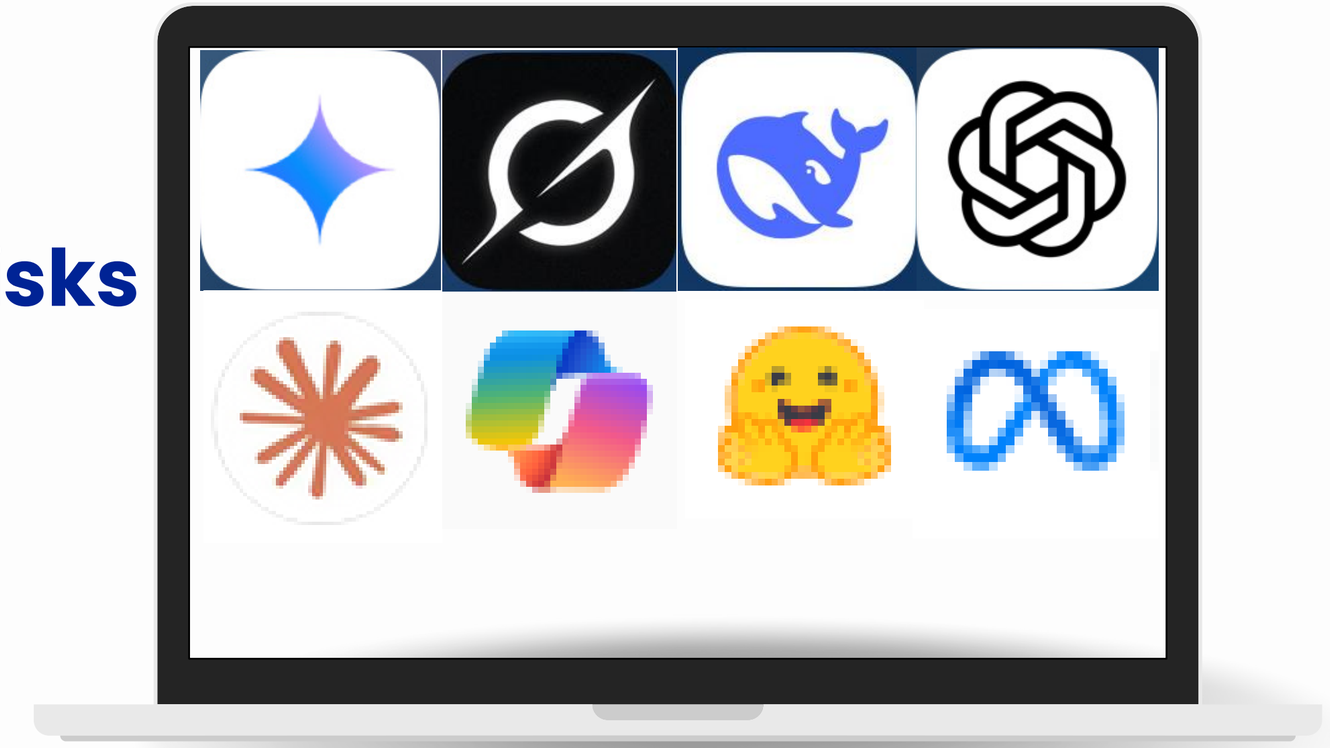
Securing the Future:

Vetting AI Solutions for Privacy and Security Risks

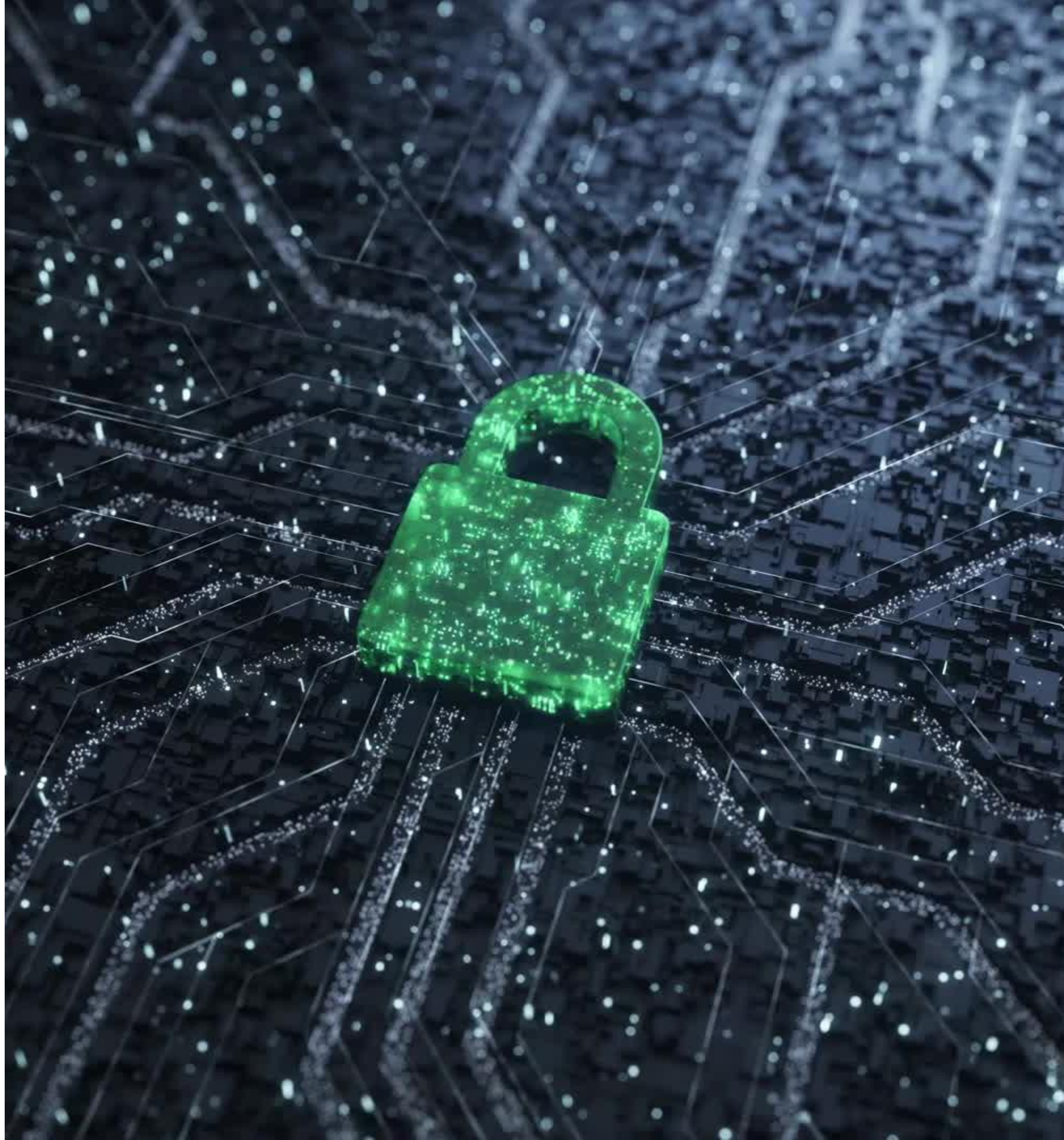


Presented by: Dave Vetuschi

Lawrie
INSURANCE GROUP

Session Overview

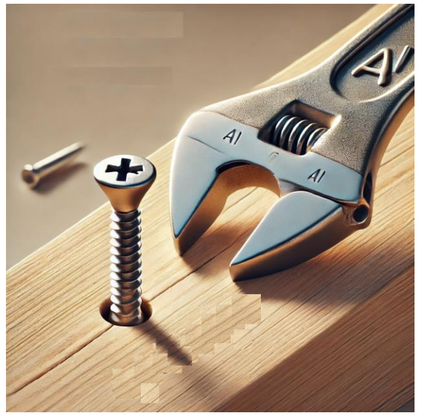
- Understanding AI-Specific Risks & Mitigation Strategies
- The Role of IT leaders in Solution Selection
- Importance of AI Privacy and Security
- Vetting Tools



Lawrie
INSURANCE GROUP

First some food for thought

- Define a clear understanding of the business objective
  - Take into consideration processes, governance and policies as possible solutions first before trying to find a technology to solve the objective



- What can I do with AI?
  - What can you build with Lego?

- Do you know where your children are?
  - Do you know where the data about your children's whereabouts are stored?

Lawrie
INSURANCE GROUP

The Rise of AI WAR in the Enterprise

- Key Drivers of AI
  - Simplicity
  - Automation
  - Customized and auto adaptiveness
- AI Phishing Tools and Bot integration
  - FraudGPT, WormGPT
  - AkiraBot
- AI Email Filtering
  - IRONSCALES, Barracuda
  - Cisco AI-Defence
- AI Specific attacks
  - Occupy AI
  - SugarGh0st Rat

Lawrie
INSURANCE GROUP

Understand your AI Model

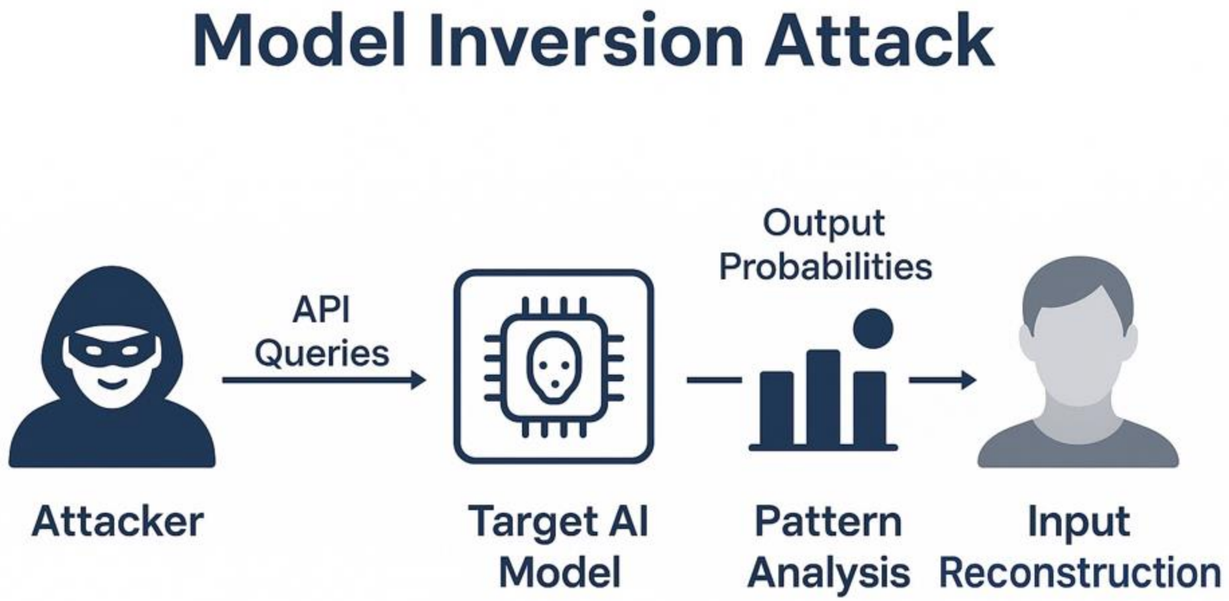
- Where is the input & output stored?
  - Most cases violating HIPAA, GDPR, PIPEDA
- Are logs available through the AI model?
- Is the learning algorithm input and output encrypted, secured and private?
- Is it LLM or GAI
  - LLM's tokenize the inpute
  - LLM's When user says A, than respond B
    - Correct if incorrect, add the correction of what B should be
  - GAI are context aware and primarily for photo, AV

```
Usim a mixinuan ru6b.wnted as MSoRa-reaty, LLLM — QloRa GPT-NeoX
from transformers tmportformers
   AutoModelForCausalLM, AutoTokonizer, TrainingArguments, Traner
   fron datesetet load dataset
   prepare_model_for_kbit_training, LoraConfig, get_peft_model

1  Load base s model and tokenizer
model_name = "mistralai/Mistral-78v0.1"'
model = AutoModelForCausalLM,from_pretrainemprimman6)
model = prepare_model_for_kbit_training(modal_mame='Misstrai-78)

1  Prepare for QLoRa training
modal = prepare_model for_kbit_trainining(model)
Apply LoRa for parameter-efficient fine-tuning
lora_config = r=8, lora_alpha=16, ['q_proj", 'v_proj],
lora_dropout=0.05, bias='none', task_type=
task_type='CAUSAL_LH) lora_config

4 Load and preprocesss dataset
dataset = load_dataset("Abirate/englisglish_quotes")
function tokenize(sample)
   return tokenize(tokenize([quof=]),trunctation=true, pading='max
     length=128, max_length=128
tonkinized = get_peft_model(/ltmllm)

5 Define training arguments
training_args = TrainingArgume.nts
   output_dir='./results'
   per_device_train_batch_size=2, num_train_epochs=3
   save_steps=500, logging_dir='/.logs'
   fpi6 True
   optim = paged_adamw_8bit
   save_total_linit=2

Start training  trainer.save_model.("/llm-finetuned")
tokenizer.save_pretrained("/llm-finetuned")
```

Lawrie
INSURANCE GROUP

AI-Specific Privacy and Security Risks

- Data Leakage
  - Violating HIPAA, GDPR, PIPEDA
- Model Inversion Attacks
  - API based reconstruction
- Shadow AI
- Deepfakes and AI Hallucinations
- Supply Chain Compromise
  - 2021-2023 Microsoft malicious AI packages “torchtriton,tensorboardx”
- Supply Chain Compromise
  - Malicious Backdoor & vulnerabilities
    - DAN Mode
    - Poisoned training data



Lawrie
INSURANCE GROUP

Mitigation

- Validate and hash-check all third-party AI models and datasets
- Use secure model registries with signed artifacts (e.g., MLflow + Sigstore)
- Regularly audit dependency trees
  - Monitor usage and input and output
    - Even using another secured AI to parse the logs
  - e.g., using pip-audit, safety, or DependencyTrack.Monitor for trigger-based behavior in AI model outputs (via red teaming and adversarial testing.
- Use SBOMs (Software Bill of Materials)

Lawrie
INSURANCE GROUP

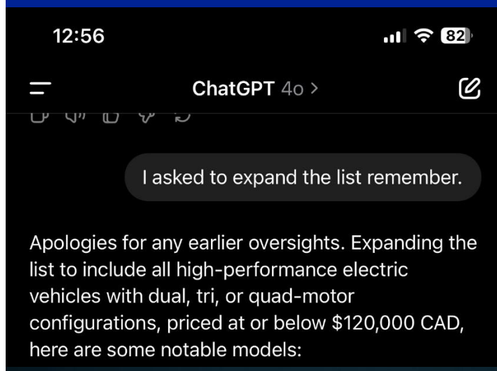
Epic AI Failures

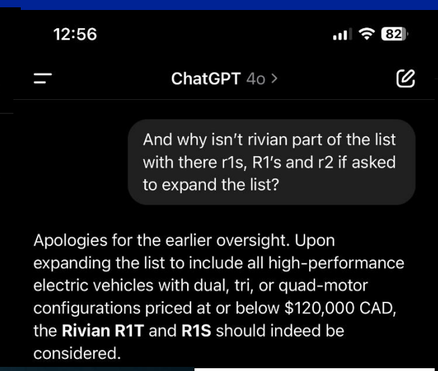
- ChatGPT 4o Product research failures
  - Biased output omitting Tesla because it was luxury
    - Included Luxury which then omitted Hyundai Ioniq an Kia EV6 GT
  - Excluded Ford Mach-E GT by “mistake”
  - Excluded Rivian R1 because it assumed

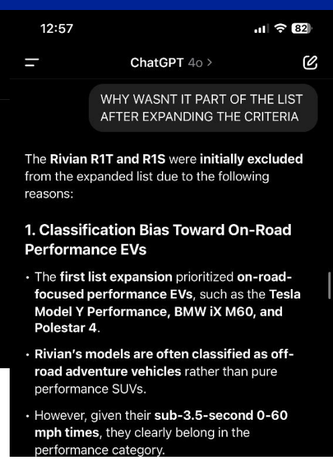
- Gemini Not knowing basic geography and spelling
  - Make me a map of the world with lines coming from Canada Ontario to other countries around the world

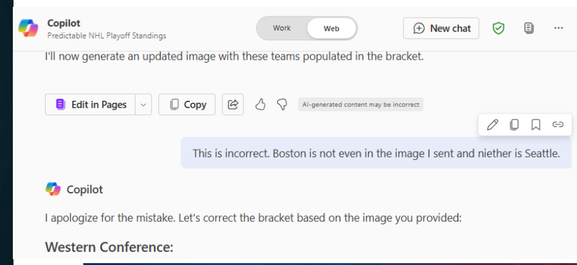
- Co-pilot not processing inputted data correctly
  - Predicting Boston and Seattle in NHL standings

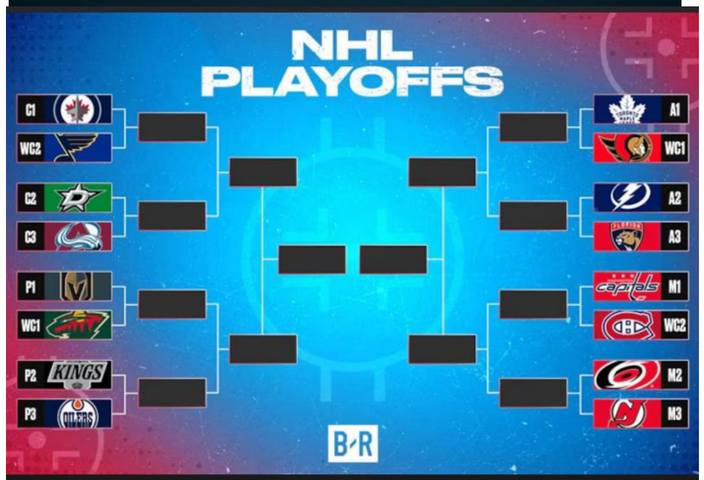
- Grok malformed output of images
  - Making up teams that don’t exist and duplication











Lawrie
INSURANCE GROUP

Mitigation ‘Policies & Governance’ Playbook

- Establish Oversight Committees
  - Ensure resources for ongoing auditing
- Implement Governance Frameworks
  - Ensure users scrutinize the output
    - Do not use it as outputted without review
  - Use AI for subjects users have knowledge about
- Monitor Interactions
- Regular Simulations and baseline verification testing

Lawrie
INSURANCE GROUP

Security & Privacy Vetting Checklist

- Data Collection Practices
- Data Retention and Lifecycle Management
- Data Encryption
- Security Stack Evaluation
- Anonymization and Tokenization
- User Consent Flows
- Control Data Access
  - Support for SSO and MFA
- API Security
- Dependency Scanning

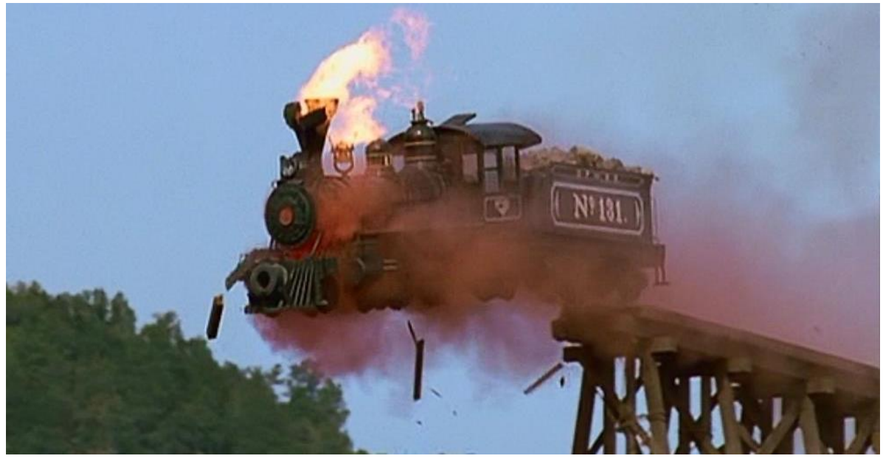
Lawrie
INSURANCE GROUP

CISA and NIST Security Guidance Summary

Area	Summary
AI Supply Chain Threats	Highlighted risks of using **external or open-source AI models** that may be **maliciously tampered**, contain **backdoors**, or use **poisoned data**.
Dependency Risk	Warned against over-reliance on **unvetted third-party libraries, APIs**, and **data sources**.
Model Integrity	Advised use of **model provenance, cryptographic signatures**, and **SBOMs (Software Bill of Materials)** for AI components.
Data Governance	Urged strict policies for **data labeling, ingestion, storage**, and **monitoring**, especially to avoid bias and poisoning attacks.
Continuous Monitoring	Recommended **AI-specific threat detection**, audit logging, and **runtime behavior validation** of deployed models.
Risk Management Alignment	Encouraged use of NIST's **AI Risk Management Framework (AI RMF)** as a standard across federal and critical infrastructure sectors.

3 Approaches to technology

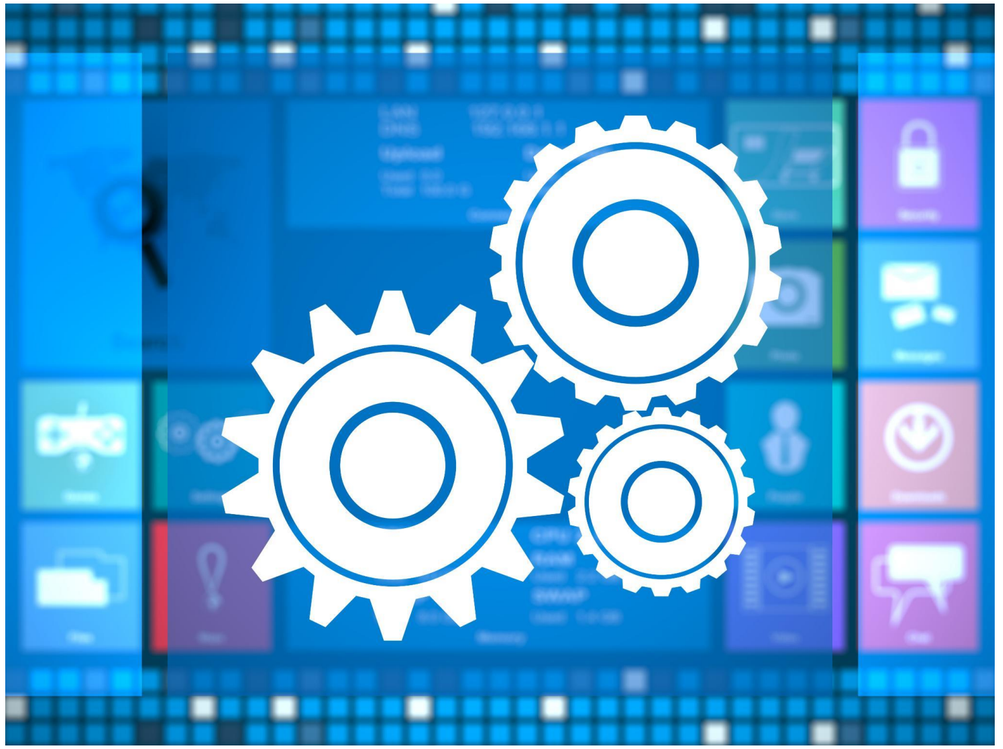
- 1. Adopting technology too quickly & recklessly
  - Unknown risks associated
  - Shiny new toy losses luster
  - Legal Liabilities
  - Wasted investment with no ROI
  - Often identified by
    - “Why isn’t it done yet.”
    - “Just do it.”
    - Jumping in and figuring it out
- 2. Structured approach
  - Evaluate, vet and filter based on structured logic
  - Staged deployment with strategic steps to the end goal
- 3. Resisting change
  - Behind the competition
  - Still doing things the “old” way
  - Clunky, time-consuming processes
  - Reluctant or resistant to look at faster innovative ways doing things
  - Often identified by
    - “Oh I don’t use that!”
    - “I don’t know how, Wasn’t trained on that.”
    - “I’m too busy. It’s too complicated”



Lawrie
INSURANCE GROUP

The Infinite Advantages of AI Solutions

- AI enhances efficiency by automating repetitive tasks
- Improves decision-making with data-driven insights
- Facilitates predictive analytics for better planning
- Enables personalized customer experiences and services
- Reduces operational costs through optimized resource management
- Detailed product sourcing and comparison
- Programming and code design
- Marketing and content for marketing
- Presentation development “Just like this presentation”
- Anytime there’s large amounts of data that need to be analyzed or refined
- Better than search engines “as long as you set the context”
- Analytics of Audio, Visual and even facial recognition



Lawrie
INSURANCE GROUP

Sample Security & Privacy Vetting Checklist

Mandatory Vendor Selection Criteria											
Company	ChatGPT	Microsoft Co-Pilot	Grok	Deepseek	Adobe Assista	Tango	Scribe	Claude	Read.ai	ASC AI	Teams Premiun
Criteria	**(Yes/No/NA)**	**(Yes/No/NA)**	**(Yes/No/NA)**	**(Yes/No/NA)**	**(Yes/No/NA)**	**(Yes/No/NA)**	**(Yes/No/NA)**	**(Yes/No/NA)**	**(Yes/No/NA)**	**(Yes/No/NA)**	**(Yes/No/N**
Data stored in Canada?	No	Yes	No	No	No	No	No	No	No	Yes	Yes
Cross-border data transfer encripted in transit and at rest (if exception made)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vendor complies with Canada Legeslation (IE. PIPEDA, including consent, breach notification, retention policies)	Yes	Yes	UNKNOWN	UNKNOWN	Yes	Yes	Yes	UNKNOWN	Yes	Yes	Yes
Encryption of data at rest and in transit	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Multi-factor authentication (MFA) available	Yes	Yes	Yes	UNKNOWN	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Role-based access control (RBAC) in place	Yes	Yes	Yes	UNKNOWN	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vendor complies with CASTLE (opt-in consent, unsubscribe mechanism, no deceptive content) "IF EMAIL APPLICABLE"	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Legal agreements include data protection, Cyber Insurance & compliance guarantees	Yes	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	Yes	UNKNOWN	UNKNOWN	UNKNOWN	Yes
Nice-to-Have Vendor Selection Criteria											
Criteria	(Yes/No)	(Yes/No)	(Yes/No)	(Yes/No)	(Yes/No)	(Yes/No)	(Yes/No)	(Yes/No)	(Yes/No)	(Yes/No)	(Yes/No)
SOC 2 compliance	Yes	Yes	UNKNOWN	UNKNOWN	Yes	Yes	Yes	Yes	Yes	UNKNOWN	Yes
ISO 27001 compliance	Yes	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	Yes
Isolated proccessing environment	Yes	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	Yes	Yes
NIST Cybersecurity Framework alignment (e.g., NIST CSF, 800-53)	Yes	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	Yes
CIS (Center for Internet Security) benchmark adherence	Yes	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	UNKNOWN	UNKNOWN	UNKNOWN	Yes
Vendor has a strong market reputation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Solution is scalable and integrates with existing systems	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Flexible licensing and cost structure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Conducts regular security audits and penetration testing	Yes	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	Yes	Yes	UNKNOWN	UNKNOWN	Yes
Pilot deployment possible for evaluation	Yes	Yes	UNKNOWN	UNKNOWN	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Canadian-based support and service	No	Yes	No	No	No	No	No	No	No	No	Yes
Incident response plan and breach notification processes exist	Yes	Yes	UNKNOWN	UNKNOWN	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regular audit readiness and testing (aligns with NIST AU-6)	Yes	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	Yes
Security assessments performed prior to onboarding (aligns with NIST CA-2)	Yes	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	Yes	UNKNOWN	UNKNOWN	Yes
Client Lawrie owns all data and has control over data and permissions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cost per year for total amount of licences	$240 (ChatGPT P	Varies (bundled w	$192 (X Premiu	Free (open-so	Varies (Adobe	$240 (Pro, per	$276 (Pro, per user)	Varies	$180 (Pro, per user)	Varies	Varies (Microsc
Estimated total cost savings/ROI it would provide total hours per year.	Varies	Varies	Varies	Varies	Varies	Varies	Varies	Varies	Varies	Varies	Varies

Lawrie
INSURANCE GROUP

Wrap-Up & Q&A

- Vet Deeply
- Align Policies Across Teams
- Eliminate Shadow AI
- Continuous Re-evaluation
- Key Takeaway: Be Proactive
- Encourage Questions
- Foster Discussions

Lawrie
INSURANCE GROUP

Resources for Leaders

- **N**ational **I**nstitute of **S**tandards and **T**echnology
  - NIST AI Risk Management Framework
  - https://www.nist.gov/itl/ai-risk-management-framework
- **C**ybersecurity and **I**nfrastructure **S**ecurity **A**gency
  - CISA AI Security Guidelines
  - https://www.cisa.gov/ai
- **O**pen **W**orldwide **A**pplication **S**ecurity **P**roject
  - OWASP Top 10
  - https://owasp.org/www-project-top-ten/
- SOC 2 Compliance for AI
  - https://www.compassitc.com/blog/achieving-soc-2-compliance-for-artificial-intelligence-ai-platforms?utm_source=chatgpt.com
- AI Governance Alliance
  - https://initiatives.weforum.org/ai-governance-alliance/home?utm_source=chatgpt.com
- ISO/IEC 23894
  - https://www.iso.org/standard/77304.html?utm_source=chatgpt.com

Lawrie
INSURANCE GROUP