls for the HNW individual in the modern game

Content

Musings:

1) Hostile state actors
   a. Digital armour to supply chain - T1, T2, T3 & T4
2) What can be done about it?
3) Future scoping

Consider this statement - The manipulation of hidden assets takes place until regulation intervenes

If you have an asset that's normal day to day operation or implementation is not in your line of sight or periphery vision then you have a blind trust in the holder or operator that allows you to assume the asset is being managed responsibly and correctly.

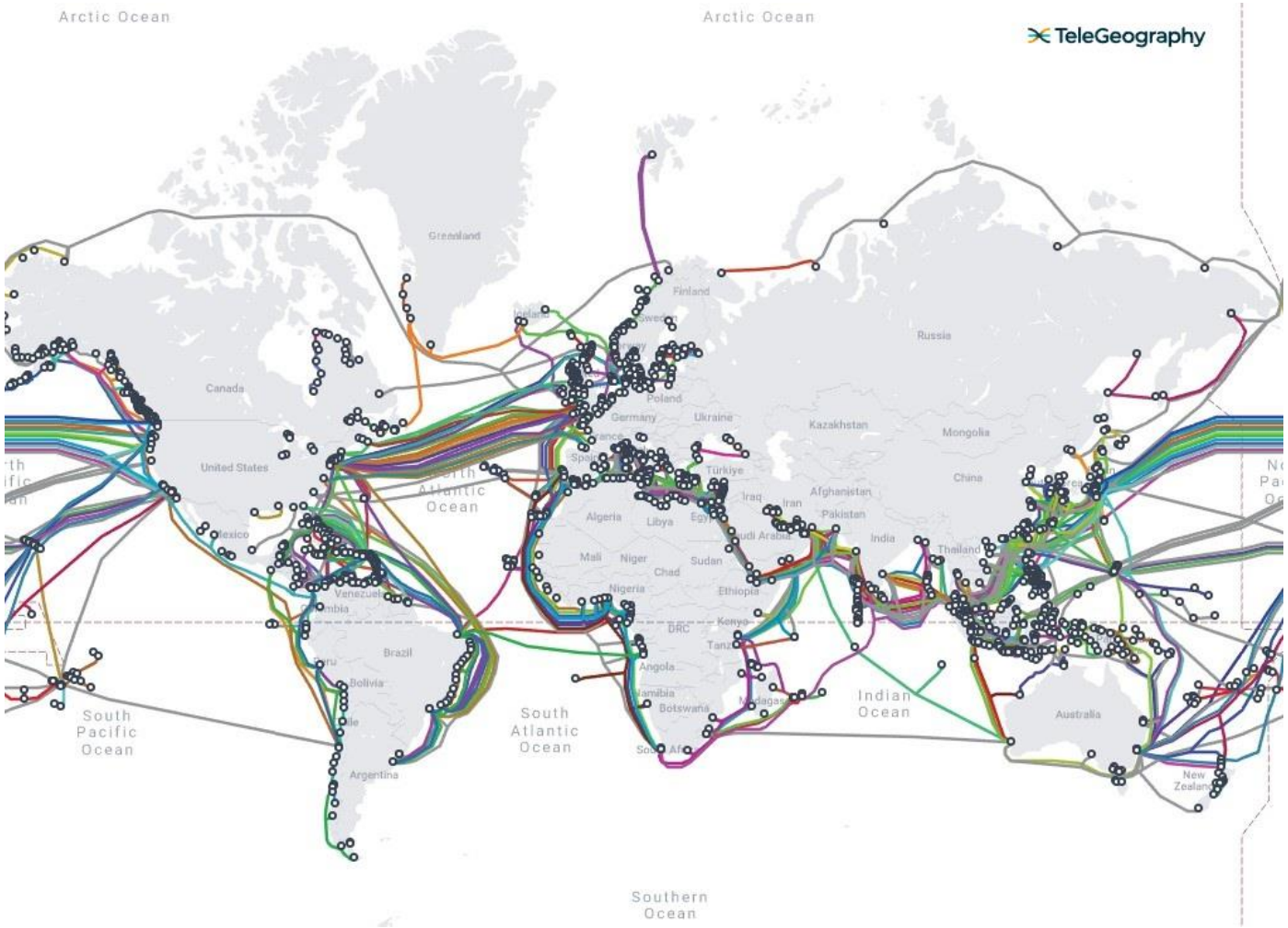
When blind trust exists without assured 3rd party regulation then it is abused to the benefit of the operator until regulation is applied.

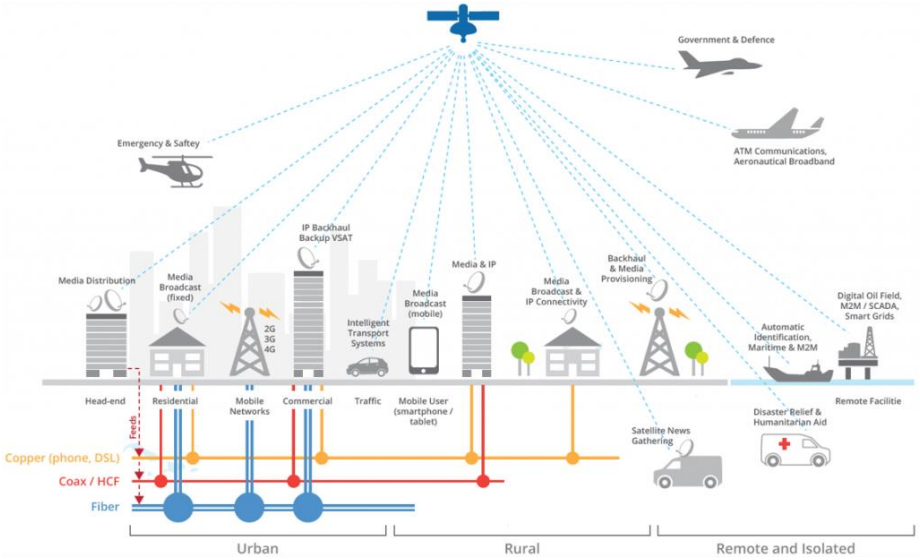
When the abuse is uncovered, outcry ensues that ensures 3rd party regulation is demanded.

Right now, your data is the asset and the laws are in place... but it is not being enforced... so infringements are rampant and often unknown by the perpetrators themselves.

Research – The Problem…

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement





DATA – Where is it?

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

A misinterpretation of guidance?

Despite these statements, Sayers said Microsoft has never given assurances that any data stored on its systems would always stay in the UK.

"People just chose to read it in that way," he said. "All Microsoft has ever done is guarantee that data would be stored at rest in a specific geography, and even then that guarantee is limited to certain services."

He continued: "In that regard, I have some limited sympathy for Microsoft, [because] users of its services perhaps haven't read the terms of service properly or conducted much in the way of due diligence before signing up to use its services. If they had done so, all this would have come into the public domain much sooner."

All the SPA did was ask Microsoft to confirm what the terms of service for its cloud products meant in practice, he continued. "Microsoft didn't duck the question – and it looks very much like the Scottish Police Authority were just the first to ask it."

Computer Weekly asked Microsoft if any government departments had ever contacted it directly for assurances about the sovereignty of data stored and processed within M365, but the company did not respond to the question.

The UK government's *Cloud guide for the public sector* document, which was jointly published in November 2023 by the Cabinet Office's technology arm, the Central Digital and Data Office (CCDO) and the Government Commercial Function, states that it is down to departments to decide where their cloud data should be hosted and, in short, their responsibility to ensure suppliers meet their requirements.

NEWS

UK government's M365 use under scrutiny after Microsoft's 'no guarantee of sovereignty' disclosure

Microsoft's hold on government IT is under scrutiny, following a disclosure to a Scottish policing body that saw the software giant advise that it cannot guarantee data sovereignty in its cloud-based Microsoft 365 suite

- https://www.computerweekly.com/news/366589468/UK-governments-M365-use-under-scrutiny-after-Microsofts-no-guarantee-of-sovereignty-disclosure

DATA – AUTHN & AUTHZ

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

okta

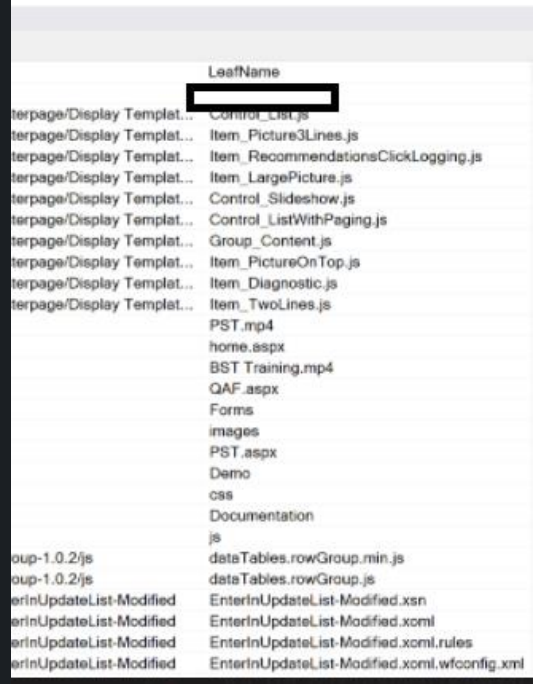
https://www.microsoft.com/en-us/security/blog/2021/06/30/the-critical-role-of-zero-trust-in-securing-our-world/

DATA – WHO HAS ACCESS?

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

This report makes two things clear:

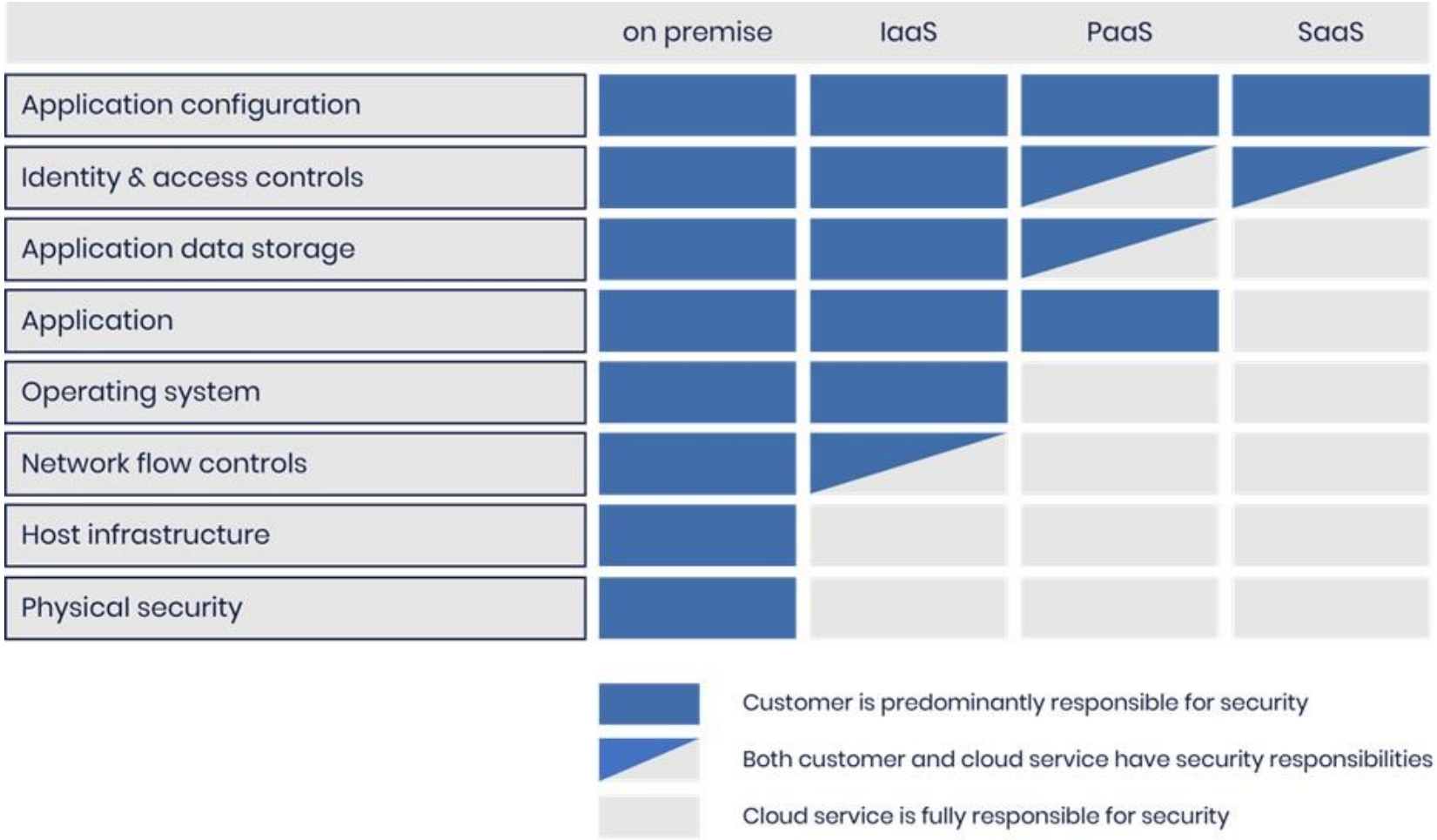
- Robust encryption must be implemented to be considered as an effective technical measure to control access in environments which have less trust due to concerns regarding law enforcement attempts to access data
- If data within the Cloud is able to be seen in the clear (e.g. the encryption mechanisms are capable of being access at will by CSP staff without the customers knowing) then any encryption deployed is not deemed to be an effective control



	LeafName
terpage/Display Templat...	[redacted]
terpage/Display Templat...	Item_Picture3Lines.js
terpage/Display Templat...	Item_RecommendationsClickLogging.js
terpage/Display Templat...	Item_LargePicture.js
terpage/Display Templat...	Control_Slideshow.js
terpage/Display Templat...	Control_ListWithPaging.js
terpage/Display Templat...	Group_Content.js
terpage/Display Templat...	Item_PictureOnTop.js
terpage/Display Templat...	Item_Diagnostic.js
terpage/Display Templat...	Item_TwoLines.js
	PST.mp4
	home.aspx
	BST Training.mp4
	QAF.aspx
	Forms
	images
	PST.aspx
	Demo
	css
	Documentation
	js
oup-1.0.2/js	dataTables.rowGroup.min.js
oup-1.0.2/js	dataTables.rowGroup.js
erInUpdateList-Modified	EnterInUpdateList-Modified.xsn
erInUpdateList-Modified	EnterInUpdateList-Modified.xoml
erInUpdateList-Modified	EnterInUpdateList-Modified.xoml.rules
erInUpdateList-Modified	EnterInUpdateList-Modified.xoml.wfconfig.xml

Shared Responsibility Model

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement



DATA – Who is responsible?

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

FIX – 1) Ask questions

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

- We should never be afraid to ask the hard questions, because most responsible CSP will often be well versed in these as they handle your data securely.

- Start simple – Ask your cloud provider the following questions:

- **Where is my Data held?** (This should include all copies of my data including ALL backups)
- **Who has access to my data and where are they located?** (This should include all copies of my data including ALL backups)
  - Why do they have access to my data?
- **Do you use any third parties to manage monitor or maintain aspects of the infrastructure that service my data?** If so:
  - **What do they have access to?**
  - **Why?**
  - **Where are they located?**

FIX – 2) MFA EVERYTHING

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

- Its is every single person's responsibility to ensure MFA is on every single account they use
- Do you do this already? Feel confident? Nice and complacent…
- What about:
  - All the accounts of the people in your immediate family that you share a wifi network with?
    - Kids Roblox account?
    - PlayStation Network?
    - Nintendo Network?
  - Great… now how about those you serve with and those under your command?
    - Are they secure?
      - Their families?
    - Social media?
    - Adult content suites?
      - Are they secure?
- Ensuring MFA is on every single-entry point is a constant battle, it has to be front and centre in EVERYONES mind, if you can not use MFA on a service then you shouldn't be using it!
- Note that in the current era MFA does not wholy guarantee security (Pass-the-cookie attacks etc.)

FIX – 3) Content is King

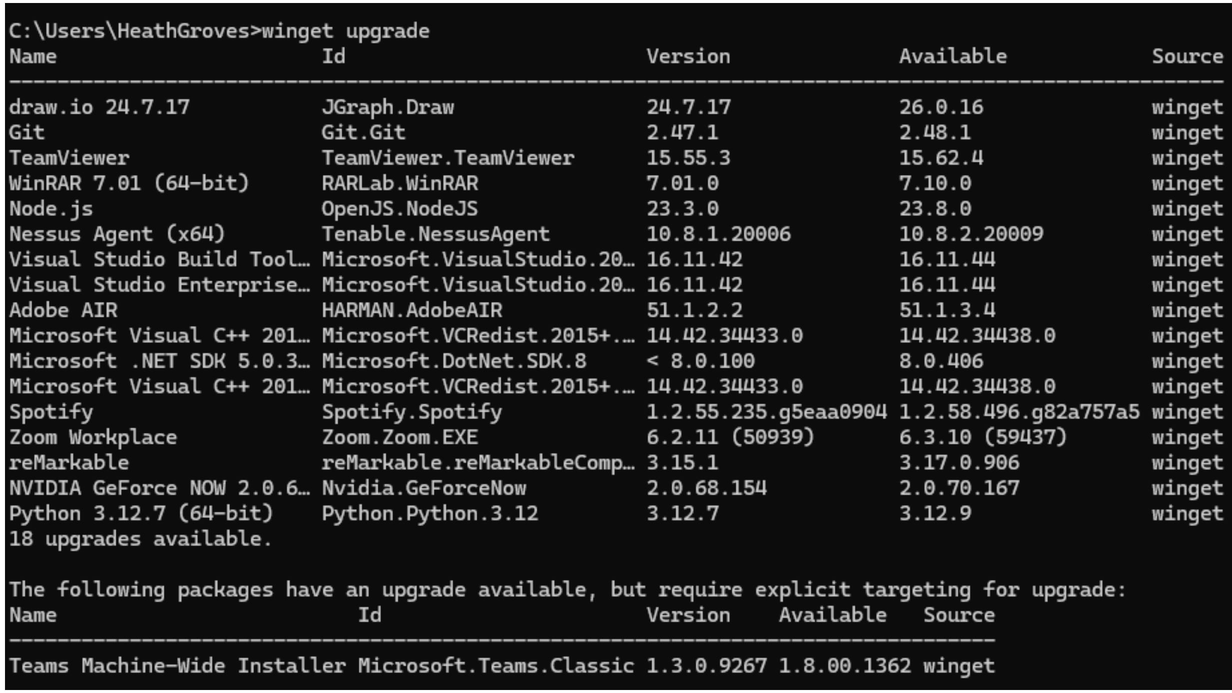
ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

- Encrypt your **CONTENT** – not just the container!
  - The container is simply the vessel that it is transported in
  - The content itself is the valuable piece
- If someone breaks into your account – they are “in” – but if the CONTENT is encrypted, then your safe!
- Consider that the threat actor may not be external but internal in either your organisation or your MSPs organisation!
- Take responsibility for you your own content security

FIX – 4) Patch Everything

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

- ***Winget Upgrade***
- Lists all third-party applications and extensions that require updating on your system
- Every one of them is a threat vector until patched
- To go one stage further and install the updates you can use the following switches:
- **-r** (Recurse – Upgrades all in the list)
- **-h** (Silent – Runs the installers in silent mode)
- For brevity, the full command is:
- ***Winget Upgrade -r -h***



```
C:\Users\HeathGroves>winget upgrade
Name                         Id                             Version            Available          Source
---------------------------------------------------------------------------------------------------------
draw.io 24.7.17              JGraph.Draw                    24.7.17            26.0.16            winget
Git                          Git.Git                        2.47.1             2.48.1             winget
TeamViewer                   TeamViewer.TeamViewer          15.55.3            15.62.4            winget
WinRAR 7.01 (64-bit)         RARLab.WinRAR                  7.01.0             7.10.0             winget
Node.js                      OpenJS.NodeJS                  23.3.0             23.8.0             winget
Nessus Agent (x64)           Tenable.NessusAgent            10.8.1.20006       10.8.2.20009       winget
Visual Studio Build Tool…    Microsoft.VisualStudio.20…     16.11.42           16.11.44           winget
Visual Studio Enterprise…    Microsoft.VisualStudio.20…     16.11.42           16.11.44           winget
Adobe AIR                    HARMAN.AdobeAIR                51.1.2.2           51.1.3.4           winget
Microsoft Visual C++ 201…    Microsoft.VCRedist.2015+.…     14.42.34433.0      14.42.34438.0      winget
Microsoft .NET SDK 5.0.3…    Microsoft.DotNet.SDK.8         < 8.0.100          8.0.406            winget
Microsoft Visual C++ 201…    Microsoft.VCRedist.2015+.…     14.42.34433.0      14.42.34438.0      winget
Spotify                      Spotify.Spotify                1.2.55.235.g5eaa0904 1.2.58.496.g82a757a5 winget
Zoom Workplace               Zoom.Zoom.EXE                  6.2.11 (50939)     6.3.10 (59437)     winget
reMarkable                   reMarkable.reMarkableComp…     3.15.1             3.17.0.906         winget
NVIDIA GeForce NOW 2.0.6…    Nvidia.GeForceNow              2.0.68.154         2.0.70.167         winget
Python 3.12.7 (64-bit)       Python.Python.3.12             3.12.7             3.12.9             winget
18 upgrades available.

The following packages have an upgrade available, but require explicit targeting for upgrade:
Name                         Id                       Version    Available    Source
-------------------------------------------------------------------------------------
Teams Machine-Wide Installer Microsoft.Teams.Classic 1.3.0.9267 1.8.00.1362 winget
```

Summary & Questions

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

- Ask the right questions
- MFA everything
- ENCRYPT your CONTENT – Not just the container!
- Patch everything!

- Any Questions?

AL & LLM Training

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

- Reviewed all current 24 active legal cases against AI firms surrounding copyright infringement
- Key theme – Illegal use of data has occurred
- Key defence – CDPA Fair use statement
- Common theme, being kicked into the long grass
- Working in the AI layer is key
  - AI Agents
  - The Agentic Layer

AL & LLM Training

ROYAL SIGNALS INSTITUTION

professional dialogue, technical achievement

Document Name	Case Number	Defendant(s)	Plaintif(s)	Type of Tool	Filed Date	Status	Court	Summary
gov.uscourts.cand.403220.135.0.pdf	4:22-cv-06823-JST	OpenAI OpCo, L.L.C. OpenAI GP, L.L.C. OpenAI Startup Fund I, L.P. OpenAI Startup Fund GP I, L.L.C.	J. Doe (Multiple)	AI technologies, specifically focusing on Copilot.	Jul-23	Active	United States District Court for the Northern District of California	A First Amended Complaint, filed against OpenAI and associated entities, alleges copyright infringement related to AI technologies such as Copilot. Plaintiffs accuse OpenAI of unlawfully using copyrighted materials in AI development. The complaint, initiated by copyright holders, highlights legal challenges at the AI-copyright intersection.
gov.uscourts.cand.414822.120.0.pdf	3:23-cv-03223-AMO	OpenAI, Inc.; OpenAI, L.P.; OpenAI OpCo, L.L.C.; OpenAI GP, L.L.C.; and related entities.	Paul Tremblay, Sarah Silverman, Christopher Golden, Richard Kadrey, Ta-Nehisi Coates, Junot Díaz, Andrew Sean Greer, David Henry Hwang, Matthew Klam, Laura Lippman, Rachel Louise Snyder, Jacqueline Woodson, and others.	ChatGPT (powered by GPT-3.5 and GPT-4)	Mar-24	Active	United States District Court, Northern District of California, San Francisco Division	A group of authors, allege that their copyrighted books were used without consent to train OpenAI's large language models. The case centers on claims of copyright infringement and seeks injunctive relief and damages
gov.uscourts.nysd.608447.74.0.pdf	1:23-cv-09152-LGS	Bloomberg L.P. and Bloomberg Finance, L.P.	Mike Huckabee et al. v. Bloomberg L.P. and Bloomberg Finance, L.P	Bloomberg's Large Language Models (LLMs) utilizing datasets such as "The Pile," including Books3	Jan-24	Active	United States District Court, Southern District of New York	They allege that Bloomberg used their copyrighted materials without authorisation to train its AI models, specifically citing the dataset Books3, which contains pirated eBooks
gov.uscourts.nysd.610699.47.0.pdf	1:23-cv-10211-SHS	OpenAI, Inc., OpenAI GP LLC, Microsoft Corporation, and related entities	Authors Guild, David Baldacci, Mary Bly, Michael Connelly, Sylvia Day, George R.R. Martin, Jodi Picoult, and others	GPT models (GPT-3, GPT-3.5, GPT-4)	Feb-24	Active	United States District Court, Southern District of New York.	Plaintiffs, comprising authors of fiction and nonfiction, claim that OpenAI and Microsoft infringed copyrights by using their works to train AI models without permission, threatening their livelihood and the value of their creative work
gov.uscourts.ded.72109.1.0_1.pdf	1:20-cv-00613-UNA	ROSS Intelligence Inc.	Thomson Reuters Enterprise Centre GmbH and West Publishing Corporation.	AI-based legal research tool	May-20	Closed	United States District Court, District of Delaware.	Plaintiffs allege that ROSS used unauthorised access to Westlaw's database to train its competing AI legal research product, constituting copyright infringement and breach of contract.
gov.uscourts.tnmd.96652.1.0.pdf	3:23-cv-01092	Anthropic PBC.	Concord Music Group Inc., Capitol CMG Inc., Universal Music Corp., and others.	Claude (AI model by Anthropic).	Oct-23	Active	United States District Court, Middle District of Tennessee.	Music publishers claim that Anthropic's AI unlawfully used copyrighted song lyrics to train its models without authorisation, violating their rights and causing harm to the music industry.
gov.uscourts.cand.407208.238.0_1.pdf	3:23-cv-00201-WHO	Stability AI Ltd., DeviantArt, Inc., MidJourney, Inc., and Runway AI, Inc.	Sarah Andersen, Kelly McKernan, Karla Ortiz, and others.	Stable Diffusion (AI image generation).	Oct-24	Active	United States District Court, Northern District of California.	Artists allege that defendants used billions of copyrighted images to train AI models without consent, credit, or compensation, damaging the art market.
gov.uscourts.ded.81407.1.0.pdf	1:23-cv-00135-UNA	Stability AI, Inc.	Getty Images (US), Inc.	Stable Diffusion (AI image generation).	Feb-23	Active	United States District Court, District of Delaware	Getty Images accuses Stability AI of copying over 12 million of its images without permission to train its AI model, violating copyright and trademark laws
gov.uscourts.cand.426188.1.0.pdf	3:24-cv-01451	Databricks, Inc.	Stewart et al.	MosaicML Pretrained Transformer (MPT)	Mar-24	Active	United States District Court, N.D. California	MosaicML allegedly used copyrighted books without consent to train AI models.
gov.uscourts.cand.426191.1.0_1.pdf	3:24-cv-01454	NVIDIA Corporation	Abdi Nazemian et al.	NeMo Megatron-GPT	Mar-24	Active	United States District Court, N.D. California	NVIDIA is accused of using copyrighted literary works for AI training without authorisation.
gov.uscourts.cand.428992.1.0.pdf	4:24-cv-02655	NVIDIA Corporation	Andre Dubus III et al.	NeMo Megatron-GPT	May-24	Active	United States District Court, N.D. California	Plaintiffs allege NVIDIA trained its AI on copyrighted works without authorisation.
gov.uscourts.mad.272063.1.0.pdf	1:24-cv-11611	Suno, Inc.	UMG Recordings et al.	Suno AI	Jun-24	Active	United States District Court, District of Massachusetts	Suno is accused of using copyrighted music recordings to train AI without obtaining permission.
gov.uscourts.nysd.612697.1.0.pdf	-	Microsoft Corporation, OpenAI Inc.	The New York Times Company	ChatGPT, Bing Chat	Apr-24	Active	United States District Court, S.D. New York	The New York Times alleges AI tools used its journalism to train models without consent, impacting revenue and credibility.
gov.uscourts.nysd.616536.1.0_1.pdf	1:24-cv-01515-UA	OpenAI Inc., Microsoft Corporation	The Intercept Media Inc.	ChatGPT	Feb-24	Active	United States District Court, S.D. New York	OpenAI allegedly trained ChatGPT using copyrighted journalism without authorisation.
gov.uscourts.nysd.620514.1.0.pdf	1:24-cv-03285	Microsoft Corporation, OpenAI Inc.	Daily News LP et al.	ChatGPT, Copilot	Apr-24	Active	United States District Court, S.D. New York	Local news publishers claim Microsoft and OpenAI used their copyrighted articles to train AI tools without consent.

AL & LLM Training

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

88. In a recent customer-support message, GitHub's support department clarified certain facts about training Copilot. First, GitHub said that "training for Codex (the model used by Copilot) is done by OpenAI, not GitHub." Second, in its support message, GitHub put forward a more detailed justification for its use of copyrighted code as training data:

> Training machine learning models on publicly available data is considered fair use across the machine learning community . . . OpenAI's training of Codex is done in accordance with global copyright laws which permit the use of publicly accessible materials for computational analysis and training of machine learning models, and do not require consent of the owner of such materials. Such laws are intended to benefit society by enabling machines to learn and understand using copyrighted works, much as humans have done throughout history, and to ensure public benefit, these rights cannot generally be restricted by owners who have chosen to make their materials publicly accessible.

The claim that training ML models on publicly available code is widely accepted as fair use is not true. And regardless of this concept's level of acceptance in "the machine learning community," under Federal law, it is illegal.

89. Former GitHub CEO Nat Friedman said in June 2021—when Copilot was released to a limited number of customers—that "training ML systems on public data is fair use."[15] Friedman's statement is pure speculation; no Court has considered the question of whether "training ML systems on public data is fair use." The Fair Use affirmative defense is only applicable to Section 501 copyright infringement. It is not a defense to violations of the DMCA, breach of contract, nor any other claim alleged herein. It cannot be used to avoid liability here. At the same time Friedman asserted "the output [of Copilot] belongs to the operator."

90. Other open-source stakeholders have made this point already. For example, in June 2021, Software Freedom Conservancy ("SFC"), a prominent open-source advocacy organization, asked Microsoft and GitHub to provide "legal references for GitHub's public legal positions." No references were provided by any of the Defendants.[16]

91. Beyond the examples above, Copilot regularly Output's verbatim copies of Licensed Materials. For example, Copilot reproduced verbatim well-known code from the game Quake III, use of which is governed by one of the Suggested Licenses—GPL-2.[17]

92. Copilot also reproduced code that had been released under a license that allowed its use only for free games and required attribution by including a copy of the license. Copilot did not mention nor include the underlying license when providing a copy of this code as Output.[18]

93. Texas A&M computer-science professor Tim Davis has provided numerous examples of Copilot reproducing code belonging to him without its license or attribution.[19]

94. 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.

95. In June 2022, Copilot had 1,200,000 users. If only 1% of users have ever received Output based on Licensed Materials and only once each, Defendants have "only" breached Plaintiffs' and the Class's Licenses 12,000 times. However, each time Copilot outputs Licensed Materials without attribution, the copyright notice, or the License Terms it violates the DMCA three times. Thus, even using this extreme underestimate, Copilot has "only" violated the DMCA 36,000 times.[20] Because Copilot constantly Outputs code as a user writes, and because nearly all of Copilot's training data was Licensed Material, this number is most likely exponentially lower than the true number of breaches and DMCA violations.

AL & LLM Training

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

https://www.youtube.com/watch?v=9vM4p9NN0Ts&t=2310s

- "These companies don't talk about how they do the data collection and also there's a copyright liability issue. They defiantly don't want to tell you that they've trained on books even though they did – because if not you can sue them."

- Yann Dubois PhD Student at Stanford – Aug 24

Potential Solutions

ROYAL SIGNALS INSTITUTION
professional dialogue, technical achievement

- Origin Attribution Page
- Retrospective Authorisation
- AI Content Removal Request
- Creative Equity Fund
- Machine Learning Amnesty

ROYAL SIGNALS INSTITUTION

promoting professional dialogue, recognising technical achievement

A professional body and a professional activity

Celebrate the successes of our talented people