CLYDE&CO

The challenges presented by AI information security

Nick Gibbons, Legal Director

AI is not synonymous with cyber:

- AI information security is not synonymous with cyber information security
- The target is different
- The attack vectors are different
- The range, type and amount of data targeted is potentially much broader

But ..

AI information security is already governed by existing legislation

AI breaches are already “secretly” covered by existing insurance policies

And….

Thid pary claims already governed by existing statute and common law

Many businesses do not understand AI risks

Proposal forms and policy wordings and insurers’ information security guidance generally focus on personal data breach and GDPR and do not address AI information security risks

Artificial intelligence is not "cyber"

- Artificial intelligence and "cyber" are not synonymous.

- The word "cyber" describes issues and events on the internet and computer networks connected to it. The internet is a means of communication and an information store that can be searched.

- AI is the field of computer science focused on creating machines or software that can perform tasks that typically require human intelligence.

- In 1950 Alan Turing published "Computing Machinery and Intelligence" and proposed the Turing Test.

- 1955, John McCarthy, a Professor of Mathematics at Dartmouth College, decided to organize a group to clarify and develop ideas about thinking machines. He picked the name 'Artificial Intelligence' for the new field.

Everybody is doing it

Businesses of every sort are making increasing use of AI for efficiency, economy and speed:

Healthcare

Law

Software Development & IT

Finance & Banking

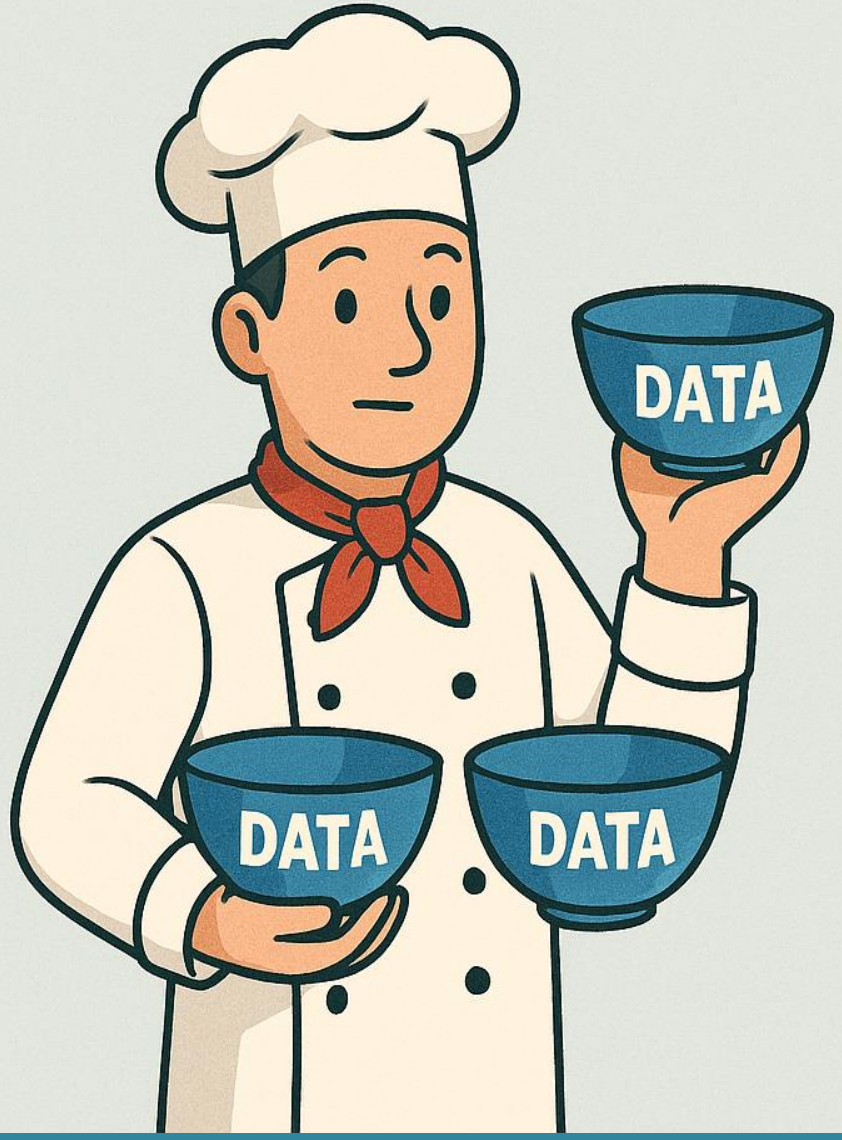
Marketing & Sales

Creative Professions (Design, Writing, Art)

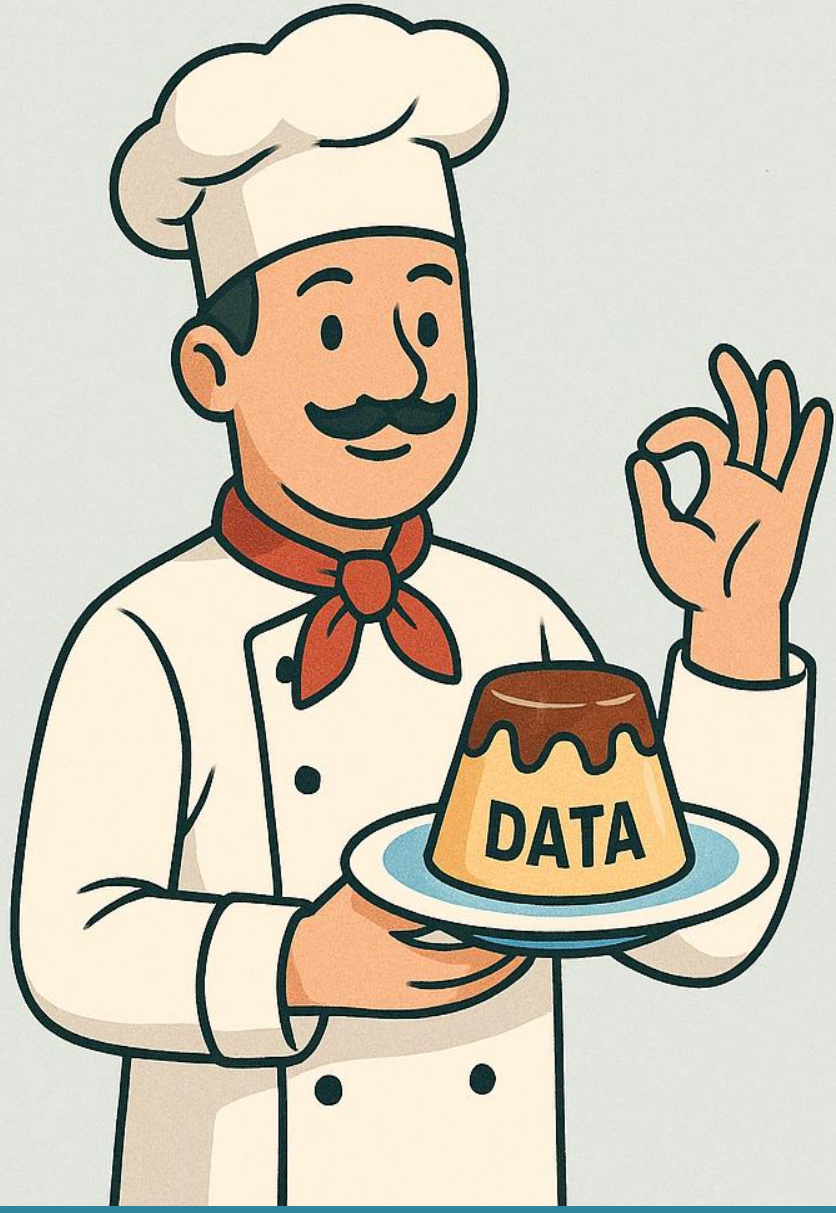
Education & Research

Engineering & Manufacturing

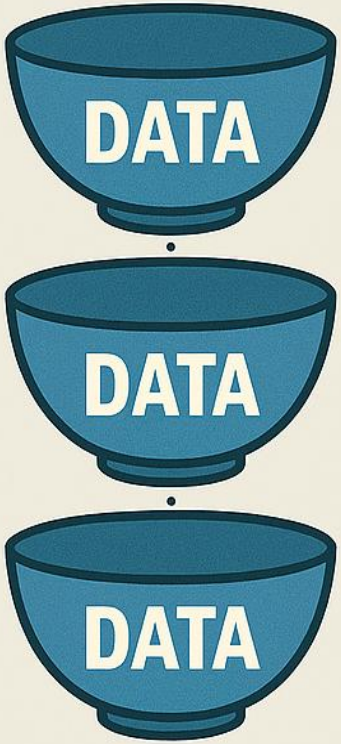
DATA IN A TRADITIONAL CYBER CONTEXT



DATA IN AN AI CONTEXT

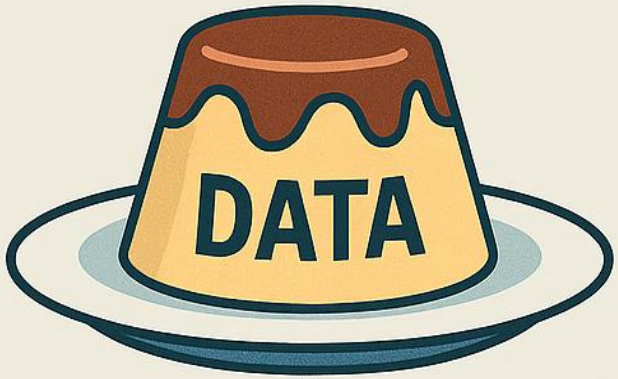


TRADITIONAL CYBER CONTEXT



SECURE EACH DATASET

AI CONTEXT



SECURE THE OVERALL MODEL

AGGREGATION TRANSFORMATION OUTPUT

- **Trational Hacking:**
- **Phishing :**Sending fraudulent emails or messages that appear legitimate to trick users into revealing sensitive information like passwords or credit card details.
- **Malware Injection**Deploying malicious software such as viruses, worms, trojans, or ransomware to compromise systems and steal or encrypt data.
- **SQL Injection**Exploiting vulnerabilities in web applications by inserting malicious SQL queries into input fields to access or manipulate databases.
- **Man-in-the-Middle (MITM) Attacks**Intercepting communication between two parties to eavesdrop or alter data without their knowledge.
- **Denial-of-Service (DoS) / Distributed DoS (DDoS)**Overloading a system or network with traffic to make it unavailable to legitimate users.

- **Traditional Hacking**

- **Password Cracking**Using brute force, dictionary attacks, or rainbow tables to guess or decrypt passwords.

- **Social Engineering**Manipulating individuals into divulging confidential information through psychological tactics rather than technical exploits.

- **Exploiting Unpatched Vulnerabilities**Taking advantage of outdated software or systems that haven't been updated with security patches.
- **Keylogging**Installing software or hardware to record keystrokes and capture sensitive information like login credentials.

AI HACKING

Adversarial Attacks

- **Description:** Crafting inputs (images, text, audio) that look normal to humans but cause AI models to misclassify or behave incorrectly.
- **Example:** Slightly altering pixels in an image so a vision model mistakes a stop sign for a speed limit sign.

2. Model Inversion

- **Description:** Using access to a trained model to reconstruct sensitive training data.
- **Example:** Recovering faces from a facial recognition model or extracting private medical records.

3. Data Poisoning

- **Description:** Injecting malicious or misleading data into the training set to corrupt the model's behavior.
- **Example:** Adding mislabeled examples so the model learns incorrect associations.

4. Prompt Injection

- **Description:** Manipulating input prompts in LLMs or AI agents to override instructions or leak sensitive data.
- **Example:** Embedding hidden instructions in user queries to make the model reveal confidential system prompts.

AI HACKING

5. Model Extraction (Stealing)

- **Description:** Querying an AI model repeatedly to approximate its parameters or replicate its functionality.
- **Example:** Building a clone of a proprietary model by observing its outputs.

. Backdoor Attacks

Description: Implanting hidden triggers during training so the model behaves normally until a specific input activates malicious behavior.

Example: A classifier that works fine but misclassifies when a certain watermark appears.

7. Membership Inference

Description: Determining whether a specific data point was part of the model's training set.

Example: Inferring if a person's medical record was used to train a health prediction model.

8. Gradient Leakage

Description: Exploiting gradients shared during collaborative training (e.g., federated learning) to reconstruct private data.

Example: Rebuilding sensitive images from gradient updates.

9. Jailbreaking AI Systems

Description: Bypassing safety filters or alignment constraints in LLMs or autonomous agents.

Example: Using clever phrasing or encoding to make an AI produce disallowed content.

10. Synthetic Identity Generation

Description: Using generative models to create realistic fake identities for fraud or impersonation.

Example: Deepfakes for social engineering or bypassing KYC checks.

The impact of AI in legal terms – all the time in the world?

- AI technology and the use of the technology **seems to be** running way ahead of specific new law and regulation
- In fact AI activities are already covered by existing non AI specific law
- Significant scope for organisations and individuals to break existing law and regulation because they don't understand and may not even have considered the interaction between AI and existing law.
- In some ways new AI legislation such as the EU AI Act is a distraction and red herring that creates the false impression of an attempt to regualate an unregulated wild west whereas, in reality, AI is already covered by very well trodden existing law and new regulation is somewhat esoteric.

AI legislation

Existing EU and UK legislation and caselaw already covers AI information security and wil sanction reaches

- GDPR
- UK NIS Regulations 2018
- Digital Operations Resilience Act (“DORA”)
- EU Artificial Intelligence Act
- Network and Information Security Directive 2
- UK common law and statute

Differences between GDPR and “new”legislation

- GDPR only concerns personal data: new legsation covers every type of data

  every type of data : commercial information, trade secrets,IP ,

  confidential information
- GDPR technical security guidance is unspecific: new legislation is much more

  detailed and specific
- GDPR generally only bites after an incident has been reported to the ICO:

  new legislation includes auditing provisions
- GDPR fines are infrequently imposed:It appears that
- New legislation will entail much more rigorous sanctions

DORA

- Digital Operations Resilence Act
- Mandatory risk reporting
- Digital Operational Resiliency testing including threat led penetration testing
- Information and intelligence sharing
- Managing ICT third party risk
- European Supervisory authority to establish arrangements with regulators in non EU countries

DORA FOOTPRINT

DORA applies to over 21,000 EU financial institutions including
Banks , credit companies, investment funds, insurers and
ICT service providers

Financial institutions outside of EU must also comply if they
Provide critical ICT processing to EU financial entities.

UK Insurers and many of their policyholders therefore need to comply

Network and Information Security (NIS) Directive 2

- Clear and precise rules
- All types of data – not just personal data
- More entities and sectors will have to take measures to protect themselves:
- “Essential sectors” such as the energy, transport, banking, health, digital infrastructure,

  public administration and space sectors will be covered by the new security provisions.
- companies, governmental and public bodies
- The new rules will also protect so-called “important sectors” such as

  postal services, waste management, chemicals, food,

  manufacturing of medical devices, electronics, machinery, motor vehicles and

  digital providers.
- All medium-sized and large companies in selected sectors will fall under legislation.

Cyber Security and Resilience (Network and Information Systems) Bill,

- introduced to Parliament on **12 November 2025**.
- Covers all types of data not just personal data
- updates and expands the existing **UK NIS Regulations 2018** (which were based on the original NIS Directive)
- designed to align closely with NIS 2 requirements while tailoring them to UK needs.

Cyber Security and Resilience (Network and Information Systems) Bill

Stricter Incident Reporting:

- Initial notification within **24 hours** of becoming aware of an incident, followed by a full report within **72 hours**.
- Incidents must also be reported to the **National Cyber Security Centre**

•**Regulatory Powers & Enforcement:**

- Increased fines (up to **£17 million or 4% of global turnover**).
- Regulators gain broader powers to impose cybersecurity requirements and share information internationally. []

•**Alignment with NIS 2:**

- The Bill mirrors NIS 2's goals of strengthening supply chain security and expanding coverage but introduces UK-specific enforcement and national security provisions.

EU Artificial Intelligence Act

Comprehensive set of rules for providers and users of AI systems, which details

transparency and reporting obligations

- expected to affect *all AI systems impacting people in the EU*, including any company

placing an AI system on the EU market or companies whose system outputs are being used

within the EU (regardless of where systems are developed or deployed).

- Large fines :

Up to 7% of global annual turnover or €35m for prohibited AI violations.

Up to 3% of global annual turnover or €15m for most other violations.

Up to 1.5% of global annual turnover or €7.5m for supplying incorrect info s**.**

"Secret" AI cover in insurance policies: why it matters

- Many businesses are using AI for everything they do: research, marketing, videos, websites, Instagram reels and accounts.

  …….....without properly checking the results.

- Many of those in business , charities do not properly understand information assets and laws : defamation; intellectual property; trade secrets and confidential information.

- AI will process information assets and produces results in seconds and may cross many lines in the process.

- Who can or will check what it produces – probably not the data protection officer, director or senior administrator.

Existing cyber exclusion clauses won’t bite

- Most computer/cyber exclusion clauses are intended to exclude liability for viruses and data breaches caused by hackers.
- These clauses won't exclude liability for unreasonable reliance on AI solutions.
- Very few insurance policies incorporate exclusion clauses covering AI risk.
- The caselaw on unreasonable reliance on AI is still in its infancy.
- A fortiori existing insurance policies are very unlikely to be construed as not covering AI risk unless there is a specific AI exclusion clause.

AI risk in a business and insurance context

Much broader than cyber risk

AI risk in a business insurance context is broader than cyber risk and cover and is likely to include :

Breach of contract

negligence;

negligent misstatement or negligent misrepresentation;

breach of a duty of care or confidence including any misuse of information which is either confidential or subject to statutory restrictions on its use;

infringement of intellectual property rights (including copyright, trademark, design, title, slogan or moral rights) or any act of passing off;

libel or slander;

breach of professional duty generally.

Impact on insurers

- Cyber insurance has been principally concerned with privacy and personal data.
- New legislation much more complex and concerns all types of online data
- Insurers and their policyholders will need to comply
- Policy wordings will need to change to accommodate changes including
- new security yardsticks which will be recognised and enforced by the courts
- Insurers and brokers will need to get their heads round hew risks abd cover
- Policy wordings will need to cover every type of data not just personal data
- Group companies and ICT service providers will need to be checked and contracts amended
- Compliance will be costly
- Ambulance chasing solicitors will have many more opportunities